Original Research Article

An Improved Product Estimator for Finite Population Mean in the presence of Nonignorable Nonresponse

Abstract

This study developed an improved estimator for finite population mean when nonignorable nonresponse is present. With the aid of auxiliary variable using stratified random sampling, the study proposed an estimator using the direct generalized ratio estimator. A simulation and empirical studies were performed on the derived statistical properties of the proposed estimator using linear and nonlinear populations. The simulation and empirical studies revealed that the proposed estimator ($\hat{R}_{DG,2,i}$) performed better in terms of bias but less efficient in terms of MSE when compared to the estimator by Ashutosh [16]. The studies also revealed that the proposed estimator ($\hat{R}_{DG,2,i}$) was found to be more consistent than that of Ashutosh [16] as the sample size increases across the various subpopulations. Furthermore, the simulation and emprircal studies indicated that the proposed estimator ($\hat{R}_{DG,2,i}$) works better with nonlinear population. Therefore, the estimator by Ashutosh (16) may prove to be useful when the focus is on inference. However, when emphasis shifts to estimation, the proposed estimator ($\hat{R}_{DG,2,i}$) is preferred.

*

1 Introduction

Most often, survey practitioners are challenged with nonignorable nonresponse and the possible bias associated with it. Nonresponse is often recorded in survey sampling as some of the subjects chosen for the sample do not contribute information or data, either by refusal or unable to reply [1]. Nonresponse can either be missing data at random, missing data completely at random or missing data not at random. The latter type of nonresponse is of grave concern to survey statisticians. Because the missing data is connected to the study variable and might increase the bias in the estimation of the finite population parameter [2]. Deleting such missingness will lead to biased estimates and wrong conclusion as well on the population parameters due to the introduction of complexities such as selection bias, undercoverage, reduced sample size, increased variability, imputation challenge, nonresponse bias, resource and cost implication [3]. The use of auxiliary variables can remove the complexities and reduce the bias as a result of nonresponse. But that comes with the challenge of availability and quality of the auxiliary Variables and model assumptions [4, 5]. Diverse statistical techniques are available, primarily relying on either model-based or design-based methods, to address the challenges and complexities presented by nonignorable nonresponse in sample survey. Although, neither of these fundamental methods can simultaneously provide both robustness and efficiency, the model based approach is able to yield a superior compromised results [6]
Ratio, product and linear regression estimators are example of model based approaches available to suitably obtain unbiased or less bias estimators when there is nonresponse [2]. However, coupled with nonlinearity and nonignorable nonresponse, linear regression is unable to adequately capture how the study variable and the auxiliary variable are related [7]. The ratio estimator is beneficial only if the study and auxiliary variables have positive relationship but performs poorly if their association is negative [8]. And Because there maybe non-linear relationship between the study variable and the auxiliary variable, a false positive or negative correlation can occur between them [9]. Joshua and Okon [10] used calibrations approach with subsampling nonresponse and proposed an estimator. But if the relationship between the study and the auxiliary variables is nonlinear, calibration approach alone may not fully enhance the performance of an estimator. The work of Makhdum et al. [11] also proposed a modified regression-cum ratio estimator. Their estimator might not work well where the the relationship between the study and the auxiliary variables is highly nonlinear.
The direct ratio estimator frequently used for estimating population mean, assumes there is a linear relationship between the auxiliary variable and the study variable [2]. And violation of this assumption, coupled with the presence of nonignorable nonresponse, can result in increased bias and inaccurate results [12]. In many industries, complex demands such as privacy concerns and the sensitivity of information often cause the presence of nonignorable nonresponse units within the strata of subpopulations during sample surveys. This situation is compounded when the auxiliary variable and the study variable are non-linearly related. Tikkwal and Ghiya [13] suggested

a generalized direct ratio estimator to get over this restriction of linearity assumption and false direction of correlation. The Tikkwal and Ghiya [13] estimator used a non-linear function of the auxiliary variable in their estimation process, thus allowing for a more flexible relationship between the auxiliary variable and the study variable. This article therefore, seeks to propose an improved estimator for finite population mean where there is high rate of nonignorable nonresponse present.

2 Methodology

2.1 Mean and Variance in the presence of Nonignorable Nonresponse

Nonignorable nonresponse sample mean estimator is defined as the average of a set of sampled data with data missing not at random present [14]

2.1.1 For the study variable

The mean of the response population of the study variable Y is

$$\bar{Y}_1 = \frac{1}{Y_{N_1}} \sum_{i=1}^{Y_{N_1}} Y_{N_{1_i}},$$

and that of nonignorable nonrespondents population of the study variable is

$$\bar{Y}_2 = \frac{1}{Y_{N_2}} \sum_{i=1}^{Y_{N_2}} Y_{N_{2_i}}.$$

Where Y_{N_1} and Y_{N_2} represent response and nonignorable nonresponse population sizes respectively. Therefore, the population mean of the study variable is

$$\overline{Y} = W_1\overline{Y}_1 + W_2\overline{Y}_2 \qquad [15]. \tag{1}$$

where; $W_1 = \frac{Y_{N_1}}{Y_N}$ and $W_2 = \frac{Y_{N_2}}{Y_N}$.
For the sampled population;
mean of the response sample of the study variable is

$$\bar{y}_1 = \frac{1}{y_{n_1}} \sum_{i=1}^{y_{n_1}} y_{y_i}.$$

Hansen and Hurwitz [16] approach of subsampling to solving the problem of nonresponse is considered and presented as follows;
1. Simple random sample without replacement of size n is selected.
2. A subsample of size $r = \frac{y_{n_{2i}}}{k}$, where $k \geq 1$ from nonignorable nonrespnse sampled population

$(y_{n_{2i}})$ units in the initial attempt of the survey.
Because $\overline{y}_{2r}$ is unbiased for $\overline{y}_2$ of the y_{n_2}, mean of the nonignorable nonresponse subsample of the study variable is considered as

$$\overline{y}_{2r} = \frac{1}{y_{n_{2r}}} \sum_{i=1}^{y_{n_{2r}}} y_{n_{2i}}. \tag{2}$$

From equation (1) the sample mean $\overline{y}^*$ of the study variable with nonignorable nonresponse present is written as

$$\overline{y}^* = w_1 \overline{y}_{n_1} + w_2 \overline{y}_{2r} \tag{3}$$

$$\text{with variance} \qquad var(\overline{y}^*) = \frac{(1-f)}{y_n} S_y^2 - \frac{W_2(k-1)}{y_n} S_{y_2}^2 \quad [15]. \tag{4}$$

2.1.2 For the auxiliary variable

The population mean of the auxiliary variable is given by

$$\overline{X} = W_1 \overline{X}_1 + W_2 \overline{X}_2 \quad [15]. \tag{5}$$

Also, from equations (2) and (5) the sample mean $(\overline{x}^*)$ of the auxiliary variable with nonignorable nonresponse present is written as

$$\overline{x}^* = w_1 \overline{x}_{n_1} + w_2 \overline{x}_{2r} \tag{6}$$

$$\text{with variance} \qquad var(\overline{x}^*) = \frac{(1-f)}{x_n} S_x^2 - \frac{W_2(k-1)}{x_n} S_{x_2}^2 \quad [15]. \tag{7}$$

where; $f = \frac{x_n}{X_n}$.

3 The Proposed Estimator

The generalized direct ratio estimator for subpopulation mean with nonignorable nonresponse present is given as

$$R_{DG,i} = \overline{y}_i^* \left(\frac{\overline{x}_i^*}{\overline{X}_{ij}} \right)^{\alpha} \quad [12]. \tag{8}$$

From equation (8) various estimators are obtained as

$$
\begin{aligned}
R_{DG,i} &= \overline{y}_i^* && \text{if} \quad \alpha = 0, \\
R_{DG,i} &= \overline{y}_i^* \frac{\overline{x}_i^*}{\overline{X}_{ij}} && \text{if} \quad \alpha = 1, \\
R_{DG,i} &= \overline{y}_i^* \frac{\overline{X}_{ij}}{\overline{x}_i^*} && \text{if} \quad \alpha = -1, \\
R_{DG,i} &= \overline{y}_i^* \left(\frac{\overline{x}_i^*}{\overline{X}_{ij}}\right)^2 && \text{if} \quad \alpha = 2, \\
R_{DG,i} &= \overline{y}_i^* \left(\frac{\overline{X}_{ij}}{\overline{x}_i^*}\right)^2 && \text{if} \quad \alpha = -2.
\end{aligned} \tag{9}
$$

From equation (9), Ashutosh [16] proposed an estimator by considering $\alpha = -1$. This study considers the case where $\alpha = 2$. Hence, the proposed estimator is

$$R_{DG,i} = \overline{y}_i^* \left(\frac{\overline{x}_i^*}{\overline{X}_{ij}}\right)^2. \tag{10}$$

3.1 Notation

Consider an independent subpopulations from a finite population U_i $(i = 1, 2, 3, \ldots, Q)$ of size N. Each of these subpopulations are stratified into j^{th} strata $(j = 1, 2, 3, \cdots, q)$ of i^{th} subpopulations U_{ij} with size N_{ij}. A random Sample S_{ij} with size n_{ij} through simple random sampling without replacement(SRWR) is drawn from the j^{th} stratum of the i^{th} subpopulation of U_{ij}. Taking (x_i, y_i) to be the auxiliary and the study variables respectively, let the population size (Y_N) of the study variable be subdivided as Y_{N_1} and Y_{N_2} for response and nonresponse population dichotomy.

Where;

Y_{1i} is the response population of the study variable and $Y_{N_{1i}}$ is the response population size of the study variable of the i^{th} subpopulation.

Y_{2i} represents nonignorable nonresponse population of the study variable and $Y_{N_{2i}}$ is the nonignorable nonresponse population size of the study variable of the i^{th} subpopulation.

The auxiliary variable is also presented as: X_{1i} is the response population of the auxiliary variable and $X_{N_{1i}}$ is the response population size of the auxiliary variable of the i^{th} subpopulation.

X_{2i} is the nonignorable nonresponse population of the auxiliary variable and $X_{N_{2i}}$ denotes the response population size of the i^{th} subpopulation for the nonignorable nonresponse.

The subpopulation means of the study and auxiliary variables for the response population are;

$\bar{Y}_{1i}$: i^{th} subpopulation mean of the $Y_{N_{1i}}$ observations.

$\bar{X}_{1i}$: i^{th} subpopulation mean of the $X_{N_{1i}}$ observations.

$\bar{Y}_{1ij}$: mean of the j^{th} stratum of the i^{th} subpopulation of $Y_{N_{1ij}}$ observations.

$\bar{X}_{1ij}$: mean of the j^{th} stratum of the i^{th} subpopulation of $X_{N_{1ij}}$ observations.

$\bar{x}_{1ij}$: sample mean of the j^{th} stratum of the i^{th} subpopulation of $x_{n_{1ij}}$ observations.
$\bar{x}_{1ij}$: sample mean of the j^{th} stratum of i^{th} subpopulation of $x_{n_{1ij}}$ observations.
The subpopulation means of the study and auxiliary variables for the nonresponse population are;
$\bar{Y}_{2i}$: i^{th} subpopulation mean of the $Y_{N_{2i}}$ observations.
$\bar{X}_{2i}$: i^{th} subpopulation mean of the $X_{N_{2i}}$ observations.
$\bar{Y}_{2ij}$:j^{th} stratum mean of the i^{th} subpopulation of $Y_{N_{2ij}}$ observations.
$\bar{X}_{2ij}$: j^{th} stratum mean of the i^{th} subpopulation of $X_{N_{2ij}}$ observations.
$\bar{y}_{2ij}$: sample mean of the j^{th} stratum of the i^{th} subpopulation of $y_{n_{2ij}}$ observations.
$\bar{x}_{2ij}$: sample mean of the j^{th} stratum of the i^{th} subpopulation of $x_{n_{2ij}}$ observations.
A subsample of size $r = \frac{y_{n_{2ij}}}{k}$, where $k \geq 1$ from nonignorable nonrespnse sampled population ($y_{n_{2ij}}$ and $x_{n_{2ij}}$) units in the initial attempt of the survey on the j^{th} stratum of the i^{th} subpopulation.
Thus, $\sum_{j=1}^{J} W_{1ij}\overline{Y}_{1ij} = \overline{Y}_{1i}, \sum_{j=1}^{J} W_{ij}\overline{X}_{ij} = \overline{X}_{1i}, W_{1ij} = \frac{Y_{N_{1ij}}}{Y_{N_{ij}}}$, and
$\sum_{j=1}^{J} W_{2ij}\overline{Y}_{2ij} = \overline{Y}_{2i}, \sum_{j=1}^{J} W_{2ij}\overline{X}_{2ij} = \overline{X}_{2i}, W_{2ij} = \frac{X_{N_{2ij}}}{X_{N_{ij}}}$.

3.2 Finite Population Mean using the Proposed Estimator

The population mean of j^{th} stratum of the study and auxiliary variables of the i^{th} subpopulation are

$$\overline{Y}_i = \sum_{j=1}^{J} W_{1ij}\overline{Y}_{1ij} + \sum_{j=1}^{J} W_{2ij}\overline{Y}_{2ij}. \tag{11}$$

$$\overline{X}_i = \sum_{j=1}^{J} W_{1ij}\overline{X}_{1ij} + \sum_{j=1}^{J} W_{2ij}\overline{X}_{2ij}. \tag{12}$$

Using the subsampling together with equations (11) and (12), the sample mean of the study variable $\overline{y}_i^*$ and auxiliary variables $\overline{x}_i^*$ with nonignorable nonresponse present is

$$\overline{y}_i^* = \sum_{j=1}^{J} W_{1ij}\overline{y}_{1ij} + \sum_{j=1}^{J} W_{2ij}\overline{y}_{n2r_{ij}}. \tag{13}$$

$$\overline{x}_i^* = \sum_{j=1}^{J} W_{1ij}\overline{x}_{1ij} + \sum_{j=1}^{J} W_{2ij}\overline{x}_{n2r_{ij}}. \tag{14}$$

Substituting equations (13) and (14) into equation (10) gives the i^{th} subpopulation mean as

$$\hat{R}_{(DG,2,i)} = \left(\sum_{j=1}^{J} W_{1ij}\overline{y}_{1ij} + \sum_{j=1}^{J} W_{2ij}\overline{y}_{n2r_{ij}}\right)\left(\frac{\sum_{j=1}^{J} W_{1ij}\overline{x}_{1ij} + \sum_{j=1}^{J} W_{2ij}\overline{x}_{n2r_{ij}}}{\overline{X}_{ij}}\right)^2. \tag{15}$$

3.3 Bias and Mean Squared Error of the Proposed Estimator

Using the large sample approximations, the following are defined:

$$
\begin{aligned}
\text{Let} \qquad e_0 &= \frac{\overline{y}_i^* - \overline{Y}_{ji}}{\overline{Y}_{ji}}, \quad e_1 = \frac{\overline{x}_i^* - \overline{X}_{ji}}{\overline{X}_{ji}}, \quad \text{so that} \\
\overline{y}_i^* &= \overline{Y}_{ji}(1+e_0) \quad \text{and} \quad \overline{x}_i^* = \overline{X}_{ji}(1+e_1), \\
E(e_0) &= 0, \quad E(e_1) = 0, \\
E(e_0^2) &= \frac{1}{\overline{Y}_{ij}^2}(\overline{y}_{ij}^* - \overline{Y}_{ij})^2, \\
&= \frac{N_{ij} - n_{ij}}{\overline{Y}_{ij}^2 N_{ij} n_{ij}} S_{ij}^2, \\
E(e_1^2) &= \frac{1}{\overline{X}_{ij}^2}(\overline{x}_{ij}^* - \overline{X}_{ij})^2, \\
&= \frac{N_{ij} - n_{ij}}{\overline{X}_{ij}^2 N_{ij} n_{ij}} S_{ij}^2, \\
E(e_0 e_1) &= \frac{1}{\overline{X}_{ij}\overline{Y}_{ij}}(\overline{x}_{ij}^* - \overline{X}_{ij})(\overline{y}_{ij}^* - \overline{Y}_{ij}), \\
&= \frac{1}{\overline{X}_{ij}\overline{Y}_{ij}}(S_{X_{ij}Y_{ij}}) \quad [17].
\end{aligned} \tag{16}
$$

Therefore, substituting equations (4) and (7) into equation (16) gives

$$
\begin{aligned}
E(e_0^2) &= \frac{N_{ij} - n_{ij}}{N_{ji} n_{ij}} C_{Y_{ij}}^2 + \frac{(k_{ij} - 1)w_{2ij}}{n_{ij}} C_{Y_{2ij}}^2, \\
E(e_1^2) &= \frac{N_{ij} - n_{ij}}{N_{ij} n_{ij}} C_{X_{ij}}^2 + \frac{(k_{ij} - 1)w_{2ij}}{n_{ij}} C_{X_{2ij}}^2, \\
E(e_0 e_1) &= \frac{N_{ij} - n_{ij}}{N_{ij} n_{ij}} C_{XY_{ij}} + \frac{(k_{ij} - 1)w_{2ij}}{n_{ij}} C_{2XY_{ij}}.
\end{aligned} \tag{17}
$$

Where;

$$
\begin{aligned}
S_{Y_{ij}}^2 &= \frac{1}{N_{ij} - 1} \sum_{k=1}^{N_{ij}} (y_{ij_k} - \overline{Y}_{ij})^2, \\
S_{X_{ij}}^2 &= \frac{1}{N_{ij} - 1} \sum_{k=1}^{N_{ij}} (x_{ij_k} - \overline{X}_{ij})^2, \\
\rho S_{Y_{ij}} S_{X_{ij}} &= S_{X_{ij}Y_{ij}}, \\
C_{Y_{ij}} &= \frac{S_{y_{ij}}}{\overline{y}_{ij}}, \\
C_{X_{ij}} &= \frac{S_{x_{ij}}}{\overline{x}_{ij}}, \\
C_{XY_{ij}} &= \frac{S_{xy_{ij}}}{\overline{x}_{ij}\overline{y}_{ij}}.
\end{aligned}
$$

Where; ρ is the population correlation coefficient between X and Y, $C_{X_{ij}}$ and $C_{Y_{ij}}$ represent coefficient of variation related to X and Y respectively.

3.3.1 Bias of the of the Proposed Estimator

Using the large sample approximations, equation (10) is now stated as

$$\begin{aligned} R_{DG,2,i} &= \overline{Y}_{ij}(1+e_0)\left(\frac{\overline{X}_{ij}(1+e_1)}{\overline{X}_{ij}}\right)^2 \\ &= \overline{Y}_{ij}(1+e_0)(1+e_1)^2. \end{aligned}$$

Assuming $|e_1| < 1$ which means $\left|\frac{(\overline{x}-\overline{X})}{\overline{X}}\right| < 1$, implies, possible estimate $\bar{x}$ of the population $\bar{X}$ lies between 0 and $2\bar{X}$. This holds if the variation in $\bar{x}$ is not large. As a result, it is assumed that the sample n is fairly large. And when the sample size is large, quantities of e_0 and e_1 are likely to be small and as such terms involving third and higher power of e_0 and e_1 will be negligible. That is, $\hat{R}_{DG,2,i} - \overline{Y}_i = \overline{Y}_i(e_0 - e_1)$ and $E(\overline{Y}_i(e_0 - e_1))$ = 0.

It can be seen that first order approximation produces unbiased ratio estimator of the population mean. Considering terms involving powers more than two negligible, then the approximation of the estimated mean is given as:

$$R_{DG,2,i} = \overline{Y}_{ij}(e_0e_1^2 + 2e_0e_1 + e_1^2 + e_0 + 2e_1 + 1). \tag{18}$$

Then the bias of $\hat{R}_{DG,2,i}$ is given by

$Bias(\hat{R}_{DG,2,i}) = E(\overline{Y}ij(e_0e_1^2 + 2e_0e_1 + e_1^2 + e_0 + 2e_1 + 1)) - \overline{Y}_i.$

Taking expectation yields

$$Bias(\hat{R}_{DG,2,i}) = \overline{Y}_i(2e_0e_1 + e_1^2). \tag{19}$$

Substituting equation (17) into Equation 19 gives

$$\begin{aligned} Bias(\hat{R}DG,2,i) = \sum_{j=1}^{J} W_{ij}\overline{Y}ij \Bigg[2\left(\frac{Nij - n_{ij}}{N_{ij}n_{ij}} C_{XY_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}} C_{XY_{2ij}}\right) \\ + \frac{N_{ij} - n_{ij}}{N_{ij}n_{ij}} C_{X_{ij}}^2 + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}} C_{X_{2ij}}^2 \Bigg]. \end{aligned} \tag{20}$$

From equation (20), bias reduces with an increasing sample size and this holds only with second order approximation.

3.3.2 Mean Squared Error of the Proposed Estimator

From equation (18), the MSE of $R_{DG,2,i}$ is given as

$$MSE(\hat{R}_{DG,2,i}) = 4\overline{Y}_i^2 e_0 e_1 + 4\overline{Y}_i^2 e_1^2 + \overline{Y}_i^2 e_0^2. \quad (21)$$

Subtituting equation (17) into equation (21) gives

$$\begin{aligned} MSE(\hat{R}_{DG,2,i}) = \sum_{j=1}^{J} W_{ij}^2 \overline{Y}_{ij}^2 \Bigg[4\left(\frac{N_{ij} - n_{ij}}{N_{ij} n_{ij}} C_{XY_{ij}} + \frac{(k_{ij} - 1) w_{2ij}}{n_{ij}} C_{XY_{2ij}} \right) \\ +4\left(\frac{N_{ji} - n_{ij}}{N_{ij} n_{ij}} C_{X_{ij}}^2 + \frac{(k_{ij} - 1) w_{2ij}}{n_{ij}} C_{X_{2ij}}^2 \right) \\ +\left(\frac{N_{ij} - n_{ji}}{N_{ij} n_{ji}} C_{Y_{ij}}^2 + \frac{(k_{ij} - 1) w_{2ij}}{n_{ij}} C_{Y_{2ij}}^2 \right) \Bigg]. \end{aligned} \quad (22)$$

Equation (22) is the mean squared error of the direct generalized ratio estimator when $\alpha = 2$. The optimum performance of equation (22) is achieved when sample size is fairly large with increased correlation between x and y.
The large sample representation of the direct generalized ratio estimator is

$$R_{DG,\alpha,i} = \overline{Y}_{ij}(1 + e_0)(1 + e_1)^{\alpha}. \quad (23)$$

Applying Binomial expansion on equation (23) up to second order of approximation gives;

$$\begin{aligned} R_{DG,\alpha,i} = \overline{Y}_{ij} + \overline{Y}_{ij}\alpha e_1 + \frac{\overline{Y}_{ij}\alpha^2 e_1^2}{2} - \frac{\overline{Y}_{ij}\alpha e_1^2}{2} + \overline{Y}_{ij} e_0 + \overline{Y}_{ij}\alpha e_0 e_1 \\ + \frac{\overline{Y}_{ij}\alpha^2 e_1^3}{2} - \frac{\overline{Y}_{ij}\alpha e_0 e_1^2}{2}. \end{aligned} \quad (24)$$

with bias

$$Bias(\hat{R}_{DG,\alpha,i}) = \frac{\overline{Y}_i \alpha(\alpha - 1) e_1^2}{2} + \overline{Y}_i \alpha e_0 e_1. \quad (25)$$

with equivalent presentation given by

$$\begin{aligned} Bias(\hat{R}_{DG,\alpha,i}) = \sum_{j=1}^{J} W_{ij} \overline{Y}_{ij} \frac{\alpha(\alpha - 1)}{2} \left[\frac{N_{ij} - n_{ij}}{N_{ij} n_{ij}} C_{X_{ij}}^2 + \frac{(k_{ij} - 1) w_{2ij}}{n_{ij}} C_{X_{2ij}}^2 \right] \\ +\alpha \sum_{j=1}^{J} W_{ij} \overline{Y}_{ij} \left[\frac{(N_{ij} - n_{ji})}{N_{ij} n_{ij}} C_{XY_{ij}} + \frac{(k_{ij} - 1) w_{2ij}}{n_{ij}} C_{XY_{2ij}} \right]. \end{aligned} \quad (26)$$

and MSE given by

$$MSE(\hat{R}_{DG,\alpha,i}) = \sum_{j=1}^{J} W_{ij}^2 \overline{Y}_{ij}^2 \Bigg[\left(\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}} C_{Y_{ij}}^2 + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}} C_{Y_{2ij}}^2 \right) + \alpha^2 \left(\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}} C_{X_{ij}}^2 + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}} C_{X_{2ij}}^2 \right) + 2\alpha \left(\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}} C_{XY_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}} C_{XY_{2ij}} \right) \Bigg]. \quad (27)$$

3.4 The Optimum α

The optimum value of α is obtain by partially differentiating Equation 25 with respect to α. Thus,

$$\frac{\partial}{\partial \alpha} \text{Bias}(\hat{R}_{DG,\alpha,i}) = \frac{\partial}{\partial \alpha} \left[\frac{\overline{Y}_i \alpha(\alpha-1)e_1^2}{2} + \overline{Y}_i \alpha e_0 e_1 \right] = \frac{\overline{Y}_i}{2} \left[2\alpha e_1^2 - e_1^2 \right] + \overline{Y}_i e_0 e_1.$$

Solving for α gives

$$\frac{\overline{Y}_i}{2} \left[2\alpha e_1^2 - e_1^2 \right] + \overline{Y}_i e_0 e_1 = 0$$

$$\implies \quad \alpha = -\left[\frac{e_0 e_1}{e_1^2} \right] + \frac{1}{2}. \quad (28)$$

Substituting equation (17) into equation (28) gives

$$\alpha_{opt} = \frac{1}{2} - \frac{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}} C_{XY_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}} C_{2XY_{ij}}}{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}} C_{X_{ij}}^2 + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}} C_{X_{2ij}}^2}. \quad (29)$$

Substituting equation (29) into equation (26) gives

$$Bias(\hat{R}_{DG,\alpha_{opt},i}) = \sum_{j=1}^{J} W_{ij}\overline{Y}_{ij}\left(\frac{1}{2} - \frac{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C_{XY_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C_{2XY_{ij}}}{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C^2_{X_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C^2_{X_{2ij}}}\right)$$
$$\times\left[\frac{\left(\frac{1}{2} - \frac{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C_{XY_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C_{2XY_{ij}}}{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C^2_{X_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C^2_{X_{2ij}}}\right) - 1}{2}\right]$$
$$\times\left[\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C^2_{X_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C^2_{X_{2ij}}\right]$$
$$+\left(\frac{1}{2} - \frac{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C_{XY_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C_{2XY_{ij}}}{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C^2_{X_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C^2_{X_{2ij}}}\right)$$
$$\times\sum_{j=1}^{J} W_{ij}\overline{Y}_{ij}\left[\frac{(N_{ij}-n_{ji})}{N_{ij}n_{ij}}C_{XY_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C_{XY_{2ij}}\right]. \qquad (30)$$

For optimum MSE, substituting equation (29) into equation (27) gives

$$MSE(\hat{R}_{DG,\alpha_{opt},i}) = \sum_{j=1}^{J} W^2_{ij}\overline{Y}^2_{ij}\left[\left(\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C^2_{Y_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C^2_{Y_{2ij}}\right)\right.$$
$$+\left(\frac{1}{2} - \frac{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C_{XY_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C_{2XY_{ij}}}{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C^2_{X_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C^2_{X_{2ij}}}\right)^2$$
$$\times\left(\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C^2_{X_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C^2_{X_{2ij}}\right)$$
$$+\left(1 - 2\left[\frac{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C_{XY_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C_{2XY_{ij}}}{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C^2_{X_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C^2_{X_{2ij}}}\right]\right)$$
$$\left.\times\left(\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C_{XY_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C_{XY_{2ij}}\right)\right]. \qquad (31)$$

From equation (31), an increase in coefficient of variation of the auxiliary variable will lead to a reduction of $MSE(\hat{R}_{DG,\alpha_{opt},i})$. Whilst in equation (22), an increase in the coefficient of variation of the auxiliary variable causes an increase in $MSE(\hat{R}_{DG,2,i})$.
To obtain the optimum mean of the i^{th} subpopulation, substituting equations (29), (13) and (14) into equation (8) gives

$$\hat{R}_{(DG,\alpha_{opt},i)} = \left(\sum_{j=1}^{J} W_{1ij}\overline{y}_{1ij} + \sum_{j=1}^{J} W_{2ij}\overline{y}_{n2r_{ij}}\right)$$
$$\times\left(\frac{\sum_{j=1}^{J} W_{1ij}\overline{x}_{1ij} + \sum_{j=1}^{J} W_{2ij}\overline{x}_{n2r_{ij}}}{\overline{X}_{ij}}\right)^{\left[\frac{1}{2} - \frac{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C_{XY_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C_{2XY_{ij}}}{\frac{N_{ij}-n_{ij}}{N_{ij}n_{ij}}C^2_{X_{ij}} + \frac{(k_{ij}-1)w_{2ij}}{n_{ij}}C^2_{X_{2ij}}}\right]}. \qquad (32)$$

4 Simulation Study

A simulation study was conducted to assess the performance of the proposed estimator. The simulation was performed using linear and nonlinear populations. The linear and nonlinear populations were population I and Population II respectively each of size 2000. Data for Population I were generated using the linear model.

$$Y_i = 1 + 2(x_i - 0.5) + e_i. \tag{33}$$

And for Population II, the nonlinear model used in generating the data is given by

$$Y_i = cos(1 + 2(x_i - 0.5)^2) + e_i. \tag{34}$$

The study considered Y as the study variable and X as the auxiliary variable. The auxiliary variable X is assumed to be uniformly distributed within a range of [0, 1]. The error term e_i is assumed as a standard normal variable, $e_i \sim \mathcal{N}(0, 1)$. Each population was divided into four subpopulations based on the 10^{th}, 20^{th}, 30^{th} and 40^{th} percentiles as $D_{10\%}$, $D_{20\%}$, $D_{30\%}$ and $D_{40\%}$. The subpopulations were each further divided into two tiers at the 40^{th} and 60^{th} percentiles as strata 1 and 2. In addition, the study considered two cases (case I and case II) of nonignorable nonresponse for each of the populations.

Case I: Approximately 30% of the population consist of nonrespondents and are available in both subpopulations and then in strata.

Case II: Two categories (20% and 40%) of nonrespondnets are respectively available in strata 1 and 2.

4.1 Population I

Table 1 presents descriptive statistics of various subpopulations under population I. The table displays the average values and variances of the study variable and auxiliary variable of the subpopulation-strata combination. The table in addition, shows the covariances and correlation coefficients between the study variable and auxiliary variable of the various subpopulations-strata combination.

Table 1: **Summary statistics for subpopulations and strata under population I**

Subpopulations	Strata	N_{ij}	n_{ij}	W_{ij}	$\overline{Y}_{ij}$	$\overline{X}_{ij}$	$S^2_{Y_{ij}}$	$S^2_{X_{ij}}$	$S_{XY_{ij}}$	$\rho_{XY_{ij}}$
$D_{10\%}$	1	80	53.3	0.4	1.0900	0.5400	2.2500	0.0900	0.3092	0.6816
	2	120	80.0	0.6	0.9600	0.4900	1.4400	0.2401	0.2209	0.6523
$D_{20\%}$	1	160	80.0	0.4	1.1000	0.5100	1.2321	0.0841	0.1678	0.5207
	2	240	120.0	0.6	0.9900	0.5300	1.3689	0.0900	0.1794	0.5105
$D_{30\%}$	1	240	96.0	0.4	0.9400	0.5000	1.2544	0.0841	0.1404	0.4351
	2	360	144.0	0.6	1.1000	0.5100	0.0841	1.2996	0.1803	0.5539
$D_{40\%}$	1	320	107.0	0.4	0.9700	0.5100	1.3689	0.0900	0.2058	0.5920
	2	480	160.0	0.6	1.0500	0.5100	1.4641	0.0841	0.1795	0.5167

Compared with auxiliary variable in Table 1, the study variable generally presents higher variances for all subpopulation-strata combination except stratum 2 of subpopulation $D_{30\%}$. Indicating largely that there is greater variability in the study variable than the auxiliary variable. The covariance values indicate the joint variability of the two variables. Positive covariance suggests that the variables have a tendency to move in tandem. In a similar vein, an inverse relationship is indicated by a negative covariance. Generally, the covariance values as presented are positive indicating the study variable and the auxiliary variable vary together in the positive direction for the various subpopulations. But the strength of association between the two variables is not solely indicated by the magnitude of the covariance; therefore, correlation coefficients are necessary. The correlation coefficients as presented, indicate generally a moderately high level of linear relationship between the study variable and the auxiliary variable.

Table 2 presents summary statistics for subpopulations and strata on nonignorable nonresponse case I for population I. It indicates the average values of nonignorable nonresponse sample and subsample of both the study and auxiliary variables for the various subpopulations. The table also contains the variances of both the study variable and auxiliary variable for case I of nonignorable nonresponse. As contained in the table, the study variable presents generally large variances across all the subpopulations compared with auxiliary variable. Indicating greater variability in the study variable than the auxiliary variable. the table further presents values on covariance with their corresponding correlation coefficient between the study variable and auxiliary variable. It can be observed from the table that there is moderately high linear correlation between the study variable and the auxiliary variable.

Table 2: **Summary statistics for subpopulation and strata for case I under population I**

Sub-population	Strata	N_{2ij}	W_{2ij}	n_{2ij}	K_{ij}	n_{2rij}	$\overline{y}_{2ij}$	$\overline{y}_{2rij}$	$\overline{x}_{2ij}$	$\overline{x}_{2rij}$	$S^2_{Y_{2ij}}$	$S^2_{X_{2ij}}$	$S_{XY_{2ij}}$	$\rho_{XY_{2ij}}$
$D_{10\%}$	1	24	0.3	20	2	10	0.5700	0.7262	0.5000	0.4703	1.6129	0.0961	0.2713	0.6804
	2	36	0.3	28	2	14	0.8900	0.5747	0.5100	0.4805	1.3456	0.0900	0.2504	0.7270
$D_{20\%}$	1	48	0.3	34	2	17	0.8900	1.1907	0.4300	0.4312	1.2996	0.0625	0.1637	0.5761
	2	72	0.3	45	2	23	1.0000	0.9470	0.5400	0.5402	1.3456	0.0900	0.1534	0.4441
$D_{30\%}$	1	72	0.3	45	2	23	0.9900	1.0249	0.5100	0.5000	1.5625	0.0784	0.1731	0.4928
	2	108	0.3	56	2	28	1.2300	1.2057	0.5300	0.5113	1.3456	0.0961	0.2063	0.5702
$D_{40\%}$	1	96	0.3	53	2	27	0.9000	0.6659	0.4600	0.3784	1.4641	0.0900	0.2172	0.6032
	2	144	0.3	65	2	33	1.0400	1.3876	0.4900	0.5374	1.5376	0.7840	0.1914	0.5478

Table 3 presents descriptive statistics for subpopulations and strata on case II of nonignorable nonresponse for population I. It displays the average values of nonignorable nonresponse samples and sub-samples for both the study variable and the auxiliary variable. It further provides information on the variances of the nonignorable nonresponse case II. Concerning variance, the study variable generally exhibits larger values compared to the auxiliary variable, indicating greater variability in the study variable than in the auxiliary variable. The covariance and correlation coefficient values between the study variable and auxiliary variable for the case II of nonignorable nonresponse suggest the study variable varies alongside the auxiliary variable, exhibit moderately high linear relationship.

Table 3: **Summary statistics for subpopulation and strata for case II under population I**

Sub-population	Strata	N_{2ij}	W_{2ij}	n_{2ij}	K_{ij}	n_{2rij}	$\overline{y}_{2ij}$	$\overline{y}_{2rij}$	$\overline{x}_{2ij}$	$\overline{x}_{2rij}$	$S^2_{Y_{2ij}}$	$S^2_{X_{2ij}}$	$S_{XY_{2ij}}$	$\rho_{XY_{2ij}}$
$D_{10\%}$	1	16	0.2	13	2	7	0.7900	0.6592	0.5600	0.3921	1.7424	0.0784	0.2215	0.6076
	2	48	0.4	28	2	14	0.7700	0.8749	0.5000	0.4989	1.4400	0.0900	0.2538	0.7064
$D_{20\%}$	1	32	0.2	23	2	12	0.8300	0.7006	0.3900	0.2970	1.1236	0.0576	0.1297	0.4994
	2	96	0.4	60	2	30	1.0001	1.0261	0.5500	0.5370	1.5376	0.0900	0.1979	0.5364
$D_{30\%}$	1	48	0.2	30	2	15	1.0400	1.1330	0.5000	0.5305	0.0784	1.6384	0.1851	0.5091
	2	144	0.4	76	2	38	1.2000	1.0886	0.5000	0.4755	1.2100	0.0900	0.1835	0.5424
$D_{40\%}$	1	64	0.2	36	2	18	0.4500	1.1325	0.8100	0.4950	1.5129	0.0900	0.2272	0.6182
	2	192	0.4	87	2	44	1.0400	1.3107	0.5000	0.4873	1.5625	0.0784	0.1937	0.5461

Table 4 displays the biases and the mean squared errors of the estimators based on subpopulations.

Table 4: **Bias and MSE of Estimators in both Case I and II for Population I Based on Subpopulation**

Case	Subpopulation	N_i	Bias		MSE	
			$\hat{R}_{DG,-1,i}$	$\hat{R}_{DG,2,i}$	$\hat{R}_{DG,-1,i}$	$\hat{R}_{DG,2,i}$
Case I	$D_{10\%}$	200	-1.0022	0.0116	0.0054	0.0356
	$D_{20\%}$	400	-1.0286	0.0077	0.0039	0.0159
	$D_{30\%}$	600	-1.0130	0.0205	0.0086	0.0479
	$D_{40\%}$	800	-1.0130	0.0112	0.0057	0.0238
Case II	$D_{10\%}$	200	-1.0019	0.0142	0.0057	0.0281
	$D_{20\%}$	400	-1.0280	0.0079	0.0041	0.0166
	$D_{30\%}$	600	-1.0196	0.0253	0.0116	0.0549
	$D_{40\%}$	800	-1.0130	0.0079	0.0038	0.0162

In Table 4, whilst the estimator by Ashutosh [16] has large negative biases, the proposed estimator ($\hat{R}_{DG,2,i}$) exhibits smaller positive biases. Furthermore, it can be observed that the MSE values of the estimator by Ashutosh (16) are generally smaller than that of the proposed estimator ($\hat{R}_{DG,2,i}$). Suggesting that the estimator by Ashutosh (16) performs better than the proposed estimator ($\hat{R}_{DG,2,i}$) as indicated by the PRE vaues.

4.2 Population II

The descriptive statistics of population II obtained using equation (34) are contained in Table 5. It indicates the study variable generally demonstrates greater variability compared to the auxiliary variable across all subpopulation-stratum combinations.

Table 5: **Summary statistics for subpopulations and strata under population II**

Subpopulations	Strata	N_{ij}	n_{ij}	W_{ij}	$\overline{Y}_{ij}$	$\overline{X}_{ij}$	$S^2_{Y_{ij}}$	$S^2_{X_{ij}}$	$S_{XY_{ij}}$	$\rho_{XY_{ij}}$
$D_{10\%}$	1	80	53.3	0.4	0.3800	0.5400	1.4161	0.0900	0.1256	0.3484
	2	120	80	0.6	0.4900	0.4900	0.8464	0.0784	0.0610	0.2336
$D_{20\%}$	1	160	80	0.4	0.5100	0.4800	0.9409	0.0841	-0.0030	-0.0108
	2	240	120	0.6	0.3100	0.5300	1.0816	0.0900	-0.0093	-0.0300
$D_{30\%}$	1	240	96	0.4	0.3300	0.5000	1.0404	0.0841	-0.0245	-0.0829
	2	360	144	0.6	0.4700	0.5100	0.9025	0.0841	0.0115	0.0426
$D_{40\%}$	1	320	107	0.4	0.3400	0.5100	0.9216	0.0900	0.0300	0.1060
	2	480	160	0.6	0.4200	0.5100	1.1236	0.0841	0.0098	0.0322

Table 5 further displayed positve covariances for some subpopuation and negative covariances for other subpopulations. Indicating that the study and auxiliary variables tend to vary in positive or negative direction depending on the subpopulation. Examining the correlation coefficients, there appears to be generally a very low level of linear relationship between the study variable and auxiliary variable.

Table 6 displays summary statistics for subpopulations and strata on nonignorable nonresponse case I for population II. It indicates the average values of nonignorable nonresponse sample and subsample for both the study and auxiliary variable. It also indicates the variances of the study variable and the auxiliary after the nonignorable nonresponse was introduced.

Table 6: **Summary statistics for subpopulation and strata for case I under population II**

Sub-population	Strata	N_{2ij}	W_{2ij}	n_{2ij}	K_{ij}	n_{2rij}	$\overline{y}_{2ij}$	$\overline{y}_{2rij}$	$\overline{x}_{2ij}$	$\overline{x}_{2rij}$	$S^2_{Y_{2ij}}$	$S^2_{X_{2ij}}$	$S_{XY_{2ij}}$	$\rho_{XY_{2ij}}$
$D_{10\%}$	1	24	0.3	20	2	10	-0.0600	-0.3156	0.5000	0.4993	0.9801	0.0961	0.0901	0.2902
	2	36	0.3	28	2	14	0.2600	0.2936	0.5100	0.4202	0.6724	0.0900	0.0635	0.2598
$D_{20\%}$	1	48	0.3	34	2	17	0.4600	0.1443	0.4300	0.4996	0.9216	0.0625	0.0479	0.1994
	2	72	0.3	45	2	23	0.2900	0.0574	0.5400	0.5122	1.2100	0.0900	-0.0334	-0.1018
$D_{30\%}$	1	72	0.3	45	2	23	0.3800	0.4093	0.5100	0.5000	1.2321	0.0784	0.0113	0.0361
	2	108	0.3	56	2	28	0.5300	0.6280	0.5300	0.6354	0.9216	0.0961	0.0043	0.0144
$D_{40\%}$	1	96	0.3	53	2	27	0.4600	0.2531	0.3400	0.3784	0.9801	0.0900	0.0430	0.1453
	2	144	0.3	65	2	33	0.4700	0.0829	0.4900	0.5144	1.1236	0.0784	0.0317	0.1064

The study variable presents generally large variances across all the subpopulations compared with auxiliary variable as presented in Table 6. Indicating greater variability in the study variable than in the auxiliary variable. Values of the covariances and the correlation coefficients as observed indicate the study variable varies positively with the auxiliary variable but with a weak linear relationship across all subpopulations.

Table 7 presents descriptive statistics for subpopulations and strata on nonignorable nonresponse case II for population II. It describes the average values of nonignorable nonresponse sample and subsample for both the study variable and the auxiliary variable. It also indicates the variances of the study variable and the auxiliary variable under case II of nonignorable nonresponse. Furthermore, the study variable presents generally large variances across all the subpopulations compared with auxiliary variable. Indicating greater variability in the study variable than the auxiliary variable.

Table 7: **Summary statistics for subpopulation and strata for case II under population II**

Sub-population	Strata	N_{2ij}	W_{2ij}	n_{2ij}	K_{ij}	n_{2rij}	$\overline{y}_{2ij}$	$\overline{y}_{2rij}$	$\overline{x}_{2ij}$	$\overline{x}_{2rij}$	$S^2_{Y_{2ij}}$	$S^2_{X_{2ij}}$	$S_{XY_{2ij}}$	$\rho_{XY_{2ij}}$
$D_{10\%}$	1	16	0.2	13	2	7	0.0800	-0.0249	0.5600	0.4237	1.2100	0.0784	0.0734	0.2410
	2	48	0.4	40	2	20	0.1500	-0.2114	0.5000	0.4566	0.7921	0.0900	0.0727	0.2745
$D_{20\%}$	1	32	0.2	23	2	12	0.4600	0.3090	0.3900	0.4350	0.0576	0.8836	0.0244	0.1062
	2	96	0.4	60	2	30	0.2800	0.2267	0.5500	0.5214	1.2100	0.0900	0.0121	0.0371
$D_{30\%}$	1	49	0.2	30	2	15	0.4400	0.5561	0.5000	0.5382	1.2769	0.0784	0.0192	0.0598
	2	144	0.4	76	2	38	0.5700	0.4867	0.5000	0.4755	0.9025	0.0900	-0.0038	-0.0133
$D_{40\%}$	1	64	0.2	36	2	18	0.2900	0.5263	0.4500	0.4950	1.0201	0.0900	0.0542	0.1802
	2	192	0.4	87	2	44	0.4300	0.7346	0.5000	0.4873	1.1236	0.0784	0.0278	0.0922

As contained in Table 7, values of the covariances and their corresponding correlation coefficients indicate same pattern of association between the study and the auxiliary variables within subpopulations except stratum 2 of $D_{30\%}$ which revealed that the study variable varies inversely with auxiliary variable.

Table 8 displays the bias and the mean squared errors of the estimators based on subpopulations to help evaluate the performance of the proposed estimator across the various subpopulations.

Table 8: **Bias and MSE of Estimators in both Case I and II for Population II Based on Subpopulation**

Case	Subpopulation	N_i	Bias		MSE	
			$\hat{R}_{DG,-1,i}$	$\hat{R}_{DG,2,i}$	$\hat{R}_{DG,-1,i}$	$\hat{R}_{DG,2,i}$
Case I	$D_{10\%}$	200	-0.4364	0.0045	0.0045	0.0071
	$D_{20\%}$	400	-0.2008	0.0011	0.0044	0.0051
	$D_{30\%}$	600	-0.4062	0.0010	0.0040	0.0046957
	$D_{40\%}$	800	-0.3811	0.0016	0.0039	0.0048250
Case II	$D_{10\%}$	200	-0.4356	0.0045	0.0045	0.0076
	$D_{20\%}$	400	-0.3814	0.0029	0.0042	0.0064
	$D_{30\%}$	600	-0.4069	0.0009	0.0040	0.0047
	$D_{40\%}$	800	-0.3811	0.0016	0.0040	0.0049

In Table 8, it can be observed that the proposed estimator ($\hat{R}_{DG,2,i}$) presents lower biases across all the subpopulations compared to the estimator by Ashutosh [16] . An increase in bias may indicate a deviation from the true values, while a decrease may imply improved accuracy. Implying the proposed estimator with less biases will present accurate estimates than the estimator by Ashutosh [16] . For the mean squared errors, the estimator by Ashutosh [16] generally present lower MSE values compared to the proposed estimator ($\hat{R}_{DG,2,i}$). The PRE values as indicated are less than 100% suggesting the estimator by Ashutosh [16] generally outperform the proposed estimator in terms of efficiency.

5 Empirical Study

The study considered California Houses datasets [18]. The datasets consist of fourteen (14) attributes but this research considered three (3) of them which include Median house value, Median income and Average household size. From the three selected attributes, two populations are obtained namely, population I and II each of size 2000. Population I is defined by considering the median house value as the study variable y and the median income as auxiliary variable x. And for Population II, the Median house value is considered as the study variable y and the Average household size as the auxiliary variable x. Each population was divided into four subpopulations based on the 10^{th}, 20^{th}, 30^{th} and 40^{th} percentiles as $D_{10\%}$, $D_{20\%}$, $D_{30\%}$ and $D_{40\%}$. The subpopulations were each further divided into two tiers at the 40^{th} and 60^{th} percentiles as strata 1 and 2. In addition, the study considered two cases (case I and case II) of nonignorable nonresponse for each of the populations.

Case I: Approximately 30% of the population consist of nonrespondents and are available in both subpopulations and then in strata.
Case II: Two categories (20% and 40%) of nonrespondnets are respectively available in strata 1 and 2.

5.1 Population I

The median house value and median income are assumed to be linearly correlated because higher-income areas often offer amenities such as good schools, lower crime rates, better infrastructure, and access to services. These factors contribute to increased demand for housing in such areas, which in turn drives up prices.
Table 9 displays the summary statistics of various subpopulations under population I. It shows the average values for the study variable and auxiliary variable.

Table 9: **Summary statistics for subpopulations and strata under population I**

Subpopulations	Strata	N_{ij}	n_{ij}	W_{ij}	$\overline{Y}_{ij}$	$\overline{X}_{ij}$	$S^2_{Y_{ij}}$	$S^2_{X_{ij}}$	$S_{XY_{ij}}$	$\rho_{XY_{ij}}$
$D_{10\%}$	1	80	53.3	0.4	209628.79	3.99	14657115014	3.02	149659.2	0.71
	2	120	80.0	0.6	200882.55	3.68	14200164896	4.59	176214.5	0.69
$D_{20\%}$	1	160	80.0	0.4	191030.01	3.61	10364471521	2.87	121352.8	0.70
	2	240	120.0	0.6	201145.86	3.71	13482448702	3.59	168479.4	0.77
$D_{30\%}$	1	240	96.0	0.4	204581.71	3.76	13894124262	3.71	166802.4	0.73
	2	360	144.0	0.6	211749.49	3.93	14021055089	3.93	166701.9	0.71
$D_{40\%}$	1	320	107.0	0.4	202657.24	3.76	13468845575	3.24	143601.5	0.69
	2	480	160.0	0.6	205026.92	3.87	13392510547	3.95	173148.7	0.75

Table 9 demonstrates that the study variable typically exhibits higher variances when contrasted with that of the auxiliary variable across all subpopulation-strata combinations. This suggests that there is generally greater variability in the study variable than in the auxiliary variable. Additionally, the covariance values between the study variable and the auxiliary variable for each stratum are generally positive indicating that the study variable and the auxiliary variable tend to vary in the same direction. And the correlation coefficients further showed a strong linear relationship between the study variable and the auxiliary variable.

Table 10 provides descriptive statistics concerning subpopulations and strata regarding nonignorable nonresponse case I for population I. It outlines the average values of both the study variable and the auxiliary variable for the nonignorable nonresponse sample and subsample. Additionally, it showcases the variances of the nonignorable nonresponse.

Table 10: **Summary statistics for subpopulation and strata for case I under population I**

Sub-population	Strata	N_{2ij}	W_{2ij}	n_{2ij}	K_{ij}	n_{2rij}	$\overline{y}_{2ij}$	$\overline{y}_{2rij}$	$\overline{x}_{2ij}$	$\overline{x}_{2rij}$	$S^2_{Y_{2ij}}$	$S^2_{X_{2ij}}$	$S_{XY_{2ij}}$	$\rho_{XY_{2ij}}$
$D_{10\%}$	1	24	0.3	20	2	10	221979.21	217030.00	4.13	4.17	16247601549	3.38	196476.8	0.84
	2	36	0.3	30	2	15	205922.28	225406.73	3.97	4.13	15284200530	6.14	227431.1	0.74
$D_{20\%}$	1	48	0.3	34	2	17	188370.85	207994.12	3.60	3.93	10813136222	3.93	172692.4	0.84
	2	72	0.3	46	2	23	206993.08	193895.65	3.77	3.54	15377696599	3.59	171855.9	0.73
$D_{30\%}$	1	72	0.3	45	2	23	215165.35	204160.91	3.78	3.58	14935426360	2.97	152112.5	0.72
	2	108	0.3	57	2	29	206615.80	216806.97	3.98	3.99	13133289195	4.15	163677.9	0.70
$D_{40\%}$	1	96	0.3	53.3	2	27	203903.19	221614.89	3.91	3.86	14268535813	3.37	147140.8	0.67
	2	144	0.3	65	2	33	204156.99	215463.70	3.94	3.92	13541676468	4.91	195007.8	0.76

Table 10 suggests that the values of the covariances and their corresponding correlation coefficients indicate a consistent pattern of association between the study and auxiliary variables within subpopulations.

Table 11 provides descriptive statistics regarding subpopulations and strata for nonignorable nonresponse case II for population II. It outlines the average values for both the study variable and the auxiliary variable for the nonignorable nonresponse sample and subsample. Additionally, it displays the variances of the nonignorable nonresponse and presents values for covariance and correlation coefficients between the study variable and auxiliary variable.

Table 11: **Summary statistics for subpopulation and strata for case II under population I**

Sub-population	Strata	N_{2ij}	W_{2ij}	n_{2ij}	K_{ij}	n_{2rij}	$\overline{y}_{2ij}$	$\overline{y}_{2rij}$	$\overline{x}_{2ij}$	$\overline{x}_{2rij}$	$S^2_{Y_{2ij}}$	$S^2_{X_{2ij}}$	$S_{XY_{2ij}}$	$\rho_{XY_{2ij}}$
$D_{10\%}$	1	16	0.2	13	2	7	238343.81	260200.14	4.42	4.61	18445348179	3.60	213307.0	0.83
	2	48	0.4	40	2	20	202377.12	210615.10	3.84	4.04	15029388410	5.29	208927.5	0.74
$D_{20\%}$	1	32	0.2	23	2	12	182606.25	171308.33	3.62	3.53	10131259960	3.00	146198.6	0.84
	2	96	0.4	60	2	30	205586.50	202066.67	3.85	3.91	15575029344	3.98	189630.3	0.76
$D_{30\%}$	1	48	0.2	30	2	15	227120.90	1860067.73	3.78	3.51	14440549286	2.64	118967.3	0.61
	2	144	0.4	76	2	38	198331.99	181955.29	3.94	3.89	12145440511	3.58	146255.1	0.70
$D_{40\%}$	1	64	0.2	36	2	18	221901.66	233972.28	4.11	4.05	16946820747	4.01	167468.8	0.64
	2	192	0.4	87	2	44	202524.00	191004.57	3.89	3.38	13269041934	4.51	187570.5	0.77

Table 11 indicates that the study variable typically demonstrates greater variances across all subpopulations in comparison to the auxiliary variable, signifying a higher level of variability in the study variable. The covariance values and their corresponding correlation coefficients consistently revealed a positive pattern of association between the study and auxiliary variables within subpopulations.

Table 12 displays the bias and the mean squared errors of the proposed estimator ($\alpha = 2$) based on subpopulations.

Table 12: **Bias and MSE of Estimators in both Case I and II for Population I Based on Subpopulation**

Case	Subpopulation	N_i	Bias		MSE	
			$\hat{R}_{DG,-1,i}$	$\hat{R}_{DG,2,i}$	$\hat{R}_{DG,-1,i}$	$\hat{R}_{DG,2,i}$
Case I	$D_{10\%}$	200	-203765.64	1501.91	36475620.19	512502610.95
	$D_{20\%}$	400	-196663.38	1079.12	22899882.64	344782432.63
	$D_{30\%}$	600	-208444.66	1033.30	27128201.09	351648481.24
	$D_{40\%}$	800	-203675.07	995.35	24731577.89	335754074.94
Case II	$D_{10\%}$	200	-203752.38	501.85	38324473.58	525148684.22
	$D_{20\%}$	400	-196862.96	1077.39	23494759.30	360651325.01
	$D_{30\%}$	600	-208613.44	993.57	27743822.89	346649336.67
	$D_{40\%}$	800	-203668.73	1012.19	25182027.80	356195394.88

In Table 12, it is evident that the proposed estimator ($\hat{R}_{DG,2,i}$) exhibit lower biases across all subpopulations compared to the estimator by Ashutosh [16] . An increase in bias may signal a deviation from true values, while a decrease may imply improved accuracy. This suggests that the proposed estimator with lower biases provide more accurate estimates than the estimator by Ashutosh [16] . Regarding MSE values, the estimator by Ashutosh [16] generally presents lower MSE values compared to the proposed estimator ($\hat{R}_{DG,2,i}$) as reduced MSE further supports enhanced accuracy.

5.2 Population II

California is known for its diverse population, with households of varying sizes and compositions. This diversity can lead to a wide range of housing preferences and choices, making it challenging to establish a strong linear correlation between house value and household size. Moreover, due to government policies and economic conditions including demographic dynamics can cause housing market experience fluctuations and trends over times. These dynamics can influence house prices independently of household size. As a results, while there may be some relationship between median house value and average household size, it is likely not strong enough to establish a clear linear correlation due to the complex interplay of various factors influencing housing markets and individual housing decisions.

Table 13 presents the summary statistics for the various subpopulations within population II. It indicates the average values for both the study and the auxiliary variables.

Table 13: **Summary statistics for subpopulations and strata under population II**

Subpopulations	Strata	N_{ij}	n_{ij}	W_{ij}	$\overline{Y}_{ij}$	$\overline{X}_{ij}$	$S^2_{Y_{ij}}$	$S^2_{X_{ij}}$	$S_{XY_{ij}}$	$\rho_{XY_{ij}}$
$D_{10\%}$	1	80	53.3	0.4	209628.79	513.77	14657115014	96167.7	4954129.0	0.13
	2	120	80.0	0.6	200882.50	477.5	14200164896	121208.8	-813996.4	-0.02
$D_{20\%}$	1	160	80.0	0.4	191030.01	517.77	10364471521	104355.5	3133480.0	0.10
	2	240	120.0	0.6	201145.86	532.21	13482448702	149952.6	951912.4	0.02
$D_{30\%}$	1	240	96.0	0.4	204581.70	521.4	10813136222	144747.5	-723630.4	-0.02
	2	360	144.0	0.6	211749.49	511.41	15377696599	212754.5	5214605.0	0.09
$D_{40\%}$	1	320	107.0	0.4	202657.24	510.38	13468845575	126969.0	-1733606.0	-0.04
	2	480	160.0	0.6	205026.92	490.62	13392510547	123877.9	1645440.0	0.04

Compared to the auxiliary variable as illustrated in Table 13, the study variable typically demonstrates higher variances across all subpopulation-strata combination. This suggests a generally greater variability in the study variable compared to the auxiliary variable. Additionally, the covariance values exhibit various patterns of variations, indicating that the study variable and the auxiliary variable tend to vary in the same direction in certain strata and inversely vary in others. However, the strength of the relationship between the two variables cannot be solely determined by the magnitude of the covariance; therefore, correlation coefficients are crucial. And the correlation coefficients generally suggest a weak linear relationship between the study and the auxiliary variables.

Table 14 displays descriptive statistics concerning subpopulations and strata for nonignorable nonresponse case I for population II. It presents the mean values of both the study variable and the auxiliary variable for case I of nonignorable nonresponse sample and subsample. It further showcases the variances of the nonignorable nonresponse and provides figures for covariance and correlation coefficients between the study variable and auxiliary variable.

Table 14: **Summary statistics for subpopulation and strata for case I under population II**

Sub-population	Strata	N_{2ij}	W_{2ij}	n_{2ij}	K_{ij}	n_{2rij}	$\overline{y}_{2ij}$	$\overline{y}_{2rij}$	$\overline{x}_{2ij}$	$\overline{x}_{2rij}$	$S^2_{Y_{2ij}}$	$S^2_{X_{2ij}}$	$S_{XY_{2ij}}$	$\rho_{XY_{2ij}}$
$D_{10\%}$	1	24	0.3	20	2	10	221979.21	217030.00	464.17	499.50	16247601549	59163.7	12099012.0	0.40
	2	36	0.3	30	2	15	205922.28	225406.73	516.67	443.47	15284200530	175481.1	1128063.0	0.02
$D_{20\%}$	1	48	0.3	34	2	17	188370.85	207994.12	551.56	579.59	10813136222	144747.5	-723630.4	-0.02
	2	72	0.3	45	2	23	206993.08	193895.65	495.69	520.78	15377696599	212754.5	5214605.0	0.09
$D_{30\%}$	1	72	0.3	45	2	23	215165.35	204160.93	521.78	522.17	14935426360	93700.8	1863152.0	0.05
	2	108	0.3	57	2	29	206615.80	216806.97	560.99	597.00	13133289195	239836.6	3326667.0	0.06
$D_{40\%}$	1	96	0.3	53.3	2	27	203903.19	221614.89	484.77	507.33	14268535813	107783.0	-1669716.0	-0.04
	2	144	0.3	65	2	33	204156.99	215463.70	500.42	500.88	13541676468	144127.9	4218243.0	0.10

As shown in Table 14, the covariance values and their corresponding correlation coefficients consistently demonstrate a different pattern of association between the study and auxiliary variables within the subpopulations.

Table 15 presents descriptive statistics regarding subpopulations and strata for nonignorable nonresponse case II for population II. It outlines the average values of both the study variable and the auxiliary variable for the nonignorable nonresponse sample and subsample. It provides the variances together with the covariance and correlation coefficients between the study variable and auxiliary variable.

Table 15: **Summary statistics for subpopulation and strata for case II under population II**

Sub-population	Strata	N_{2ij}	W_{2ij}	n_{2ij}	K_{ij}	n_{2rij}	$\overline{y}_{2ij}$	$\overline{y}_{2rij}$	$\overline{x}_{2ij}$	$\overline{x}_{2rij}$	$S^2_{Y_{2ij}}$	$S^2_{X_{2ij}}$	$S_{XY_{2ij}}$	$\rho_{XY_{2ij}}$
$D_{10\%}$	1	16	0.2	13	2	7	238343.81	260200.14	443.56	405.14	18445348179	27782.66	14342649	0.63
	2	48	0.4	40	2	20	202377.12	210615.10	485.73	574.35	15029388410	138933.8	-104170.4	-0.00
$D_{20\%}$	1	32	0.2	23	2	12	182606.25	171308.33	555.84	659.75	10131259960	167951.9	-3522880	-0.09
	2	96	0.4	60	2	30	205586.50	202066.67	493.55	569.9	15575029344	191625.8	1918172	0.04
$D_{30\%}$	1	49	0.2	30	2	15	227120.90	186006.73	530.77	615.47	14440549286	106406.7	5827205	0.15
	2	144	0.4	76	2	38	198331.99	181955.29	563.76	486.526	12145440511	222220.6	5817673	0.11
$D_{40\%}$	1	64	0.2	36	2	18	221901.66	233972.28	505.77	428.89	16946820747	133645.3	-3639414	-0.08
	2	192	0.4	87	2	44	202524.00	191004.57	495.68	449.21	13269041934	142343.9	5140143	0.12

Table 15 indicates that the study variable typically demonstrates greater variances across all subpopulations compared to the auxiliary variable, suggesting a higher degree of variability in the study variable. The covariance values and their corresponding correlation coefficients consistently reveal a distinct pattern of association between the study and auxiliary variables within the subpopulations.

Table 16 presents bias and mean squared errors of the proposed estimators ($\alpha = 2$) based on subpopulations.

Table 16: **Bias and MSE of Estimators under both Case I and II for Population II Based on Subpopulation**

Case	Subpopulation	N_i	Bias		MSE	
			$\hat{R}_{DG,-1,i}$	$\hat{R}_{DG,2,i}$	$\hat{R}_{DG,-1,i}$	$\hat{R}_{DG,2,i}$
Case I	$D_{10\%}$	200	-203577.06	1068.36	157070996.22	493707767.66
	$D_{20\%}$	400	-196447.82	897.87	127830133.44	406184014.56
	$D_{30\%}$	600	-208010.64	642.10	157543983.40	528781839.47
	$D_{40\%}$	800	-203416.20	753.61	123546824.58	361943686.79
Case II	$D_{10\%}$	200	-203566.23	980.88	159475697.11	474920107.78
	$D_{20\%}$	400	-196439.50	883.97	137564989.04	414817468.76
	$D_{30\%}$	600	-208179.43	1170.08	159146261.42	557542998.86
	$D_{40\%}$	800	-203409.86	789.82	127859678.59	365081408.60

Table 16 clearly demonstrates that the proposed estimator ($\hat{R}_{DG,2,i}$) exhibit lower biases across all subpopulations compared to the estimator by Ashutosh [16]. Regarding MSE values, the estimator by Ashutosh [16] generally presents lower MSE values compared to the proposed estimator as reduced MSE further supports enhanced accuracy.

6 Discussion

This sections discusses results on the bias and the mean squared error (MSE). Large bias, renders an estimator poor and an estimator with good MSE properties exhibits minimal combined variance and bias. In this studies, the referenced estimator is the estimator by Ashutosh [16]. The biases of the proposed estimator ($\hat{R}_{DG,2,i}$) and the estimator by Ashutosh [16] for populations I and II of the simulation studies are presented in Table 4 and Table 8 respectively. The estimator by Ashutosh [16] has large negative biases indicating an average systematic underestimation. For the proposed estimator ($\hat{R}_{DG,2,i}$), it exhibits smaller positive biases suggesting small and likely negligible systematic errors on average. It can be observed that there is reduction in the biases of the estimator by Ashutosh [16] and the proposed estimator ($\hat{R}_{DG,2,i}$) as the sample size increases across the various subpopulations under both cases I and II. Whilst there is fluctuations in the biases of the estimator by Ashutosh [16] at $D_{20\%}$ (400) under both cases I and II of nonresponse, it occurs at $D_{30\%}$ (600) for the the proposed estimator ($\hat{R}_{DG,2,i}$). This suggest that the accuracy of the estimator is sensitive to changes in sample size. An increase in bias may indicate a deviation from the true values, while a reduction in bias may suggest an improved accuracy. Generally, the proposed estimator ($\hat{R}_{DG,2,i}$) have smaller biases across the various subpopulations compared to the estimator by Ashutosh [16] suggesting that the proposed estimator is better in terms of bias under population I. The results for population II is not different from that of population I except that comparing the two populations under the simulation studies, the biases under Population II appeared more consistent compared to those under population I as the sample size increase across the various subpopulations for the proposed estimator ($\hat{R}_{DG,2,i}$).

Tables 12 and 16 respectively present the biases of the proposed estimator ($\hat{R}_{DG,2,i}$) and that of Ashutosh [16] under the empirical studies for populations I and II. The results from the empirical studies are not generally different from the simulation studies. Thus, concerning the bias, proposed estimator ($\hat{R}_{DG,2,i}$) outperformed the estimator by Ashutosh [16] under both simulation studies and empirical studies.

With regareds to the MSE values for populations I and II under the simulation studies, it can be observed that the MSE values of the estimator by Ashutosh [16] are generally smaller than that of the proposed estimator ($\hat{R}_{DG,2,i}$). Suggesting that the estimator by Ashutosh [16] performed better than the proposed estimator ($\hat{R}_{DG,2,i}$). For population I, the MSE values fluctuate as the sample size increases across the various subpopulations under both cases I and II.

However, for population II, the MSE values get reduced for the proposed estimator ($\hat{R}_{DG,2,i}$) as the sample size increases across the various subpopulations. Suggesting that the proposed estimator ($\hat{R}_{DG,2,i}$) and the estimator by Ashutosh [16] are consistent under population II. Comparing the reduction in the MSE values, the values under population II of the proposed estimator ($\hat{R}_{DG,2,i}$) deteriorate faster than that of the estimator by Ashutosh [16] as the sample size increases across the various subpopulations. As MSE is the combination of variance and bias, the MSE values of the the proposed estimator ($\hat{R}_{DG,2,i}$) are higher than that of Ashutosh [16] because of high variance in the study variable. For the empirical studies, though the values are generally large, the implications of the results are not different from that of the simulation studies for both populations I and II.

7 Conclusions

An improved estimator for mean of a finite population in the presence of nonignorable nonresponse has been developed. Theoretical properties of the proposed estimator were derived. The simulation and empirical studies revealed that the proposed estimator under the population I and population II performed better compared to the estimator by Ashutosh [16] in terms of bias. The biases of both the proposed estimator ($\hat{R}_{DG,2,i}$) and the estimator by Ashutosh [16] were found to be generally consistent for both population I and population II under both simulation and empirical studies. However, the MSE values indicated that the the estimator by Ashutosh [16] is marginally efficient than the proposed estimator ($\hat{R}_{DG,2,i}$). And this could be as a result of large variances in the study variable across the subpopulations. Moreover, the MSE values suggest that the proposed estimator ($\hat{R}_{DG,2,i}$) may be less bias and more efficient at higher sample sizes for nonlinear population when there is nonignorable nonresponse present. Therefore, when the focus is on estimation, the proposed estimator is preferred. However, when the emphasis shifts to inference, the estimator by Ashutosh [16] may prove to be beneficial.

Declarations

Additional information
No available additional information.

Data Availability
Data used in this study are published data and hence publicly available.

References

[1] Stoop, I. A. 2005; *The hunt for the last respondent: Nonresponse in sample surveys* (Vol. 200508). Sociaal en Cultureel Planbu.

[2] Cochran, W. G. Sampling techniques. John Wiley and Sons; 1977.

[3] Heeringa, S. G., West, B. T., and Berglund, P. A. Applied Survey Data Analysis (2nd ed.). Chapman and Hall/CRC; 2017.

[4] Sarndal, C.E., Swensson, B., and Wretman, J. 2003. Model-Assisted Survey Sampling. Springer.

[5] Lohr, S. L. Sampling: design and analysis. CRC press; 2021.

[6] Kikechi, C. B. 2019; Model based robust estimation of finite population total using the procedure of local linear regression (Doctoral dissertation, University of Nairobi).

[7] Ahmed, S. and Shabbir, J. 2019; Model based estimation of population total in presence of non-ignorable non-response. *PLOS ONE*, *14*(10), e0222701.

[8] Yadav, R. and Tailor, R. 2020; Estimation of finite population mean using two auxiliary variables under stratified random sampling. *Statistics in Transition new series*, *21*(1).

[9] Pernet, C. R., Wilcox, R., and Rousselet, G. A. 2013; Robust correlation analyses: false positive and power validation using a new open source matlab toolbox. *Frontiers in psychology*, *3*(606).

[10] Joshua, I. M., and Okon, B. M. 2022; Calibration estimators for population mean with subsampling the nonrespondents under stratified sampling. *Science Journal of Applied Mathematics and Statistics*, *10*(4), 45 − 56.

[11] Makhdum, M., Sanaullah, A., and Hanif, M. 2020; A modified regression-cum-ratio estimator of population mean of a sensitive variable in the presence of non-response in simple random sampling. *Journal of Statistics and Management Systems*, 23, 495 − 510.

[11] Kim, S. and Carver, R. P. 2019; Nonresponse bias in a study on the relationship between income and educational attainment. *Journal of Applied Social Science Research*, *42*(2), 87 − 105.

[12] Tikkiwal, G. C. and Ghiya, A. 2000; A generalized class of synthetic estimators with application to crop acreage estimation for small domains. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *42*(7), 865 − 876.

[14] Sarndal, C. E. and Lundstrom, S. Estimation in surveys with nonresponse. John Wiley and Sons;2005.

[15] Singh, H. P. and Kumar, S. 2008; A regression approach to the estimation of the finite population mean in the presence of non-response. *Australian & New Zealand Journal of Statistics*,

[16] Ashutosh, A. 2021; Estimator of Domain Mean Using Stratified Sampling in the Presence on Non-Response. *Sri Lankan Journal of Applied Statistics*, *22*(1), 13 − 29.

[17] Hansen, M. H. and Hurwitz, W. N. 1946; The problem of non-response in sample surveys. *Journal of the American Statistical Association*, *41*(236), 517 − 529.

[18] Sampath, S. Sampling theory and methods. Alpha Science Int'l Ltd; 2005.

[19] Pace, R. Kelley, and Ronald Barry. "Sparse spatial autoregressions." Statistics and Probability Letters 33.3 1997 : 291-297