Open Journal of Statistics, 2022, 12, 175-187
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

A Modified Regression Estimator for Single Phase Sampling in the Presence of Observational Errors

Nujayma M. A. Salim, Christopher O. Onyango

Department of Mathematics and Actuarial Science, Kenyatta University, Nairobi, Kenya
Email: nujayma.m.a@gmail.com

How to cite this paper: Salim, N.M.A. and Onyango, C.O. (2022) A Modified Regression Estimator for Single Phase Sampling in the Presence of Observational Errors. *Open Journal of Statistics*, **12**, 175-187.
https://doi.org/10.4236/ojs.2022.122012

Received: January 19, 2022
Accepted: April 11, 2022
Published: April 14, 2022

Abstract

In this paper, a regression method of estimation has been used to derive the mean estimate of the survey variable using simple random sampling without replacement in the presence of observational errors. Two covariates were used and a case where the observational errors were in both the survey variable and the covariates was considered. The inclusion of observational errors was due to the fact that data collected through surveys are often not free from errors that occur during observation. These errors can occur due to over-reporting, under-reporting, memory failure by the respondents or use of imprecise tools of data collection. The expression of mean squared error (MSE) based on the obtained estimator has been derived to the first degree of approximation. The results of a simulation study show that the derived modified regression mean estimator under observational errors is more efficient than the mean per unit estimator and some other existing estimators. The proposed estimator can therefore be used in estimating a finite population mean, while considering observational errors that may occur during a study.

Keywords

Estimate, Regression, Covariates, Single Phase Sampling, Observational Errors, Mean Squared Error

1. Introduction

In sample survey, there are two types of errors, namely; sampling and non-sampling errors. Sampling error occurs when there is a difference between an unrevealed parameter of population and its estimate, determined using data from a sample and not the entire population. The different kinds of errors that may occur during data collection, processing, and estimation are bounded by non-sampling errors, which are: coverage error, frame error, response/non-response error, observational error, and processing error (Baker [1]).

In sample survey, an assumption is often made that all observations on the elements under research are quantified correctly. However, in practice, this supposition is often violated when observational errors occur during the survey. This error comes about when the measured quantity differs from the true value. The justification for this is stated below: The variable is distinctly specified although it is demanding to take accurate observations at least with techniques presently available or due to other types of practical drawbacks, for example. the test done on the level of hypoglycaemia in a human being is not exactly the perfect measure of the level of sugar in the body. The variable is ideally well described but observations can only be obtained on some strongly resembling substitutes known as surrogates. Also, the measurement of economic status of a person is based on a person's income or occupation, thus, these are the surrogates. The variable is totally comprehensible but it is not instinctively defined, for instance, brilliance and aggressiveness (Singh [2]).

If the size of observational errors is small, then they can be said to be incorporated in the disturbance or error term and hence will not have an impact on the statistical inferences made. The random error term broadly stands for the influence of various explanatory variables that have not actually been included in the relation. Conversely, if they are large in size, they will then result in erroneous and inaccurate statistical inferences. These observational errors will hence make the derived result invalid. If they are very minimal, they can be disregarded; hence the statistical inferences made from data observed remain reliable (Srivasatava *et al.* [3]).

Srivasatava *et al.* [3] studied two regression co-efficient estimators, one of which emerged from direct regression and the other from reverse regression when predisposed to observational errors with information on a single covariate. The comparison study showed that the presence of observational errors influences the quality of the efficiencies of the proposed estimators such that the sample mean could be surpassing the two proposed regression estimators in regard to the variances. The study recommended the analysis of observational errors on the efficiencies of estimators with information on more than one covariate.

In sample surveys, auxiliary information is utilized at both selection and estimation stages to enhance the efficiency of the estimators. When auxiliary information is to be utilized at approximation stage, the ratio, product and regression estimators are broadly applied during several cases. In a single-phase sampling scheme, a sample is selected using a sampling method and data are collected from each unit in the sample. A researcher, then, observes both the survey variable and covariates in the sample. Single-phase sampling is typically used when auxiliary information from the previous census is known.

In practice for customary sample surveys, it is common that only sampling error is considered when finding the estimate of the MSE since it is quite simple. There is then an implication that the sampling error has predominance over observational errors and all the other round-off errors. This supposition is not credible when it comes to administrative data or a complete enumeration. It is utilized for most sample surveys because of its expediency and not its credibility.

The issue of observational errors tainting data and its effects has been studied by several authors like Cochran [4], Fuller [5], Alwin [6], Gregoire and Salas [7], Kumar *et al.* [8], Singh and Karpe [9] and Singh *et al.* [10]. The indistinguishable conclusion was that the observational errors have a notable effect on the estimators and that they are worthy of scrutiny during parameter estimation. Singh and Karpe [11] studied the estimation of population average in the case where observational errors arise in the survey. They later observed the properties of the bias and mean squared error of these estimators and concluded that the presence of observational errors raises the variances of the estimators.

Allen *et al.* [12] suggested a class of estimators of the population average with multi covariate information and later examined the estimators in the presence of observational errors. It was observed that there was an increase in the variance of the estimator due to the inclusion of imperfect measurements of the study variates and covariates.

Kumar [13] proposed an estimator where observational errors were introduced during the estimation stage and showed that his estimator performed best after an empirical study was performed on four populations.

Tum *et al.* [14] proposed a consistent and efficient modified regression type estimator that combined regression estimator and ratio-product type exponential estimator then compared it to the mean per unit estimator, ratio estimator, product estimator, regression estimator, exponential ratio estimator, exponential product estimator and ratio-product type exponential estimator. The estimator was used to estimate a population mean or total using covariates that had either positive or negative correlation with the main variable. It was established that the estimator was biased in small samples but the bias was found to be negligible in large samples.

Authors such as Singh and Espejo [15], Hanif *et al.* [16], Tailor and Sharma [17] and Shukla *et al.* [18] have worked on estimating the mean using a single-phase sampling scheme and tested the efficiencies of their estimators, however, little attention was given to estimation in the presence of observational errors in the data. Quite a number of researchers have, however, emphasized on the importance of including non-sampling errors.

In this paper, we develop the mean squared error of the Tum *et al.* [14] estimator in single-phase sampling while assuming the presence of observational errors in both the survey variable and the covariates. Some existing estimators are reviewed and later the mean squared error of the proposed estimator is derived. An empirical study is later done to compare the performance of the proposed estimator with the reviewed existing estimators (in works of Tum *et al.*

[14], Singh and Espejo [15], Kumar [19] and Bahl and Tuteja [20]) in terms of their mean squared errors and their relative efficiency.

In the paper by Tum *et al.* [14], the emergence of errors in the variables wasn't taken into consideration, specifically observational errors. We've briefly discussed the effects observational errors can have on the final results of statistical research hence the reason why the work done by TUM had to be extended.

2. Review of Existing Estimators

The estimators mentioned below were used during the simulation study for comparison with the proposed regression estimator.

The ratio estimator:

$$t_1 = \bar{y}\frac{\mu_x}{\bar{x}} \tag{1}$$

where $\bar{y}$ and $\bar{x}$ are the sample means of the main variable and covariate respectively and μ_x the population mean of the covariate.

The MSE of the estimator t_1 is given in Kumar *et al.* [19] as,

$$\text{MSE}(t_1) = \frac{1}{n}\left\{\left(\sigma_y^2 - 2R\rho_{yx}\sigma_y\sigma_x + R^2\sigma_x^2\right) + \left(\frac{\mu_y^2}{\mu_x^2}\sigma_v^2 + \sigma_u^2\right)\right\}$$

where $R = \frac{\mu_y}{\mu_x}$.

The product estimator:

$$t_2 = \bar{y}\frac{\bar{x}}{\mu_x} \tag{2}$$

where $\bar{y}$ and $\bar{x}$ are the sample means of the main variable and covariate respectively and μ_x the population mean of the covariate.

The MSE of the estimator t_2 is given in Kumar *et al.* [19] as:

$$\text{MSE}(t_2) = \frac{1}{n}\left\{\left(\sigma_y^2 + 2R\rho_{yx}\sigma_y\sigma_x + R^2\sigma_x^2\right) + \left(\frac{\mu_y^2}{\mu_x^2}\sigma_v^2 + \sigma_u^2\right)\right\}$$

The regression estimator:

$$t_3 = \bar{y} + \beta\left(\mu_x - \bar{x}\right) \tag{3}$$

where $\bar{y}$ and $\bar{x}$ are the sample means of the main variable and covariate respectively and μ_x the population mean of the covariate.

The MSE of the estimator t_3 is given in Kumar *et al.* [19] as:

$$\text{MSE}(t_3) = \theta\mu_y^2\left(1+\frac{\sigma_u^2}{\sigma_y^2}\right)C_y^2\left(1-\rho^2\Big/\left(\left(1+\frac{\sigma_u^2}{\sigma_y^2}\right)\left(1+\frac{\sigma_v^2}{\sigma_x^2}\right)\right)\right)$$

where $k = \mu_y\rho C_y \Big/ \mu_x\left(1+\frac{\sigma_v^2}{\sigma_x^2}\right)C_x$.

The exponential ratio type estimator proposed by Bahl and Tuteja [20]

$$t_4 = \bar{y}\exp\left(\frac{\mu_x - \bar{x}}{\mu_x + \bar{x}}\right) \tag{4}$$

where $\bar{y}$ and $\bar{x}$ are the sample means of the main variable and covariate respectively and μ_x the population mean of the covariate.

The MSE of the estimator t_4 is given

$$\text{MSE}(t_4) = \frac{\sigma_y^2}{n}\left(1 - \frac{C_x}{C_y}\left(\rho_{yx} - \frac{C_x}{4C_y}\right)\right) + \frac{1}{n}\left(\frac{\mu_y^2}{4\mu_x^2}\sigma_v^2 + \sigma_u^2\right)$$

The exponential product type estimator proposed by Bahl and Tuteja [20]

$$t_5 = \bar{y}\exp\left(\frac{\bar{x} - \mu_x}{\bar{x} + \mu_x}\right) \tag{5}$$

where $\bar{y}$ and $\bar{x}$ are the sample means of the main variable and covariate respectively and μ_x the population mean of the covariate.

The MSE of the estimator t_5 is given

$$\text{MSE}(t_5) = \frac{\sigma_y^2}{n}\left(1 + \frac{C_x}{C_y}\left(\rho_{yx} - \frac{C_x}{4C_y}\right)\right) + \frac{1}{n}\left(\frac{\mu_y^2}{4\mu_x^2}\sigma_v^2 - \sigma_u^2\right)$$

The exponential ratio-product type estimator proposed by Singh and Espejo [15]

$$t_6 = \bar{y}\exp\left(\alpha\left(\frac{\mu_x - \bar{x}}{\mu_x + \bar{x}}\right) + (1-\alpha)\left(\frac{\bar{x} - \mu_x}{\bar{x} + \mu_x}\right)\right) \tag{6}$$

where $\bar{y}$ and $\bar{x}$ are the sample means of the main variable and covariate respectively and μ_x the population mean of the covariate.

The MSE of the estimator t_6 is given

$$\text{MSE}(t_6) = \theta\bar{Y}^2\left(\left(C_y^2 - \gamma\rho C_x C_y + \frac{1}{4}\gamma C_x^2\right) + \left(\frac{S_U^2}{\bar{Y}^2} + \frac{1}{4}\mu_O^2\frac{S_V^2}{\bar{X}^2}\right)\right) + \lambda\bar{Y}^2\left(\frac{S_U^2}{\bar{Y}^2} + \frac{1}{4}\mu_O^2\frac{S_V^2}{\bar{X}^2}\right)$$

Lastly, the modified regression estimator proposed by Tum *et al.* [14], given by

$$t_7 = \left(\bar{y} + \beta\left(\bar{X} - \bar{x}\right)\right)\left(\alpha\exp\left(\frac{\bar{Z} - \bar{z}}{\bar{Z} + \bar{z}}\right) + (1-\alpha)\exp\left(\frac{\bar{Z} - \bar{z}}{\bar{Z} + \bar{z}}\right)\right) \tag{7}$$

where $\bar{X}$ and $\bar{Z}$ are the population means of the covariates, $\bar{x}$ and $\bar{z}$ are the sample means of the covariates and $\bar{y}$ the sample mean of the main variable.

The MSE of the estimator t_7 is given,

$$\text{MSE}(t_7) = \theta\bar{Y}^2 C_y^2\left(1 - \rho_{y.xz}^2\right)\rho_{y.xz}$$

where, $\theta = \frac{N-n}{Nn}$, $\bar{Y}$ is the mean of the main variable, C_y is the coefficient of variation and $\rho_{y.xz}$ is the multiple correlations between the main variable and the covariates.

The above estimators are consistent but biased in small samples. However, the bias is negligible in large samples. The estimators are more efficient than mean per unit estimator in large samples.

3. The Proposed Modified Regression Estimator in Single Phase Sampling

Let (x_i, y_i) be the observed values and (X_i, Y_i) be the true values on two characteristics (x, y) respectively corresponding to the $i^{th}(i = 1, 2, \cdots, n)$ sample unit. Let the observational errors be

$$U_i = y_i - Y_i$$
$$V_i = x_i - X_i$$
$$Q_i = z_i - Z_i$$

where U_i is for the study variable and V_i and Q_i for the covariates.

The observational errors are assumed to be random in nature and they are uncorrelated with mean zero and variances, S_U^2, S_V^2 and S_Q^2 respectively. The actual values of the variables Y and X are assumed to be independent of observational errors. An additional assumption is that the observational errors for variable Y, X and Z are independent.

Let S_Y^2, S_X^2 and S_Z^2 denote the variances of the study variable of interest Y and the auxiliary variables X and Z respectively for the population. Let ρ_{YX}, ρ_{YZ} and ρ_{XZ} be the coefficients of correlation between the variable Y and X, Y and Z and X and Z for the population.

$$\text{Let, } W_X = \sum_{i=1}^{n}\left(X_i - \bar{X}\right),\ W_X = \sum_{i=1}^{n}\left(X_i - \bar{X}\right) \text{ and } W_Z = \sum_{i=1}^{n}\left(Z_i - \bar{Z}\right) \quad (8)$$

$$W_U = \sum_{i=1}^{n} U_i,\ W_V = \sum_{i=1}^{n} V_i \text{ and } W_Q = \sum_{i=1}^{n} Q_i. \quad (9)$$

$$[\ \bar{y} = \bar{Y} + \frac{1}{n}\left(W_Y + W_U\right),\ \bar{x} = \bar{X} + \frac{1}{n}\left(W_X + W_V\right) \text{ and } \bar{z} = \bar{Z} + \frac{1}{n}\left(W_Z + W_Q\right)\ ] \quad (10)$$

Let $Y_i - \bar{Y}$, $X_i - \bar{X}$ and $Z_i - \bar{Z}$ be the deviations of the true values from the population means of the study variable and auxiliary variable for the i^{th} unit and these deviations are summed over the sample size. Further, assumed that the average of the main variable Y is unknown and the average of the covariate X is known.

Taking expectation of corresponding summations of (8) and (9) divided by n and squaring we get the following

$$\begin{cases} E\left(\dfrac{W_Y + W_U}{n}\right)^2 = \theta\left(S_Y^2 + S_U^2\right) = \theta\bar{Y}^2\left(C_Y^2 + C_U^2\right) \\ E\left(\dfrac{W_X + W_V}{n}\right)^2 = \theta\left(S_X^2 + S_V^2\right) = \theta\bar{Y}^2\left(C_X^2 + C_V^2\right) \\ E\left(\dfrac{W_Z + W_Q}{n}\right)^2 = \theta\left(S_Z^2 + S_Q^2\right) = \theta\bar{Y}^2\left(C_Z^2 + C_Q^2\right) \end{cases} \quad (11)$$

Further combining pairs in (11), we obtain

$$\begin{cases} E\left(\left(\frac{W_Y+W_U}{n}\right)\left(\frac{W_X+W_V}{n}\right)\right)=\theta\rho_{XY}S_YS_X=\theta\bar{Y}\bar{X}\rho_{XY}C_YC_X \\ E\left(\left(\frac{W_Y+W_U}{n}\right)\left(\frac{W_Z+W_Q}{n}\right)\right)=\theta\rho_{YZ}S_YS_Z=\theta\bar{Y}\bar{Z}\rho_{YZ}C_YC_Z \\ E\left(\left(\frac{W_X+W_V}{n}\right)\left(\frac{W_Z+W_Q}{n}\right)\right)=\theta\rho_{XZ}S_XS_Z=\theta\bar{X}\bar{Z}\rho_{XZ}C_XC_Z \end{cases} \tag{12}$$

The estimator from (7) is utilised, however, the study variable Y and the two covariates X and Z are assumed to contain observational errors.

Taking (10) in (7), the proposed estimator becomes

$$\begin{aligned} t_{\text{Proposed}} &= \left(\bar{Y}+\frac{1}{n}(W_Y+W_U)+\beta\left(\bar{X}-\bar{X}-\frac{1}{n}(W_X+W_V)\right)\right) \\ &\times\left(\alpha\exp\left(\frac{\bar{Z}-\bar{Z}-\frac{1}{n}(W_Z+W_Q)}{\bar{Z}+\bar{Z}+\frac{1}{n}(W_Z+W_Q)}\right)\right. \\ &\left.+(1-\alpha)\exp\left(\frac{\bar{Z}+\frac{1}{n}(W_Z+W_Q)-\bar{Z}}{\bar{Z}+\bar{Z}+\frac{1}{n}(W_Z+W_Q)}\right)\right) \end{aligned} \tag{13}$$

Using first-order approximation, ignoring the second and higher terms for each expansion of (13) and after simplification, we can write

$$t_{\text{Proposed}}=\bar{Y}+\bar{Y}\left(\frac{1}{2}-\alpha\right)\frac{1}{\bar{Z}n}(W_Z+W_Q)+\frac{1}{n}(W_Y+W_U)-\beta\frac{1}{n}(W_X+W_V) \tag{14}$$

The mean squared error of (14) is given by

$$\begin{aligned} &\text{MSE}(t_{\text{Proposed}})\approx \text{E}(t_{\text{Proposed}}-\bar{Y})^2 \\ &\approx \text{E}\left(\bar{Y}+\bar{Y}\left(\frac{1}{2}-\alpha\right)\frac{1}{\bar{Z}n}(W_Z+W_Q)+\frac{1}{n}(W_Y+W_U)-\beta\frac{1}{n}(W_X+W_V)-\bar{Y}\right)^2 \end{aligned} \tag{15}$$

Squaring the terms in (15) and taking expectation we find

$$\begin{aligned} \text{MSE}(t_{\text{Proposed}}) &= \bar{Y}^2\left(\frac{1}{2}-\alpha\right)^2\frac{1}{\bar{Z}^2}E\left(\frac{W_Z+W_Q}{n}\right)^2 \\ &+E\left(\frac{W_Y+W_U}{n}\right)^2+\beta^2E\left(\frac{W_X+W_V}{n}\right)^2 \\ &+2\bar{Y}\left(\frac{1}{2}-\alpha\right)\frac{1}{\bar{Z}}E\left(\left(\frac{W_Z+W_Q}{n}\right)\left(\frac{W_Y+W_U}{n}\right)\right) \\ &-2\bar{Y}\left(\frac{1}{2}-\alpha\right)\frac{1}{\bar{Z}}\beta E\left(\left(\frac{W_Z+W_Q}{n}\right)\left(\frac{W_X+W_V}{n}\right)\right) \\ &-2\beta E\left(\left(\frac{W_Y+W_U}{n}\right)\left(\frac{W_X+W_V}{n}\right)\right) \end{aligned} \tag{16}$$

Substituting (11) and (12) in (16) we find

$$\begin{aligned}\text{MSE}\left(t_{\text{Proposed}}\right) &= \bar{Y}^2\left(\frac{1}{2}-\alpha\right)^2\frac{1}{\bar{Z}^2}\theta\bar{Z}^2\left(C_Z^2+C_Q^2\right)+\theta\bar{Y}^2\left(C_Y^2+C_U^2\right)\\&+\beta^2\theta\bar{X}^2\left(C_X^2+C_V^2\right)+2\bar{Y}\left(\frac{1}{2}-\alpha\right)\frac{1}{\bar{Z}}\theta\bar{Y}\bar{Z}\rho_{YZ}C_YC_Z\\&-2\bar{Y}\left(\frac{1}{2}-\alpha\right)\frac{1}{\bar{Z}}\beta\theta\bar{X}\bar{Z}\rho_{XZ}C_XC_Z-2\beta\theta\bar{Y}\bar{X}\rho_{XY}C_XC_Y\end{aligned}$$

On expansion and simplification, the $\text{MSE}\left(t_{\text{Proposed}}\right)$ can be divided into two sections:

MSE Independent of observational errors given by

$$\begin{aligned}\text{MSE}\left(t_{\text{Proposed}}\right) &= \frac{1}{4}\theta\bar{Y}^2C_Z^2-\alpha\theta\bar{Y}^2C_Z^2+\alpha^2\theta\bar{Y}^2C_Z^2+\theta\bar{Y}^2C_Y^2+\beta^2\theta\bar{X}^2C_X^2\\&+\theta\bar{Y}^2\rho_{YZ}C_YC_Z-2\alpha\theta\bar{Y}^2\rho_{YZ}C_YC_Z-\bar{Y}\beta\theta\bar{X}\rho_{XZ}C_ZC_X\\&+2\alpha\bar{Y}\beta\theta\bar{X}\rho_{XZ}C_ZC_X-2\beta\theta\bar{Y}\bar{X}\rho_{XY}C_YC_X\end{aligned}\tag{17}$$

And MSE is dependent on observational errors given by

$$\text{MSE}\left(t_{\text{Proposed}}\right)=\frac{1}{4}\theta\bar{Y}^2C_Q^2-\alpha^*\theta\bar{Y}^2C_Q^2+\alpha^{*2}\theta\bar{Y}^2C_Q^2+\theta\bar{Y}^2C_U^2+\beta^{*2}\theta\bar{X}^2C_V^2\tag{18}$$

Differentiating (18) with respect to α^* and β^* then equating to zero,

$$-\theta\bar{Y}^2C_Q^2+2\alpha^*\theta\bar{Y}^2C_Q^2=0\quad\text{then}\quad\alpha^*=\frac{1}{2}$$

$$2\beta^*\theta\bar{X}^2C_V^2=0\quad\beta^*=\frac{0}{2\theta\bar{X}^2C_V^2}=0$$

Differentiating (17) with respect to β and α then putting it in matrix form to find the optimum values of α and β

$$\begin{pmatrix}1&\rho_{XZ}\\\rho_{XZ}&1\end{pmatrix}\begin{pmatrix}-\bar{Y}C_Z+2\alpha\bar{Y}C_Z\\2\beta\bar{X}C_X\end{pmatrix}=2\bar{Y}C_Y\begin{pmatrix}\rho_{ZY}\\\rho_{XY}\end{pmatrix}$$

Let $\begin{pmatrix}1&\rho_{XZ}\\\rho_{XZ}&1\end{pmatrix}$ be denoted by a square matrix $R_{2\times2}$. Then

$$2\bar{Y}C_Y\begin{pmatrix}\rho_{ZY}\\\rho_{XY}\end{pmatrix}=R_{2\times2}\begin{pmatrix}-\bar{Y}C_Z+2\alpha\bar{Y}C_Z\\2\beta\bar{X}C_X\end{pmatrix}$$

$$\begin{pmatrix}-\bar{Y}C_Z+2\alpha\bar{Y}C_Z\\2\beta\bar{X}C_X\end{pmatrix}=2\bar{Y}C_YR_{2\times2}^{-1}\begin{pmatrix}\rho_{ZY}\\\rho_{XY}\end{pmatrix}=2\bar{Y}C_Y\frac{AdjR_{2\times2}}{\left|R_{2\times2}\right|}\begin{pmatrix}\rho_{ZY}\\\rho_{XY}\end{pmatrix}$$

$$\begin{pmatrix}-\bar{Y}C_Z+2\alpha\bar{Y}C_Z\\2\beta\bar{X}C_X\end{pmatrix}=\frac{2\bar{Y}C_Y}{\left|R_{2\times2}\right|}\begin{pmatrix}(-1)^{1+1}\left|R_{2\times2}\right|\\(-1)^{2+1}\left|R_{2\times2}\right|\end{pmatrix}$$

$$-\bar{Y}C_Z+2\alpha\bar{Y}C_Z=\frac{2\bar{Y}C_Y}{\left|R_{2\times2}\right|}(-1)^{1+1}\left|R_{yz}\right|_{yxz}$$

$$\alpha=\frac{1}{2}+\frac{C_Y}{C_Z\left|R_{2\times2}\right|}(-1)^{1+1}\left|R_{yz}\right|_{yxz}\tag{19}$$

$$2\beta\bar{X}C_X = \frac{2\bar{Y}C_Y}{|R_{2\times 2}|}(-1)^{2+1}|R_{yx}|_{yxz}$$

$$\beta = \frac{\bar{Y}C_Y}{\bar{X}C_X|R_{2\times 2}|}(-1)^{2+1}|R_{yx}|_{yxz} \tag{20}$$

Using normal equations that are used to find the optimum values of α *and* β $\text{MSE}(t_{\text{Proposed}})$ can be written in simplified form as:

$$\text{MSE}(t_{\text{Proposed}}) \approx E\left(\frac{1}{n}(W_Y+W_U)\left(\frac{1}{n}(W_Y+W_U)+\bar{Y}\left(\frac{1}{2}-\alpha\right)\frac{1}{\bar{\bar{Z}}n}(W_Z+W_Q)\right.\right.$$
$$\left.\left.-\beta\frac{1}{n}(W_X+W_V)\right)\right).$$

Taking expectation and using (11) and (12)

$$\text{MSE}(t_{\text{Proposed}}) \approx \theta\bar{Y}^2(C_Y^2+C_U^2)+\bar{Y}\left(\frac{1}{2}-\alpha\right)\frac{1}{\bar{Z}}\theta\bar{Y}\bar{Z}\rho_{YZ}C_YC_Z-\beta\theta\bar{Y}\bar{X}\rho_{XY}C_YC_X$$

Substituting the optimum values of α and β from (19) and (20) and simplifying

$$\text{MSE}(t_{\text{Proposed}}) = \frac{\theta\bar{Y}^2C_Y^2}{|R_{2\times 2}|}|R|_{yxz}+2\theta\bar{Y}^2C_U^2$$

4. Simulation Study

A simulation study was performed using R-programming language to compare the performance of the proposed modified regression type estimator in the presence of observational errors with already existing estimators in finite population, developed by Kumar *et al.* [10], Bahl and Tuteja [3], Singh and Espejo [15] and Tum *et al.* [14].

4.1. Simulated Population

1) The true study variable $Y_i \sim N(75,10)$. The observational error is $u \sim N(3,1.5)$. The observed (Y_i) in presences of observational error $y_i = Y_i + u_i$.

2) The characteristics of the auxiliary variable for Ratio estimator are; N = 500, n = 60 mean = 60.72 standard deviation = 9.69. The observational error is $q \sim N(2,1)$. The observed (z_i) in presences of observational error is $z_i = Z_i + q_i$.

3) The characteristics of the auxiliary variable for Regression estimator are; N = 500, n = 60 mean = 23.16, standard deviation = 2.59. The observational error is $v \sim N(2.5,1)$. The observed (x_i) in presences of observational error $x_i = X_i + v_i$.

4) The characteristics of the auxiliary variable for Product estimator are; N = 500, n = 60 mean = 40.59, standard deviation = 7.52. The observational error is $q \sim N(1,0.5)$. The observed (z_i) in presences of observational error $z_i = Z_i + q_i$.

4.2. Simulation Results

A summary of the results from Tables 1-4

1) The mean squared error of the estimators decreases as the sample size increases from n = 30 to n = 90. This shows that the efficiency of the estimators

increases with an increase in sample size.

2) Out of the eight estimators the proposed estimator ($\bar{y}_{\text{MRE(proposed)}}$) is the most efficient followed by the exponential ratio product estimator ($\bar{y}_{ERP}$) with covariates positively correlated to main variable, regression estimator ($\bar{y}_{RE}$), exponential ratio estimator ($\bar{y}_{ER}$), exponential product estimator ($\bar{y}_{EP}$), exponential ratio product estimator ($\bar{y}_{ERP}$) with covariates negatively correlated to main variable, ratio estimator ($\bar{y}_{R}$) and the least efficient being the product estimator ($\bar{y}_{P}$). This arrangement stays true with increasing sample size.

3) The proposed estimator ($\bar{y}_{\text{MRE(proposed)}}$) is less efficient than the estimator by Tum *et al.* [14] $\bar{y}_{\text{MRE(TUM)}}$.

Table 1. Mean squared error of existing estimators and the proposed estimator.

Estimators	Mean squared errors, MSE for varying sample sizes, n		
N	$n = 30$	$n = 60$	$n = 90$
$\text{var}(\bar{y})$	3.94	1.78	1.11
Product estimator, $\bar{y}_P$	3.17	1.67	0.72
Ratio estimator, $\bar{y}_R$	2.06	0.92	0.59
Regression estimator, $\bar{y}_{RE}$	1.52	0.67	0.39
Exponential product estimator, $\bar{y}_{EP}$	1.59	0.78	0.58
Exponential ratio estimator, $\bar{y}_{ER}$	1.35	0.66	0.45
Exponential ratio-product estimator, $\bar{y}_{ERP}$	1.63	0.82	0.59
Exponential ratio-product estimator $\bar{y}_{ERP}$	1.25	0.62	0.42
Modified regression estimator (Proposed) $\bar{y}_{\text{MRE(proposed)}}$	1.05	0.44	0.29
Modified regression estimator (Proposed) $\bar{y}_{\text{MRE(proposed)}}$	0.92	0.41	0.25

Table 2. Relative efficiency of existing and proposed estimator with respect to mean per unit estimator in single phase sampling in the presence of observational error.

Estimators	Relative percent efficiency with respect to mean per unit
$\text{var}(\bar{y})$	100
Product estimator, $\bar{y}_P$	106
Ratio estimator, $\bar{y}_R$	192
Regression estimator, $\bar{y}_{RE}$	264
Exponential product estimator, $\bar{y}_{EP}$	225
Exponential ratio estimator, $\bar{y}_{ER}$	268
Exponential ratio-product estimator, $\bar{y}_{ERP}$	216
Exponential ratio-product estimator, $\bar{y}_{ERP}$	278
Modified regression estimator (Proposed), $\bar{y}_{\text{MRE(proposed)}}$	398
Modified regression estimator (Proposed), $\bar{y}_{\text{MRE(proposed)}}$	429

Table 3. Mean Squared Error of the proposed Modified regression estimator in the presence of observational errors and the estimator by Tum *et al.* [14].

Estimators	Mean squared errors		
N	$n = 30$	$n = 60$	$n = 90$
Modified regression estimator (Proposed) $\bar{y}_{\text{MRE(proposed)}}$	0.96	0.47	0.20
Modified regression estimator (TUM) $\bar{y}_{\text{MRE(TUM)}}$	0.55	0.34	0.14

Table 4. Relative efficiency of the proposed modified regression estimator in the presence of observational errors and the estimator by Tum *et al.* [14].

Estimators	Relative percent efficiency for $n = 90$
Modified regression estimator (Proposed) $\bar{y}_{\text{MRE(proposed)}}$	100
Modified regression estimator (TUM), $\bar{y}_{\text{MRE(TUM)}}$	145

The proposed estimator was seen to be less efficient than the Tum *et al.* [14] estimator. This is owing to the fact that the proposed estimator has taken into account observational errors that have in turn led to its higher variance. Similar results have been seen by Srivastava and Shalabh [3], Allen *et al.* [12], Singh and Karpe [9] and Kumar *et al.* [13].

5. Conclusions

This study dealt with the problem of observational errors that occur in a survey and derived the mean squared error of a modified regression estimator in single-phase sampling while assuming the presence of observational errors in both the main variable and the covariates.

A comparison of the obtained MSE with an increasing sample size shows a decrease in the MSE with an increase in sample sizes. Also, the proposed estimator had the least mean squared error and highest efficiency compared to the other existing estimators, except for the Tum *et al.* [14] estimator. This was a result of the consideration of observational errors that have in turn led to its higher variance when compared to the Tum *et al.* [14] estimator.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Baker, B. (2011) Micro Integration. Statistical Methods Paper 201108. Statistics Netherlands, The Hague/Heerlen.

[2] Singh, S., Jain, K. and Sharma, S. (2014) Replicated Measurement Error Model under Exact Linear Restrictions. *Statistical Papers*, **55**, 253-274. https://doi.org/10.1007/s00362-012-0469-7

[3] Srivasatava, A.K. and Shalabh (2001) Effect of Measurement Error on the Regression Method of Estimation in Survey Sampling. *Journal of Statistical Research*, **35**, 35-44.

[4] Cochran, W.G. (1968) Errors of Measurement in Statistics. *Technometrics*, **10**, 637-666. https://doi.org/10.2307/1267450

[5] Fuller, W.A. (1995) Estimation in the Presence of Measurement Error. *International Statistical Review*/*Revue Internationale de Statistique*, **63**, 121-141. https://doi.org/10.2307/1403606

[6] Alwin, D.F. (2007) Margins of Error: A Study of Reliability in Survey Measurement. John Wiley & Sons, Hoboken. https://doi.org/10.1002/9780470146316

[7] Gregoire, T.G. and Salas, C. (2009) Ratio Estimation with Measurement Error in the Auxiliary Variate. *Biometrics*, **65**, 590-598. https://doi.org/10.1111/j.1541-0420.2008.01110.x

[8] Kumar, M., Singh, R., Sawan, N. and Chauhan, P. (2011) Exponential Ratio Method of Estimators in the Presence of Observational Errors. *International Journal of Agricultural and Statistical Sciences*, **7**, 457-461.

[9] Singh, H.P. and Karpe, N. (2008) Ratio-Product Estimator for Population Mean in Presence of Observational Errors. *Journal of Applied Statistical Science*, **16**, 437-452.

[10] Singh, R., Malik, S. and Kohshnivesan, M. (2014) An Alternative Estimator for Estimating the Finite Population Mean in Presence of Measurement Errors with the View to Financial Modeling. *Science Journal of Mathematics and Statistics*, **2**, 107-111.

[11] Singh, H.P. and Karpe, N. (2010) Estimation of Mean, Ratio and Product Using Auxiliary Information in the Presence of Measurement Errors in Sample Surveys. *Journal of Statistical Theory and Practice*, **4**, 111-136 https://doi.org/10.1080/15598608.2010.10411976

[12] Allen, J., Singh, H.P. and Smarandache, F. (2003) A Family of Estimators of Population Mean Using Multiauxiliary Information in Presence of Measurement Errors. *International Journal of Social Economics*, **30**, 837-848. https://doi.org/10.1108/03068290310478775

[13] Kumar, S., Bhougal, S., Nataraja, N.S. and Viswanathaiah, M. (2015) Estimation of Population Mean in the Presence of Non-Response and Measurement Error. *Revista Colombiana de Estadística*, **38**, 145-161.

[14] Tum, E.C., Kung'u, J. and Odongo, L. (2014) A New Regression Type Estimator with Two Covariates for Single-Phase Sampling. *Open Journal of Statistics*, **4**, 789-796. https://doi.org/10.4236/ojs.2014.49074

[15] Singh, H.P. and Espejo, M.R. (2003) On Linear Regression and Ratio-Product Estimation of a Finite Population Mean. *Journal of the Royal Statistical Society*: *Series D* (*The Statistician*), **52**, 59-67. https://doi.org/10.1111/1467-9884.00341

[16] Hanif, M., Hamad, N. and Shahbaz, M.Q. (2009) A Modified Regression Type Estimator in Survey Sampling. *World Applied Sciences Journal*, **7**, 1559-1561.

[17] Tailor, R., Verma, M.R. and Sharma, B. (2010) An Alternative Ratio-Cum-Product Estimator of Population Mean Using a Coefficient of Kurtosis for Two Auxiliary Variates. *Data Science Journal*, **9**, 93-99.

[18] Shukla, D., Pathak, S. and Thakur, N.S. (2012) Estimation of Population Mean Using Two Auxiliary Sources in Sample Surveys. *Statistics in Transition New Series*, **13**, 21-36.

[19] Kumar, M., Singh, R., Singh, A.K. and Smarandache, F. (2011) Some Ratio Type Estimators under Observational Errors. *World Applied Sciences Journal*, **14**, 272-276.

[20] Bahl, S. and Tuteja, R.K. (1991) Ratio and Product Type Exponential Estimators. *Journal of Information and Optimization Sciences*, **12**, 159-164. https://doi.org/10.1080/02522667.1991.10699058