Modified Calibration Variance Estimators in the Presence of Non- Response and Measurement Error

Abstract

In sample survey, there are many problems encountered by survey statisticians in estimating the population variance of the study variable. These problems include: presence of outliers in data collected for analysis, non-response and measurement errors occurrence during survey. Shahzad et al. [26] developed variance estimators by addressing the problem of outliers using L-moment and calibration approach. However, they do not consider the situation of non-response and measurement errors. This paper addresses these problems by proposing modified variance estimators in the presence of non-response and measurement errors. The properties (Biases and MSEs) were derived up to the first order of approximation using Taylor series approach. The efficiency conditions of the modified estimators over the existing estimators considered in the study were established. The result of simulation studies revealed that the estimators are efficient.

Keywords: MSE; Bias; Calibration variance estimator; measurement error; non-response.

UNDER PEER REVIEW

1 Introduction

The use of auxiliary information is very important in estimation, because it enhance the performance of estimators, many researchers have developed variance estimators for the estimation of population variance of the study variable using auxiliary information, authors such as Arnab and Singh [1], Audu and Singh [2], Audu et al. [3], Cekim and Kadilar [4], Das and Tripathi [5], Isaki [6], Kadilar and Cingi [7], Adejumobi and Yunusa [8,9], Ozel et al. [10], Singh et al. [11], Upadhyaya and Singh [12], Yadav et al. [13], Yunusa et al. [14,15], have use auxiliary information in the development of estimators under simple and stratified random sampling schemes. Non-response and measurement errors are two common non sampling errors that normally occur during the conduct of sample survey. These errors affect estimation strategies properties and such estimation strategies may give unreliable estimates, apart from these errors, other factor such as the presence of outliers in the data can distort the results obtained from estimation. Hanse and Hurwitz [16], were the first to addressed the problem of non-response while authors such as Cochran [17], Misra et al. [18], Audu et al. [19,20], have addressed the issues of non-response and measurement errors in estimation.

Calibration estimation is a general method for improving the original weight of an estimator while minimizing a particular distance measure using auxiliary variable and a set of calibration constraints. The construction of new weight requires two basic components: a distance measure and a set of calibration constraints. Calibration provides a method for systematically incorporating auxiliary data into the workflow. Hence, it becomes a widely used procedure of estimation in sample survey. In the existence of auxiliary variables, when the sample sum of the weighted auxiliary variable equals the known population total for that auxiliary variable, the calibrated weight may produce flawless estimators. Authors such as Deville and Sarndal [21], Estevao and Sarndal [22], Kim and Park [23], Koyuncu and Kadilar [24], Audu et al. [25] have used calibration approach in developing estimators. Shahzad et al. [26] developed variance estimators by addressing the problem of outliers using L-moment and calibration approach. However, they do not consider the situation of non-response and measurement errors.

This study is limited to modification of Shahzad et al. [26] L-Moments based calibrated variance estimators to capture the situation of non-response and measurement error under stratified random sampling.

Assume a finite population $U=(u_1,u_2,u_3...u_N)$ of size N, and let y and x respectively, be the study and auxiliary variables associated with each unit $u_i;(i=1,2,...,N)$ of population. Let the population size N be stratified into L strata with hth stratum containing N_h units, where $h=1,2,...L$ such that $\sum_{i=h}^{L} N_h = N$. A simple random sample of size n_h is drawn without replacement from the hth stratum such that $\sum_{h=1}^{L} n_h = n$. Let (y_{hi},x_{hi}) be the observed values of the variables y and x on jth of the hth stratum, where $i=1,2,...,N$ and $h=1,2,....,L$ before discussing about the existing estimators we will write the nomenclatures to use in this study

2 Variance estimators and Calibration variance estimators in the literature

The unbiased variance estimator under stratified random sampling is giving by

$$t_1=\sum_{h=1}^{L}\frac{W_h^2}{n_h}s_{yh}^2 \tag{2.1}$$

The variance of t_1 is given as;

$$Var(t_1)=\sum_{h=1}^{L}\frac{W_h^4}{n^3}S_{yh}^4\,(\lambda_{40h}-1) \tag{2.2}$$

Prasad and Singh [27] proposed unbiased estimator of finite population variance using auxiliary information in sample survey as:

$$t_2=\sum_{h=1}^{L}\frac{W_h^2}{n_h}\left[s_{yh}^2-\frac{s_{xh}^2}{S_{xh}^2}+1\right] \tag{2.3}$$

The mean square error of the estimator is given by

$$MSE(t_2)=\sum_{h=1}^{L}\frac{(W_hS_{yh})^4}{n_h^3}\left[(\lambda_{40h}-1)+\frac{(\lambda_{04h}-1)}{S_{yh}^4}-\frac{2(\lambda_{22h}-1)}{S_{yh}^2}\right] \tag{2.4}$$

UNDER PEER REVIEW

Ozel et al. [9] suggested separate ratio estimator for population variance as

$$t_3 = \sum_{h=1}^{L} W_h \, \frac{s_{yh}^2}{s_{xh}^2} S_{xh}^2 \tag{2.5}$$

$$t_4 = \sum_{h=1}^{L} W_h \frac{s_{yh}^2}{s_{xh}^2 + C_{xh}} (S_{xh}^2 + C_{xh}) \tag{2.6}$$

$$t_5 = \sum_{h=1}^{L} W_h \, s_{yh}^2 \left[\frac{S_{xh}^2 + \beta_{Xh}}{s_{xh}^2 + \beta_{Xh}} \right] \tag{2.7}$$

$$t_6 = \sum_{h=1}^{L} W_h s_{yh}^2 \left[2 - \left[\frac{S_{xh}^2 + \beta_{xh}}{s_{xh}^2 + \beta_{xh}} \right]^{Q_h} \right] \tag{2.8}$$

The mean square errors of the estimators are giving by;

$$MSE(t_3) = \sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^4 \left[(\lambda_{40h} - 1) + (\lambda_{04h} - 1) - 2(\lambda_{22h} - 1) \right] \tag{2.9}$$

$$MSE(t_4) = \sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^4 \left[(\lambda_{40h} - 1) + R_{1h}^2 (\lambda_{04h} - 1) - 2R_{1h} (\lambda_{22h} - 1) \right] \tag{2.10}$$

$$MSE(t_5) = \sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^4 \left[(\lambda_{40h} - 1) + R_{2h}^2 (\lambda_{04h} - 1) - 2R_{2h} (\lambda_{22h} - 1) \right] \tag{2.11}$$

$$MSE(t_6) = \sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^4 \left[(\lambda_{40h} - 1) + Q_h^2 R_{2h}^2 (\lambda_{04h} - 1) - 2Q_h R_{2h} (\lambda_{22h} - 1) \right] \tag{2.12}$$

where $R_{1h} - \frac{S_{xh}^2}{S_{xh}^2 + C_{xh}}, R_{2h} - \frac{S_{xh}^2}{S_{xh}^2 + \beta_{Xh}}$ and $Q_h = \frac{(\lambda_{22h} - 1)}{R_{1h}(\lambda_{04h} - 1)}$

Shahzad et al. [26] defined the traditional variance estimator under double stratified random sampling based on traditional moment as

$$V_a = \sum_{h=1}^{L} W_h \, s_{yh}^2 \tag{2.13}$$

UNDER PEER REVIEW

Shahzad et al. [26] proposed L-moments based calibrated variance estimators as

$$V_{ai} = \sum_{h=1}^{L} \Phi_h \, s_{ymh}^2 \qquad (2.14)$$

Where Φ_h in the calibrated weight are selected to minimize the measure of chi-square distance

$$\left.\begin{aligned} &Min\, z = \sum_{h=1}^{L} \frac{(\Phi_h - W_h)}{W_h \theta_h} \\ &st. \sum_{h=1}^{L} \Phi_h = \sum_{h=1}^{L} W_h \\ &\sum_{h=1}^{L} \Phi_h \, D_{xmh} = \sum_{h=1}^{L} W_h D_{xmh}^{d} \end{aligned}\right\} \qquad (2.15)$$

Where $D_{xmh}^{(d)} = \left[\bar{x}_h^{(d)} = l_{1xl}, \, c_{xmh}^{(d)} = \frac{l_{2xl}^{(d)}}{l_{1xl}^{(d)}} \, , \;\; s_{xmh}^{2(d)} = l_{2xl}^{2(d)} \right]$ is the first stage L-location, L-cv, and L-variance.

$D_{xmh} = \left(\bar{x}_h = l_{1xl}, \, c_{xmh} = \frac{l_{2xl}}{l_{1xl}}, \, s_{xmh}^2 = l_{2xl}^2 \right)$ is the second stage L-location, L-cv and L-variance of X.

$$\begin{aligned} \Phi_h = W_h + W_h \, \theta_h & \left[\frac{ -\left(\sum_{h=1}^{K} W_h (D_{xmh}^{d} - D_{xmh}) \left(\sum_{h=1}^{L} W_h \theta_h \, D_{xmh} \right) \right) }{ \left(\sum_{h=1}^{L} W_h \theta_h D_{xmh}^2 \right) \left(\sum_{h=1}^{L} W_h \theta_h \right) - \left(\sum_{h=1}^{L} W_h \theta_h D_{xmh} \right)^2 } \right] \\ + W_h \theta_h D_{xmh} & \left[\frac{ \left(\sum_{h=1}^{L} W_h \, (D_{xmh}^{d} - D_{xmh}) \left(\sum_{h=1}^{L} W_h \theta_h \right) \right) }{ \left(\sum_{h=1}^{L} W_h \theta_h D_{xmh}^2 \right) \left(\sum_{h=1}^{L} W_h \, \theta_h \right) - \left(\sum_{h=1}^{L} W_h \theta_h D_{xmh} \right)^2 } \right] \end{aligned} \qquad (2.16)$$

The L-moment calibration variance estimator is given by

$$Vai = \sum_{h=1}^{L} W_h s_{ymh}^2 + \hat{\beta}_{cs} \sum_{h=1}^{L} W_h \left(D_{xmh}^{d} - D_{xmh} \right) \qquad (2.17)$$

where $\hat{\beta}_{cs} = \left[\frac{ \left(\sum_{h=1}^{L} W_h \theta_h \right) \left(\sum_{h=1}^{L} W_h \theta_h D_{xml} s_{yml}^2 \right) - \left(\sum_{h=1}^{L} W_h \theta_h D_{xml} \right) \left(\sum_{h=1}^{L} W_h \theta_h s_{yml}^2 \right) }{ \left(\sum_{h=1}^{L} W_h \theta_h D_{xml}^2 \right) \left(\sum_{h=1}^{L} W_h \theta_h \right) - \left(\sum_{h=1}^{L} W_h \theta_h D_{xml} \right)^2 } \right]$

3 Proposed Estimators

UNDER PEER REVIEW

Having studied the work of Shahzad et al. [26] estimators and point out the weakness of their work, the following class of estimators in the presence of measurement errors and non-response under two phase sampling stratified sampling were proposed

3.1 Class of proposed calibration Estimators

$$T_{(d)ai} = \sum_{h=1}^{L} \varphi_h^{*(ai)} s_{ymh(e)}^{*2} \tag{3.1}$$

$$\left.\begin{aligned} &\min Z = \frac{\sum_{h=1}^{L}\left(\varphi_h^{*(ai)} - W_h\right)^2}{\lambda_h W_h} \\ &\textit{subject to} \\ &\qquad \sum_{h=1}^{L} \varphi_h^{*(ai)} = \sum_{h=1}^{L} W_h \\ &\qquad \sum_{h=1}^{L} \varphi_h^{*(ai)} D_{xmh(e)}^{*} = \sum_{h=1}^{L} W_h D_{xmh}^{d} \end{aligned}\right\} \tag{3.2}$$

$$D_{xmh}^{d} = \left[\bar{x}_h^{(d)} = l_{1xl}^{(d)}, c_{xmh}^{(d)} = \frac{l_{2xl}^{(d)}}{l_{1xl}^{(d)}}, \; s_{xmh}^{2(d)} = l_{2xl}^{(d)} \right]$$

$$D_{xmh(e)}^{*} = \left[\bar{x}_{h(e)}^{*} = l_{1xl(e)}^{*}, c_{xmh(e)}^{*} = \frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}}, \; s_{xmh(e)}^{*2} = l_{2xl(e)}^{*2} \right]$$

$$s_{ymh(e)}^{*2} = \frac{(n_{1h}-1)s_{ymh(e)}^2 + n_{2h}s_{yh2mh(e)}^2}{n_{1h}+n_{2h}-1}, \quad s_{xmh(e)}^{*2} = \frac{(n_{1h}-1)s_{xmh(e)}^2 + n_{2h}s_{xh2mh(e)}^2}{n_{1h}+n_{2h}-1}$$

$$\vec{y}_{(e)h}^{*} = \frac{n_{1h}\bar{y}_{1(e)h} + n_{2h}\bar{y}_{h2(e)h}}{n_{1h}+n_{2h}}, \quad \bar{x}_{(e)h}^{*} = \frac{n_{1h}\bar{x}_{1h(e)} + n_{2h}\bar{x}_{h2(e)h}}{n_{1h}+n_{2h}}$$

$D_{xmh(e)}^{*}$ *and* D_{xmh}^{d} are the sample and population characteristics of the auxiliary variable in the second and the first stage.

The biases of the estimator $\hat{T}_{(d)i}$ will be obtained using function in (3.3)

$$Bias(T) = 2^{-1}\left[\sum_{i=1}^{q}\sum_{j=1}^{q} D_{ij(h)}\, E\left(\hat{\theta}_{i(h)} - \theta_{i(h)}\right) E\left(\hat{\theta}_{j(h)} - \theta_{j(h)}\right)\right] \qquad (3.3)$$

Where q is the number of sample variances in the estimators, q=2

$$\theta_{1h} = s^{*2}_{y(e)mh},\ \theta_{2(h)} = s^{*2}_{x(e)mh}\ \theta_{1(h)} = S^{2}_{y(h)}\ \theta_{2(h)} = S^{2}_{x(h)}$$

$$\Delta_{ij} = \frac{\partial^2 T}{\partial\theta_{i(h)}\,\partial\theta_{j(h)}} \Bigg/ S^{2}_{y(h)}, S^{2}_{x(h)}$$

The MSEs of the estimators will be obtained using function in (3.4)

$$\text{MSE(T)} = \Delta_h \sum \Delta_h^T \qquad (3.4)$$

Where $\Delta_h = \left[\frac{\partial T}{\partial s^{*2}_{y(e)mh}}\ \frac{\partial T}{\partial s^{*2}_{x(e)mh}}\right] S^{2}_{y(h)}, S^{2}_{x(h)}, B_{rg}$

That is , $S^{2}_{y(h)}, S^{2}_{x(h)}, Brg$ are substituted for $s^{*2}_{y(e)mh}$ *and* $B_{rg(e)mh}$ *and* $s^{*2}_{x(e)mh}$

$$\sum = \begin{bmatrix} Var\left(s^{*2}_{y(e)mh}\right) & Cov\left(s^{*2}_{y(e)mh,}\ s^{*2}_{x(e)mh}\right) \\ Cov\left(s^{*2}_{x(e)mh,}\ s^{*2}_{y(e)mh}\right) & Var\left(s^{*2}_{x(e)mh}\right) \end{bmatrix}$$

$$Var(s^{*2}_{y(e)mh}) = \sum_{h=1}^{L} \frac{(W_h\, S_{y(h)})^4}{n^3{}_h}\left[K_{1(h)}\, H_{y(h)} + K_{2(h)}\, H_{y(h_2)h}\right]$$

$$H_{y(h)} = \lambda_{40(h)} + \gamma_{40(h)}\, S^{4}_{u(h)} S^{-4}_{y(h)} + 2\left(1 + S^{2}_{u(h)}\, S^{-2}_{y(h)}\right)^2,$$

$$H_{y(2)h} = \lambda_{40(2)h} + \gamma_{40(2)h}\, S^{4}_{u(2)h} S^{-4}_{y(2)h} + 2\left(1 + S^{2}_{u(2)h}\, S^{-2}_{y(2)h}\right)^2$$

$$Var\left(s^{*2}_{x(e)mh}\right) = \sum_{h=1}^{L} \frac{(W_h\, S_{y(h)})^4}{n^3{}_h}\left[K_{1(h)}\, H_{y(h)} + K_{2(h)}\, H_{y(h_2)h}\right]$$

$$H_{x(h)} = \lambda_{04(h)} + \gamma_{40(h)}\, S^{4}_{v(h)} S^{-4}_{x(h)} + 2\left(1 + S^{2}_{v(h)}\, S^{-2}_{x(h)}\right)^2$$

UNDER PEER REVIEW

$$H_{x(2)h} = \lambda_{04(2)h} + \gamma_{04(2)h}\, S^{4}_{v(2)h} S^{-4}_{x(2)h} + 2\left(1 + S^{2}_{v(2)h}\, S^{-2}_{x(2)h}\right)^{2}$$

$$Cov\left(s^{*2}_{y(e)mh},\, s^{*2}_{x(e)mh}\right) = \sum_{h=1}^{L}\left(K_{1h}\, \lambda_{22(h)} + K_{2(h)}\, \lambda_{22(2)h}\right)$$

$$K_{1(h)} = \left(n_h^{-1} - N_h^{-1}\right) \quad K_{2(h)} = \frac{W_{2h}\left(f_h - 1\right)}{n_h}, \quad W_{2h} = \left[\frac{N_2}{N}\right]_h, \quad f_h = \frac{n_h}{N_h}$$

$$Cov\left(s^{*2}_{x(e)mh},\, \bar{x}^{*}_{(e)mh}\right) = \sum_{h=1}^{L}\frac{(W_h\, S_{y(h)})^4}{n^3{}_h}\left(K_{1h}\, \lambda_{12(h)}\, C_{x(h)} + K_{2(h)}\, \lambda_{12(2)h} C_{x(2)h}\right)$$

$$Cov\left(s^{*2}_{y(e)mh},\, \bar{y}^{*}_{(e)mh}\right) = \sum_{h=1}^{L}\frac{(W_h\, S_{y(h)})^4}{n^3{}_h}\left(K_{1h}\, \lambda_{12(h)}\, C_{y(h)} + K_{2(h)}\, \lambda_{12(2)h} C_{y(2)h}\right)$$

$$Var\left(\bar{x}^{*}_{(e)mh}\right) = \sum_{h=1}^{L}\frac{(W_h\, S_{y(h)})^4}{n^3{}_h}\left(K_{1h} C^2_{xh}\left(1 + S^2_{v(h)} S^{-2}_{x(h)}\right) + K_{2(h)}\, C^2_{x(2)h}\left(1 + S^2_{v(2)h} S^{-2}_{x(2)h}\right)\right)$$

To obtain the calibration weight of the estimators and properties of the estimators $T_{d(ai)}$, we define the Lagrange function as

$$L_{ai} = \sum_{h=1}^{L}\frac{\left(\phi_h^{*(ai)} - W_h\right)^2}{2\lambda_h W_h} - g_1\left(\sum_{h=1}^{L}\phi_h^{*(ai)} - \sum_{h=1}^{L} W_h\right) - g_2\left(\sum_{h=1}^{L}\phi_h^{*(ai)} D^{*}_{xmh(e)} - \sum_{h=1}^{L} W_h D^{d}_{xmh}\right) \tag{3.5}$$

Where g_1 and g_2 are Lagrange's multipliers, Differentiate (3.5) partially with respect to $\phi_h^{*(ai)}$, g_1 and g_2 respectively and equate to zero to obtain (3.26), (4.3) and (4.4) after simplification.

$$\phi_h^{*(ai)} = W_h + g_1 \lambda_h W_h + g_2 \lambda_h W_h D^{*}_{xmh(e)} \tag{3.6}$$

$$\sum_{h=1}^{L}\phi_h^{*(ai)} = \sum_{h=1}^{L} W_h \tag{3.7}$$

UNDER PEER REVIEW

$$\sum_{h=1}^{L}\phi_h^{*(ai)}D^{*}_{xmh(e)}=\sum_{h=1}^{L}W_h\Delta^{d}_{xmh(h)} \tag{3.8}$$

Substitute (3.6) into (3.7) and (3.8) and simplify to generate two simultaneous equations in (3.9) as

$$\begin{pmatrix}\sum_{h=1}^{L}\lambda_h W_h & \sum_{h=1}^{L}\lambda_h W_h D^{*}_{xmh(e)} \\ \sum_{h=1}^{L}\lambda_h W_h D^{*}_{xmh(e)} & \sum_{h=1}^{L}\lambda_h W_h D^{*2}_{xmh(e)}\end{pmatrix}\begin{pmatrix}g_1\\ g_2\end{pmatrix}=\begin{pmatrix}0\\ \sum_{h=1}^{L}W_h\left(D^{d}_{xmh}-D^{*}_{xmh(e)}\right)\end{pmatrix} \tag{3.9}$$

Solving equations (3.9) simultaneously, the results obtained are,

$$g_1=\frac{-\sum_{h=1}^{L}W_h\lambda_h D^{*}_{xmh(e)}\left(\sum_{h=1}^{L}W_h D^{d}_{xmh}-\sum_{h=1}^{L}W_h D^{*}_{xmh(e)}\right)}{\left(\sum_{h=1}^{L}W_h\lambda_h\right)\left(\sum_{h=1}^{L}W_h\lambda_h D^{*2}_{xmh(e)}\right)-\left(\sum_{h=1}^{L}\lambda_h W_h D^{*}_{xmh(e)}\right)^2} \tag{3.10}$$

$$g_2=\frac{\left(\sum_{h=1}^{L}\lambda_h W_h\right)\left(\sum_{h=1}^{L}W_h D^{d}_{xmh}-\sum_{h=1}^{L}W_h D^{*}_{xmh(e)}\right)}{\left(\sum_{h=1}^{L}W_h\lambda_h\right)\left(\sum_{h=1}^{L}W_h\lambda_h D^{*2}_{xmh(e)}\right)-\left(\sum_{h=1}^{L}{}_h\,\lambda_h W_h D^{*}_{xmh(e)}\right)^2} \tag{3.11}$$

Substituting (3.10) and (3.11) into (3.6) and simplify, we obtained the calibration weights as,

$$\phi_h^{*(ai)}=W_h+\lambda_h W_h\frac{\left[D^{*}_{xmh(e)}\sum_{h=1}^{L}W_h\lambda_h-\sum_{h=1}^{L}W_h\lambda_h D^{*}_{xmh(e)}\right]\left[\sum_{h=1}^{L}W_h\left(D^{d}_{xmh}-D^{*}_{xmh(e)}\right)\right]}{\left[\left(\sum_{h=1}^{L}W_h\lambda_h D^{*2}_{xmh(e)}\right)\left(\sum_{h=1}^{L}W_h\lambda_h\right)-\left(\sum_{h=1}^{L}W_h\lambda_h D^{*}_{xmh(e)}\right)^2\right]} \tag{3.12}$$

By substituting (3.12) into calibration schemes defined in (3.1), we obtained the proposed estimators $T_{(d)ai,}$ i= 1, 2, 3, …,10 as

UNDER PEER REVIEW

$$T_{(d)ai} = \sum_{h=1}^{L} W_h s_{ymh(e)}^{*2} + \pi_i \sum_{h=1}^{L} W_h \left(D_{xmh}^{d} - D_{xmh(e)}^{*} \right) \tag{3.13}$$

$$Where, \pi_i = \frac{\left(\sum_{h=1}^{L} W_h \lambda_h D_{xmh(e)}^{*} s_{ymh(e)}^{*2}\right)\left(\sum_{h=1}^{L} W_h \lambda_h\right) - \left(\sum_{h=1}^{L} W_h \lambda_h s_{ymh(e)}^{*2}\right)\left(\sum_{h=1}^{L} W_h \lambda_h D_{xmh(e)}^{*}\right)}{\left(\sum_{h=1}^{L} W_h \lambda_h D_{xmh(e)}^{*2}\right)\left(\sum_{h=1}^{L} W_h \lambda_h\right) - \left(\sum_{h=1}^{L} W_h \lambda_h D_{xmh(e)}^{*}\right)^2}$$

Table 1. Members of the first proposed estimators T$_{d(ai)}$,

I	$T_{(d)ai}$	D_{xmh}^{d}	$D_{xmh(e)}^{*}$	Estimators
1	1	l_{1xl}^{d}	$l_{1xl(e)}^{*}$	$T_{(d)a1} = \sum_{h=1}^{L} W_h \left(s_{ymh(e)}^{*2} + \pi_1 \left(l_{1xl}^{d} - l_{1xl(e)}^{*} \right) \right)$
2	$\left(l_{1xl(e)}^{*}\right)^{-1}$	l_{1xl}^{d}	$l_{1xl(e)}^{*}$	$T_{(d)a2} = \sum_{h=1}^{L} W_h \left(s_{ymh(e)}^{*2} + \pi_2 \left(l_{1xl}^{d} - l_{1xl(e)}^{*} \right) \right)$
3	$\left(l_{2xl(e)}^{*}\right)^{-1}$	l_{1xl}^{d}	$l_{1xl(e)}^{*}$	$T_{(d)a3} = \sum_{h=1}^{L} W_h \left(s_{ymh(e)}^{*2} + \pi_3 \left(l_{1xl}^{d} - l_{1xl(e)}^{*} \right) \right)$
4	$\left(l_{2xl(e)}^{*2}\right)^{-1}$	l_{1xl}^{d}	$l_{1xl(e)}^{*}$	$T_{(d)a4} = \sum_{h=1}^{L} W_h \left(s_{ymh(e)}^{*2} + \pi_4 \left(l_{1xl}^{d} - l_{1xl(e)}^{*} \right) \right)$
5	$\left(\frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}}\right)^{-1}$	l_{1xl}^{d}	$l_{1xl(e)}^{*}$	$T_{(d)a5} = \sum_{h=1}^{L} W_h \left(s_{ymh(e)}^{*2} + \pi_5 \left(l_{1xl}^{d} - l_{1xl(e)}^{*} \right) \right)$
6	1	$\left(\frac{l_{2xl}^{d}}{l_{1xl}^{d}}\right)$	$\left(\frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}}\right)$	$T_{(d)a6} = \sum_{h=1}^{L} W_h \left(s_{ymh(e)}^{*2} + \pi_6 \left(\frac{l_{2xl}^{d}}{l_{1xl}^{d}} - \frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}} \right) \right)$
7	$\left(l_{1xl(e)}^{*}\right)^{-1}$	$\left(\frac{l_{2xl}^{d}}{l_{1xl}^{d}}\right)$	$\left(\frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}}\right)$	$T_{(d)a7} = \sum_{h=1}^{L} W_h \left(s_{ymh(e)}^{*2} + \pi_7 \left(\frac{l_{2xl}^{d}}{l_{1xl}^{d}} - \frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}} \right) \right)$
8	$\left(l_{2xl(e)}^{*}\right)^{-1}$	$\left(\frac{l_{2xl}^{d}}{l_{1xl}^{d}}\right)$	$\left(\frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}}\right)$	$T_{(d)a8} = \sum_{h=1}^{L} W_h \left(s_{ymh(e)}^{*2} + \pi_8 \left(\frac{l_{2xl}^{d}}{l_{1xl}^{d}} - \frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}} \right) \right)$
9	$\left(l_{2xl(e)}^{*2}\right)^{-1}$	$\left(\frac{l_{2xl}^{d}}{l_{1xl}^{d}}\right)$	$\left(\frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}}\right)$	$T_{(d)a9} = \sum_{h=1}^{L} W_h \left(s_{ymh(e)}^{*2} + \pi_9 \left(\frac{l_{2xl}^{d}}{l_{1xl}^{d}} - \frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}} \right) \right)$
10	$\left(\frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}}\right)^{-1}$	$\left(\frac{l_{2xl}^{d}}{l_{1xl}^{d}}\right)$	$\left(\frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}}\right)$	$T_{(d)a10} = \sum_{h=1}^{L} W_h \left(s_{ymh(e)}^{*2} + \pi_{10} \left(\frac{l_{2xl}^{d}}{l_{1xl}^{d}} - \frac{l_{2xl(e)}^{*}}{l_{1xl(e)}^{*}} \right) \right)$

UNDER PEER REVIEW

The expected value of an estimator minus the parameter is called Bias. To obtain the bias of the estimators $T_{d(ai)}$, we take expectation of (4.9), we have

$$E\left(T_{(d)ai}\right)=\sum W_h E\left(s^{*2}_{ymh(e)}\right)+\pi_i \sum W_h\left(D^d_{xmh}-E\left(D^*_{xmh(e)}\right)\right) \qquad (3.14)$$

$$Since, E\left(D^*_{xmh(e)}\right)=D^d_{xmh}$$

$$E\left(T_{(d)ai}\right)=\sum_{h=1}^{L} W_h S^2_{yh}+\pi_i \sum_{h=1}^{L} W_h\left(D^d_{xmh}-D^d_{xmh}\right) \qquad (3.15)$$

$$E\left(T_{(d)ai}\right)=\sum_{h=1}^{L} W_h S^2_{yh} \qquad (3.16)$$

Subtract S^2_y from both sides, we have

$$E\left(T_{(d)ai}\right)-S^2_{yh}=\sum_{h=1}^{L} W_h S^2_{yh}-S^2_y \qquad (3.17)$$

$$Bias\left(T_{(d)ai}\right)=S^2_y-S^2_y \qquad (3.18)$$

$$Bias\left(T_{(d)ai}\right)=0 \qquad (3.19)$$

UNDER PEER REVIEW

The biases of the proposed estimators are zero, this shows that they are unbiased estimators

Differentiating $T_{(d)ai}$ i = 1, 2,…,10 partially with respect to $s^{*2}_{ymh(e)}$ and $D^*_{xmh(e)}$, we obtained

$$\frac{\partial T_{d(ai)}}{\partial s^{*2}_{ymh(e)}}=\sum_{h=1}^{L} W_h=1 \qquad (3.20)$$

$$\frac{\partial T_{d(ai)}}{\partial D^{*}_{xmh(e)}} = -\pi_i \sum_{h=1}^{L} W_h = -\pi_i \qquad (3.21)$$

Using the expression in chapter three, we have that

$$\Delta_h = \begin{bmatrix} 1 - \pi_i \end{bmatrix} \quad \Delta_h^T = \begin{bmatrix} 1 \\ -\pi_i \end{bmatrix} \qquad (3.22)$$

The mean square errors of the estimators $T_{(d)ai}$, i = 1, 2,…,15. are obtained as

$$MSE\left(T_{(d)ai}\right) = \Delta_h \sum \ \Delta_h^T \qquad (3.23)$$

$$MSE\left(T_{(d)ai}\right) = \begin{bmatrix} 1 & -\pi_i \end{bmatrix} \begin{bmatrix} Var\left(s^{*2}_{ymh(e)}\right) & Cov\left(s^{*2}_{ymh(e)} \quad D^{*}_{xmh(e)}\right) \\ Cov\left(D^{*}_{xmh(e)} \quad s^{*2}_{ymh(e)}\right) & Var\left(D^{*}_{xmh(e)}\right) \end{bmatrix} \begin{bmatrix} 1 \\ -\pi_i \end{bmatrix} \qquad (3.24)$$

$$MSE\left(T_{(d)ai}\right) = Var\left(s^{*2}_{ymh(e)}\right) - 2\pi_i Cov\left(s^{*2}_{ymh(e)h} \quad D^{*}_{xmh(e)}\right) + \pi_i^2 Var\left(D^{*}_{xmh(e)}\right) \qquad (3.25)$$

4.4 Efficiency Comparison

In this section, conditions for the efficiency of the new estimators over existing estimators under double stratified sampling are established.

i. The proposed estimators $T_{d(ai)}$ are more efficient than Sample variance estimator and Ozel et al. (2014) estimators if

$$MSE\left(T_{(d)ai}\right) < MSE\left(V_a\right) \quad i = 1, 2, ..., 10 \qquad (4.1)$$

$$MSE\left(T_{(d)ai}\right) < MSE\left(t_j\right) \quad i = 1, 2, ..., 10,\ j = 3, 4, 5, 6 \qquad (4.2)$$

Then,

$$\begin{bmatrix} Var\left(s^{*2}_{ymh(e)}\right) - 2\pi_i Cov\left(s^{*2}_{ymh(e)}, \quad D^{*}_{xmh(e)h}\right) \\ +\pi_i^2 Var\left(D^{*}_{xmh(e)h}\right) \end{bmatrix} < \sum_{h=1}^{L} W_h{}^2 S_{yh}^4 \left(\lambda_{40h} - 1\right) \tag{4.3}$$

$$\begin{bmatrix} Var\left(s^{*2}_{ymh(e)}\right) - 2\pi_i Cov\left(s^{*2}_{ymh(e)}, \quad D^{*}_{xmh(e)h}\right) \\ +\pi_i^2 Var\left(D^{*}_{xmh(e)h}\right) \end{bmatrix} < \sum_{h=1}^{L} W_h{}^2 S_{yh}^4 \begin{bmatrix} \left(\lambda_{40h} - 1\right) + \left(\lambda_{04h} - 1\right) \\ -2\left(\lambda_{22h} - 1\right) \end{bmatrix} \tag{4.4}$$

$$\begin{bmatrix} Var\left(s^{*2}_{ymh(e)}\right) - 2\pi_i Cov\left(s^{*2}_{ymh(e)}, \quad D^{*}_{xmh(e)h}\right) \\ +\pi_i^2 Var\left(D^{*}_{xmh(e)h}\right) \end{bmatrix} < \sum_{h=1}^{L} W_h{}^2 S_{yh}^4 \begin{bmatrix} \left(\lambda_{40h} - 1\right) + R_{1h}^2\left(\lambda_{04h} - 1\right) \\ -2R_{1h}\left(\lambda_{22h} - 1\right) \end{bmatrix} \tag{4.5}$$

$$\begin{bmatrix} Var\left(s^{*2}_{ymh(e)}\right) - 2\pi_i Cov\left(s^{*2}_{ymh(e)}, \quad D^{*}_{xmh(e)h}\right) \\ +\pi_i^2 Var\left(D^{*}_{xmh(e)h}\right) \end{bmatrix} < \sum_{h=1}^{L} W_h{}^2 S_{yh}^4 \begin{bmatrix} \left(\lambda_{40h} - 1\right) + R_{2h}^2\left(\lambda_{04h} - 1\right) \\ -2R_{2h}\left(\lambda_{22h} - 1\right) \end{bmatrix} \tag{4.6}$$

$$\begin{bmatrix} Var\left(s^{*2}_{ymh(e)}\right) - 2\pi_i Cov\left(s^{*2}_{ymh(e)}, \quad D^{*}_{xmh(e)h}\right) \\ +\pi_i^2 Var\left(D^{*}_{xmh(e)h}\right) \end{bmatrix} < \sum_{h=1}^{L} W_h{}^2 S_{yh}^4 \begin{bmatrix} \left(\lambda_{40h} - 1\right) + Q_h^2 R_{1h}^2\left(\lambda_{04h} - 1\right) \\ -2Q_h R_{1h}\left(\lambda_{22h} - 1\right) \end{bmatrix} \tag{4.7}$$

4.5 Empirical study using simulated Data

In this section, simulation studies to assess the performance of the proposed estimators $s^{*2}_{xmh(e)}$ and $s^{*2}_{xmh(e)}$ with respect to existing estimators were conducted. Data of size 1000 units were generated for study population using functions defined in Table 4.3, sample of size 100 were selected by method of Simple random sampling without replacement (SRSWOR) 1000 times. The biases, MSEs and PREs of the considered estimators were computed using (4.8), (4.9) and (4.10).

$$Bias\left(T\right) = \frac{1}{1000}\sum_{k=1}^{1000}\left(T - S_y^2\right) \tag{4.8}$$

$$MSE\left(T\right) = \frac{1}{1000}\sum_{k=1}^{1000}\left(T - S_y^2\right)^2 \tag{4.9}$$

UNDER PEER REVIEW

$$PRE(T)=\frac{MSE(t_1)}{MSE(T)}\times 100 \qquad (4.10)$$

Where T are any of the proposed or existing estimators.

Table 2. Population used for Simulation Study

Populations	**Auxiliary Variable (X)**	**Study Variable (Y)**
1	$X_h \sim N(N_h, \mu_h, \sigma_h)$ $\mu_1 = 10, \sigma_1 = 40, \mu_2 = 30, \sigma_2 = 70,$ $\mu_3 = 30, \sigma_3 = 50$	
2	$X_h \sim beta(N_h, b_h, c_h)$ $b_1 = 1.1, c_1 = 2, b_2 = 1.2, c_2 = 3,$ $b_3 = 1.3, c_3 = 5$	$Y_h = 0.5X_h + 0.5X_h^2 + e_h$ $Where, e_h \sim N(0,1)$
3	$X_h \sim weibull(N_h, z_h, s_h)$ $z_1 = 1.8, s_1 = 4, z_2 = 2.2, s_2 = 3,$ $z_3 = 1.3, s_3 = 5$	
4	$X_h \sim pois(N_h, \Theta_h)$ $\Theta_1 = 1.8, \Theta_h = 2.2, \Theta_h = 1.3$	

UNDER PEER REVIEW

TABLE 3. Biases, MSEs and PREs of the proposed and existing estimators using population 1 data

Estimators	**Biases**	**MSEs**	**PREs**
Sample variance V_a	-6761673	4.572022e+13	100
Ozel et al. (2014)			
t_3	-7837635.9	6.142854e+13	74.429
t_4	-7932256.3	6.292069e+13	72.663
t_5	-7941887.4	6.315348e+13	71.325

t_6	-6781541.2	4.598921e+13	99.412
Proposed estimators			
T_{a1}	-837635.9	1.120984e+12	4078.579
T_{a2}	-837656.1	1.121018e+12	4078.456
T_{a3}	-840887.7	1.126442e+12	4058.816
T_{a4}	59772154	3.57313e+15	1.279557
T_{a5}	-837636	1.120984e+12	4078.579
T_{a6}	-833775.2	1.114531e+12	4102.193
T_{a7}	-833276.2	1.113699e+12	4105.257
T_{a8}	-840903.5	1.126469e+12	4058.721
T_{a9}	2799693776	7.838286e+18	0.0005832936
T_{a10}	-833701.5	1.114408e+12	4102.646

TABLE 4. Biases, MSEs and PREs of the proposed and existing estimators using population 2 data

Estimators	**Biases**	**MSEs**	**PREs**
Sample variance V_a	9276.539	86054180	100
Ozel et al. (2014)			
t_3	9537.9	89865432.39	95.762
t_4	9783.3	94653421.32	91.483
t_5	9818.2	95654323.41	90.472
t_6	9291.54	86331021.93	99.677
Proposed estimators			
T_{a1}	3272.078	10706493	803.757
T_{a2}	3251.958	10575233	813.7332
T_{a3}	20.32421	413.0742	20832619
T_{a4}	60613062	3.673943e+15	2.342284e-06
T_{a5}	3272.006	10706022	803.7923
T_{a6}	7132.817	50877077	169.1414
T_{a7}	7631.87	58245443	147.7441
T_{a8}	4.540115	20.61355	417464055
T_{a9}	2800534685	7.842995e+18	1.097211e-09
T_{a10}	7206.539	51934207	165.6985

TABLE 5. Biases, MSEs and PREs of the proposed and existing estimators using population 3 data

Estimators	**Biases**	**MSEs**	**PREs**
Sample variance V_a	9113.095	83048501	100
Ozel et al. (2014)			
t_3	9371.242	87821341.4	94.5711
t_4	9451.161	89324001.32	92.9723

UNDER PEER REVIEW

t_5	9663.112	93265210.47	89.0432
t_6	9118.312	83144223.49	99.8851
Proposed estimators			
T_{a1}	3260.769	10632899	781.0523
T_{a2}	3240.65	10502094	790.7804
T_{a3}	9.015919	363.3342	22857330
T_{a4}	60613050	3.673942e+15	2.260474e-06
T_{a5}	3260.697	10632430	781.0867
T_{a6}	7121.509	50716167	163.7515
T_{a7}	7620.562	58073246	143.0065
T_{a8}	-6.768171	327.8556	25330820
T_{a9}	2800534673	7.842994e+18	1.058888e-09
T_{a10}	7195.231	51771630	160.4131

TABLE 6. Biases, MSEs and PREs of the proposed and existing estimators using population 4 data

Estimators	**Biases**	**MSEs**	**PREs**
Sample variance V_a	9261.88	85782426	100
Ozel et al. (2014)			
t_3	9511.45	90043284.22	95.2645
t_4	9731.33	93732115.25	91.5082
t_5	9944.91	98501251.89	87.0566
t_6	9284.18	86194834.92	99.522
Proposed estimators			
T_{a1}	3269.786	10691499	802.3424
T_{a2}	3249.666	10560331	812.3081
T_{a3}	18.41932	340.1172	25221429
T_{a4}	57061931	3.256084e+15	2.634528e-06
T_{a5}	3267.836	10678753	803.3
T_{a6}	5953.661	35500590	241.6366
T_{a7}	6085.703	37096624	231.2405
T_{a8}	1.452488	2.982374	2876313588
T_{a9}	1432584139	2.068194e+18	4.147697e-09
T_{a10}	5964.567	35630744	240.754

5 Results and Discussion

Table 3-6 shows the biases, MSEs and PREs for various existing and the proposed estimators under the simultaneous influence of non-response and measurement errors using the four simulated populations as defined in Table 2. The findings indicate that with the exception of estimators T_{a9}, all other proposed estimators are more efficient than the traditional variance estimator V_a, by Shahzad et al. [26], Ozel et al. [10], t_3, t_4, t_5, and t_6 with evidence of minimum

UNDER PEER REVIEW

mean square errors and higher percentage relative efficiencies. Hence proposed estimators, are highly efficient.

6 Conclusion

In the current study, we have suggested modified variance estimators in the presence of non-response and measurement errors for estimation of population variance under stratified random sampling scheme. From the empirical results, the results showed that the proposed estimators were more efficient than the existing ones considered in the study. Hence, we recommend the proposed estimators for theoretical and real life applications.

References

[1] Arnab R, Singh S. A note on variance estimation for the generalized regression predictor. Australian and New Zealand journal of statistics. 2005;47(2):231-234.

[2] Audu A, Singh RVK. Improved exponential type estimators for estimating population variance in survey sampling. International journal of advanced research. 2015;3(4):1-16.

[3] Audu A, Adewara AA, Singh RVK. Improved class of ratio estimators for finite population variance. Global journal of Science frontier research for Mathematics and Decision Sciences. 2016;16(2):23117-24.

[4] Cekim HO, Kadilar C. Ln type estimator for the population variance in stratified random sampling. Communication in Statistics-Simulation and Computation. 2019;1-13.

[5] Das AK, Tripathi TP. Use of auxiliary information in estimating the finite population variance. Sankhya. 1978; 40:139-148.

[6] Isaki CT. Variance estimation using auxiliary information. Journal of Statistical Association. 1983; 78:117-123.

[7] Kadilar C, Cingi H. Improvent of in variance estimation using auxiliary information. Journal of Mathematics and Statistics. 1983; 35:111-115.

[8] Adejumobi A, Yunusa MA. An enhanced variance estimator with power transformation. Discovery. 2023; 59:1-5.

[9] Adejumobi A, Yunusa MA. Some improved class of ratio estimators for finite population variance with the use of known parameters. LC International Journal of STEM. 2022; 3(3)35-45.

UNDER PEER REVIEW

[10] Ozel G, Cingi H, Ogus, M. Separate ratio estimators for the population variance in stratified random sampling. Communication in Statistics-Theory and Methods. 2014; 43(2):4766-4779.

[11] Singh R, Chauhan P, Sawan, N, Smarandache, F. Improved exponential estimator for population variance using two auxiliary variables. Italian Journal of Pure and Applied Mathematics. 2011; 28:101-108.

[12] Upadhyaya LN, Singh HP. Use of auxiliary information in the estimation of population variance. Mathematical Forum. 1999; 4:33-36.

[13] Yadav SK, Mushram SS, Khanal, SP. Variance estimator using tri-mean and interquartile range of auxiliary variable. Nepalese Journal of Statistics. 2017; 1:83-91.

[14] Yunusa MA, Audu A, Muili, JO. On the Arithmetic mean estimators of a family of estimators of finite populationof ratio type. Oriental Journal of Physical Sciences. 2022; 7(2):67-73.

[15] Yunusa MA, Adejumobi A, Audu, A. A modified ratio estimator in the presence of tri-mean and interquartile range for the estimation of population variance. LC International Journal of STEM. 2022; 3(4):40-50.

[16] Hansen MN, Hurwitz WN. The problem of non-response in sample surveys. Journal of the American Statistical Association. 1946; 41:517-529.

[17] Cochran WG. Errors of measurement in statistics. Technometric. 1968; 10:637-666.

[18] Misra S, Dipika K, Yadav DK. Some improved estimators for estimating population variance in the presence of measurement error. Journal of Statistics Application and Probability. 2016; 5(2):311-320.

[19] Audu A, Singh RVK, Ishaq OO, Khare S, Singh R, Adewara AA. On the estimation of finite population variance for a mail survey design in the presence of non-response using new conventional and calibration estimators. Communication in Statistics-Theory and Methods. 2016; 53(3):848-864.

[20] Audu A, Singh R, Khare S, Dauran NS. Almost unbiased variance estimators under the simultaneous influenceof non-response and measurement errors. Journal of Statistical Theory and Practice. 2016; 16(15):1-14.

[21] Deville JC, Sarndal CE. Calibration estimators in survey sampling. Journal of the American Statistical Association. 1992; 49(4):437-440.

[22] Estevao VM, Sarndal CE. Survey estimates by calibration on complex auxiliary information. International Statistical Review. 2006; 74:127-147.

UNDER PEER REVIEW

[23] Kim JK, Park M. Calibration estimation in survey sampling. International Statistical Review. 2010; 78(1):21-39.

[24] Koyuncu N, Kadilar C. Calibration weighting in stratified random sampling. Communication in Statistics Simulation and Computation. 2016; 2267-2275.

[25] Audu A, Singh R, Khare S. Developing calibration estimators for population mean using robust measures of dispersion under stratified random sampling. Statistics in Transition New Series. 2021; 22(2):125-142.

[26] Shahzad U, Ahmad I, Almanjahie IM, Al-Noor N. L-moment calibration variance estimators using double stratified sampling. Computers,Materials and Continua. 2021; 68(3):3411-3430.

[27] Prasad B, Singh HP. Unbiased estimators of finite population variance using auxiliary information in sample surveys. Communication in Statistics. 1992; 21(5):1367-1376.

UNDER PEER REVIEW