Current Applied Science and Technology Vol. 23 No. 23 No. 2 (March-April 2023)

Research article

New Generalized Regression Estimators Using a Ratio Method and Its Variance Estimation for Unequal Probability Sampling without Replacement in the Presence of Nonresponse

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

[1]*Department of Mathematics, 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*

Received: 12 January 2022, Revised: 17 June 2022, Accepted: 12 August 2022

DOI: 10.55003/cast.2022.02.23.007

Abstract

One of the most important problems for planning in economics is that reliable data can be difficult to obtain either because it has not been recorded or because of nonresponse in surveys. This paper is aimed at proposing new generalized regression estimators using the ratio method of estimation for estimating population mean and population total and also variance estimators of the proposed generalized regression estimators in the presence of uniform nonresponse of a study variable. We show in theory that the proposed estimators are almost unbiased under unequal probability sampling without replacement when nonresponse occurs in the study. In the simulation studies, the performances of the proposed estimators were better when compared to the existing ones in terms of minimum relative bias and relative root mean square error. In an application to Thai maize in Thailand with 2019 data, we can see that the proposed estimators gave smaller variance estimates when compared to the existing estimators.

Keywords

ratio estimator;

generalized regression estimator;

automated linearization approach;

response probabilities;

nonresponse

1. Introduction

The Thai economy is weak due in part to lack of investment. Therefore, the Royal Thai government has a policy that targets ten industries in order to improve growth of the Thai economy under the 'Thailand 4.0' initiative. Production efficiency and competitiveness have become a major problem in Thailand's industrial economic structure. The nonresponse issue can lead to poor planning and decision making in business and economics as decisions are being made based on incomplete data. We need to address the problem of nonresponse before data can be used effectively in financial

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

planning. Hansen and Hurwitz [1] first pointed out the existence of the issue of nonresponse in mail surveys and then introduced a subsampling technique. Särndal and Lundström [2] proposed the generalized regression (GREG) estimator and investigated the variance of the new estimator in the presence of nonresponse under a two-phase framework whereby the selected sample was considered in the first phase and nonresponse was studied in the second phase. The GREG estimator for estimating population total, which was developed from the Horvitz and Thompson estimator [3], incorporates the weighting method, which helps to reduce nonresponse bias. The GREG estimator is a form of nonlinear estimator, so its properties such as expectation and variance can be obtained using the Taylor linearization approach. However, this approach requires each estimator to be derived separately. Estevao and Särndal [4] proposed the application of an automated linearization approach to estimate the variance of the GREG estimator. Chauvet [5] proposed a variance estimator for estimators from a 2006 French housing survey. Their proposal involved unit nonresponse and calibration and was applied to real data from the city of Rennes. Complementary samples were selected from a basic national sample that was obtained from a multistage sampling design.

Lawson and Ponkaew [6] proposed a new GREG estimator for estimating total population using Lawson's estimator [7], which proposed a new population total estimator in the form of a nonlinear ratio estimator using unequal probability sampling without replacement. They also proposed the variance of the new GREG estimator under a reverse framework where the nonresponse mechanism was missing completely at random (MCAR). In the reverse framework that was introduced by Fay [8], the order of the first and second phases from the two-phase framework was reversed. Lawson and Ponkaew [6] studied the scenario of having a small sampling fraction in which the response probabilities were uniform. Recently, Lawson and Panich [9] proposed a new GREG estimator that was made by adjusting Lawson and Ponkaew's estimator [6] using different nonresponse mechanisms when the response probability was non-uniform and the sampling fraction was large and could not be omitted.

The efficiency of population mean or population total estimators can be improved by having a known auxiliary variable, a variable that is positively related to the study variable, using the ratio estimator which was pointed out by Cochran [10]. The ratio estimator is very popular in research because it is highly efficient. The ratio estimator is biased, but the bias becomes less noticeable for large sample sizes. Many available parameters for auxiliary variables have been applied to the ratio estimators to increase their efficiency in estimating population mean. Bacanli and Kadilar [11] suggested a new ratio estimator made by replacing the usual population total estimator with the Horvitz and Thompson estimator under unequal probability sampling without replacement. Later Ponkaew and Lawson [12] proposed a new ratio estimator based on the Bacanli and Kadilar [11] and Särndal and Lundström [2] estimators for estimating population total where nonresponse existed with MCAR mechanism and small sampling fraction.

We proposed new ratio GREG estimators based on the Ponkaew and Lawson [12] and Lawson and Ponkaew [6] estimators that used a ratio estimator to create more efficient estimators and almost unbiased estimators in the presence of nonresponse using unequal probability sampling. The previous estimators used the MCAR mechanism where the sampling fraction was negligible. Despite the proposed GREG estimators being studied under the same circumstances as the preceding estimators, they were developed to be used when the sampling fraction was large and could not be omitted, and could not be used for all sampling fractions. We also suggest variance estimation methods for the proposed GREG estimators for use in the case of uniform nonresponse, for both when the response probabilities are known and for when the response probabilities are unknown. This work can be improved by using a known auxiliary variable to increase the efficiency of the estimator when it is being used where non-response exists, which might make it suitable for use as an estimator and be stringent enough for economic forward planning.

2. Materials and Methods

2.1 Basic setup

Consider a finite population $U=\{1,2,\ldots,N\}$ of size N. Let y be a study variable, and y_i be the value of y for a unit labeled i for all $i\in U$. We aim to estimate the population total of y defined by $Y=\sum_{i\in U} y_i$. Suppose we have information about three auxiliary variables denoted by x, k and w. Let $\boldsymbol{X}_N=(\boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_N)'$ be the $N\times(q+1)$ matrix of values x and $\boldsymbol{x}_i=(1\ x_{i1}\ \cdots\ x_{iq})'$ is the $(q+1)\times 1$ vector of values of the q variates for all unit $i\in U$. The auxiliary variables x were used as calibration variables. The vector of values of auxiliary variables k are $(k_1\ k_2\ \cdots\ k_N)'$ and they are used to determine values of first and joint inclusion probabilities under unequal probability sampling without replacement. The vector $(w_1\ w_2\ \cdots\ w_N)'$ defines the value of the auxiliary variables w for constructing the ratio estimator.

Under unequal probability sampling without replacement (UPWOR), a sample s of size n was selected. Let $\mathcal{F}$ be the set of all possible subsets of U and sampling design $P(\bullet)$ be the probability measure for possible s, i.e. $P(s)\geq 0$ for all $s\in\mathcal{F}$. 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 sample s of size n, it is assumed that the information of $n\times(q+1)$ matrix of values x or $\boldsymbol{X}_n=(\boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_n)'$ is known for all $\boldsymbol{x}_i$ when $i\in s$. We also define $E_S(\bullet)$ and $V_S(\bullet)$ as the expectation and variance operators, respectively, with respect to UPWOR sampling design.

In the presence of nonresponse, let subscript R and r_i be the nonresponse mechanism and nonresponse indicator variable of y_i which $r_i=1$ if unit i responds to item y, 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 uniform nonresponse. Let $E_R(\bullet)$ and $V_R(\bullet)$ be the expectation and variance operators with respect to the nonresponse mechanism.

2.2 The existing estimators

2.2.1 The Ponkaew and Lawson estimator

Ponkaew and Lawson [12] proposed an adjusted ratio estimator which is an almost unbiased estimator for estimating population total, following the Bacanli and Kadilar [11] and Särndal and Lundström [2] estimators, where nonresponse occurs in the study. They considered circumstances under the uniform nonresponse mechanism where the sampling fraction is negligible. The Ponkaew and Lawson [12] estimators for estimating the population total is given by

$$\hat{Y}_R^* = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\sum_{i\in s}\frac{w_i}{\pi_i}} \sum_{i\in U} w_i = \frac{\hat{Y}_r^*}{\hat{w}_{HT}} W , \tag{1}$$

where $\hat{Y}_r^* = \sum_{i\in s}\frac{r_i y_i}{\pi_i p}$, $\hat{w}_{HT} = \sum_{i\in s}\frac{w_i}{\pi_i}$, $W = \sum_{i\in U} w_i$.

Then the ratio estimator for estimating population mean is defined as

$$\hat{\bar{Y}}_R^* = \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\frac{1}{N}\sum_{i\in s}\frac{w_i}{\pi_i}} \frac{1}{N}\sum_{i\in U} w_i = \frac{\hat{\bar{Y}}_r^*}{\hat{\bar{w}}_{HT}} \bar{W} , \tag{2}$$

where $\hat{\bar{Y}}_r^* = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}$, $\hat{\bar{w}}_{HT} = \frac{1}{N}\sum_{i\in s}\frac{w_i}{\pi_i}$, $\bar{W} = \frac{1}{N}\sum_{i\in U} w_i$.

2.2.2 The Lawson and Ponkaew estimator

Lawson and Ponkaew [6] proposed new GREG estimators and variance estimators for estimating population mean and population total using unequal probability sampling without replacement under a reverse framework. The nonresponse mechanism was uniform, and the sampling fraction was negligible. The Lawson and Ponkaew [6] estimators are almost unbiased estimators, and they are given by

$$\hat{\bar{Y}}_{GREG.LP} = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i}}{\sum_{i\in s}\frac{r_i}{\pi_i}} + \left(\bar{\boldsymbol{X}} - \frac{\sum_{i\in s}\frac{r_i \boldsymbol{x}_i}{\pi_i}}{\sum_{i\in s}\frac{r_i}{\pi_i}}\right)' \left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i \boldsymbol{x}_i'}{\pi_i}\right)^{-1} \left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i y_i}{\pi_i}\right) = \hat{\bar{Y}}_r + \left(\bar{\boldsymbol{X}} - \hat{\bar{\boldsymbol{X}}}_r\right)' \hat{\boldsymbol{\beta}}_r , \tag{3}$$

$$\hat{Y}_{GREG.LP} = N\hat{\bar{Y}}_{GREG.LP} = N\left[\hat{\bar{Y}}_r + \left(\bar{\boldsymbol{X}} - \hat{\bar{\boldsymbol{X}}}_r\right)' \hat{\boldsymbol{\beta}}_r\right], \tag{4}$$

where $\hat{\bar{Y}}_r = \sum_{i\in s}\frac{r_i y_i}{\pi_i} \Big/ \sum_{i\in s}\frac{r_i}{\pi_i}$, $\hat{\bar{\boldsymbol{X}}}_r = \sum_{i\in s}\frac{r_i \boldsymbol{x}_i}{\pi_i} \Big/ \sum_{i\in s}\frac{r_i}{\pi_i}$, $\hat{\boldsymbol{\beta}}_r = \left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i \boldsymbol{x}_i'}{\pi_i}\right)^{-1} \left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i y_i}{\pi_i}\right)$,

$\bar{\boldsymbol{X}} = \frac{1}{N}\sum_{i\in U} \boldsymbol{x}_i$.

An automated linearization approach was used to find the variance of $\hat{Y}_{GREG.LP}$ under the reverse framework, and the overall sampling fraction was negligible. The Lawson and Ponkaew [6] variance estimators are shown in equations (5) and (6):

$$V_1(\hat{Y}_{GREG.LP}) \approx \frac{1}{p}\sum_{i\in U} D_i e_i^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} D_{ij} e_i e_j \,, \tag{5}$$

$$V_2(\hat{Y}_{GREG.LP}) \approx \frac{1}{p}\sum_{i\in U} D_i (e_i - \overline{e})^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} D_{ij} (e_i - \overline{e})(e_j - \overline{e}) \,, \tag{6}$$

where $D_i = (1-\pi_i)\pi_1^{-1}$, $D_{ij} = (\pi_i\pi_j - \pi_{ij})(\pi_i\pi_j)^{-1}$, $e_i = (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})$, $\overline{e} = \frac{1}{N}\sum_{i\in U} e_i$.

The estimators of $V_1(\hat{Y}_{GREG.LP})$ and $V_2(\hat{Y}_{GREG.LP})$ are obtained respectively by

$$\hat{V}_1(\hat{Y}_{GREG.LP}) \approx \left(\frac{N}{\sum_{i\in s}\frac{r_i}{\pi_i}}\right)^2 \left[\sum_{i\in s} \hat{D}_i r_i \hat{e}_i^2 + \sum_{i\in s}\sum_{j/\{i\}\in s} \hat{D}_{ij} r_i \hat{e}_i r_i \hat{e}_j \right], \tag{7}$$

$$\hat{V}_2(\hat{Y}_{GREG.LP}) \approx \left(\frac{N}{\sum_{i\in s}\frac{r_i}{\pi_i}}\right)^2 \left[\sum_{i\in s} \hat{D}_i r_i (\hat{e}_i - \hat{\overline{e}}_r)^2 + \sum_{i\in s}\sum_{j/\{i\}\in s} \hat{D}_{ij} r_i (\hat{e}_i - \hat{\overline{e}}_r)^2 r_j (\hat{e}_j - \hat{\overline{e}}_r)^2 \right], \tag{8}$$

where $\hat{D}_i = (1-\pi_i)\pi_1^{-2}$, $\hat{D}_{ij} = (\pi_i\pi_j - \pi_{ij})(\pi_{ij}\pi_i\pi_j)^{-1}$, $\hat{e}_i = (y_i - \boldsymbol{x}_i\hat{\boldsymbol{\beta}}_r)$ and $\hat{\overline{e}}_r = \frac{\sum_{i\in s} r_i\hat{e}_i/\pi_i}{\sum_{i\in s} r_i/\pi_i}$.

Under the reverse framework, the variance of GREG estimator $\hat{Y}_{GREG.LP}$ is defined by

$$V(\hat{Y}_{GREG.LP}) = E_R V_S(\hat{Y}_{GREG.LP} \,|\, \boldsymbol{R}) + V_R E_S(\hat{Y}_{GREG.LP} \,|\, \boldsymbol{R}) = V_1 + V_2 \,, \tag{9}$$

where $V_1 = E_R V_S(\hat{Y}_{GREG.LP} \,|\, \boldsymbol{R})$ and $V_2 = V_R E_S(\hat{Y}_{GREG.LP} \,|\, \boldsymbol{R})$.

Lawson and Ponkaew [6] studied specifically the scenario of a small sampling fraction, so the overall sampling fraction is negligible. The formulas from equations (5)-(8) were derived from equation (9) when the second component (V_2) was omitted. Then, we suggested the Lawson and Ponkaew [6] under a situation where the sampling fraction was large, and the second component could not be omitted. From equation (9) V_2 is defined by

$$V_2 = V_R E_S(\hat{Y}_{GREG.LP} \,|\, \boldsymbol{R}) \,. \tag{10}$$

However, $\hat{Y}_{GREG.LP}$ is nonlinear so we use the automated linearization approach to transform $\hat{Y}_{GREG.LP}$ to a simple form which is defined by

$$\hat{Y}_{GREG.LP} \approx \bar{\boldsymbol{X}}'\boldsymbol{\beta} + \frac{1}{\sum_{i\in s}\frac{r_i}{\pi_i p}}\sum_{i\in s}\frac{r_i e_i}{\pi_i p}, \tag{11}$$

where $e_i = (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})$.

Substituting equation (11) into equation (10), then

$$V_2 \approx V_R E_S\left(\bar{\boldsymbol{X}}'\boldsymbol{\beta} + \frac{1}{\sum_{i\in s}\frac{r_i}{\pi_i p}}\sum_{i\in s}\frac{r_i e_i}{\pi_i p}\middle|\boldsymbol{R}\right)$$

$$\approx V_R\left(\bar{\boldsymbol{X}}'\boldsymbol{\beta} + \frac{1}{\sum_{i\in U}\frac{r_i}{p}}\sum_{i\in U}\frac{r_i e_i}{p}\right) = V_R\left(\frac{\sum_{i\in U}\frac{r_i e_i}{p}}{\sum_{i\in U}\frac{r_i}{p}}\right)$$

Therefore,

$$V_2 \approx V_R\left(\frac{\sum_{i\in U}\frac{r_i e_i}{p}}{\sum_{i\in U}\frac{r_i}{p}}\right) \tag{12}$$

From equation (12), we see that $\sum_{i\in U}\frac{r_i e_i}{p}\Big/\sum_{i\in U}\frac{r_i}{p}$ is nonlinear. By using the Taylor linearization approach to transform it, we may write

$$\frac{\sum_{i\in U}\frac{r_i e_i}{p}}{\sum_{i\in U}\frac{r_i}{p}} \approx \bar{e} + \frac{1}{N}\sum_{i\in U}\frac{r_i}{p}(e_i - \bar{e}), \tag{13}$$

where $e_i = (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})$ and $\bar{e} = \frac{\sum_{i\in U} e_i}{N}$.

Substituting equation (13) into equation (12),

$$V_2 \approx V_R\left(\bar{e} + \frac{1}{N}\sum_{i\in U}\frac{r_i}{p}(e_i - \bar{e})\right) = \frac{1}{N^2}\sum_{i\in U}\frac{V_R(r_i)}{p^2}(e_i - \bar{e})^2$$

$$=\frac{1}{N^2}\sum_{i\in U}\frac{p(1-p)}{p^2}(e_i-\overline{e})^2$$
$$=\frac{1}{N^2}\frac{(1-p)}{p}\sum_{i\in U}(e_i-\overline{e})^2$$

Therefore,

$$V_2\approx\frac{1}{N^2}\frac{(1-p)}{p}\sum_{i\in U}(e_i-\overline{e})^2 \tag{14}$$

Then, new variance of the estimator of Lawson and Ponkaew [6] is shown as follows:

$$V_1(\hat{Y}_{GREG.LP})\approx\frac{1}{p}\sum_{i\in U}D_ie_i^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}D_{ij}e_ie_j+\frac{(1-p)}{p}\sum_{i\in U}(e_i-\overline{e})^2\,,$$

$$V_2(\hat{Y}_{GREG.LP})\approx\frac{1}{p}\sum_{i\in U}\frac{(1-\pi_i)}{\pi_i}(e_i-\overline{e})^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}D_{ij}(e_i-\overline{e})(e_j-\overline{e})+\frac{(1-p)}{p}\sum_{i\in U}(e_i-\overline{e})^2\,,$$

where $D_i=(1-\pi_i)\pi_1^{-1}$, $D_{ij}=(\pi_i\pi_j-\pi_{ij})(\pi_i\pi_j)^{-1}$, $e_i=(y_i-\boldsymbol{x}_i'\boldsymbol{\beta})$ and $\overline{e}=\frac{1}{N}\sum_{i\in U}e_i$

The estimators of $V_1(\hat{Y}_{GREG.LP})$ and $V_2(\hat{Y}_{GREG.LP})$, respectively, are obtained, thus

$$\hat{V}_1(\hat{Y}_{GREG.LP})\approx\left(\frac{N}{\sum_{i\in s}\frac{r_i}{\pi_i}}\right)^2\left[\sum_{i\in s}\hat{D}_ir_i\hat{e}_i^2+\sum_{i\in s}\sum_{j/\{i\}\in s}\hat{D}_{ij}r_i\hat{e}_ir_i\hat{e}_j\right]+\frac{(1-p^*)}{(p^*)^2}\sum_{i\in s}\frac{r_i}{\pi_i}(\hat{e}_i-\hat{\overline{e}})^2 \tag{15}$$

$$\hat{V}_2(\hat{Y}_{GREG.LP})\approx\left(\frac{N}{\sum_{i\in s}\frac{r_i}{\pi_i}}\right)^2\left[\sum_{i\in s}\hat{D}_ir_i(\hat{e}_i-\hat{\overline{e}}_r)^2+\sum_{i\in s}\sum_{j/\{i\}\in s}\hat{D}_{ij}r_i(\hat{e}_i-\hat{\overline{e}}_r)^2r_j(\hat{e}_j-\hat{\overline{e}}_r)^2\right]$$
$$+\frac{(1-p^*)}{(p^*)^2}\sum_{i\in s}\frac{r_i}{\pi_i}(\hat{e}_i-\hat{\overline{e}})^2\,, \tag{16}$$

where $\hat{D}_i=(1-\pi_i)\pi_1^{-2}$, $\hat{D}_{ij}=(\pi_i\pi_j-\pi_{ij})(\pi_{ij}\pi_i\pi_j)^{-1}$, $\hat{e}_i=(y_i-\boldsymbol{x}_i\hat{\boldsymbol{\beta}}_r)$, $\hat{\overline{e}}_r=\frac{\sum_{i\in s}r_i\hat{e}_i/\pi_i}{\sum_{i\in s}r_i/\pi_i}$ and

$p^*=p$ if p is known, and otherwise $p^*=\hat{p}=\sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$

3. Results and Discussion

3.1 The proposed GREG estimators and associated variance estimators

3.1.1 The proposed GREG estimators

We proposed new ratio GREG estimators for estimating the population mean and population total based on the estimators of Ponkaew and Lawson [12] and Lawson and Ponkaew [6] using a known auxiliary variable to improve the efficiency of the estimators. We considered the proposed estimators under the same conditions as the Ponkaew and Lawson [12] and Lawson and Ponkaew [6] estimators, where the nonresponse mechanism was uniform and the sampling fraction was small. We also extended the new estimators to be able to be used when the sampling fraction was large. First, we made three assumptions as follows.

(A_1) The response mechanism is missing completely at random (MCAR),

(A_2) $\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta} = O_p(n_r^{-\frac{1}{2}})$,

(A_3) $V\left(\sum_{i\in s}\frac{b_i}{\pi_i}\right)\rightarrow 0$ as $n\rightarrow\infty$ where $b_i = w_i$ or r_i

As we mentioned in section 1, the ratio estimator is an efficient estimator that can be used to estimate population total or population mean and the Ponkaew and Lawson [12] estimator, $\hat{\bar{Y}}_R^*$ is in the form of a ratio estimator. We used it to modify the estimator of Lawson and Ponkaew [6] to make it more efficient than the existing ones. We proposed to replace $\hat{\bar{Y}}_r$ in equation (3) with $\hat{\bar{Y}}_R^*$ in equation (2), and we then obtained the proposed GREG estimator for estimating population mean in the presence of nonresponse, which can be shown as follows.

$$\hat{\bar{Y}}_{GREG.R}^* = \hat{\bar{Y}}_R^* + \left(\bar{\boldsymbol{X}} - \hat{\bar{\boldsymbol{X}}}_r\right)' \hat{\boldsymbol{\beta}}_r \tag{17}$$

Furthermore, the proposed GREG estimator for estimating population total can be shown as

$$\hat{Y}_{GREG.R}^* = N\hat{\bar{Y}}_{GREG.R}^* = N\left[\hat{\bar{Y}}_R^* + \left(\bar{\boldsymbol{X}} - \hat{\bar{\boldsymbol{X}}}_r\right)' \hat{\boldsymbol{\beta}}_r\right], \tag{18}$$

where $\hat{\bar{Y}}_R^* = \dfrac{\hat{\bar{Y}}_r^{(1)}}{\hat{\bar{w}}_{HT}}\bar{W}$, $\hat{\bar{\boldsymbol{X}}}_r = \sum_{i\in s}\dfrac{r_i\boldsymbol{x}_i}{\pi_i}\bigg/\sum_{i\in s}\dfrac{r_i}{\pi_i}$, $\hat{\boldsymbol{\beta}}_r = \left(\sum_{i\in s}\dfrac{r_i q_i \boldsymbol{x}_i\boldsymbol{x}_i'}{\pi_i}\right)^{-1}\left(\sum_{i\in s}\dfrac{r_i q_i \boldsymbol{x}_i y_i}{\pi_i}\right)$,

$\bar{\boldsymbol{X}} = \dfrac{1}{N}\sum_{i\in U}\boldsymbol{x}_i$.

Theorem 1. Assume that assumptions (A_1) to (A_3) hold under the reverse framework, and with unequal probability sampling without replacement, the proposed GREG estimators

$\hat{Y}^*_{GREG.R}$ and $\hat{\bar{Y}}^*_{GREG.R}$ are almost unbiased estimators of Y and $\bar{Y}$, respectively.

Proof.
Recall from (17) and (18) we have

$$\hat{\bar{Y}}^*_{GREG.R} = \hat{\bar{Y}}^*_R + \left(\bar{\boldsymbol{X}} - \hat{\bar{\boldsymbol{X}}}_r\right)' \hat{\boldsymbol{\beta}}_r ,$$

$$\hat{Y}^*_{GREG.R} = N\hat{\bar{Y}}^*_{GREG.R} = N\left[\hat{\bar{Y}}^*_R + \left(\bar{\boldsymbol{X}} - \hat{\bar{\boldsymbol{X}}}_r\right)' \hat{\boldsymbol{\beta}}_r\right],$$

Then, the overall expectation of $\hat{Y}^*_{GREG.R}$ is given by

$$E\left(\hat{Y}^*_{GREG.R}\right) = E_R E_S\left(N\hat{\bar{Y}}^*_{GREG.R}\middle|\boldsymbol{R}\right) = NE_R E_S\left(\hat{\bar{Y}}^*_{GREG.R}\middle|\boldsymbol{R}\right). \tag{19}$$

However, $\hat{\bar{Y}}^*_{GREG.R}$ in (19) is in the form of nonlinear estimator, then we use an automated linearization approach to transform the value of $\hat{\bar{Y}}^*_{GREG.R}$, which can be defined as

$$\hat{\bar{Y}}^*_{GREG.R} \approx \hat{\bar{Y}}^*_R + \left(\bar{\boldsymbol{X}} - \hat{\bar{\boldsymbol{X}}}_r\right)' \boldsymbol{\beta} = \bar{\boldsymbol{X}}\boldsymbol{\beta}' + \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\frac{1}{N}\sum_{i\in s}\frac{w_i}{\pi_i}}\frac{1}{N}\sum_{i\in U} w_i - \frac{\sum_{i\in s}\frac{r_i \boldsymbol{x}_i'\boldsymbol{\beta}}{\pi_i}}{\sum_{i\in s}\frac{r_i}{\pi_i}}. \tag{20}$$

Replace (20) into (19), then

$$E\left(\hat{Y}^*_{GREG.R}\right) \approx NE_R E_S\left(\bar{\boldsymbol{X}}\boldsymbol{\beta}' + \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\frac{1}{N}\sum_{i\in s}\frac{w_i}{\pi_i}}\frac{1}{N}\sum_{i\in U} w_i - \frac{\sum_{i\in s}\frac{r_i \boldsymbol{x}_i'\boldsymbol{\beta}}{\pi_i}}{\sum_{i\in s}\frac{r_i}{\pi_i}}\middle|\boldsymbol{R}\right)$$

$$\approx N\left(\bar{\boldsymbol{X}}\boldsymbol{\beta}' + \frac{\frac{1}{N}\sum_{i\in U} y_i}{\frac{1}{N}\sum_{i\in U} w_i}\frac{1}{N}\sum_{i\in U} w_i - \frac{\sum_{i\in U}\boldsymbol{x}_i'\boldsymbol{\beta}}{N}\right)$$

$$= N\left(\bar{\boldsymbol{X}}\boldsymbol{\beta}' + \bar{Y} - \bar{\boldsymbol{X}}\boldsymbol{\beta}'\right) = N\bar{Y} = Y .$$

Then, $E\left(\hat{Y}^*_{GREG.R}\right) \approx Y$, i.e. $\hat{Y}^*_{GREG.R}$, is an almost unbiased estimator of Y. Furthermore, we can conclude that $\hat{\bar{Y}}^*_{GREG.R}$ is also an almost unbiased estimator of $\bar{Y}$.

3.1.2 The variance of the proposed estimators

In this section, the variance of the proposed estimators was studied under a reverse framework, where the nonresponse mechanism was MCAR. Recall from equation (18), the proposed GREG estimator for estimating population total was defined by $\hat{Y}^*_{GREG.R} = N\hat{\bar{Y}}^*_{GREG.R}$. Therefore, the variance of the population total estimator $\hat{Y}^*_{GREG.R}$ is given by

$$V(\hat{Y}^*_{GREG.R}) = V(N\hat{\bar{Y}}^*_{GREG.R}) = N^2 V(\hat{\bar{Y}}^*_{GREG.R}) \tag{21}$$

Under the reverse framework the value of $V(\hat{\bar{Y}}^*_{GREG.R})$ can be obtained by

$$V(\hat{\bar{Y}}^*_{GREG.R}) = E_R V_S(\hat{\bar{Y}}^*_{GREG.R} \mid \boldsymbol{R}) + V_R E_S(\hat{\bar{Y}}^*_{GREG.R} \mid \boldsymbol{R}) = V_1 + V_2, \tag{22}$$

where $V_1 = E_R V_S(\hat{\bar{Y}}^*_{GREG.R} \mid \boldsymbol{R})$, $V_2 = V_R E_S(\hat{\bar{Y}}^*_{GREG.R} \mid \boldsymbol{R})$.

Next, we investigated the value of $V_1 = E_R V_S(\hat{\bar{Y}}^*_{GREG.R} \mid \boldsymbol{R})$. The sampling variance of $\hat{\bar{Y}}^*_{GREG.R}$ assumes that the vector of response probability $\boldsymbol{R}$ can be derived using the modified automated linearization approach, which was proposed by Lawson and Ponkaew [6]. First, an automated linearization approach was used to transform the value of $\hat{\bar{Y}}^*_{GREG.R}$, which is defined by

$$\hat{\bar{Y}}^*_{GREG.R} \approx \hat{\bar{Y}}^*_R + \left(\bar{\boldsymbol{X}} - \hat{\bar{\boldsymbol{X}}}_r\right)' \boldsymbol{\beta}, \tag{23}$$

where $\boldsymbol{\beta} = \left(\sum_{i \in U} q_i \boldsymbol{x}_i \boldsymbol{x}'\right)^{-1} \left(\sum_{i \in s} q_i \boldsymbol{x}_i y_i\right)$.

From equation (23) we can rewrite $\hat{\bar{Y}}^*_{GREG.R}$ as

$$\hat{\bar{Y}}^*_{GREG.R} \approx \bar{\boldsymbol{X}}\boldsymbol{\beta}' + \frac{\frac{1}{N}\sum_{i \in s} \frac{r_i y_i}{\pi_i p}}{\frac{1}{N}\sum_{i \in s} \frac{w_i}{\pi_i}} \frac{1}{N}\sum_{i \in U} w_i - \frac{\sum_{i \in s} \frac{r_i \boldsymbol{x}_i' \boldsymbol{\beta}}{\pi_i}}{\sum_{i \in s} \frac{r_i}{\pi_i}} \tag{24}$$

Next, we use two estimation methods proposed by Lawson and Ponkaew [6] to transform $\hat{\tilde{Y}}^{*}_{GREG.R}$ into a linear estimator as follows:

Method 1: Substituting $\sum_{i\in s}\frac{w_i}{\pi_i}$ by $\sum_{i\in U} w_i$ and $\sum_{i\in s}\frac{r_i}{\pi_i}$ by $\sum_{i\in U} r_i$

Let

$$\hat{\tilde{Y}}^{*}_{GREG.R(1)} \approx \bar{\boldsymbol{X}}\boldsymbol{\beta}' + \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\frac{1}{N}\sum_{i\in U} w_i}\frac{1}{N}\sum_{i\in U} w_i - \frac{\sum_{i\in s}\frac{r_i \boldsymbol{x}_i'}{\pi_i}}{\sum_{i\in U} r_i}\boldsymbol{\beta} = \bar{\boldsymbol{X}}\boldsymbol{\beta}' + \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p} - \frac{\sum_{i\in s}\frac{r_i \boldsymbol{x}_i'}{\pi_i}}{\sum_{i\in U} r_i}\boldsymbol{\beta} \tag{25}$$

In Method 1, we can approximate the variance of $\hat{\tilde{Y}}^{*}_{GREG.R}$ using the continuous mapping theorem to transform $\hat{\tilde{Y}}^{*}_{GREG.R}$ to a linear form. Recall from assumption (A_3) $V\left(\sum_{i\in s}\frac{b_i}{\pi_i}\right)\to 0$ as $n\to\infty$ where $b_i = w_i$ or r_i then $\sum_{i\in s}\frac{b_i}{\pi_i}\xrightarrow{P}\sum_{i\in U} b_i$ because

$$P\left(\left|\sum_{i\in s}\frac{b_i}{\pi_i}-\sum_{i\in U} b_i\right|>\varepsilon\right)\le \varepsilon^{-2}V\left(\sum_{i\in s}\frac{b_i}{\pi_i}\right)=0.$$

By using continuous mapping, we can conclude that $\hat{\tilde{Y}}^{*}_{GREG.R}\xrightarrow{p}\hat{\tilde{Y}}^{*}_{GREG.R(1)}$, and then the term $V_S(\hat{\tilde{Y}}^{*}_{GREG.R}\,|\,\boldsymbol{R})$ can be approximated as

$$\begin{aligned}
V_S(\hat{\tilde{Y}}^{*}_{GREG.R}\,|\,\boldsymbol{R}) &\approx V_S(\hat{\tilde{Y}}^{*}_{GREG.R(1)}\,|\,\boldsymbol{R}) \\
&= V_S\left(\bar{\boldsymbol{X}}\boldsymbol{\beta}' + \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p} - \frac{\sum_{i\in s}\frac{r_i \boldsymbol{x}_i'}{\pi_i}}{\sum_{i\in U} r_i}\boldsymbol{\beta}\,\middle|\,\boldsymbol{R}\right) = V_S\left(\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p} - \frac{\sum_{i\in s}\frac{r_i \boldsymbol{x}_i'}{\pi_i}}{\sum_{i\in U} r_i}\boldsymbol{\beta}\,\middle|\,\boldsymbol{R}\right) \\
&= V_S\left(\frac{1}{\sum_{i\in U} r_i}\sum_{i\in s}\frac{r_i}{\pi_i}\left(\frac{\sum_{i\in U} r_i y_i}{Np} - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\middle|\,\boldsymbol{R}\right) = V_S\left(\frac{1}{N_r}\sum_{i\in s}\frac{r_i}{\pi_i}\left(\frac{N_r y_i}{Np} - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\middle|\,\boldsymbol{R}\right) \\
&= V_S\left(\sum_{i\in s}\frac{Z_{1i}}{\pi_i}\,\middle|\,\boldsymbol{R}\right),
\end{aligned} \tag{26}$$

where $Z_{1i} = r_i\left(\frac{N_r y_i}{Np} - \boldsymbol{x}_i'\boldsymbol{\beta}\right)$ and $N_r = \sum_{i\in U} r_i$.

Then,

$$V_S(\hat{\hat{Y}}^*_{GREG.R} \mid \boldsymbol{R}) \approx V_S\left(\sum_{i\in s}\frac{Z_{1i}}{\pi_i}\middle|\boldsymbol{R}\right) \text{and } E_R V_S(\hat{\hat{Y}}^*_{GREG.R} \mid \boldsymbol{R}) \approx E_R V_S\left(\sum_{i\in s}\frac{Z_{1i}}{\pi_i}\middle|\boldsymbol{R}\right) \tag{27}$$

Finally, in Method 1 the term V_1 in equation (22) can be approximated by

$$V_{11} = E_R V_S(\hat{\hat{Y}}^*_{GREG.R} \mid \boldsymbol{R})$$

$$\approx E_R V_S\left(\sum_{i\in s}\frac{Z_{1i}}{\pi_i}\middle|\boldsymbol{R}\right) = E_R\left(\sum_{i\in U} D_i Z_{1i}^2 + \sum_{i\in U}\sum_{j\setminus\{i\}\in U} D_{ij} Z_{1i} Z_{1j}\right)$$

$$\approx \sum_{i\in U} D_i E_R\left(Z_{1i}^2\right) + \sum_{i\in U}\sum_{j\setminus\{i\}\in U} D_{ij} E_R\left(Z_{1i}\right) E_R\left(Z_{1j}\right),$$

Then,

$$\approx \sum_{i\in U} D_i E_R\left(Z_{1i}^2\right) + \sum_{i\in U}\sum_{j\setminus\{i\}\in U} D_{ij} E_R\left(Z_{1i}\right) E_R\left(Z_{1j}\right) V_{11} \tag{28}$$

where $D_i = (1-\pi_i)\pi_i^{-1}$ and $D_{ij} = (\pi_{ij} - \pi_i\pi_j)(\pi_i\pi_j)^{-1}$.

Method 2: Using the Taylor linearization approach

We apply the Taylor linearization approach to transform $\hat{\hat{Y}}^*_{GREG.R}$ in equation (24) into a linear estimator which is defined by Method 2, and it is equal to

$$\hat{\hat{Y}}^*_{GREG.R(2)} \approx \text{Constant} + \sum_{i\in s}\frac{1}{\pi_i}\left(\frac{1}{N}\left(\frac{r_i y_i}{p} - \frac{\frac{1}{N}\sum_{i\in U} r_i y_i}{\bar{\bar{W}}} w_i\right) - \frac{r_i}{N_r}\left(\boldsymbol{x}_i'\boldsymbol{\beta} - \frac{1}{N_r}\sum_{i\in U} r_i \boldsymbol{x}_i'\boldsymbol{\beta}\right)\right). \tag{29}$$

We can rewrite the function of $\hat{\hat{Y}}^*_{GREG.R(2)}$ as follows:

$$\hat{\hat{Y}}^*_{GREG.R(2)} \approx \text{Constant} + \sum_{i\in s}\frac{Z_{2i}}{\pi_i}, \tag{30}$$

where $Z_{2i} = \frac{1}{N}\left(\frac{r_i y_i}{p} - \frac{\frac{1}{N}\sum_{i\in U} r_i y_i}{\bar{\bar{W}}} w_i \right) - \frac{r_i}{N_r}\left(\boldsymbol{x}_i'\boldsymbol{\beta} - \frac{1}{N_r}\sum_{i\in U} r_i \boldsymbol{x}_i'\boldsymbol{\beta} \right)$.

Under Method 2, the term $V_S(\hat{\bar{Y}}^*_{GREG.R} \mid \boldsymbol{R})$ can be estimated by

$$V_S(\hat{\bar{Y}}^*_{GREG.R} \mid \boldsymbol{R}) \approx V_S(\hat{\bar{Y}}^*_{GREG.R(2)} \mid \boldsymbol{R})$$
$$= V_S\left(\text{Constant} + \sum_{i\in s} \frac{Z_{2i}}{\pi_i} \middle| \boldsymbol{R} \right) = V_S\left(\sum_{i\in s} \frac{Z_{2i}}{\pi_i} \middle| \boldsymbol{R} \right)$$

Then,

$$V_S(\hat{\bar{Y}}^*_{GREG.R} \mid \boldsymbol{R}) \approx V_S\left(\sum_{i\in s} \frac{Z_{2i}}{\pi_i} \middle| \boldsymbol{R} \right) \text{ and } E_R V_S(\hat{\bar{Y}}^*_{GREG.R} \mid \boldsymbol{R}) \approx E_R V_S\left(\sum_{i\in s} \frac{Z_{2i}}{\pi_i} \middle| \boldsymbol{R} \right) \tag{31}$$

Finally, in Method 2, the term V_1 in equation (22) can be estimated by

$$V_{12} = E_R V_S(\hat{\bar{Y}}^*_{GREG.R} \mid \boldsymbol{R}) \approx E_R V_S\left(\sum_{i\in s} \frac{Z_{2i}}{\pi_i} \middle| \boldsymbol{R} \right) = E_R\left(\sum_{i\in U} D_i Z_{2i}^2 + \sum_{i\in U} \sum_{j\backslash\{i\}\in U} D_{ij} Z_{2i} Z_{2j} \right)$$
$$\approx \sum_{i\in U} D_i E_R\left(Z_{2i}^2\right) + \sum_{i\in U} \sum_{j\backslash\{i\}\in U} D_{ij} E_R\left(Z_{2i}\right) E_R\left(Z_{2j}\right).$$

Then,

$$\approx \sum_{i\in U} D_i E_R\left(Z_{2i}^2\right) + \sum_{i\in U} \sum_{j\backslash\{i\}\in U} D_{ij} E_R\left(Z_{2i}\right) E_R\left(Z_{2j}\right)\ V_{12} \tag{32}$$

From equation (27) and equation (32), we may write

$$m = 1,2\,,\ \approx \sum_{i\in U} D_i E_R\left(Z_{2i}^2\right) + \sum_{i\in U} \sum_{j\backslash\{i\}\in U} D_{ij} E_R\left(Z_{2i}\right) E_R\left(Z_{2j}\right)\ V_{1m} \tag{33}$$

Next, we investigate the value of $V_2 = V_R E_S(\hat{\bar{Y}}^*_{GREG.R} \mid \boldsymbol{R})$ in equation (22). Recall from equation (24) that the value of $\hat{\bar{Y}}^*_{GREG.R}$ is equal to

$$\hat{\bar{Y}}^{*}_{GREG.R} \approx \bar{\boldsymbol{X}}\boldsymbol{\beta}' + \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\frac{1}{N}\sum_{i\in s}\frac{w_i}{\pi_i}}\frac{1}{N}\sum_{i\in U} w_i - \frac{\sum_{i\in s}\frac{r_i \boldsymbol{x}'_i}{\pi_i}}{\sum_{i\in s}\frac{r_i}{\pi_i}}\boldsymbol{\beta}. \tag{34}$$

Then,

$$V_2 = V_R E_S(\hat{\bar{Y}}^{*}_{GREG.R} \,|\, \boldsymbol{R}) \approx V_R E_S\left(\bar{\boldsymbol{X}}\boldsymbol{\beta}' + \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\frac{1}{N}\sum_{i\in s}\frac{w_i}{\pi_i}}\frac{1}{N}\sum_{i\in U} w_i - \frac{\sum_{i\in s}\frac{r_i \boldsymbol{x}'_i\boldsymbol{\beta}}{\pi_i}}{\sum_{i\in s}\frac{r_i}{\pi_i}}\middle|\, \boldsymbol{R}\right)$$

$$\approx V_R\left(\bar{\boldsymbol{X}}\boldsymbol{\beta}' + \frac{\frac{1}{N}\sum_{i\in U}\frac{r_i y_i}{p}}{\frac{1}{N}\sum_{i\in U} w_i}\frac{1}{N}\sum_{i\in U} w_i - \frac{\sum_{i\in U} r_i \boldsymbol{x}'_i\boldsymbol{\beta}}{\sum_{i\in U} r_i}\right) = V_R\left(\frac{1}{N}\sum_{i\in U}\frac{r_i y_i}{p} - \frac{\sum_{i\in U} r_i \boldsymbol{x}'_i\boldsymbol{\beta}}{\sum_{i\in U} r_i}\right).$$

Therefore,

$$V_2 \approx V_R\left(\frac{1}{N}\sum_{i\in U}\frac{r_i y_i}{p} - \frac{\sum_{i\in U} r_i \boldsymbol{x}'_i\boldsymbol{\beta}}{\sum_{i\in U} r_i}\right). \tag{35}$$

From equation (35), we see that the function of parameter in $V_R(.)$ is nonlinear, so we use Taylor linearization to transform this value to a linear function that is defined by

$$\frac{1}{N}\sum_{i\in U}\frac{r_i y_i}{p} - \frac{\sum_{i\in U} r_i \boldsymbol{x}'_i\boldsymbol{\beta}}{\sum_{i\in U} r_i} \approx \text{Constant} + \frac{1}{Np}\sum_{i\in U} r_i(e_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta}), \tag{36}$$

where $e_i = y_i - \boldsymbol{x}'_i\boldsymbol{\beta}$ and $\bar{\boldsymbol{X}}'\boldsymbol{\beta} = \frac{1}{N}\sum_{i\in U}\boldsymbol{x}'_i\boldsymbol{\beta}$.

Substituting equation (36) into equation (35),

$$V_2 \approx V_R\left(\text{Constant} + \frac{1}{Np}\sum_{i\in U} r_i(e_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta})\right) = \frac{1}{N^2}\frac{(1-p)}{p}\sum_{i\in U}\left(e_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2. \tag{37}$$

Therefore,

$$V_2 \approx \frac{1}{N^2}\frac{(1-p)}{p}\sum_{i\in U}\left(e_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2. \tag{38}$$

From equation (33) and equation (38), we can conclude that the variance of the proposed GREG estimator for estimating population mean is given by

$$V_1(\hat{\bar{Y}}^*_{GREG.R}) \approx V_{1m} + V_2$$

$$= \sum_{i\in U} D_i E_R\left(Z^2_{mi}\right) + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} D_{ij} E_R\left(Z_{mi}\right) E_R\left(Z_{mj}\right) + \frac{1}{N^2}\frac{(1-p)}{p}\sum_{i\in U}\left(e_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2, \quad (39)$$

where $m=1,2$ $Z_{1i} = r_i\left(\frac{N_r y_i}{Np} - \boldsymbol{x}'_i\boldsymbol{\beta}\right)$,

$$Z_{2i} = \frac{1}{N}\left(\frac{r_i y_i}{p} - \frac{\frac{1}{N}\sum_{i\in U} r_i y_i}{\bar{W}} w_i\right) - \frac{r_i}{N_r}\left(\boldsymbol{x}'_i\boldsymbol{\beta} - \frac{1}{N_r}\sum_{i\in U} r_i \boldsymbol{x}'_i\boldsymbol{\beta}\right).$$

The variance of the proposed GREG estimator for estimating population mean can be obtained by replacing $V(\hat{\bar{Y}}^*_{GREG.R})$ in equation (21) with $V_1(\hat{\bar{Y}}^*_{GREG.R})$ in equation (39). The variance of $\hat{Y}^*_{GREG.R}$ is given by

$$V_m(\hat{Y}^*_{GREG.R}) \approx N^2 V_m(\hat{\bar{Y}}^*_{GREG.R})$$

$$= N^2\left(\sum_{i\in U} D_i E_R\left(Z^2_{mi}\right) + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} D_{ij} E_R\left(Z_{mi}\right) E_R\left(Z_{mj}\right) + \frac{1}{N^2}\frac{(1-p)}{p}\sum_{i\in U}\left(e_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2\right), \quad (40)$$

where $m=1,2$, $D_i = (1-\pi_i)\pi_i^{-1}$, $D_{ij} = (\pi_{ij} - \pi_i\pi_j)(\pi_i\pi_j)^{-1}$, $e_i = y_i - \boldsymbol{x}'_i\boldsymbol{\beta}$, $\bar{\boldsymbol{X}}'\boldsymbol{\beta} = \frac{1}{N}\sum_{i\in U}\boldsymbol{x}'_i\boldsymbol{\beta}$

Finally, in equation (41) we can show that the estimated value of the variance estimator $V_m(\hat{Y}^*_{GREG.R})$ can be obtained by

$$\hat{V}_m(\hat{Y}^*_{GREG.R}) \approx N^2\left(\sum_{i\in s}\hat{D}_i\hat{Z}^2_{mi} + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{mi}\hat{Z}_{mj} + \frac{1}{N^2}\frac{(1-p^*)}{(p^*)^2}\sum_{i\in s}\frac{r_i}{\pi_i}\left(\hat{e}_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2\right), \quad (41)$$

where $m=1,2$, $\hat{Z}_{mi}$ is the estimator of Z_{mi} for all $i\in s$, $\hat{D}_i = (1-\pi_i)\pi_i^{-2}$, $\hat{D}_{ij} = (\pi_i\pi_j - \pi_{ij})\left(\pi_i\pi_j\pi_{ij}\right)^{-1}$, $p^* = p$ if p is known, otherwise $p^* = \hat{p} = \sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$,

$\hat{e}_i = (y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r)$, $\hat{\boldsymbol{\beta}}_r = \left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i \boldsymbol{x}_i'}{\pi_i p^*}\right)^{-1}\left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i y_i}{\pi_i p^*}\right) = \left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i \boldsymbol{x}_i'}{\pi_i}\right)^{-1}\left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i y_i}{\pi_i}\right)$ and $\bar{\boldsymbol{X}}'\boldsymbol{\beta} = \frac{1}{N}\sum_{i\in s}\frac{r_i \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r}{p^*\pi_i}$

The variance and associated variance estimators of the proposed GREG estimator for estimating population total are shown in Theorem 2 and Theorem 3.

Theorem 2. Assuming that (A_1) to (A_3) are satisfied under the reverse framework with unequal probability sampling without replacement. Let $o_i = w_i \bar{Y}\bar{W}^{-1} - \bar{\boldsymbol{X}}'\boldsymbol{\beta}$, then the variance of the proposed GREG estimators for estimating population total is defined as follows.

(1) $V_1(\hat{Y}^*_{GREG.R})$ is given by

$$V_1(\hat{Y}^*_{GREG.R}) \approx \sum_{i\in U}\left(D_i (Npe_i)^2 + (1-p)p^{-1}(e_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta})^2\right) + (Np)^2 \sum_{i\in U}\sum_{j\backslash\{i\}\in U} D_{ij} e_i e_j \,. \quad (42)$$

(2) $V_2(\hat{Y}^*_{GREG.R})$ is obtained by

$$V_2(\hat{Y}^*_{GREG.R}) \approx \sum_{i\in U}\left(D_i (e_i - o_i)^2 + (1-p)p^{-1}(e_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta})^2\right) + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} D_{ij}(e_i - o_i)(e_j - o_i)\,. \quad (43)$$

Proof.
From equation (40), the variance of the proposed estimators is equal to

$$V_m(\hat{Y}^*_{GREG.R}) \approx N^2 V_m(\hat{\bar{Y}}^*_{GREG.R})$$

$$= N^2\left(\sum_{i\in U} D_i E_R\left(Z_{mi}^2\right) + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} D_{ij} E_R\left(Z_{mi}\right) E_R\left(Z_{mj}\right) + \frac{1}{N^2}\frac{(1-p)}{p}\sum_{i\in U}\left(e_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2\right) \quad (44)$$

where $Z_{1i} = r_i\left(\frac{y_i \sum_{i\in U} r_i}{Np} - \boldsymbol{x}_i'\boldsymbol{\beta}\right)$

and $Z_{2i} = \frac{1}{N}\left(\frac{r_i y_i}{p} - \frac{\frac{1}{N}\sum_{i\in U} r_i y_i}{\bar{\bar{W}}} w_i\right) - \frac{r_i}{N_r}\left(\boldsymbol{x}_i'\boldsymbol{\beta} - \frac{1}{N_r}\sum_{i\in U} r_i \boldsymbol{x}_i'\boldsymbol{\beta}\right)$

(1) From equation (44) if $m=1$, then

$$V_1(\hat{Y}^*_{GREG.R}) \approx N^2\left(\sum_{i\in U} D_i E_R\left(Z_{1i}^2\right)+\sum_{i\in U}\sum_{j\backslash\{i\}\in U} D_{ij}E_R\left(Z_{1i}\right)E_R\left(Z_{1j}\right)+\frac{1}{N^2}\frac{(1-p)}{p}\sum_{i\in U}\left(e_i+\bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2\right). \quad (45)$$

Recall from equation (44), $Z_{1i}=r_i\left(\frac{y_i\sum_{i\in U} r_i}{Np}-\boldsymbol{x}_i'\boldsymbol{\beta}\right)$, then

$$E_R(Z_{1i})=E_R\left(r_i\left(\frac{y_i\sum_{i\in U} r_i}{Np}-\boldsymbol{x}_i'\boldsymbol{\beta}\right)\right)\approx p\left(\frac{y_i Np}{Np}-\boldsymbol{x}_i'\boldsymbol{\beta}\right)=p(y_i-\boldsymbol{x}_i'\boldsymbol{\beta})=pe_i\,.$$

Therefore,

$$E_R(Z_{1i})\approx pe_i\,, \quad (46)$$

where $e_i=y_i-\boldsymbol{x}_i'\boldsymbol{\beta}$.
Substituting equation (46) in equation (45), we have

$$\begin{aligned}V_1(\hat{Y}^*_{GREG.R}) &\approx N^2\left(\sum_{i\in U} D_i(pe_i)^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U} D_{ij}pe_ipe_j+\frac{1}{N^2}(1-p)p^{-1}\sum_{i\in U}\left(e_i+\bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2\right)\\ &=\sum_{i\in U} D_i(Npe_i)^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}(Np)^2D_{ij}e_ie_j+(1-p)p^{-1}\sum_{i\in U}\left(e_i+\bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2\\ &=\sum_{i\in U}\left(D_i(Npe_i)^2+(1-p)p^{-1}\left(e_i+\bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2\right)+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}(Np)^2D_{ij}e_ie_j\end{aligned}$$

Then,

$$V_1(\hat{Y}^*_{GREG.R})\approx\sum_{i\in U}\left(D_i(Npe_i)^2+(1-p)p^{-1}\left(e_i+\bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2\right)+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}(Np)^2D_{ij}e_ie_j \quad (47)$$

(2) The proof of (2) is similar to (1).

Theorem 3. Assuming that (A_1) to (A_3) are satisfied under the reverse framework with unequal probability sampling without replacement, the estimators of variance of the proposed GREG estimators for estimating population total derived as follows.

(1) The estimators of $V_1(\hat{Y}^*_{GREG.R})$ are given by

$$\hat{V}_1(\hat{Y}^*_{GREG.R})=\begin{cases}\hat{E}_{1p}+\dfrac{(1-p)}{p^2}\sum_{i\in s}\dfrac{r_i}{\pi_i}\left(\hat{e}_i+\dfrac{1}{Np}\sum_{i\in s}\dfrac{r_i\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)^2, & when\ p\ is\ known\\ \hat{E}_{1\hat{p}}+\dfrac{(1-\hat{p})}{\hat{p}^2}\sum_{i\in s}\dfrac{r_i}{\pi_i}\left(\hat{e}_i+\dfrac{1}{N\hat{p}}\sum_{i\in s}\dfrac{r_i\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)^2, & when\ p\ is\ unknown\end{cases}, \quad (48)$$

where $\hat{p}=\sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$, $\hat{Z}_{1ip}=r_i\left(\frac{\hat{N}_r y_i}{Np}-\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}\right)$, $\hat{Z}_{1i\hat{p}}=r_i\left(\frac{\hat{N}_r y_i}{N\hat{p}}-\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}\right)$,

$\hat{e}_i=y_i-\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r$, $\hat{N}_r=\sum_{i\in s}\frac{r_i}{\pi_i}$, $\hat{E}_{1p}=N^2\left[\sum_{i\in s}\hat{D}_i\hat{Z}^2_{1ip}+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{1ip}\hat{Z}_{1jp}\right]$,

$\hat{E}_{1\hat{p}}=N^2\left[\sum_{i\in s}\hat{D}_i\hat{Z}^2_{1i\hat{p}}+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{1i\hat{p}}\hat{Z}_{1j\hat{p}}\right]$.

(2) The estimators of $V_2(\hat{Y}^*_{GREG.R})$ are given by

$$\hat{V}_2(\hat{Y}^*_{GREG.R})=\begin{cases}\hat{E}_{1p}+\dfrac{(1-p)}{p^2}\sum_{i\in s}\dfrac{r_i}{\pi_i}\left(\hat{e}_i+\dfrac{1}{N}\sum_{i\in s}\dfrac{r_i\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)^2, & when\ p\ is\ known\\ \hat{E}_{2\hat{p}}+\dfrac{(1-\hat{p})}{\hat{p}^2}\sum_{i\in s}\dfrac{r_i}{\pi_i}\left(\hat{e}_i+\dfrac{1}{N}\sum_{i\in s}\dfrac{r_i\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)^2, & when\ p\ is\ unknown\end{cases}, \quad (49)$$

where $\hat{p}=\sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$, $\hat{Z}_{2ip}=\frac{1}{N}\left(\frac{r_i y_i}{p}-\frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i}}{\bar{W}}w_i\right)-\frac{r_i}{\hat{N}_r}\left(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}-\frac{1}{\hat{N}_r}\sum_{i\in s}\frac{r_i\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)$,

$\hat{Z}_{2i\hat{p}}=\frac{1}{N}\left(\frac{r_i y_i}{\hat{p}}-\frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i}}{\bar{W}}w_i\right)-\frac{r_i}{\hat{N}_r}\left(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}-\frac{1}{\hat{N}_r}\sum_{i\in s}\frac{r_i\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)$,

$\hat{E}_{2p}=N^2\left(\sum_{i\in s}\hat{D}_i\hat{Z}^2_{2ip}+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{2ip}\hat{Z}_{2jp}\right)$,

$$\hat{E}_{2\hat{p}} = N^2\left(\sum_{i\in s}\hat{D}_i\hat{Z}_{2i\hat{p}}^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{2i\hat{p}}\hat{Z}_{2j\hat{p}}\right)$$

Proof.

Recall from equation (41) that the estimators of $V_m(\hat{Y}^*_{GREG.R})$ are defined by

$$\hat{V}_m(\hat{Y}^*_{GREG.R}) \approx N^2\left(\sum_{i\in s}\hat{D}_i\hat{Z}_{mi}^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{mi}\hat{Z}_{mj} + \frac{1}{N^2}\frac{(1-p^*)}{(p^*)^2}\sum_{i\in s}\frac{r_i}{\pi_i}\left(\hat{e}_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2\right), \quad (50)$$

where $m=1,2$, $\hat{Z}_{mi}$ is the estimator of Z_{mi} for all $i\in s$, $\hat{D}_i = (1-\pi_i)\pi_i^{-2}$,

$\hat{D}_{ij} = (\pi_i\pi_j - \pi_{ij})\left(\pi_i\pi_j\pi_{ij}\right)^{-1}$, $p^* = p$ if p is known otherwise $p^* = \hat{p} = \sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$,

$\hat{e}_i = (y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r)$, $\hat{\boldsymbol{\beta}}_r = \left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i\boldsymbol{x}_i'}{\pi_i p^*}\right)^{-1}\left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i y_i}{\pi_i p^*}\right) = \left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i\boldsymbol{x}_i'}{\pi_i}\right)^{-1}\left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i y_i}{\pi_i}\right)$ and $\bar{\boldsymbol{X}}'\boldsymbol{\beta} = \frac{1}{Np}\sum_{i\in s}\frac{r_i\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}$

(1) If $m=1$ and p is known then $p^* = p$ and the estimator of Z_{1i} is given by

$$\hat{Z}_{1i} = r_i\left(\frac{y_i\hat{N}_r}{Np} - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r\right), \quad (51)$$

$$\hat{\boldsymbol{\beta}}_r = \left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i\boldsymbol{x}_i'}{\pi_i}\right)^{-1}\left(\sum_{i\in s}\frac{r_i q_i \boldsymbol{x}_i y_i}{\pi_i}\right), \hat{N}_r = \sum_{i\in s}\frac{r_i}{\pi_i} \text{ and } \hat{e}_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r \quad (52)$$

Substituting equation (51) and equation (52) into equation (50), then

$$\hat{V}_1(\hat{Y}^*_{GREG.R}) \approx N^2\left(\sum_{i\in s}\hat{D}_i\hat{Z}_{1i}^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{1i}\hat{Z}_{1j} + \frac{1}{N^2}\frac{(1-p)}{p^2}\sum_{i\in s}\frac{r_i}{\pi_i}\left(\hat{e}_i + \bar{\boldsymbol{X}}'\boldsymbol{\beta}\right)^2\right)$$

$$= \hat{E}_{1p} + \frac{(1-p)}{p^2}\sum_{i\in s}\frac{r_i}{\pi_i}\left(\hat{e}_i + \frac{1}{Np}\sum_{i\in s}\frac{r_i\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)^2, \quad (53)$$

where $\hat{E}_{1p} = N^2\left(\sum_{i\in s}\hat{D}_i\hat{Z}_{1i}^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{1i}\hat{Z}_{1j}\right)$

If $m=1$ and p is known, then

$$\hat{V}_1(\hat{Y}^*_{GREG.R}) \approx \hat{E}_{1p} + \frac{(1-p)}{p^2}\sum_{i\in s}\frac{r_i}{\pi_i}\left(\hat{e}_i + \frac{1}{Np}\sum_{i\in s}\frac{r_i\mathbf{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)^2 \tag{54}$$

Next, if $m=1$ and p is unknown, then

$$\hat{V}_1(\hat{Y}^*_{GREG.R}) \approx \hat{E}_{1\hat{p}} + \frac{(1-\hat{p})}{\hat{p}^2}\sum_{i\in s}\frac{r_i}{\pi_i}\left(\hat{e}_i + \frac{1}{N}\sum_{i\in s}\frac{r_i\mathbf{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)^2, \tag{55}$$

where $\hat{p} = \sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$.

From equation (54) and equation (55), we may write

$$\hat{V}_1(\hat{Y}^*_{GREG.R}) = \begin{cases} \hat{E}_{1p} + \frac{(1-p)}{p^2}\sum_{i\in s}\frac{r_i}{\pi_i}\left(\hat{e}_i + \frac{1}{Np}\sum_{i\in s}\frac{r_i\mathbf{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)^2 & \text{when } p \text{ is known} \\ \hat{E}_{1\hat{p}} + \frac{(1-\hat{p})}{\hat{p}^2}\sum_{i\in s}\frac{r_i}{\pi_i}\left(\hat{e}_i + \frac{1}{N\hat{p}}\sum_{i\in s}\frac{r_i\mathbf{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)^2 & \text{when } p \text{ is unknown} \end{cases} \tag{56}$$

(2) The proof of (2) is similar to (1).

If the sampling fraction is small, term V_2 in equation (22) can be omitted. The variance of the proposed GREG estimators $V_1\left(\hat{Y}^*_{GREG.R}\right)$ and $V_2\left(\hat{Y}^*_{GREG.R}\right)$ can be obtained in Corollary 3 and the estimator of the $V_1\left(\hat{Y}^*_{GREG.R}\right)$ and $V_2\left(\hat{Y}^*_{GREG.R}\right)$ can be obtained in Corollary 4.

Corollary 4. Under the reverse framework with unequal probability sampling without replacement. Assuming that (A_1) to (A_3) are satisfied and the sampling fraction is negligible. The variance of the proposed GREG estimators is given by

$$V_1(\hat{Y}^*_{GREG.R}) \approx (Np)^2\left(\sum_{i\in U}D_i(e_i)^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U}D_{ij}e_ie_j\right), \tag{57}$$

$$V_2(\hat{Y}^*_{GREG.R}) \approx (Np)^2\left(\sum_{i\in U} D_i(e_i-\overline{e})^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U} D_{ij}(e_i-\overline{e})(e_j-\overline{e})\right), \tag{58}$$

where $e_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$ and $\overline{e} = \frac{1}{N}\sum_{i\in U} e_i$.

Corollary 5. Under the reverse framework with unequal probability sampling without replacement, assuming that (A_1) to (A_3) are satisfied and the sampling fraction is negligible. The variance estimators of the proposed GREG estimators are given as follows.

(1) The estimators of $V_1(\hat{Y}^*_{GREG.R})$ are given by

$$\hat{V}_1(\hat{Y}^*_{GREG.R}) = \begin{cases} N^2\left[\sum_{i\in s}\hat{D}_i\hat{Z}^2_{1ip}+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{1ip}\hat{Z}_{1jp}\right], & \textit{when } p \textit{ is known} \\ N^2\left[\sum_{i\in s}\hat{D}_i\hat{Z}^2_{1i\hat{p}}+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{1i\hat{p}}\hat{Z}_{1j\hat{p}}\right], & \textit{when } p \textit{ is unknown} \end{cases}, \tag{59}$$

where $\hat{p} = \sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$, $\hat{Z}_{1ip} = r_i\left(\frac{\hat{N}_r y_i}{Np} - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}\right)$, $\hat{Z}_{1i\hat{p}} = r_i\left(\frac{\hat{N}_r y_i}{N\hat{p}} - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}\right)$,

$\hat{e}_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r$ and $\hat{N}_r = \sum_{i\in s}\frac{r_i}{\pi_i}$.

(2) The estimators of $V_2(\hat{Y}^*_{GREG.R})$ are given by

$$\hat{V}_2(\hat{Y}^*_{GREG.R}) = \begin{cases} N^2\left(\sum_{i\in s}\hat{D}_i\hat{Z}^2_{2ip}+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{2ip}\hat{Z}_{2jp}\right), & \textit{when } p \textit{ is known} \\ N^2\left(\sum_{i\in s}\hat{D}_i\hat{Z}^2_{2i\hat{p}}+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{D}_{ij}\hat{Z}_{2i\hat{p}}\hat{Z}_{2j\hat{p}}\right), & \textit{when } p \textit{ is unknown} \end{cases}, \tag{60}$$

where $\hat{p} = \sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$, $\hat{Z}_{2ip} = \frac{1}{N}\left(\frac{r_i y_i}{p} - \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i}}{\overline{W}} w_i\right) - \frac{r_i}{\hat{N}_r}\left(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\boldsymbol{r}} - \frac{1}{\hat{N}_r}\sum_{i\in s}\frac{r_i\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i}\right)$

and

$$\hat{Z}_{2i\hat{p}} = \frac{1}{N}\left(\frac{r_i y_i}{\hat{p}} - \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i}}{\bar{W}} w_i \right) - \frac{r_i}{\hat{N}_r}\left(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{\boldsymbol{r}} - \frac{1}{\hat{N}_r}\sum_{i\in s}\frac{r_i \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_r}{\pi_i} \right).$$

3.2 Simulation studies

To see the performance of the proposed GREG estimators compared to the Lawson and Ponkaew [6] estimators. A linear model $y_i = \beta_0 + \beta_1 x_i + \beta_2 k_i + \beta_3 w_i + \varepsilon_i$ was used to generate the study variable y_i with population size N = 5000, where $x_i \sim N(150,5)$, $k_i \sim N(100,5)$, $w_i \sim N(180,10)$, $\varepsilon_i \sim N(0,1)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_4)' = (250,\ 3.10,\ 1.5, -4.21)'$ and $i = 1,2,\ldots,N$. Samples of sizes $n=100$, $n=150$ and $n=500$ were selected using unequal probability sampling without replacement with Midzuno's [13] scheme. Under Midzuno's [13] scheme, the first and second order of inclusion probabilities are defined by

$$\pi_i = \frac{k_i}{K}\left(\frac{N-n}{N-1}\right) + \frac{n-1}{N-1}, \tag{61}$$

$$\pi_{ij} = \left(\frac{k_i + k_j}{K}\right)\left(\frac{N-n}{N-1}\right)\left(\frac{n-1}{N-2}\right) + \left(\frac{n-1}{N-1}\right)\left(\frac{n-2}{N-2}\right). \tag{62}$$

We considered three levels of response; 50%, 70% and 85% in the simulation study and repeated the study 10,000 times (M=10,000). The relative bias (RB) and the relative root mean square error (RRMSE) were used as criteria to compare the performance of the proposed estimators with the Lawson and Ponkaew [6] estimators. The relative bias and the relative root mean square error of the GREG estimator ($\hat{Y}_{GREG,m}$) and the variance estimator ($V_m(\hat{Y}_{GREG})$) are given as follows.

$$RB(\hat{Y}) = \frac{E(\hat{Y}_{GREG,m}) - Y}{Y},\ m = 1,2,\ldots,M, \tag{63}$$

$$RRMSE(\hat{Y}) = \frac{\sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{Y}_{GREG,m} - Y)^2}}{Y}, \tag{64}$$

$$RB(\hat{V}_m(\hat{Y}_{GREG})) = \frac{E\ (\hat{V}_m(\hat{Y}_{GREG})) - V(\hat{Y}_{GREG})}{V(\hat{Y}_{GREG})},\ m = 1,2,\ldots,M, \tag{65}$$

$$RRMSE(\hat{V}_m\left(\hat{Y}_{GREG}\right)) = \frac{\sqrt{\frac{1}{M-1}\sum_{m=1}^{M}\left(\hat{V}_m\left(\hat{Y}_{GREG}\right) - V(\hat{Y}_{GREG})^2\right)}}{V(\hat{Y})}. \quad (66)$$

The results are shown in Tables 1 to 3. Table 1 shows that the proposed GREG estimators with the benefit of the ratio estimator performed well in terms of both minimum relative bias and relative root mean square error for all situations. The proposed GREG estimator, $\hat{Y}^*_{GREG.R}$ gave a smaller relative bias and relative root mean square error for both when p is either known or unknown. Tables 2 and 3 showed similar results to Table 1 where the proposed variance estimators both $\hat{V}_1(\hat{Y}^*_{GREG.R})$ and $\hat{V}_2(\hat{Y}^*_{GREG.R})$ performed well. They gave a smaller relative bias and a smaller relative root mean square error when compared to the existing Lawson and Ponkaew [6] estimator in all situations. We can see that using an auxiliary variable that was related to the study variable increased the efficiency of the estimator by using the ratio estimator for estimating population total.

Table 1. The relative bias and relative root mean square error of the GREG estimators

Response rate (%)	Sample size	The response probability p	Relative bias		Relative root mean square error	
			Lawson and Ponkaew	Proposed	Lawson and Ponkaew	Proposed
0.5	100	p is known	0.0283	0.0274	0.0544	0.0432
		p is unknown	1.0208	1.0002	1.0444	1.0051
	150	p is known	0.0211	0.0207	0.0434	0.0428
		p is unknown	1.0137	0.9991	1.0277	1.0041
	500	p is known	0.0200	0.0190	0.0421	0.0410
		p is unknown	1.0128	0.9981	1.0056	1.0038
0.7	100	p is known	0.0201	0.0181	0.0467	0.0421
		p is unknown	.04400	0.4325	0.4716	0.4385
	150	p is known	0.0200	0.0172	.00371	0.0369
		p is unknown	0.4369	0.4270	0.4388	0.4312
	500	p is known	0.0012	0.0010	0.0171	0.0169
		p is unknown	0.4296	0.4265	0.4319	0.4298
0.85	100	p is known	0.0032	0.0028	0.0171	0.0165
		p is unknown	0.4296	0.4285	0.4319	0.4298
	150	p is known	0.0027	0.0021	0.0118	0.0110
		p is unknown	0.1780	0.1767	0.1840	0.1815
	500	p is known	0.0027	0.0021	0.0118	0.0110
		p is unknown	0.1780	0.1767	0.1840	0.1815

Table 2. The relative bias of the variance estimators

Response rate (%)	Sample size	The response probability p	Relative bias			
			Method 1		Method 2	
			Lawson and Ponkaew	Proposed	Lawson and Ponkaew	Proposed
0.5	100	p is known	1.9244	1.1574	1.9161	1.1472
		p is unknown	1.9305	1.1595	1.9338	1.1587
	150	p is known	1.3357	1.1350	1.4447	1.1345
		p is unknown	1.3956	1.1374	1.4624	1.1372
	500	p is known	1.1345	1.0210	1.3387	1.1287
		p is unknown	1.1983	1.0251	1.3950	1.1290
0.7	100	p is known	0.9726	0.4842	1.8690	0.4601
		p is unknown	0.9941	0.4848	1.8780	0.4613
	150	p is known	0.7713	0.4601	1.0420	0.4542
		p is unknown	0.7912	0.4613	1.0457	0.4548
	500	p is known	0.6780	0.4514	1.0396	0.4317
		p is unknown	0.6896	0.4583	1.0420	0.4320
0.85	100	p is known	0.4038	0.1965	1.1195	0.1965
		p is unknown	0.4164	0.1967	1.1207	0.1967
	150	p is known	0.3150	0.1850	1.0411	0.1890
		p is unknown	0.3280	0.1890	1.0418	0.1898
	500	p is known	0.2352	0.1775	1.0318	0.1681
		p is unknown	0.2370	0.1784	1.0325	0.1690

3.3 Application to real data

To see the performance of the proposed estimators, we used the data from the Thai maize agricultural industry in Thailand in 2019 from the Office of the Agricultural Economics. A sample size of 25 provinces was selected according to the unequal probability sampling without replacement method using Midzuno's [14] scheme from a total of 63 provinces with a 70 percent response. Then, $p_i = p = 0.7$ for all $i = 1, 2, \ldots, 63$ and we generated the nonresponse indicator r_i using $r \sim rbern(p)$ in the R program (R Core Team (2021)). We computed the value of the estimators and their variance estimators for $n = 25$ and then the value of $\hat{p} = \left(\sum_{i \in s} r_i \pi_i^{-1}\right)\left(\sum_{i \in s} \pi_i^{-1}\right)^{-1}$ was equal to 0.71. The study variable y was the yield of maize in 2019, the auxiliary variables x and w represented the cultivated area and the harvest area in 2019, respectively, and the size variable k was the cultivated area in 2018. The average cultivated area and harvest area in 2019 were 1,510 and 1,492 acres, respectively. The results are displayed in Table 4.

Table 3. The relative root mean square error of the variance estimators

Response rate (%)	Sample size	The response probability p	Relative root mean square error			
			Method 1		Method 2	
			Lawson and Ponkaew	Proposed	Lawson and Ponkaew	Proposed
0.5	100	p is known	1.8925	1.3449	1.3410	1.3249
		p is unknown	1.9957	1.3534	1.3899	1.3531
	150	p is known	1.4426	1.2399	1.2564	1.2009
		p is unknown	1.4962	1.2397	1.2685	1.2081
	500	p is known	1.3867	1.1228	1.1118	1.1115
		p is unknown	1.3900	1.1320	1.1290	1.1277
0.7	100	p is known	1.0084	0.5239	1.0580	0.5270
		p is unknown	1.0190	0.5278	1.1865	0.5304
	150	p is known	0.8334	0.5200	1.0457	0.6329
		p is unknown	0.8356	0.5218	1.1193	0.5245
	500	p is known	0.7341	0.5122	1.0385	0.5108
		p is unknown	0.7966	0.5191	1.0393	0.5140
0.85	100	p is known	0.3668	0.2577	0.8957	0.2577
		p is unknown	0.3748	0.2587	0.8964	0.2581
	150	p is known	0.3270	0.2389	0.8527	0.2378
		p is unknown	0.3380	0.2404	0.8643	0.2400
	500	p is known	0.3168	0.2175	0.7438	0.2142
		p is unknown	0.3245	0.2287	0.7548	0.2179

Table 4. The total yield of maize estimates for all provinces and variance estimates for the total yield

Estimator	Total yield of maize estimates for all provinces	Variance estimates
1. Lawson and Ponkaew estimator	544,317	$\hat{V}_1(\hat{Y}_{GREG.LP})=584{,}868{,}293$
		$\hat{V}_2(\hat{Y}_{GREG.LP})=523{,}112{,}394$
2. Proposed estimator	525,124	$\hat{V}_1(\hat{Y}^*_{GREG.R})=476{,}754{,}210$
		$\hat{V}_2(\hat{Y}^*_{GREG.R})=455{,}864{,}312$

Table 4 shows the estimated total yield of maize for all provinces in Thailand and the estimated variance for the total yield of maize. We can obviously see that both of the proposed variance estimators, $\hat{V}_1(\hat{Y}^*_{GREG.R})$ and $\hat{V}_2(\hat{Y}^*_{GREG.R})$, gave a lot smaller variance when compared to those of the Lawson and Ponkaew estimators in this yield of maize data set. The proposed estimators performed very well in terms of minimum variance.

4. Conclusions

Using a ratio to estimate population total and population mean increased estimator efficiency. New generalized regression estimators for estimating population total using ratios were proposed under unequal probability sampling without replacement in the presence of nonresponse. The variances of the proposed GREG estimators were also studied under two different methods. We followed the method of the Lawson and Ponkaew [6] estimator whereby the GREG estimator was considered under a uniform response mechanism. The proposed GREG estimators using ratios to estimate population total outperformed the existing Lawson and Ponkaew [6] estimators in both simulation studies and application to real data.

For simplicity, the nonresponse mechanism is uniform although it is quite restricted because it increases the efficiency of population total estimation. Overall, it outperforms existing estimators. The proposed GREG estimators and variance estimation can be useful in many areas of study and can help with forward planning. The estimators can improve estimates and thus decision making and yield, e.g., total yield, total profit, total number of unemployed people, and total number of patients infected by a virus. Effective decision making can improve economic wealth for the whole country. In future studies, we intend to extend the proposed GREG estimators using a ratio method of estimation for estimating population total and population mean for use in more flexible situations, e.g., when the nonresponse mechanism is not uniform, and we also plan to apply them in more complex survey designs in the presence of nonresponse.

5. Acknowledgements

This research was funded by National Science, Research and Innovation Fund (NSRF), and King Mongkut's University of Technology North Bangkok, under Contract no. KMUTNB-FF-65-28.

References

[1] Hansen, M.H. and Hurwitz, W.N., 1946. The problem of nonresponse in sample surveys. *Journal of the American Statistical Association,* 41, 517-529.

[2] Särndal, C.E. and Lundström, S., 2005. *Estimation in Surveys with Nonresponse*. New York: John Wiley.

[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.

[4] Estevao, V.M. and Särndal, C.E., 2006. Survey estimates by calibration on complex auxiliary information. *International Statistical Review*, 72(2), 127-147.

[5] Chauvet, G., 2016. Variance estimation for the 2006 French housing survey. *Mathematical Population Studies*, 23(3), 147-163.

[6] 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.

[7] 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, pp. 50-54.

[8] Fay, R.E., 1991. Design-based perspective on missing data variance. *Proceedings of the 1991 Annual Research Conference: Bureau of the Census*, Virginia, USA, March 17-20, 1991, pp. 17-20.

[9] Lawson, N. and Siripanich, P., 2022. 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.

[10] Cochran, W.G., 1977. *Sampling Techniqu*es. 3rd ed. New York: John Wiley.

[11] Bacanli, S. and Kadilar, C., 2008. Ratio estimators with unequal probability designs. *Pakistan Journal of Statistics*, 24(3), 167-172.

[12] 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): Asian Conference on Fixed Point Theory and Optimization 2018*, 417-429.

[13] Midzuno, H., 1952. On the sampling system with probability proportional to sum of sizes. *Annals of the Institute of Statistical Mathematics*, 4(3), 99-107.

[14] R Core Team, 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.