Abstract

In this study, we have proposed a new generalized estimator using auxiliary information for the estimation of population variance in presence and absence of measurement error. The expressions approximate bias and mean square error of the proposed generalized estimator are derive up to the first-order. Several new and existing estimators are found as the special cases of proposed estimator and expressed on various values of optimizing and generalized constants. The proposed estimator is compared mathematically with some existing estimators under certain conditions. The performance of proposed estimator is observed by simulation and real data application under different sample sizes. It is observed that the proposed estimator performed well than other competing estimators in the presence and absence of measurement error.

KEYWORDS: Auxiliary Information, Measurement Errors, Variance Estimation, Mean Square Error and Percentage Relative Efficiency.

1. Introduction

A good understanding about the variation is vitally important for obtaining the efficient results in different walks of life. In sample surveys, we need the efficient estimators for the estimation of different characteristic, such as population mean, total, variance, etc. In literature, various authors have proposed many efficient estimators of population parameters based on the auxiliary information. Variance estimation of population is very important issue where variability control is difficult in its application. The main interest of researchers is to control the
variation in objects/subjects, such as; in health matters, body temperature, blood pressure and pulse rate are the basic diagnosis monitors when the treatment is designed to control the variation. In agricultural fields, sufficient knowledge about the climatic variation, rainfall and area is required to formulate suitable plan for cultivating a crop. In industries, a manufacturer requires regular information of variation in the reaction of people’s behavior towards his product so that he will be able to know whether to improve the quality of the product or to increase or decrease the price of the product.

The estimation problem of population variance is to measure the variability in study variable which received a significant interest of the statisticians in survey sampling. [1] discussed the ratio and regression estimators to estimate finite population variance. [2-4] suggested exponential ratio-type, exponential product-type and generalized exponential estimators of population variance. [5] recommended ratio-cum-dual type estimator when the population parameters of auxiliary variable are known for the estimation of population variance. Recently, the problem of the variance estimation has discussed by various authors such as; [6] proposed population variance estimator using two auxiliary variable and generalized class of estimator under simple random sampling without replacement (SRSWOR). [7] proposed chain ratio-type and chain ratio-type exponential estimators along with their generalized version. The properties of the proposed estimators were also discussed and the conditions are also obtained in which proposed estimators performed well than the existing estimators. [8] suggested a new generalized ratio-product-type estimator for population variance of study variable utilizing the information obtained from two auxiliary variables. [9] developed an estimator for population variance by introducing the linear combination of coefficient of kurtosis and decile mean of the auxiliary variable to achieve the efficiency of the proposed estimator. [5] concerned with the estimation of population variance of the study variable using tri-mean and third quartile of the auxiliary variable. [10] proposed a class of ratio estimators for the estimation of population variance of the study variable by using the coefficient of quartile deviation of auxiliary variable. [11] considered the inter quartile range and population correlation coefficient of the auxiliary variable for the populations of different characteristics.

A very common assumption in statistical data is: the values observed during the survey are correctly measure for their corresponding true values and there is no error in collected observations. But this assumption in practice does not met usually. Generally different types of errors contaminated the data due to many inevitable causes like respondents non response, faulty questionnaires preparation, flawed collection of sampling units, interview techniques inappropriateness or/and combination of some or all these. “Measurement errors ($MEs$) are defined as the difference between the observations obtained during the survey and true observations of the study variable, [12]”. $MEs$ include observational error, instrument error, respondent error, etc. It has severe effects on the population parameters estimation in terms of increase in bias and variation. Thus, it is essential to study the impact of measurement error to
develop better estimation techniques and to get more reliable and efficient estimates of the parameters in the presence of measurement errors.

Many authors have used the auxiliary information in the presence of MEs to estimate population mean, which includes [13-19] have studied the effects of MEs on estimation of population parameters. Many researchers, like [20-25] contributed in the variance estimation of the variable of interest in the presence of MEs.

The main objective of this study is to propose a generalized variance estimator using single auxiliary variable for the estimation of population variance in the presence and absence of MEs. The proposed estimator may produce some families and sub-families of the estimators as the special cases using different suitable choices of the scalar constants. After brief introduction and literature review, the rest of the paper is as follow. The sampling methodology with some existing estimators and basic notations for variance estimation in presence and absence of MEs are described in Section 2. The approximate mathematical expressions of bias and mean square error of proposed estimator are derived in presence and absence of MEs in Section 3. In the same section, some special cases of the proposed generalized estimator are also expressed on various values of optimize and generalize constants. In Section 4, the mathematical conditions are derived in which the proposed estimator have least mean square error than other existing estimators. In Section 5, the performance of suggested estimators is checked by conducting a numerical study in the presence and absence of MEs on the data based on three different artificial populations generated by normal distribution under different sample sizes. An application to real population is presented in Section 6. Final discussion on the paper is given in Section 7.

2. Sampling Methodology, Notations and Basic Estimators

In this section, we have described the sampling strategy of simple random sampling along with essential notations and some associated estimators based on population variance. We have also defined the measurement errors for both the study and the auxiliary variables.

2.1 Sampling Procedure

Suppose, it is considered that Y and X be the study variable and the auxiliary variable which are defined on N identifiable but distinct units of a finite population, \( U = \{U_1, U_2, U_3, \ldots, U_N\} \). Let \( n \) pair of observations are obtained using simple random sampling without replacement (SRSWOR) on two variables Y and X. Suppose a situation where both variables Y and X are observed with some considerable error. Let for the \( i^{th} \) sampling unit, \( y_i \) and \( x_i \) are observed instead of true observations \( Y_i \) and \( X_i \) where \( (i = 1, 2, \ldots, n) \). The MEs may be defined as

\[
u_i = (y_i - Y_i).
\]

and
where $u_i$ and $v_i$ be stochastic in nature which are respective associated $MEs$ with constant or zero mean and known variances $S_u^2$ and $S_v^2$ respectively. Following [26], it is assumed that the error $u_i$ and $v_i$ are independent from each other and also independent from $Y_i$ and $X_i$, that implies, $COV(X,Y)\neq 0$ and $COV(X,U) = COV(X,V) = COV(U,Y) = 0$ and $COV(U,V) = 0$. It is also assumed that the finite population correction is neglected. Let $(\bar{Y}, \bar{X})$ be the finite population means and variances of the variable of interest and auxiliary variable $(Y, X)$ respectively and $\rho_{yx}$ be the correlation coefficient between the subscripts.

2.2 Notations

Let $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ be the sample mean estimators which are unbiased of the population means $\{\bar{Y}, \bar{X}\}$ respectively. But under $MEs$, $s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$ and $s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ are not unbiased estimator of $(S_y^2, S_x^2)$ respectively, where expected values of $s_y^2$ and $s_x^2$ in the presence of $MEs$ are
\[
E(s_y^2) = S_y^2 + S_u^2.
\]
\[
E(s_x^2) = S_x^2 + S_v^2.
\]

As error variances and $S_y^2$ associated with respective study variable and the auxiliary variable are known. In such situations, the unbiased estimators of $S_y^2$ and $S_x^2$ are respectively given by
\[
\hat{s}_y^2 = S_y^2 - S_u^2 > 0.
\]
\[
\hat{s}_x^2 = S_x^2 - S_v^2 > 0.
\]

Now, let us define
\[
\hat{\sigma}_y^2 = S_y^2 (1 + e_o).
\]
\[
\hat{\sigma}_x^2 = S_x^2 (1 + e_1).
\]

Such that $E(e_o) = E(e_1) = 0, E(e_o^2) = \frac{A_x}{n}, E(e_1^2) = \frac{A_x}{n}$ and $E(e_o e_1) = \frac{\delta - 1}{n}$.
where \( A_y = \gamma_2 + \gamma_2u \frac{S_u^4}{S_y^2} + 2 \left( 1 + \frac{S_u^2}{S_y^2} \right)^2 \), \( A_x = \gamma_2 + \gamma_2 \frac{S_x^4}{S_x^2} + 2 \left( 1 + \frac{S_y^2}{S_x^2} \right)^2 \),

\[
\gamma_2 = \beta_2 - 3, \beta_2 = \frac{\mu_{zz}}{\mu_{zz}}, \mu_{zz} = E(z_i - \mu_z)^2, \theta_x = \frac{S_x^2}{S_x^2 - S_y^2} \text{ and } \delta = \frac{\mu_{yy}(X, Y)}{S_x^2 S_y^2},
\]

where \( z = Y, X, U \) and \( V \).

If there is no measurement error, then \( E(e_o) = E(e_i) = 0, E(e_o^2) = \frac{\beta_2 - 1}{n} = V_{o}, \)

whereas \( E\left(e_i^2\right) = \frac{\beta_{2x} - 1}{n} = V_{o1}, E(e_o e_i) = \frac{\mu_{xx} - 1}{n} = V_{20}. \)

2.3 Basic Estimators

In this sub-section, we have mentioned some classical estimators for population variance in the presence and absence of MEs

The unbiased estimator \( \hat{s}_y^2 \) for population variance in presence of MEs is

\[
t_o = \hat{s}_y^2.
\]

The variances of \( t_o \) with and without MEs are respectively given as

\[
Var(t_o) = \frac{S_y^4}{n} A_y. \]

and

\[
var(t_o) = S_y^2 V_{40}.
\]

[1] classical ratio estimator under measurement errors is defined as

\[
t_1 = \frac{s_y^2}{s_x^2} \left( \frac{S_x^2}{s_y^2} \right).
\]

The expressions of approximate bias and MSE of \( t_1 \) in the presence and absence of MEs are respectively given as

\[
Bias(t_1) = \frac{S_y^2}{n} \left( 1 + A_x - \delta \right).
\]
\[ Bias(t_1) \equiv S_y^2 (V_{04} - V_{22}) \]

and

\[ MSE(t_1) \equiv \frac{S_y^4}{n} \left( 2 + A_x + A_x - 2\delta \right). \]

\[ MSE(t_1) \equiv S_y^4 \left( V_{40} + V_{04} - 2V_{22} \right). \]

[2] proposed modified ratio-type estimator in presence of MEs is

\[ t_2 = \frac{S_y^2}{S_x} \left( \frac{S_x^2 + C_x}{S_y^2 + C_x} \right). \]

The expressions approximate bias and MSE of \( t_2 \) in the presence and absence of MEs are respectively given as

\[ Bias(t_2) \equiv \frac{S_y^2}{n} B \left( 1 + BA_x - \delta \right). \]

\[ Bias(t_2) \equiv S_y^2 B \left( BV_{04} - V_{22} \right). \]

and

\[ MSE(t_2) \equiv \frac{S_y^4}{n} \left( A_x + B^2 A_x - 2B(\delta - 1) \right). \]

\[ MSE(t_2) \equiv S_y^4 \left( V_{40} + B^2 V_{04} - 2BV_{22} \right), \]

where \( B = \frac{S_y^2}{S_x} \left( S_x^2 + C_x \right). \)

The exponential ratio estimator suggested by [4] in the presence of MEs is

\[ t_3 = \frac{S_y^2}{S_x} \exp \left( \frac{S_y^2 - S_x^2}{S_x^2 + S_x^2} \right). \]

The equations of bias and MSE of \( t_3 \) with and without ME are respectively given as

\[ Bias(t_3) \equiv \frac{S_y^2}{n} \left( \frac{3A_x}{4} - \delta - 1 \right). \]
\[ \text{Bias}(t_3) \approx \frac{S^2_y}{2} \left( \frac{3V_{04}}{4} - V_{22} \right). \]

and

\[ \text{MSE}(t_3) \approx \frac{S^4_y}{n} \left( A_y + \frac{A_y}{4} - (\delta - 1) \right). \]

\[ \text{MSE}(t_3) \approx S^4_y \left( V_{40} + \frac{V_{04}}{4} - V_{22} \right). \]

3. Proposed Generalized Estimator

In this section, we have proposed a generalized estimator for population variance in presence of MEs using simple random sampling technique. The proposed estimator produces many special cases (summarized in Appendix A) as the family of the proposed estimator in which the \([7]\) is a particular case for different values of suitably chosen constant. The \([7]\) is a less generalized estimator for population variance as compared to suggested estimator here and it was without considering the MEs. The proposed estimator is defined as

\[ t_{pi} = \hat{s}_y^2 \left[ M_1 + M_2 \left( S^2_x - \hat{s}_x^2 \right) \right] \left( \frac{aS^2_x + b}{aS^2_x + b} \right)^\alpha \exp \left[ \beta \left( \frac{a(S^2_x - \hat{s}_x^2)}{a(S^2_x - \hat{s}_x^2) + 2b} \right) \right], \]  

where \( \alpha \) and \( \beta \) are suitable constants which can assume \{(0, 0), (1, 1), (-1, -1), (1, 0), (0, 1), (0, -1), (-1, 0), (1, -1), (-1, 1)\}, whereas \((a \neq 0; b)\) are real numbers or some known parameters of the auxiliary variable and \( M_1 \) and \( M_2 \) are the optimize constants which minimize the mean square error of the proposed estimators \( t_{pi} \).

To obtain the mathematical expressions for bias and \( \text{MSE} \) of \( t_{pi} \), we may re-write Eq. (3) in terms \( e \)'s as

\[ t_{pi} = S^2_y \left[ 1 + e_o \right] \left[ M_1 + M_2 \left( S^2_x - S^2_y \left( 1 + e_1 \right) \right) \right] \left( 1 + ke_1 \right)^{-\alpha} \exp \left( \frac{-\beta ke_1}{2} \left( 1 + \frac{ke_1}{2} \right)^{-1} \right), \]  

where \( k = \frac{aS^2_x}{aS^2_x + b} \).

Expanding Eq. (4) using Taylor series, we have
\[ t_{pi} = S_y^2 \left( 1 + e_o \right) \left[ M_1 + M_2 S_x^2 e_1 \right] \left( 1 - \alpha k e_1 + \frac{\alpha (\alpha + 1) k^2 e_1^2}{2} \right) \exp \left( -\frac{\beta k e_1}{2} \left( 1 + \frac{ke_1}{2} \right) \right). \]  \hspace{1cm} (5)

On simplification of the Eq. (5), we get

\[
(t_{pi} - S_y^2) \approx S_y^2 \left[ (M_1 - 1) + M_1 e_o - e_1 \left( M_2 S_x^2 + M_1 \alpha k + \frac{M_1 \beta k}{2} \right) - e_o e_1 \left( M_2 S_x^2 + M_1 \alpha k + \frac{M_1 \beta k}{2} \right) + e_1^2 \right]. \hspace{1cm} (6)
\]

Simplifying and taking expectations on both sides of Eq. (6), we get the bias of the estimator \( t_{pi} \) in the presence of the \( \text{MSE} \). Here \( c = \alpha + (\beta / 2) \).

\[
\text{Bias} \left( t_{pi} \right) \approx S_y^2 \left[ (M_1 - 1) + \frac{A_c}{n} \left( M_2 S_x^2 k c + \frac{M_1 c (c + 1) k^2}{2} \right) - \frac{(\delta - 1)}{n} \left( M_2 S_x^2 + M_1 k c \right) \right]. \hspace{1cm} (7)
\]

If \( \text{ME} \) is zero than the expression of bias of the estimator \( t_{pi}' \) will be

\[
\text{Bias} \left( t_{pi}' \right) \approx S_y^2 \left[ (M_1 - 1) + V_{04} \left( M_2 S_x^2 k c + \frac{M_1 c (c + 1) k^2}{2} \right) - V_{22} \left( M_2 S_x^2 + M_1 k c \right) \right]. \hspace{1cm} (8)
\]

To get the \( \text{MSE} \) of the proposed estimator, squaring and applying Taylor series on Eq. (5), we have

\[
(t_{pi} - S_y^2)^2 \approx S_y^4 \left[ (M_1 - 1)^2 + M_1 e_o - e_1 \left( M_2 S_x^2 + M_1 \alpha k + \frac{M_1 \beta k}{2} \right) - e_o e_1 \left( M_2 S_x^2 + M_1 \alpha k + \frac{M_1 \beta k}{2} \right) + e_1^2 \right] \left[ \left( \frac{A_c}{n} \left( M_2 S_x^2 k c + \frac{M_1 c (c + 1) k^2}{2} \right) - \frac{(\delta - 1)}{n} \left( M_2 S_x^2 + M_1 k c \right) \right) \right]. \hspace{1cm} (9)
\]

On simplification of Eq. (6), we have final expression of \( \text{MSE} \) in the presence of \( \text{ME} \) of the proposed estimator is

\[
\text{MSE} \left( t_{pi} \right) \approx S_y^4 \left[ 1 + M_1 \left( 1 + \frac{A_c}{n} + \frac{A_c k^2 c (2c + 1)}{n} - 4 \left( \frac{\delta - 1}{n} \right) k c \right) + M_2^2 S_x^4 A_c \frac{A_c}{n} + (4M_1 M_2 - 2M_2^2) \right]. \hspace{1cm} (10)
\]
Or,
\[
MSE(t_{pi}) \cong S^4_y \left[1 + M_1^2 C_1 + M_2^2 C_2 + (4M_1M_2 - 2M_2) C_3 - 2M_1 C_4 \right],
\]
where,
\[
C_1 = \left\{ 1 + \frac{A_1}{n} + \frac{A_1 k^2 c (2c+1)}{n} - 4 \left( \frac{\delta - 1}{n} \right) k c \right\}, \quad C_2 = S^4_y \frac{A_1}{n}, \quad C_3 = S^2_y \left\{ \frac{A_1}{n} k c - \left( \frac{\delta - 1}{n} \right) \right\}
\]
and
\[
C_4 = \left\{ 1 + \frac{A_1 k^2 c (c+1)}{2n} - \left( \frac{\delta - 1}{n} \right) k c \right\}.
\]

If \( ME \) is assume to be negligible than Eq. (11) will be
\[
MSE(t_{pi}) \cong S^4_y \left[1 + M_1^2 \left\{ 1 + V_{40} + V_{04} k^2 c (2c+1) - 4V_{22} k c \right\} + M_2^2 S^4_y V_{04} + \right.
\]
\[
\left. (4M_1M_2 - 2M_2) S^2_y \left\{ V_{04} k c - V_{22} \right\} - 2M_1 \left\{ 1 + \frac{V_{04} k^2 c (c+1)}{2} - V_{22} k c \right\} \right].
\]

Or,
\[
MSE(t_{pi}) \cong S^4_y \left[1 + M_1^2 C_1^\prime + M_2^2 C_2^\prime + (4M_1M_2 - 2M_2) C_3^\prime - 2M_1 C_4^\prime \right],
\]
where \( C_1^\prime = \left\{ 1 + V_{40} + V_{04} k^2 c (2c+1) - 4V_{22} k c \right\}, \quad C_2^\prime = S^4_y V_{04}, \quad C_3^\prime = S^2_y \left\{ V_{04} k c - V_{22} \right\} \) and
\[
C_4^\prime = \left\{ 1 + \frac{V_{04} k^2 c (c+1)}{2} - V_{22} k c \right\}.
\]

For the optimum values of \( M_1 \) and \( M_2 \), we partially differentiate Eq. (13) with respect to \( M_1 \) and \( M_2 \) and equate to zero
\[
\frac{\partial MSE(t_{pi})}{\partial M_1} = S^4_y [2M_1 C_1 - 2C_4 + 4M_2 C_3] = 0.
\]
\[
M_1 C_1 + 2M_2 C_3 = C_4.
\]

Similarly,
\[
\frac{\partial MSE(t_{pi})}{\partial M_2} = S^4_y [2M_2 C_2 - 2C_3 + 4M_1 C_4] = 0.
\]
\[
2M_1 C_3 + M_2 C_2 = C_3.
\]

Solving both Eq. (14) and Eq. (15) simultaneously, we have
\[
\begin{bmatrix}
C_1 & 2C_3 \\
2C_3 & C_2
\end{bmatrix}
\begin{bmatrix}
M_1 \\
M_2
\end{bmatrix}
= \begin{bmatrix}
C_4 \\
C_3
\end{bmatrix}.
\]

The optimum values of \(M_1\) and \(M_2\) are

\[
M_1 = \frac{C_4C_2 - 2C_3^2}{C_1C_2 - 4C_3^2} \quad \text{and} \quad M_2 = \frac{C_4C_3 - 2C_4C_3}{C_1C_2 - 4C_3^2}.
\]

We substitute the values of \(M_1\) and \(M_2\) in Eq. (11) and get the minimum MSE of the proposed estimator as

\[
\text{MSE}(t_{pi})_{\text{min}} \approx S_y^2 \left[1 - A\right],
\]

where \(A = \frac{C_1C_3^2 + C_2C_4 - 4C_4C_3^2}{C_1C_2 - 4C_3^2}. \quad \text{(with ME)}\)

Similarly, if ME is negligible than the MSE of the estimator \(t'_{pi}\) is

\[
\text{MSE}(t'_{pi})_{\text{min}} \approx S_y^2 \left[1 - A'\right],
\]

where \(A' = \frac{C_1'C_3'^2 + C_2'C_4'^2 - 4C_4'C_3'^2}{C_1'C_2'^2 - 4C_3'^2}. \quad \text{(without ME)}\)

Many special cases of the proposed estimators are obtained and presented in Appendix Table A.1

Some special cases of estimator \(t_{pi}\) are:

**Case 1:** For \(M_2 = 0\), the estimator is reduced to [7]

\[
t_{pi1} = M_{11}S_y^2 \left(aS_y^2 + b\right)^{\alpha} \exp \left[\beta \left(\frac{a(S_y^2 - S_x^2)}{a(S_y^2 - S_x^2) + 2b}\right)\right].
\]

The approximate bias and MSE of the estimator \(t_{pi1}\) defined in Eq. (13) under ME are

\[
\text{Bias}(t_{pi1}) \approx S_y^2 \left[(M_{11} - 1) + \frac{M_{11}kc}{n}\left(\frac{A_1(c + 1)k}{2} - (\delta - 1)\right)\right].
\]

and
The $MSE$ of the estimator $t'_{pi1}$ without $ME$ is

$$MSE(t'_{pi1}) \approx S_y^4 \left[ 1+ M_{i1}^2 \left( 1+ \frac{A_y}{n} + \frac{A_k k^2 c (2c+1)}{n} - 4 \left( \frac{\delta - 1}{n} \right) kc \right) - 2M_{i1} \left( 1+ \frac{A_k k^2 c (c+1)}{2n} - \left( \frac{\delta - 1}{n} \right) kc \right) \right].$$ \hspace{1cm} (20)

The $MSE$ of the estimator $t'_{pi1}$ without $ME$ is

$$MSE(t'_{pi1}) \approx S_y^4 \left[ 1+ M_{i1}^2 \left( 1+ V_{a0} + V_{a0} k^2 c (2c+1) - 4V_{22} kc \right) - 2M_{i1} \left( 1+ \frac{V_{a0} k^2 c (c+1)}{2} - V_{22} kc \right) \right].$$ \hspace{1cm} (21)

**Case 2:** For $M_{i1} = 0$, the estimator is:

$$t_{pi2} = M_{i2} S_y^2 \left( S_x^2 - \hat{s}_x^2 \right) \left( \frac{aS_x^2 + b}{a\hat{s}_x^2 + b} \right)^{\beta} \exp \left[ \frac{a(S_x^2 - \hat{s}_x^2)}{a(S_x^2 - \hat{s}_x^2) + 2b} \right].$$ \hspace{1cm} (22)

The approximate bias and $MSE$ of the estimator $t_{pi2}$ defined in Eq. (14) under $ME$ are

$$Bias(t_{pi2}) \approx S_y^2 \left[ \frac{M_{i2} S_y^2}{n} \left\{ A_k k c - (\delta - 1) \right\} - 1 \right].$$ \hspace{1cm} (23)

and

$$MSE(t_{pi2}) \approx S_y^4 \left[ 1+ M_{i2}^2 S_y^2 \left( \frac{A_y}{n} - \frac{2M_{i2} S_y^2}{n} \right) \left\{ A_k k c - (\delta - 1) \right\} \right].$$ \hspace{1cm} (24)

The $MSE$ of estimator $t'_{pi2}$ without $ME$ is

$$MSE(t'_{pi2}) \approx S_y^4 \left[ 1+ M_{i2}^2 S_y^2 \left\{ V_{a0} k c - V_{22} \right\} \right].$$ \hspace{1cm} (25)

**4. Efficiency Comparisons**

In this section, we have mathematically compared the $MSE$ of the proposed estimator with the $MSE$s of some existing estimators ($t_0$, $t_1$, $t_2$ and $t_3$). By definition, the proposed estimator would be more efficient than the existing estimators, if the $MSE$ of $t_{pi}$ will be smaller than the $MSE$s of existing estimators. The optimum conditions for $t_{pi}$ from Eq. (11) with respect to $t_o$, $t_1$, $t_2$ and $t_3$ are respectively given by

$$MSE(t_o) - MSE(t_{pi})_{\min} > 0.$$ 

If $$S_y^2 \left[ A' + \frac{A_y}{n} - 1 \right] > 0.$$
If $S^4 \left[ A' + \frac{1}{n} \left\{ A_y + A_x - 2(\delta - 1) \right\} - 1 \right] > 0$.

If $S^4 \left[ A' + \frac{1}{n} \left\{ A_y + B^2A_x - 2B(\delta - 1) \right\} - 1 \right] > 0$.

If $S^4 \left[ A' + \frac{1}{n} \left\{ A_y + \frac{A}{4} -(\delta - 1) \right\} - 1 \right] > 0$.

5. Simulation Study

In this section, the efficiency of the proposed estimator is numerically examined over the existing estimators mentioned in this paper. It is carried through R-software developed by R [27]. The samples are generated using SRSWOR from the three finite populations each of size $N = 5000$. The populations are generated using multivariate normal distribution with same $ME$ variances as $S^2_U = S^2_V = 9$. The detail of population means and variances are:

**Population I:**

$\mu_{yx} = \left[ 6 \quad 5 \right], S^2_y = 125, S^2_x = 100, \rho_{yx} = 0.898$.

**Population II:**

$\mu_{yx} = \left[ 6 \quad 5 \right], S^2_y = 157, S^2_x = 121, \rho_{yx} = 0.881$.

**Population III:**

$\mu_{yx} = \left[ 6 \quad 5 \right], S^2_y = 149, S^2_x = 100, \rho_{yx} = 0.826$.

For each population, different sample sizes are considered as $n = 50, 100, 300$ and $500$. The percent relative efficiencies ($PRE$) can be computed for all the estimators by using the following formula.

$$PREs(t_i, t_o) = \frac{MSE(t_o)}{MSE(t_i)} \times 100,$$

where $t_i = t_1, t_2, t_3, t_{p1}, t_{p12}, t_{p13}$.
The results of \( MSE \) and \( PRE \) of the proposed estimator along with the existing estimators are computed in the presence and absence of \( ME \) and summarized in Table-1 to Table-3.

The results of \( MSE \) and \( PRE \) of different estimators in presence and absence of \( ME \) for population-I to population-III are summarized in Table-2 to Table-4 reveal that the \( MSEs \) of the mentioned estimators decrease the \( PREs \) increase by increase the sample sizes. Bold values indicate the least \( MSE \) and high \( PRE \) of the estimators under study. It is also observed that for all three populations, the suggested estimator \( t'_{pi} \) has minimum \( MSE \) and largest \( PRE \) than the estimators \( t_o, t_1, t_2, t_3, t'_{pi1} \) and \( t'_{pi2} \) with and without \( ME \).

6. An Application to Real Population

In this section, we considered a hypothetical data taken from [28]. In the data set, the study variable \( Y \) is true consumption expenditure and the auxiliary variable \( X \) is the true income. Further, the variables are contaminated with \( ME \) and are taken as measured consumption (\( y \)) expenditure and measured income (\( x \)) as shown in Table-4.

The parameters of real-data set are presented in Table-5

A sample of size 4 is taken for the computation of \( MSEs \) and \( PREs \) of different variance estimators in presence and absence of \( ME \). The results are presented in Table-6.

The results of \( MSE \) and \( PRE \) based on real data presented in Table-6 shows that the efficiency of the proposed estimator is better than other existing estimators in the presence and absence of \( ME \). The amount of \( MSE \) of the proposed estimator is the most least among all the competing estimators. It is also evident that the proposed estimator has least amount of \( ME \).

7. Conclusions

A new estimator is proposed in this paper for the estimation of population variance in presence and absence of \( ME \). Many special cases are obtained on various choices of optimize and generalize constants of the proposed estimator including [7] and many other product-type, ratio-type and exponential ratio-type estimators. The expressions of approximate bias and \( MSE \) are derived for proposed estimator. The conditions are also obtained for which the proposed estimator is mathematically efficient than the competing estimators. The role of sample size is significant in simulation study to judge the performance of proposed estimator, as the sample size increases from 50 to 500 the \( MSEs \) decreased and \( PRE \) increases for each sample size in the presence and absence of \( ME \). From the numerical findings, it is displayed both mathematically and numerically that the proposed estimator performed well than the other mentioned estimators in presence and absence of \( ME \) for all the artificial as well as real population data-sets.
Further expansions of the present effort merit additional consideration. In this present study, one auxiliary variable is considered for variance estimation in the presence of $ME$ under SRS design. A fruitful area for future research would be to incorporate multi-auxiliary variable during the multi estimation phases under different sampling designs, such as stratified sampling, systematic sampling etc. This could, in principle, be carried out by using multivariable extension of the proposed generalized estimator.

Acknowledgement

The authors are thankful to the Editor-in-Chief, Prof. S.T.A. Niaki, and anonymous referees for their valuable suggestions that helped to improve the article.

REFERENCES


**Table A.1: Some Known Cases of Proposed Generalized Estimator**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$a$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_0 = \hat{s}_y^2$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$t_1 = \hat{s}_y^2 \left( \frac{S_x^2}{\hat{s}_x^2} \right)$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$t_2 = \hat{s}_y^2 \left( \frac{S_x^2 + C_x}{\hat{s}_x^2 + C_x} \right)$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>$C_x$</td>
</tr>
<tr>
<td>$t_3 = \hat{s}_y^2 \exp \left( \frac{S_x^2 - \hat{s}_x^2}{S_x^2 + \hat{s}_x^2} \right)$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$t_4 = \hat{s}_y^2 \exp \left( \frac{S_x^2 - \hat{s}_x^2}{S_x^2 + \hat{s}_x^2} \right)$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$t_5 = \hat{s}_y^2 \left( \frac{S_x^2 + \text{Med}}{\hat{s}_x^2 + \text{Med}} \right)$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>$\text{Med}$</td>
</tr>
<tr>
<td>$t_6 = \hat{s}_y^2 \left[ k_1 + k_2 \left( S_x^2 - \hat{s}_x^2 \right) \right] \exp \left( g \frac{a(S_x^2 - \hat{s}_x^2)}{a(S_x^2 + \hat{s}_x^2) + 2b} \right)$</td>
<td>$k_1$</td>
<td>$k_2$</td>
<td>1</td>
<td>$g$</td>
<td>$a$</td>
<td>$b$</td>
</tr>
<tr>
<td>$t_7 = \hat{s}_y^2 \left( \frac{S_x^2}{\hat{s}_x^2} \right)^{1/2}$</td>
<td>1</td>
<td>0</td>
<td>1/2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
\[ t_8 = s_y^2 \left( \frac{S_x^2}{S_x^2} \right)^2. \] [7]
\[ t_9 = s_y^2 \left( \frac{S_x^2 + D}{S_x^2 + D} \right). \] [9]
\[ t_{10} = s_y^2 \left( \frac{S_x^2 + D_{AM}b_{2x}}{S_x^2 + D_{AM}b_{2x}} \right). \] [10]
\[ t_{11} = s_y^2 \left( \frac{S_x^2Q^2_1 + Q_d}{S_x^2Q^2_c + Q_d} \right). \] [9]
\[ t_{12} = s_y^2 \left( \frac{S_x^2Q^2_c + Q_d}{S_x^2Q^2_c + Q_c} \right). \] [32]
\[ t_{13} = s_y^2 \left( \frac{S_x^2Q^2_c + Q_d}{S_x^2Q^2_c + Q_c} \right). \] [32]
\[ t_{14} = s_y^2 \left( \frac{S_x^2Q^2_c + Q_d}{S_x^2Q^2_c + Q_c} \right). \] [32]
\[ t_{15} = s_y^2 \left( \frac{S_x^2(TM + Q_3)}{S_x^2(TM + Q_3)} \right). \] [11]
\[ t_{16} = s_y^2 \left( \frac{S_x^2Q + Q_r}{S_x^2Q + Q_r} \right). \] [11]

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Estimator</th>
<th>Without ME</th>
<th>With ME</th>
<th>Amount of ME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MSE</td>
<td>PRE</td>
<td>MSE</td>
</tr>
<tr>
<td>50</td>
<td>( t_1 )</td>
<td>271.451</td>
<td>233.48</td>
<td>343.396</td>
</tr>
<tr>
<td></td>
<td>( t_2 )</td>
<td>264.880</td>
<td>239.27</td>
<td>331.397</td>
</tr>
<tr>
<td></td>
<td>( t_3 )</td>
<td>290.793</td>
<td>217.95</td>
<td>317.135</td>
</tr>
<tr>
<td></td>
<td>( t_{pi1} )</td>
<td>220.498</td>
<td>287.43</td>
<td>321.311</td>
</tr>
<tr>
<td></td>
<td>( t_{pi2} )</td>
<td>173.262</td>
<td>365.8</td>
<td>239.157</td>
</tr>
<tr>
<td></td>
<td>( t_{pi} )</td>
<td>145.515</td>
<td>435.54</td>
<td>213.646</td>
</tr>
<tr>
<td>100</td>
<td>( t_1 )</td>
<td>129.395</td>
<td>234.15</td>
<td>159.491</td>
</tr>
<tr>
<td></td>
<td>( t_2 )</td>
<td>126.567</td>
<td>239.38</td>
<td>154.725</td>
</tr>
<tr>
<td></td>
<td>( t_3 )</td>
<td>137.914</td>
<td>219.69</td>
<td>149.984</td>
</tr>
<tr>
<td></td>
<td>( t_{pi1} )</td>
<td>102.712</td>
<td>294.98</td>
<td>142.543</td>
</tr>
<tr>
<td></td>
<td>( t_{pi2} )</td>
<td>81.3528</td>
<td>372.43</td>
<td>107.344</td>
</tr>
</tbody>
</table>

Table 1: The MSE and PRE (with and without ME) of Different Estimators for Population-I
### Table 2: The MSE and PRE (with and without ME) of Different Estimators for Population-II

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Estimator</th>
<th>Without ME</th>
<th>With ME</th>
<th>Amount of ME</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>( t_1 )</td>
<td>477.972</td>
<td>207.67</td>
<td>583.918</td>
</tr>
<tr>
<td>50</td>
<td>( t_2 )</td>
<td>467.370</td>
<td>212.38</td>
<td>566.401</td>
</tr>
<tr>
<td>50</td>
<td>( t_3 )</td>
<td>473.727</td>
<td>209.53</td>
<td>509.240</td>
</tr>
<tr>
<td>50</td>
<td>( t_{p1} )</td>
<td>385.912</td>
<td>257.21</td>
<td>531.197</td>
</tr>
<tr>
<td>50</td>
<td>( t_{p2} )</td>
<td>410.978</td>
<td>241.52</td>
<td>543.840</td>
</tr>
<tr>
<td>50</td>
<td>( t_p )</td>
<td>271.322</td>
<td>365.84</td>
<td>376.590</td>
</tr>
<tr>
<td>50</td>
<td>( t_{p1} )</td>
<td>258.044</td>
<td>192.39</td>
<td>312.217</td>
</tr>
<tr>
<td>50</td>
<td>( t_{p2} )</td>
<td>252.729</td>
<td>196.44</td>
<td>303.598</td>
</tr>
<tr>
<td>50</td>
<td>( t_p )</td>
<td>248.851</td>
<td>199.5</td>
<td>267.039</td>
</tr>
<tr>
<td>50</td>
<td>( t_{p1} )</td>
<td>193.878</td>
<td>256.06</td>
<td>262.217</td>
</tr>
<tr>
<td>50</td>
<td>( t_{p2} )</td>
<td>201.501</td>
<td>246.38</td>
<td>263.138</td>
</tr>
<tr>
<td>50</td>
<td>( t_p )</td>
<td>134.665</td>
<td>368.66</td>
<td>183.967</td>
</tr>
<tr>
<td>100</td>
<td>( t_1 )</td>
<td>69.784</td>
<td>216.61</td>
<td>83.186</td>
</tr>
<tr>
<td>100</td>
<td>( t_2 )</td>
<td>68.303</td>
<td>221.31</td>
<td>80.898</td>
</tr>
<tr>
<td>100</td>
<td>( t_3 )</td>
<td>70.440</td>
<td>214.59</td>
<td>75.407</td>
</tr>
<tr>
<td>100</td>
<td>( t_{p1} )</td>
<td>54.210</td>
<td>278.84</td>
<td>71.319</td>
</tr>
<tr>
<td>300</td>
<td>( t_1 )</td>
<td>41.0270</td>
<td>248.14</td>
<td>50.766</td>
</tr>
<tr>
<td>300</td>
<td>( t_2 )</td>
<td>40.039</td>
<td>254.26</td>
<td>49.116</td>
</tr>
<tr>
<td>300</td>
<td>( t_3 )</td>
<td>44.017</td>
<td>231.28</td>
<td>47.907</td>
</tr>
<tr>
<td>300</td>
<td>( t_{p1} )</td>
<td>33.805</td>
<td>301.15</td>
<td>47.062</td>
</tr>
<tr>
<td>300</td>
<td>( t_{p2} )</td>
<td>27.277</td>
<td>373.22</td>
<td>35.984</td>
</tr>
<tr>
<td>300</td>
<td>( t_p )</td>
<td>22.496</td>
<td>452.53</td>
<td>31.448</td>
</tr>
</tbody>
</table>
Table 3: The MSE and PRE (with and without ME) of Different Estimators for Population-III

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Estimator</th>
<th>Without ME</th>
<th>With ME</th>
<th>Amount of ME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MSE</td>
<td>PRE</td>
<td>MSE</td>
</tr>
<tr>
<td>50</td>
<td>$t_{p_1}$</td>
<td>60.141</td>
<td>251.34</td>
<td>76.014</td>
</tr>
<tr>
<td></td>
<td>$t_{p_2}$</td>
<td>38.298</td>
<td>394.69</td>
<td>50.695</td>
</tr>
<tr>
<td></td>
<td>$t_1$</td>
<td>45.042</td>
<td>219.76</td>
<td>53.527</td>
</tr>
<tr>
<td></td>
<td>$t_2$</td>
<td>44.335</td>
<td>223.26</td>
<td>52.391</td>
</tr>
<tr>
<td></td>
<td>$t_3$</td>
<td>48.331</td>
<td>204.8</td>
<td>51.734</td>
</tr>
<tr>
<td></td>
<td>$t_{p1}$</td>
<td>32.321</td>
<td>306.25</td>
<td>42.532</td>
</tr>
<tr>
<td></td>
<td>$t_{p2}$</td>
<td>34.747</td>
<td>284.86</td>
<td>43.827</td>
</tr>
<tr>
<td></td>
<td>$t_1$</td>
<td>22.365</td>
<td>442.58</td>
<td>29.675</td>
</tr>
<tr>
<td>100</td>
<td>$t_{p_1}$</td>
<td>293.146</td>
<td>153.07</td>
<td>380.159</td>
</tr>
<tr>
<td></td>
<td>$t_{p_2}$</td>
<td>286.550</td>
<td>156.59</td>
<td>367.643</td>
</tr>
<tr>
<td></td>
<td>$t_1$</td>
<td>254.039</td>
<td>176.64</td>
<td>277.923</td>
</tr>
<tr>
<td></td>
<td>$t_2$</td>
<td>74.340</td>
<td>172.14</td>
<td>81.350</td>
</tr>
<tr>
<td></td>
<td>$t_3$</td>
<td>210.265</td>
<td>213.41</td>
<td>317.218</td>
</tr>
<tr>
<td></td>
<td>$t_{p1}$</td>
<td>172.485</td>
<td>260.15</td>
<td>254.692</td>
</tr>
<tr>
<td></td>
<td>$t_{p2}$</td>
<td>130.666</td>
<td>343.41</td>
<td>200.531</td>
</tr>
<tr>
<td>300</td>
<td>$t_1$</td>
<td>86.309</td>
<td>148.27</td>
<td>109.388</td>
</tr>
<tr>
<td></td>
<td>$t_2$</td>
<td>84.556</td>
<td>151.34</td>
<td>106.247</td>
</tr>
<tr>
<td></td>
<td>$t_3$</td>
<td>74.340</td>
<td>172.14</td>
<td>81.350</td>
</tr>
<tr>
<td></td>
<td>$t_{p1}$</td>
<td>59.570</td>
<td>214.82</td>
<td>86.192</td>
</tr>
<tr>
<td></td>
<td>$t_{p2}$</td>
<td>45.187</td>
<td>283.2</td>
<td>64.062</td>
</tr>
<tr>
<td></td>
<td>$t_1$</td>
<td>36.440</td>
<td>351.18</td>
<td>53.556</td>
</tr>
<tr>
<td>500</td>
<td>$t_2$</td>
<td>53.067</td>
<td>137.3</td>
<td>66.874</td>
</tr>
<tr>
<td></td>
<td>$t_3$</td>
<td>51.876</td>
<td>140.46</td>
<td>64.833</td>
</tr>
<tr>
<td></td>
<td>$t_{p1}$</td>
<td>42.665</td>
<td>170.78</td>
<td>46.625</td>
</tr>
</tbody>
</table>
Table 4: Hypothetical Population Data

<table>
<thead>
<tr>
<th>Y</th>
<th>X</th>
<th>Y</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>75.4666</td>
<td>80</td>
<td>67.6011</td>
<td>80.094</td>
</tr>
<tr>
<td>74.9801</td>
<td>100</td>
<td>75.4438</td>
<td>91.5721</td>
</tr>
<tr>
<td>102.8242</td>
<td>120</td>
<td>109.6956</td>
<td>112.1406</td>
</tr>
<tr>
<td>125.7651</td>
<td>140</td>
<td>129.4159</td>
<td>145.5969</td>
</tr>
<tr>
<td>106.5035</td>
<td>160</td>
<td>104.2388</td>
<td>168.5579</td>
</tr>
<tr>
<td>131.4318</td>
<td>180</td>
<td>125.8319</td>
<td>171.4793</td>
</tr>
<tr>
<td>149.3693</td>
<td>200</td>
<td>153.9926</td>
<td>203.5366</td>
</tr>
<tr>
<td>143.8628</td>
<td>220</td>
<td>152.9208</td>
<td>222.8533</td>
</tr>
<tr>
<td>177.5218</td>
<td>240</td>
<td>176.3344</td>
<td>232.9879</td>
</tr>
<tr>
<td>182.2748</td>
<td>260</td>
<td>174.5252</td>
<td>261.1813</td>
</tr>
</tbody>
</table>

Table 5: The parameters of real-data

<table>
<thead>
<tr>
<th>Parameters</th>
<th>N</th>
<th>( \bar{Y} )</th>
<th>( \bar{X} )</th>
<th>( S_Y^2 )</th>
<th>( S_X^2 )</th>
<th>( S_U^2 )</th>
<th>( S_V^2 )</th>
<th>( \rho_{xy} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>10</td>
<td>127</td>
<td>170</td>
<td>1420</td>
<td>3666.667</td>
<td>36</td>
<td>41.24</td>
<td>0.964</td>
</tr>
</tbody>
</table>

Table 6: The MSE and PRE (with and without ME) of Different Estimators

<table>
<thead>
<tr>
<th>Estimators</th>
<th>With ME</th>
<th>Without ME</th>
<th>Amount of ME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>PRE</td>
<td>MSE</td>
</tr>
<tr>
<td>( t_1 )</td>
<td>159542</td>
<td>320.34</td>
<td>155341</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>159461</td>
<td>320.50</td>
<td>155261</td>
</tr>
<tr>
<td>( t_3 )</td>
<td>210428</td>
<td>242.87</td>
<td>203026</td>
</tr>
<tr>
<td>( t_{pi1} )</td>
<td>152056</td>
<td>336.11</td>
<td>151744</td>
</tr>
<tr>
<td>( t_{pi2} )</td>
<td>147175</td>
<td>347.26</td>
<td>146889</td>
</tr>
<tr>
<td>( t_{pi} )</td>
<td>135669</td>
<td>376.71</td>
<td>135398</td>
</tr>
</tbody>
</table>

Biographies

Muhammad Umair Tariq obtained his PhD degree in Statistics from National College of Business Administration and Economics, Lahore, Pakistan. He has been working as an Assistant
Professor at the Department of Statistics, Government Graduate College Sahiwal. His research interests include Sampling Theory, Quality Control and Biostatistics.

Muhammad Nouman Qureshi holds the position of Assistant Professor of Statistics from National College of Business Administration and Economics, Lahore, Pakistan. He has more than 20 research publications in well-reputed journals. His areas of interest are survey sampling, probability distributions, and time series analysis.

Muhammad Hanif received his MPhil from New South Wales Australia, and PhD in Statistics from the Department of Statistics, University of the Punjab, Pakistan. He is currently working as Pro Rector and Professor of Statistics at National College of Business Administration and Economics, Lahore, Pakistan. His Research interests include Advanced Survey Sampling, Quality Control, Probability Distribution, Biostatistics and Regression Analysis.