บทความวิจัย

SMT KPRU Journal of KPRU Science Mathematics and Technology

ปีที่ 3 ฉบับที่ 2 กรกฎาคม – ธันวาคม 2567

วารสาร มรภ.กพ. วิทยาศาสตร์ คณิตศาสตร์ และเทคโนโลยี

Journal of KPRU Science Mathematics and Technology

The Resampling Method for Estimating Variance of the Generalized Regression Estimator in the Presence of Nonresponse

Chugiat Ponkaew*

Department Applied Mathematics and Statistics, Faculty of Science and Technology

Phetchabun Rajabhat University, Phetchabun, Thailand

Abstract

This paper aims to propose new variance estimators for the generalized regression estimator to estimate the population mean under a reverse framework, where nonresponse occurs in the study variable, as proposed by Ponkaew [1] in 2018. The proposed variance estimators are investigated using two resampling techniques, namely the Jackknife and the Rao-Wu bootstrap. The new variance estimator does not require joint inclusion probabilities, which differs from the variance estimator proposed by Ponkaew [1]. The effectiveness of the suggested estimators is examined through simulation experiments and an application to air pollution data from Phetchabun province, Thailand. The findings demonstrate that, in comparison to other estimators, the proposed Rao-Wu bootstrap variance estimator achieves the highest precision, producing the smallest the root mean square error.

Keywords : Jackknife Method, Missing Data, Variance, Estimation

Introduction

In 1952, the estimator of Horvitz and Thompson [2] is the best for estimating the population mean of the study variable in the full response scenario and unequal probability sampling without replacement design (UPWOR). Later, in 1994 Deville and Särndal [3] devised the generalized regression (GREG) estimator to estimate the population mean when auxiliary variable is provided and there

*Corresponding author : chugiat.pon@pcru.ac.th

Received : 17/07/2024 Revised : 9/12/2024 Accepted : 17/12/2024

is a correlation with the study variable. The GREG is equal to the estimator developed by Horvitz and Thompson [2] with an additional adjustment term derived from the auxiliary variable. In practice, when the chosen distance function is the Chi-square distance, the GREG estimator is a particular instance of the calibration estimator was discussed by Deville and Sarndal [4] . In 2005, non-response is present according to Särndal and Lundström [5] suggestion the GREG estimator can estimate the population mean under the assumption that the response probability is known. A two-phase framework was used to explore variance and related estimators. In 2018, Ponkaew [1] suggested a linear GREG estimator for estimating the population mean under the presumption that the response probability is the same for all sample units. Ponkaew [1] also explored variance and its estimator of linear GREG estimator under the reverse framework and negligible sampling fraction. Joint inclusion probabilities are needed in the computation of the value for the variance estimator of Ponkaew [1]. Under the UPWOR sampling design, it can be difficult to calculate the joint inclusion probabilities, though. As a result, it is challenging to establish the value of variance estimator of Ponkaew [1] in actual use.

The GREG estimator takes the form of a nonlinear function, its characteristics, such as expectation or variance, can be discovered by applying the Taylor linearization approach to convert the GREG estimator into a linear function. As a result, this linear function can be used to approximate the expectation or variance of the GREG estimator. Since the GREG estimator function consists of four estimators, the Taylor linearization approach necessitates a distinct derivation for each individual estimator. Additionally, the variance estimator of the GREG estimator under the UPWOR sampling design requires joint inclusion probability, which may be challenging to obtain for complex designs.

In contrast to Taylor linearization procedures, a resampling method is an alternative strategy for investigating the GREG estimator's variance estimator because it does not call for joint inclusion probabilities or separate derivations for each individual estimator. Resampling techniques like Jackknife and Bootstrap are used with data from the economy or education. The use of the Jackknife produce to reduce estimator bias and standard error was initially discussed by Quenouille [6] in 1949. If the estimator is not smooth, the Jackknife could, however, fail disastrously. The Bootstrap approach was first presented by Efron [7] in 1979, and it was further modified by Efron and Tibshirani [8] in 1994. In contrast to the Jackknife method, this one does not demand that the estimator be smooth. In the presence of nonresponse Haziza [9] discussed the Jackknife and Bootstrap procedure to estimated variance of population mean estimator. However, Haziza [9] discussed a method to estimate the variance of the population mean estimator using only simple random sampling. In this paper, the author extends the method of Haziza [9] to estimate the variance of the GREG estimator on the UPWOR sampling design.

Materials and Methods

1. Notation and assumption

Let $U = \{1, 2, ..., N\}$ be a finite population of size N and $y_i; i = 1, 2, ..., N$ be the value of study variable y . The aim of this study is to investigate the estimator of variance of the GREG estimator in the presence of nonresponse and sampling fraction is negligible under a reverse framework. The sample s of size n , selected with probability $p(s)$ according to sampling design s .

Suppose that information on auxiliary and size variable exists, which are related to study variable y as denoted by x and k respectively. It is assumed that the values of x and k are all greater than zero and known for all units in U . Let π_i and π_{ij} be the inclusion probabilities for the ith , and ith and jth unit of U . Let $E_s(\bullet)$ and $V_s(\bullet)$ denote the expectation and variance operators with respect to the sampling design.

Under nonresponse, let $n_r(\subset n)$ denotes the total number of respondents in response sample $s_r(\subset s)$. Let subscript R denote the nonresponse mechanism and r_i denote the response indicator variable of y_i as defined by, $r_i = 1$ if y_i is observed otherwise $r_i = 0$. Let P_i denote the response probability as defined by $p_i = p(r_i = 1)$. The author assumes that $p_i = p$ for all units in U . Let $E_R(\bullet)$ and $V_R(\bullet)$ denote the expectation and variance operators with respect to the nonresponse mechanism and and denote the overall expectation and variance operators. So, $E_R(r_i) = P(r_i = 1) = p$ and $V_R(r_i) = p(1-p)$.

2. Midzuno scheme

In order to select samples of size from a population of size , Midzuno [10] proposed the UPWOR design in 1952 as follows:

(i) Use a probability proportional to size to select one unit.

(ii) Use the SRSWOR design to select units from the remaining units of the population. In this scheme, the is given by,

$$p(s) = \begin{cases} \frac{K}{K}\frac{1}{\binom{N}{n}} & , n(s) = n \\ \text{otherwise} & \end{cases}$$

where $K = \frac{N}{n}\sum_{i \in s} K_f$. In Midzuno [10] scheme the π_i can be obtained by, $\pi_i = \frac{N-n}{N-1}\frac{K_i}{K} + \frac{n-1}{N-1}$.

3. The GREG estimator in the presence of nonresponse

In a full response case, Deville and Särndal [3] proposed their GREG estimator based on sample elements in s . In the presence of nonresponse, Ponkaew [1] proposed GREG estimator based on response set s_r defined by:

$$\widehat{Y}_{GREG}=\sum_{i\in s_r}\frac{d_iy_i}{P_i}+\left[\sum_{i\in U}x_i-\sum_{i\in s_r}\frac{d_ix_i}{P_i}\right]'\left(\sum_{i\in s_r}\frac{d_ic_ix_iy_i}{P_i}\right)\left(\sum_{i\in s_r}\frac{d_ic_ix_ix_i'}{P_i}\right)^{-1}=\widehat{Y}_r+[x-\widehat{x}_r]'\widehat{\beta}_r,$$

where

$$\widehat{Y}_r=\sum_{i\in s_r}\frac{d_iy_i}{P_i},\widehat{x}_r=\sum_{i\in s_r}\frac{d_ix_i}{P_i},\widehat{\beta}_r=\left(\sum_{i\in s_r}\frac{d_ic_ix_iy_i}{P_i}\right)\left(\sum_{i\in s_r}\frac{d_ic_ix_ix_i'}{P_i}\right)^{-1},x=\sum_{i\in U}x_i,d_i=1/\pi_i$$

is design weight, c_i are specified constants and $x_i=(1,x_{i1},...,x_{ij},...,x_{iJ})'$ are column vector of auxiliary variable with $J\geq 1$. Later, Ponkaew [1] proposed the linear GREG estimator to estimate population total under reverse framework and $p_i=p$ for all defined by,

$$\hat{\dot{Y}}_{GREG}=\sum_{i\in s}\frac{r_id_iy_i}{P}+\left[\sum_{i\in U}x_i-\sum_{i\in s}\frac{d_ix_i}{P}\right]'\left(\sum_{i\in s}\frac{r_id_ic_ix_iy_i}{P}\right)\left(\sum_{i\in s}\frac{d_ic_ix_ix_i'}{P}\right)^{-1}$$

$$=\sum_{i\in s}\frac{r_id_iy_i}{P}+\left[\sum_{i\in U}x_i-\sum_{i\in s}\frac{d_ix_i}{P}\right]'\left(\sum_{i\in s}r_id_ic_ix_iy_i\right)\left(\sum_{i\in s}d_ic_ix_ix_ix'\right)^{-1}$$

$$=\hat{\dot{Y}}_r+\left[X-\hat{\dot{X}}_r\right]'\hat{\dot{\beta}}_r,$$

where

$$\hat{\dot{Y}}_r=\sum_{i\in s}\frac{r_id_iy_i}{P},\hat{\dot{X}}_r=\sum_{i\in s}\frac{d_ix_i}{P}$$

and

$$\hat{\dot{\beta}}_r=\left(\sum_{i\in s}r_id_ic_ix_iy_i\right)\left(\sum_{i\in s}d_ic_ix_ix_ix'\right)^{-1}$$

Then, the linear GREG estimator to estimate population mean is equal to

$$\hat{\dot{\overline{Y}}}_{GREG} = \frac{1}{N}\sum_{i\in s}\frac{r_i d_i y_i}{p} + \frac{1}{N}\left[\sum_{i\in U} x_i - \sum_{i\in s}\frac{d_i x_i}{P}\right]' \left(\sum_{i\in s} r_i d_i c_i x_i y_i\right)\left(\sum_{i\in s} d_i c_i x_i x_i x'\right)^{-1}$$

$$= \hat{\dot{\overline{Y}}}_r + \left[\overline{X} - \hat{\overline{X}}_r\right]' \hat{\dot{\beta}}_r, \quad (1)$$

where

$$\hat{\dot{\overline{Y}}}_r = \frac{1}{N}\sum_{i\in s}\frac{r_i d_i y_i}{p}, \hat{\overline{X}}_r = \frac{1}{N}\sum_{i\in s}\frac{d_i x_i}{p}.$$

Under, reverse framework the variance of $\hat{\dot{\overline{Y}}}_{GREG}$ can be obtained by, $V\left(\hat{\dot{\overline{Y}}}_{GREG}\right) = E_R V_S\left(\hat{\dot{\overline{Y}}}_{GREG}|R\right) + V_R E_S\left(\hat{\dot{\overline{Y}}}_{GREG}|R\right)$. Haziza [9] state that, under reverse framework with sampling fraction is negligible the variance of $\hat{\dot{\overline{Y}}}_{GREG}$ can be approximated by $V\left(\hat{\dot{\overline{Y}}}_{GREG}\right) \cong E_R V_S\left(\hat{\dot{\overline{Y}}}_{GREG}|R\right)$. From equation (1) we see that $\hat{\dot{\overline{Y}}}_{GREG}$ has form of nonlinear function because it consisting of four estimators that is $\frac{1}{N}\sum_{i\in s}\frac{r_i d_i y_i}{p}$, $\frac{1}{N}\sum_{i\in s}\frac{d_i x_i}{p}$, $\sum_{i\in s} r_i d_i c_i x_i y_i$ and $\sum_{i\in s} d_i c_i x_i x_i x'$ then the variance of $\hat{\dot{\overline{Y}}}_{GREG}$ can be obtained by using Taylor linearization approach or an automated linearization approach to transform this estimator into a linear function. The linear form of $\hat{\dot{\overline{Y}}}_{GREG}$ by using automated linearization approach is $\hat{\dot{\overline{Y}}}_{GREG,lin} = \hat{\overline{X}}'\beta + \frac{1}{N}\sum_{i\in s}\frac{ei}{\pi_i}$ where $e_i = \frac{r_i}{p}(y_i - x_i{}'\beta)$. Then, the variance of $\hat{\dot{\overline{Y}}}_{GREG}$ can be approximated by,

$$V\left(\hat{\dot{\overline{Y}}}_{GREG}\right) \cong E_R V_S\left(\hat{\dot{\overline{Y}}}_{GREG}|R\right) = \frac{1}{N^2}\left[\sum_{i\in U}\frac{1}{p}D_i e_i^2 + \sum_{i\in U}\sum_{i\backslash\{j\}\in U} D_{ij} e_i e_j\right],$$

where $D_i = (1-\pi_i)\pi_i{}^{-1}, D_{ij} = (\pi_{ij} - \pi_i\pi_j)(\pi_i\pi_j)^{-1}, e_i = (y_i - x_i{}'\beta)$

and $\beta = \left(\sum_{i\in U} c_i x_i x_i{}^{-1}\right)\left(\sum_{i\in U} c_i x_i y_i\right)$. The estimator of $V\left(\hat{\dot{\overline{Y}}}_{GREG}\right)$ is given by,

$$\hat{V}\left(\hat{\dot{\overline{Y}}}_{GREG}\right) = \frac{1}{N^2 P^2}\left[\sum_{i\in s} r_i \hat{D}_i \hat{e}_i^2 + \sum_{i\in U}\sum_{i\backslash j\in U} r_i r_j \hat{D}_{ij}\hat{e}_i\hat{e}_j\right], \quad (2)$$

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

$\hat{\hat{\beta}}_r = \left(\sum_{i\in s} r_i d_i c_i x_i y_i\right)\left(\sum_{i\in s} d_i, c_i, x_i x'\right)^{-1}$ and $\hat{e}_i = (y_i - x_i'\hat{\hat{\beta}}_r)$

Due to the fact that the technique for calculating variance estimator under UPWOR requires joint inclusion probabilities π_{ij} then the variance estimator of $\hat{\dot{\bar{Y}}}_{GREG}$ are difficult to calculate. Since, the joint inclusion probabilities are probability function that two difference units in population include in sample that is $\pi_{ij} = \sum_{s\supset\{i,j\}} P(s)$. In 2003 Berger [11] stated that size of joint inclusion probabilities set is $n(n-1)/2$ where n denotes sample size then the computational value of estimator variance may be inconvenient when n is large. Then, Hájek [12] approximated π_{ij} under rejective sampling defined by, $\hat{\pi}'_{ij} = \pi_i\pi_j\left[1 - \frac{(1-\pi_i)(1-\pi_j)}{c}\right]$, where c = $\sum_{i\in U} c_i$ and $c_i = \pi_i(1-\pi_i)$. Let $\hat{D}'_{ij} = (\hat{\pi}'_{ij} - \pi_i\pi_j)(\pi'_{ij}\pi_i\pi_j)^{-1}$then $\hat{V}\left(\hat{\dot{\bar{Y}}}_{GREG}\right)$ in (2) can be approximated by,

$$\hat{V}^{(1)}\left(\hat{\dot{\bar{Y}}}_{GREG}\right) = \frac{1}{N^2p^2}\left[\sum_{i\in s} r_i\hat{D}_i\hat{e}_i^2 + \sum_{i\in U}\sum_{i\backslash j\in U} r_i r_j \hat{D}'_{ij}\hat{e}_i\hat{e}_j\right]$$

If is unknown $\hat{V}^{(1)}\left(\hat{\dot{\bar{Y}}}_{GREG}\right)$ can be calculated from,

$$\hat{V}^{(1)}\left(\hat{\dot{\bar{Y}}}_{GREG}\right) = \frac{1}{N^2\hat{p}^2}\left[\sum_{i\in s} r_i\hat{D}_i\hat{e}_i^2 + \sum_{i\in U}\sum_{i\backslash j\in U} r_i r_j \hat{D}'_{ij}\hat{e}_i\hat{e}_j\right] \tag{3}$$

where $\hat{p} = \sum_{i\in s} r_i d_i / \sum_{i\in s} d_i$.

4. The resampling method for estimating variance

Since $\hat{\dot{\bar{Y}}}_{GREG}$ has form of nonlinear function then its variance can be approximated by using automated linearization approach or Taylor linearization approach. The second approach requires separate derivation for each particular estimator then it is difficult to determine a linear form of this estimator. Furthermore, the variance estimator of $\hat{\dot{\bar{Y}}}_{GREG}$ by using automated linearization approach and Taylor linearization approach requires π_{ij} in the step to compute the value that may be difficult to obtain for complex designs such as the UPWOR sampling design. Then, Haziza [9] introduces an alternative method namely the resampling method to estimate the variance of the estimator when the estimator has form of a nonlinear function and sampling design is simple random sampling without replacement (SRSWOR). The resampling method discussed by Hazza [9] consists of the jackknife and the Rao-Wu bootstrap method as follows.

4.1 The jackknife method

Under SRSWOR sampling design, let $\bar{Y} = \frac{1}{N}\sum_{i\in U} y_i = \frac{1}{N}\sum_{i=1}{}^{N} y_i, s = \{1,2,...,n\}$ is a sample set size n , $\bar{y} = \frac{1}{N}\sum_{i\in s} d_i y_i$ is the estimator of $\bar{Y}, j = 1, \tilde{s} = \{j\}$ and $d_i = \frac{N}{n}$. The jackknife variance estimator is obtained as follows:

(i) $\acute{s} = s - \tilde{s}$.

(ii) Adjust the design weights d_i to obtain the so-called jackknife weights $d_{i(j)}$ where is given by $d_{i(j)} = \frac{N}{n}\frac{n}{n-1} = \frac{N}{n-1}$ if $i \neq j$ otherwise $d_{i(j)} = 0$.

(iii) Compute $\overline{y}_{(j)} = \frac{i}{N}\sum_{i \in s'} d_{i(j)} y_i$ and let $j = j + 1$.

(iv) Repeat the steps (i)-(iii) if $j \neq n$ otherwise stop process.

The the jackknife variance estimator of $\overline{y}$ is defined by,

$$\widehat{V}_j = \frac{n-1}{n}\sum_{i \in s}\left(\overline{y}_{(j)} - \overline{y}\right)^2. \tag{4}$$

4.2 The Rao-Wu bootstrap method

Rao-Wu bootstrap was proposed by Rao and Wu [13] in 1988. This method may be described as follows. Let $s = \{1, 2, ..., n\}$ is a sample set size n , $\overline{y} = \frac{1}{n}\sum_{i \in s} y_i$ is the estimator of $\overline{Y}$, $\tilde{n}$ is bootstrap sample size, $B = \{1, 2, ..., \tilde{B}\}$ and $b = 1$. We note that $\tilde{n}$ may be different from n .

(i) Compute $z_i = \overline{y} + \sqrt{C}(y_i - \overline{y})$ where $i = 1, 2, ..., n, C = \frac{\tilde{n}(1-f)}{n-1}, f = \frac{n}{N}$ and let $Z = \{z_1, z_2, ..., z_n\}$.

(ii) Draw a simple random sample $\tilde{Z} = \{\tilde{Z}_1, \tilde{Z}_2, ..., \tilde{Z}_{\tilde{n}}\}$ with replacement from $Z = \{z_1, z_2, ..., z_n\}$ and let $\tilde{s} = \{1, 2, ..., \tilde{n}\}$. Compute $\widehat{\overline{Z}}_{(b)} = \frac{1}{\tilde{n}}\sum_{i \in \tilde{s}} \tilde{Z}_i = \frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}} \tilde{Z}_i$ and let $b = b + 1$.

(iii) If $b = \tilde{B}$ stop process otherwise repeat (iii).

The Rao-Wu bootstrap variance estimator is given by,

$$\widehat{V}_{RW} = \frac{1}{\tilde{B}}\sum_{b \in B}\left(\widehat{\overline{Z}}_{(b)} - \widehat{\overline{Z}}_{(\bullet)}\right)^2, \tag{5}$$

where

$$\widehat{\overline{Z}}_{(\bullet)} = \frac{1}{\tilde{B}}\sum_{b \in B}\widehat{\overline{Z}}_{(b)}$$

Results and Discussion

1. The proposed variance estimators in the presence of nonresponse

This section, we aim to propose variance estimators of liner GREG estimator by using resampling method. Recall from (1) the linear GREG estimator is given by

$$\hat{\dot{\bar{Y}}}_{GREG} = \frac{1}{N}\sum_{i\in s}\frac{r_i d_i y_i}{p} + \frac{1}{N}\left[\sum_{i\in U} x_i - \sum_{i\in s}\frac{d_i x_i}{p}\right]'\left(\sum_{i\in s} r_i d_i c_i x_i y_1\right)\left(\sum_{i\in s} d_i c_i x_i x_i'\right)^{-1} \tag{6}$$

Then, we modified the Jackknife variance to the estimated variance of the linear GREG estimator as follows, from sample $s = \{1, 2, ..., n\}$ of size was selected according to the UPWOR sampling design.

(i) Remove the unit $j = 1$ from the sample.

(ii) Adjust the design weights $d_i = 1/\pi_i$ to obtain the so-called jackknife weights $\tilde{d}_{i(j)}$ and defined by, $\tilde{d}_{i(j)} = \frac{n}{n-1}d_i$ if $i \neq j$ otherwise $\tilde{d}_{i(j)} = 0$ and compute $\widehat{Y}_{GREG(j)}$ in equation (7),

$$\hat{\dot{\bar{Y}}}_{GREG(j)} = \frac{1}{N}\sum_{i\in s}\frac{r_i \tilde{d}_{i(j)} y_i}{\widehat{p}_{(j)}} + \frac{1}{N}\left[\sum_{i\in U} x_i - sum_{i\in s}\frac{\tilde{d}_{i(j)} x_i}{\hat{p}_{(j)}}\right]'\left(\sum_{i\in s} r_i \tilde{d}_{i(j)} c_i x_i y_i\right)\left(\sum_{i\in s} \tilde{d}_{i(j)} c_i x_i x_i{}'\right)^{-1}, \tag{7}$$

where $\hat{p}_{(j)} = \sum_{i\in s} r_i \tilde{d}_{i(j)} / \sum_{i\in s} \tilde{d}_{i(j)}$

(iii) Insert back unit $j = 1$ delected in step (i) .

(iv) Repeat the steps (i)-(iii) for units $j = 1, 2, ...n$.

The estimator of variance for the estimator of [5] is obtained as follows:

$$\hat{V}^{(2)}\left(\hat{\dot{\bar{Y}}}_{GREG}\right) = \frac{n-1}{n}\sum_{i\in s}\left(\hat{\dot{\bar{Y}}}_{GREG(j)} - \hat{\dot{\bar{Y}}}_{GREG}\right)^2 \tag{8}$$

where $\hat{\dot{\bar{Y}}}_{GREG(j)}$ defined in equation (7) and $\hat{\dot{\bar{Y}}}_{GREG}$ defined in equation (6).

The variance estimator of the linear GREG estimator based on the Rao-Wu bootstrap.

(i) Compute

$$\hat{\dot{\bar{Y}}}_{GREG} = \frac{1}{N}\sum_{i\in s}\frac{r_i d_i y_i}{p} + \frac{1}{N}\left[\sum_{i\in U} x_i - \sum_{i\in s}\frac{d_i x_i}{p}\right]'\left(\sum_{i\in s} r_i d_i c_i x_i y_1\right)\left(\sum_{i\in s} d_i c_i x_i x_i'\right)^{-1}$$

(ii) $z_i = \hat{\dot{\hat{Y}}}_{GREG} + \sqrt{C}(y_i - \hat{\dot{\hat{Y}}}_{GREG})$ where $i = 1, 2, ..., n, C = \frac{\tilde{n}(1-f)}{n-1}, f = \frac{n}{N}$ and let $Z = \{z_1, z_2, ..., z_n\}$

(ii) Draw a simple random sample $\tilde{Z} = \{\tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_{\tilde{n}}\}$ with replacement from $Z = \{z_1, z_2, ..., z_n\}$ and let $\tilde{s} = \{1, 2, ..., \tilde{n}\}$ .Compute,

$$\hat{\hat{Z}}_{GREG(b)} = \frac{1}{N}\sum_{i\in\tilde{s}}\frac{r_i d_i \tilde{z}_i}{p} + \frac{1}{N}\left[\sum_{i\in U} x_i - \sum_{i\in\tilde{s}}\frac{d_i x_i}{p}\right]'\left(\sum_{i\in\tilde{s}} r_i d_i c_i x_i \tilde{z}_i\right)\left(\sum_{i\in\tilde{s}} d_i c_i x_i x_i'\right)^{-1}$$

and let $b = b + 1$.

(iii) If $b = \tilde{B}$ stop process otherwise repeat (iii).

The Rao-Wu bootstrap variance estimator is given by,

$$\hat{V}^{(3)}\left(\hat{\dot{\hat{Y}}}_{GREG}\right) = \frac{1}{\tilde{B}}\sum_{b\in B}\left(\hat{\hat{Z}}_{GREG(b)} - \hat{\hat{Z}}_{GREG(.)}\right)^2. \tag{9}$$

where $\hat{\hat{Z}}_{GREG(.)} = \frac{1}{\tilde{B}}\sum_{b\in B}\hat{\hat{Z}}_{GREG(b)}$.

2. Simulation studies

In this section, the author have compared the efficiency of the variance estimator of linear GREG estimator with estimated joint inclusion probability $\hat{V}^{(1)}$, Jackknife variance $\hat{V}^{(2)}$ and Rao-Wu Bootstrap $\hat{V}^{(3)}$ through simulation studies. First of all, the study variable of population size 5,000 is obtained by adjusting the model of Cheng, Slud and Hogue [14] as follows:

$$y_i = 5x_i^2 + 2x_i + 20k_i + e_i; i = 1, 2, 3, ..., 5000, \tag{10}$$

where x and k are auxiliary and size variables respectively. We generate x_i from Gamma (9,3/2) k_i from $U(1,5)$ and e_i from $N(0,9)$. Next, the response probability or p will be consequently defined as constant 0.5, 0.7 and 0.9. In each p , $A_i \sim U(0,1)$ is generated and it is compared with the p . A result r_i is 1 if $A_i \le p$; otherwise r_i is 0 . Then, the sample s of size $n =$ 5%, 10%, 20% and 30% are selected according to Midzuno [10] scheme. Finally, the efficiency of the proposed variance estimator with the existing variance estimator are compared by using the relative root mean square error defined in Definition 1.

Definition 1. Let $\hat{\theta}$ be the GREG estimators of a population total, $V(\hat{\theta})$ is the variance of $\hat{\theta}$ and $\hat{V}(\hat{\theta})$ are the estimators of . We define $\hat{V}(\hat{\theta})_{(m)}$ as the variance estimator in iteration m^{th} and M is the

Table 1: The root mean square error of the existing variance estimator and proposed variance estimators

P	n	RRMSE of the variance estimators		
		$\widehat{V}^{(1)}(\widehat{\dot{\bar{Y}}}_{GREG})$	$\widehat{V}^{(2)}(\widehat{\dot{\bar{Y}}}_{GREG})$	$\widehat{V}^{(3)}(\widehat{\dot{\bar{Y}}}_{GREG})$
0.5	5%	0.9838	0.6494	***0.4846***
	10%	0.9815	0.5006	***0.3156***
	20%	0.9801	0.4853	***0.2276***
	30%	0.9799	0.4825	***0.2003***
0.7	5%	0.9703	0.5535	***0.4592***
	10%	0.9693	0.4892	***0.3223***
	20%	0.9578	0.4026	***0.2201***
	30%	0.9476	0.3862	***0.1876***
0.9	5%	0.94702	0.4454	***0.3577***
	10%	0.9447	0.4213	***0.2757***
	20%	0.9434	0.2990	***0.2050***
	30%	0.9327	0.2808	***0.1663***

number of replications. The relative root mean square error of $(PRMSE)$ given by

$$PRMSE(\widehat{V}(\widehat{\theta})) = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(\widehat{V}(\widehat{\theta})_M - V(\widehat{\theta}))^2}$$

Table 1 demonstrated that the proposed Rao-Wu Bootstrap variance estimator performed admirably at all levels of response probabilities and sample sizes because it has the smallest root mean square error compared to the Jackknife variance estimator and the variance estimator with estimated joint inclusion probability.

3.3 An application to real data

We used air pollution data from Thailand's Phetchabun province to apply proposed variance estimators in this section. The data are from the Air Quality and Noise Management Division Bureau during March 2023. The study variable y is PM2.5 (Mg/m^3), the temperature (Celsius) is auxiliary variable x and The size variable k is fine particulate matter 10 micrometers or less in diameter (PM10 : Mg/m3). The estimated value of PM2.5 is 63.34 Mg/m^3. The nonresponse rate for this study is 2%. A sample of size 300 records is chosen from a population of size 741 records using the scheme of [14]. The results are displayed in Table 2.

Table 2: The estimated mean and variance of PM2.5 in Phetchabun province

Estimator	Estimated mean of PM2.5	Variance estimators	Estimated variance of mean of PM2.5
$\hat{\dot{\bar{Y}}}_{GREG}$	61.46	$\widehat{V}^{(1)}(\hat{\dot{\bar{Y}}}_{GREG})$	1.5333
		$\widehat{V}^{(2)}(\hat{\dot{\bar{Y}}}_{GREG})$	1.0494
		$\widehat{V}^{(3)}(\hat{\dot{\bar{Y}}}_{GREG})$	0.7350

The results from the air pollution data in Thailand's Phetchabun province from Table 2 demonstrate that the Rao-Wu bootstrap variance estimator's estimated variance of the mean of PM 2.5 is the smallest value compared to other estimators.

Conclusion

Jackknife and bootstrap methods are resampling techniques used to estimate the variance of nonlinear estimators, such as the GREG estimator. The jackknife method is more suitable for small datasets, while the bootstrap method produces better results for data with skewed distributions. This paper, the jackknife and Rao-Wu bootstrap methods were used to evaluate the variance estimators of the GREG estimator under UPWOR. The calculated joint inclusion probability or real value are not necessary for the proposed variance estimators. The suggested estimator using Rao-Wu bootstrap outperformed the other variance estimators, according to the simulation results and application to the air pollution data. However, the proposed variance estimators are studied under the assumption that the nonresponse mechanism is MCAR, which is generally unrealistic in practice. Therefore, future work should extend these estimators to account for the MAR mechanism.

Reference

[1] Ponkaew, C. 2018. Estimation for population total in the presence of nonresponse. [Doctoral thesis, Graduate college king Mongkut's university of technology north Bangkok].

[2] Horvitz, D.G., and Thompson D.J., A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47(1952), 663-685.

[3] Deville, J.-C., and Särndal, C.-E. ,Variance Estimation for the Regression Imputed Horvitz Thompson Estimator. Journal of Official Statistics, 10(1994), 381-394.

[4] Deville, J.-C. and Sarndal, C.-E. (1992) Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 87, 376. https://doi.org/10.1080/01621459.1992.10475217

[5] Särndal, C.E., and Lundström,S., Estimation in Surveys with Nonresponse. Wiley, New York, 2005

[6] Quenouille, M. H. (1949). Approximate Tests of Correlation in Time Series. Journal of the Royal Statistical Society B 11, 68–84.

[7] Efron, B. 1979. Bootstrap Methods: Another Look at the Jackknife. Annals of Statistics, 7, 1–26.

[8] Efron, B. and Tibshirani, R.J. 1994. An Introduction to the Bootstrap. Chapman and Hall, New York.

[9] Haziza, D. 2010. Resampling methods for variance estimation in the presence of missing survey data. Proceedings of the Annual Conference of the Italian Statistical Society.

[10] Midzuno, H., On sampling system with probability proportional to sum of sizes. Ann. Inst.Statist. Math, 3 (1952), 99-107.

[11] Berger, Y. G. 2003. A simple variance estimator for unequal probability sampling without replacement. Journal of Applied Statistics. Vol.31 : 305-315.

[12] Hajek, J. 1981. Sampling from finite population. New York, Marcel Dekker.

[13] Rao, J. N. K. and Wu, C.F.J. (1988), Resampling inference for complex survey data. Journal of American Statistical Association, 83, 231-241.

[14] Cheng, Y., Slud, E., Hogue, C. , Variance Estimation for Decision-based Estimators with Application to the Annual Survey of Public Employment and, JSM Proceedings,(2010).