

Performance of a Predictive Ratio Estimator in Two-Stage Sampling with Auxiliary Information

*Akanyi George Orkuma, Ikughur Jonathan Atsua and Uba Tersoo

¹Department of Statistics, Joseph Sarwuan Tarka University, Makurdi, Benue State, Nigeria.

*Corresponding authors' email: georgeorkuma@gmail.com

ABSTRACT

This study examines the relative performance of a predictive ratio estimator against the conventional ratio estimator within a two-stage sampling design incorporating double sampling for the auxiliary variable. Fertilizer application quantity was used as the auxiliary characteristic, with crop yield as the study variable. Estimator performance was assessed through the mean squared error (MSE), standard error (SE), coefficient of variation (CV), and relative efficiency (RE). The predictive ratio estimator had a lower MSE of 0.1542, a SE of 0.3927 and a CV of 2.065% against the MSE of 0.1560, a SE of 0.3927 and a CV of 2.092% for the classical ratio estimator. The predictive ratio estimator also yielded a slightly higher relative efficiency of 101.17% compared with classical ratio estimator 100%. This efficiency gain is attributable to the estimator's capacity to exploit the strong auxiliary–study variable correlation captured in the first-phase sample. The study recommends the adoption of predictive ratio estimators in two-stage agricultural surveys with reliable auxiliary data.

Received: 29 Sep. 2026

Accepted: 14 Oct. 2026

Published: 20 Oct. 2026

Keywords:

Predictive Ratio Estimator, Two-Stage Sampling, Double Sampling, Auxiliary Information, Relative Efficiency

INTRODUCTION

Survey sampling methods increasingly make use of auxiliary information because information that is strongly associated with the study variable can improve the precision of population estimates. However, in the situation where survey resources are limited and the study variable is costly or difficult to measure, predictive estimators extend this approach by using observed information and auxiliary variables to predict values for non-sampled units (Kumar & Saini, 2020; Khan et al., 2025). In this context, prediction refers to estimating unobserved finite-population values rather than forecasting future observations.

Auxiliary information has long been used in ratio and regression estimation, particularly where the study and auxiliary variables are positively associated. The use of auxiliary information can also be extended to two-phase or double sampling, where an initial sample is used to obtain information that is not readily available for all sampled units, followed by a second-phase sample in which the study variable is observed (Cochran, 1977; Hidiroglou & Särndal, 1998). Predictive approaches have subsequently been developed for ratio, regression, and product estimators in two-phase and multistage settings (Saini, 2013; Yadav & Mishra, 2015; Kumar & Saini, 2020).

Previous studies have examined modified and predictive estimators under different sampling designs. For example, Ozturk (2019) considered two-stage cluster sampling with ranked set sampling, while Michael and Mbanefo (2022) examined two-stage cluster sampling with unequal probability sampling in the first stage. Other studies have investigated predictive estimation and improved estimators using auxiliary information in two-phase sampling (Panda, 2015; Yadav et al., 2018; Lohr, 2021).

Accurate estimation of population parameters can be challenging when direct measurement is costly, time-consuming, or otherwise difficult. The problem is compounded when complete auxiliary information is not available at the level required for estimation (Arbia et al.,

2020). In such circumstances, double sampling provides a practical means of obtaining auxiliary information in an initial phase before observing the study variable in a second phase. This study aims to compare the efficiency of classical and predictive ratio estimators under a two-stage, two-phase sampling design using strongly correlated auxiliary information.

MATERIALS AND METHODS

This study reproduces and relies upon the methodological framework proposed by Saini (2013). All theoretical derivatives and structural equations used throughout this analysis are drawn directly from Saini's model.

Procedure and Notations

Let the finite population U be partitioned into N first Stage Units (fsu) denoted by $(U_1, \dots, U_N)$ such that the number of second stage units (ssu) in U_i is M_i and $M = \sum_{i=1}^N M_i$. Let y_{ij} and x_{ij} be the values of the study variable y correlated with auxiliary character x for the j^{th} sampling units of U_i ($j = 1, 2, \dots, M_i; i = 1, 2, \dots, N$).

Here, the auxiliary information is used at estimation stage to improve the precision of an estimator but when the required auxiliary information is not available at ssu level, a double sampling procedure is used as an alternative for the estimation of population mean. Assume a sample s of n fsu is drawn from N fsu by using SRSWOR, then we select a sample s'_i of m'_i ($< n$) ssu from the i^{th} selected fsu U_i by using SRSWOR. At this stage, information on an auxiliary variable is collected. Further from the selected m'_i ssu, a sample of m_i ssu is selected by using SRSWOR, y the character of interest is observed at this stage.

Define:

M_i = the number of selected fsu's in the sample n .

m'_i = First-phase ssu sample size selected from the number of fsu's in the sample

m_i = the Second-phase ssu sample size selected from the first-phase sample, where both x and y are observed.

$M = \sum_i^{M_i} M_i$ is the number of elements in the population
 Y_{ij} = value of the character y on the j^{th} ssu's in the i^{th} (selected)
 fsu in the population, $j=1,2,\dots,M_i$, $i=1,2,\dots,N$.

y_{ij} = value of the character under study on the j^{th} observation
 in the sample from the i^{th} fsu $j=1,2,\dots,m_i$; $i=1,2,\dots,n$

$Y_i = \sum_j^{M_i} Y_{ij}$ is study variable population total for the i^{th}
 selected fsu.

$y_{ij} = \sum_j^{m_i} y_{ij}$ is sample total for the i^{th} selected fsu.

$\bar{y}_i = \frac{1}{m_i} \sum_j^{m_i} y_{ij}$ is the sample mean for the i^{th} selected fsu.

$\bar{Y}_i = \frac{Y_i}{M_i}$ is study variable population mean for the selected
 fsu.

$S_y^2 = \frac{1}{N-1} \sum_i^N (w_i \bar{Y}_i - \bar{Y})^2$ is population variance of the
 selected ssu for the study variable

$S_{iy}^2 = \frac{1}{M_i-1} \sum_j^{M_i} (y_{ij} - \bar{Y})^2$ is variance of the selected fsu for
 the study variables

$S_x^2 = \frac{1}{N-1} \sum_i^N (w_i \bar{X}_i - \bar{X})^2$ is population variance of the
 selected fsu for the auxiliary variable

$S_{ix}^2 = \frac{1}{M_i-1} \sum_j^{M_i} (x_{ij} - \bar{X})^2$ is the variance of the selected fsu
 for the auxiliary variable

$S_{yx} = \frac{1}{N-1} \sum_i^N (w_i \bar{Y}_i - \bar{Y})(w_i \bar{X}_i - \bar{X})$ is population
 covariance of the study and auxiliary variable y and x
 respectively

$S_{iyx} = \frac{1}{M_i-1} \sum_j^{M_i} (y_{ij} - \bar{Y})(x_{ij} - \bar{X})$ Covariance of the
 selected fsu for the study and auxiliary variable

$\rho = \frac{s_{xy}}{s_x s_y}$ is correlation of the study variable y and auxiliary
 variable x

$R = \frac{\bar{y}}{\bar{x}}$ and $R_i = \frac{\bar{y}_i}{\bar{x}_i}$ is the total population and Cluster i
 population ratio respectively

$\bar{X}_i = \frac{1}{M_i} \sum_j^{M_i} x_{ij}$ is auxiliary variable of the variables in the
 population.

$\bar{Y} = \frac{1}{N} \sum_i^N w_i \bar{Y}_i$, $\bar{X} = \frac{1}{N} \sum_i^N w_i \bar{X}_i$ as the overall population
 means of the study and auxiliary variables respectively,
 where, $w_i = \frac{M_i}{M} = \frac{NM_i}{M}$ and $\bar{M} = \frac{1}{N} \sum_i^N M_i$

$f = \frac{n}{N}$, $f_i' = \frac{m_i'}{M_i}$, $f_i = \frac{m_i}{M_i}$

$t = \frac{\bar{y}}{\bar{x}}$ is sample ratio statistic

$T = \frac{\bar{Y}}{\bar{X}}$ is theoretical ratio coefficient, $T = R$

$t_i = \frac{\bar{y}_i}{\bar{x}_i}$ is within-fsu sample ratio

$T_i = \frac{\bar{Y}_i}{\bar{X}_i}$ is within-PSU population ratio, available because the
 finite population is known.

$\bar{Y}_{ir}$ is actual mean of Y for first-phase units not included in
 second-phase sample

$\bar{Y}_{r'}$ is the actual mean of Y for units not selected in the first-
 phase sample

$\bar{Y}_r$ is actual mean of Y for non-sampled fsus

Z_r is the predictor of $\bar{Y}_r$

Z_{ir} is the Predictor of $\bar{Y}_{ir}$

$Z_{r'}$ is the predictor of $\bar{Y}_{r'}$

Z_r^* is the combined predictor for all units not selected in the
 second phase within fsu i .

Ratio Estimator

When information on auxiliary character for U is unknown and collected by using double sampling, the classical two-stage
 ratio estimator of $\bar{Y}$ is given in Equation (1);

$$\hat{\bar{Y}}_{R,d} = \frac{\bar{y}}{\bar{x}} \bar{x}' \quad (1)$$

Where $\bar{x} = \frac{1}{n} \sum_{i \in S} w_i \bar{x}_i$, and $\bar{x}_i = \frac{1}{m_i} \sum_{j \in S_i} x_{ij}$, $\bar{y} = \frac{1}{n} \sum_{i \in S} w_i \bar{y}_i$, and $\bar{y}_i = \frac{1}{m_i} \sum_{j \in S_i} y_{ij}$,

also $\bar{x}' = \frac{1}{n} \sum_{i \in S'} w_i \bar{x}_i'$, and $\bar{x}_i' = \frac{1}{m_i'} \sum_{j \in S_i'} x_{ij}'$

Mean Square Error of Ratio Estimator

The ratio estimator is bias, owing to the complexity of estimation of the bias and mean square error of ratio estimation.

Denote $\varepsilon_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}}$, $\varepsilon_1 = \frac{\bar{x} - \bar{X}}{\bar{X}}$, $\varepsilon_2 = \frac{\bar{x}' - \bar{X}}{\bar{X}}$

It is assumed that $E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_2) = 0$

Therefore, the Mean Square Error of $\hat{\bar{Y}}_{R,d}$ is defined as

$$\begin{aligned} MSE(\hat{\bar{Y}}_{R,d}) &= E(\hat{\bar{Y}}_{R,d} - \bar{Y})^2 \\ &= \bar{Y}^2 E(\varepsilon_0 + \varepsilon_1 + \varepsilon_2)^2 \text{ (Retaining the terms up to order two)} \\ &= \bar{Y}^2 E[\varepsilon_0^2 + \varepsilon_1^2 + \varepsilon_2^2 + 2\varepsilon_0\varepsilon_1 - 2\varepsilon_0\varepsilon_2 - 2\varepsilon_1\varepsilon_2] \\ &= \bar{Y}^2 E[\varepsilon_0^2 + \varepsilon_1^2 + \varepsilon_2^2 + 2\varepsilon_0\varepsilon_2 - 2\varepsilon_0\varepsilon_1 - 2\varepsilon_1^2] \end{aligned}$$

Therefore,

$$MSE(\hat{\bar{Y}}_{R,d}) = \left(\frac{1-f}{n}\right) \{S_y^2 - 2RS_{xy} + R^2 S_x^2\} + \frac{1}{nN} \sum_i^N w_i^2 \left(\frac{1-f_i'}{m_i'}\right) S_{iy}^2 + \frac{1}{nN} \sum_i^N w_i^2 \left(\frac{1-f_i}{m_i}\right) \{S_{iy}^2 - 2RS_{iyx} + R^2 S_{ix}^2\} \quad (2)$$

Predictive Ratio Estimator

Predictive Ratio Estimator Criteria and their Properties

Using the predictive set-up of Saini (2013), it possible to express for a given non empty set S , by partitioning the estimation
 of population total in two stages sampling as,

$$\begin{aligned} \bar{Y} &= \frac{\sum_i^N \sum_j^{M_i} y_{ij}}{\sum_i^N M_i} = \frac{\sum_i^N M_i \bar{Y}_i}{\sum_i^N M_i} \\ &= \frac{1}{M} [\sum_i^N M_i \bar{Y}_i] \\ &= \frac{1}{M} [\sum_{i \in S'} M_i \bar{Y}_i + \sum_{i \in S} M_i \bar{Y}_i] \\ &= \frac{1}{M} [\sum_{i \in S'} \{ \sum_{i \in S} M_i \bar{Y}_i + \sum_{i \in S} M_i \bar{Y}_i \} + \sum_{i \in S} M_i \bar{Y}_i] \end{aligned}$$

$$= \frac{1}{M} [\sum_{i \in s'} \{ \sum_{i \in s} (\sum_{j \in s_i} y_{ij} + \sum_{j \in \bar{s}_i} y_{ij}) \} + \sum_{i \in \bar{s}} M_i \bar{Y}_i + \sum_{i \in \bar{s}'} M_i \bar{Y}_i] \quad (3)$$

where $\bar{s}'$, denotes the set of $(N - n')$ first phase of fsu's in U which are not included in s' , $\bar{s}$ denoted the set of $(n' - n)$ second phase of fsu's in U which are not included in s and $\bar{s}_i$ the set of $(M_i - m_i)$ ssus of U_i which are not included in s_i , $i = 1, 2, \dots, n$

$$\bar{Y} = \frac{1}{M} [\sum_{i \in s'} \{ \sum_{i \in s} (m_i \bar{Y}_i + (M_i - m_i) \bar{Y}_{ir}) \} + \frac{1}{M} \sum_{i \in s} M_i \bar{Y}_i + \frac{1}{M} \sum_{i \in \bar{s}'} M_i \bar{Y}_i] \quad (4)$$

$$\text{Let } \bar{Y}_{ir'} = \frac{1}{M_i - m_i'} \sum_{i \in \bar{s}'} w_i \bar{Y}_i$$

$$\bar{Y}_r = \frac{1}{N - n} \sum_{i \in \bar{s}_i} w_i \bar{Y}_i$$

$$\bar{Y}_{ir} = \frac{1}{m_i' - m_i} \sum_{j \in \bar{s}_i} Y_{ij}$$

Then

$$\bar{Y} = \frac{1}{M} [\sum_{i \in s'} \{ \sum_{i \in s} (m_i \bar{Y}_i + (m_i' - m_i) \bar{Y}_{ir'}) \} + (M_i - m_i') \bar{Y}_{ir} + \frac{N-n}{N} \bar{Y}_r] \quad (5)$$

Where,

$$\bar{Y}_{ir'} = \frac{M_i \bar{Y}_i - m_i' \bar{Y}_i'}{M_i - m_i'}$$

$$\bar{Y}_r = \frac{N \bar{Y} - n \bar{y}}{N - n}$$

$$\bar{Y}_{ir} = \frac{m_i' \bar{Y}_i' - m_i \bar{Y}_i}{m_i' - m_i}$$

To estimate $\bar{Y}$ requires the prediction of $\bar{Y}_{ir}$, $\bar{Y}_r$, and $\bar{Y}_{ir'}$ from the sample data because the first component of the right hand side of Equation (5) is already known. Using $Z_{ir'}$, Z_r and Z_{ir} as their respective predictors,

Then predictive estimator of the population mean $\bar{Y}$ is

$$\hat{Y}_{pre} = \frac{1}{M} \left[\sum_{i \in s} \{ \sum_{j \in s} (m_i \bar{Y}_i + (m_i' - m_i) Z_{ir}) + (M_i - m_i) Z_{ir}' \} + \frac{N-n}{N} Z_r \right] \quad (6)$$

where Z_r is predictors of $\bar{Y}_r$ of for the first stage using double sampling, $Z_{ir'}$ and Z_{ir} is the predictor of $\bar{Y}_{r'}$ and $\bar{Y}_{ir}$ for second stage unit respectively.

In Equation (6) combine the last two terms i.e Non-sampled part of the first phase and second phase in the first stage unit because in the first stage using double sampling when we go from one phase to second sampling unit is same and non-sampled units are

$$M_i - m_i' + m_i' - m_i = M_i - m_i$$

Equation (6) can be written as

$$\hat{Y}_{pre} = \frac{1}{M} [\sum_{i \in s} (m_i \bar{Y}_i + (M_i - m_i) Z_{ir}^*)] + \frac{N-n}{N} Z_r \quad (7)$$

Define, Z_r^* and Z_{ir} as the class of estimators using single auxiliary variable x for first stage unit using double sampling and second stage unit as

$$\begin{aligned} Z_r &= \bar{y}_r + t(\bar{X}_r - \bar{x}) \\ Z_{ir}^* &= \bar{y}_{ir} + t_i(\bar{x}'_{ir} - \bar{x}_i) \end{aligned} \quad (8)$$

From the above predictive criterion, t and t_i is suitably chosen statistics for defining ratio estimator.

The first order approximation of the MSE of the predictive ratio are given respectively.

Put the different values if T and T_i , the following equations for the ratio estimator are obtained

$$M(Z_r) = \left(\frac{1-f}{n} \right) \{ S_y^2 - 2T\rho S_y S_x + T^2 S_x^2 \} \quad (9)$$

$$M(Z_{ir}^*) = \left(\frac{1-f_i'}{m_i'} \right) S_{iy}^2 + \left(\frac{1-f_i}{m_i} \right) \{ S_{iy}^2 - 2T\rho S_{iy} S_{ix} + T_i^2 S_{ix}^2 \} \quad (10)$$

where Equation (9) is the mean square error for first stage unit using double sampling and Equation (10) is the mean square error for second stage unit respectively.

So that the predictive estimator of the population mean for ratio estimator is

$$\hat{Y}_{pre,R} = \frac{1}{M} \left[\sum_{i \in s} \sum_{j \in s} (M_i (\bar{y}_i + t_i (\bar{x}'_i - \bar{x}_i))) + (\bar{y} + t(\bar{X} - \bar{x})) \right] \quad (11)$$

It is assume that $E(t) = E_1 E_2(t) = T$ or $E(t) \approx T$ and $E(t_i) = T_i$ or $E(t_i) \approx T_i$. From Table 1 after some algebraic manipulations $\hat{Y}_{pre}$ turns out to be ratio estimator. We define the Predictive ratio estimator in two stages sampling using double sampling as;

$$d_{o1} = \hat{Y}_{R,d} + \frac{1}{N} \sum_i w_i \left(\frac{\bar{y}_i}{\bar{x}_i} - \frac{\bar{y}}{\bar{x}} \right) \bar{x}' \quad (12)$$

Mean Square Error of Predictive Ratio Estimator

After some simplification, we have the first order of Mean Square Error of $\hat{Y}_{pre}(d_{o1})$ as follows:

$$M(\hat{Y}_{pre,R}) = \left(\frac{1-f}{n} \right) \{ S_y^2 + T^2 S_x^2 - 2T S_{xy} \} + \frac{1}{nN} \sum_i^N w_i^2 \left(\frac{1-f_i'}{m_i'} \right) S_{iy}^2 + \frac{1}{nN} \sum_i^N w_i^2 \left(\frac{1-f_i}{m_i} \right) \{ S_{iy}^2 + \alpha_i^2 S_{ix}^2 - 2\alpha_i S_{ixy} \} \quad (13)$$

where $\alpha_i = T - f(T - T_i)$

The optimum value of T and T_i which minimize $M(\hat{Y}_{pre,R})$ are given respectively as

$$T_{opt} = \frac{S_{yx}}{S_x^2} = B_b$$

and

$$T_{i\text{opt}} = \frac{S_{iyx}}{S_{ix}^2} = B_{ib}$$

Then the Minimum mean square error of $\hat{Y}$ is

$$M(\hat{Y})_{\text{opt}} = \left(\frac{1-f}{n}\right)\{S_y^2 + T_{\text{opt}}^2 S_x^2 - 2T_{\text{opt}} S_{xy}\} + \frac{1}{nN} \sum_i^N w_i^2 \left(\frac{1-f_i'}{m_i'}\right) S_{iy}^2 + \frac{1}{n} \sum_i^N w_i^2 \left(\frac{1-f_i}{m_i}\right) \{S_{iy}^2 + \alpha_{i\text{opt}}^2 S_{ix}^2 - 2\alpha_{i\text{opt}} S_{ixy}\} \quad (14)$$

The minimum mean square errors of predictive Ratio estimator up to first order approximations is given as;

$$M(d_{01}) = \left(\frac{1-f}{n}\right)\{S_y^2 + R^2 S_x^2 - 2RS_{xy}\} + \frac{1}{nN} \sum_i^N w_i^2 \left(\frac{1-f_i'}{m_i'}\right) S_{iy}^2 + \frac{1}{nN} \sum_i^N w_i^2 \left(\frac{1-f_i}{m_i}\right) \{S_{iy}^2 + \alpha_i^2 S_{ix}^2 - 2\alpha_i S_{ixy}\} \quad (15)$$

Where $\alpha_i = R - f(R - R_i)$, $R = T$

Subtracting the Equation (15) from the equation (2), we get

Difference: $D_j = MSE(d_j) - MSE(d_{0j})$ $j = 1, 2$ i.e

$$= MSE(\hat{Y}_{R,d}) - MSE(d_{01})$$

$$D_1 = \frac{1}{nN} \sum_i^N \left(\frac{1-f_i}{m_i}\right) \{(R^2 - \alpha_i^2) S_{ix}^2 - 2(R - \alpha_i) S_{ixy}\}$$

Thus d_{01} will be more efficient than $\hat{Y}_{R,d}$ if,

$$\beta_{iyx} \leq \frac{R + \alpha_i}{2} \quad i = 1, 2, \dots, N \quad (16)$$

where $\beta_{iyx} = \frac{S_{iyx}}{S_{ix}^2}$ be the regression confident of y on x in U_i

Relative Efficiency (RE) of Estimators

$$\text{Relative Efficiency} = \frac{\text{MSE Classical Ratio Estimator}}{\text{MSE Predictive Ratio Estimator}} * 100\% \quad (17)$$

- If $RE < 100\%$, the classical Estimator is more efficient
- If $RE > 100\%$, the Predictive Estimator is more efficient

Sampling and Sampling Design

The sampling design employed a two-stage cluster sampling approach for population of 2,093 farmers distributed across 48 clusters (cooperative societies).

Firstly, a preliminary sample of 24 clusters were selected using simple random sampling without replacement (SRSWOR), representing 50% of the total clusters. These are the Primary Sampling Units (PSU) that are used to obtain information on the auxiliary variable x .

In the Second Stage, 20% of Cluster units (elements) from the uneven clusters were chosen in the first phase. The selection of element within clusters was done using proportional allocation to size, ensuring larger clusters contribute more unit to the sample than smaller ones. The sampled data was used to obtain information on both auxiliary and study variables at this stage.

Mathematically, the number of element selected from the i^{th} cluster in the second stage is given by Equation (18):

$$n_i = 0.20 * M_i \quad (18)$$

where:

n_i = number of sampled elements in cluster i

The proportional allocation ensure that the sample accurately reflect the underlying population structure thereby improving precision of estimates This design allows for efficient estimation of population parameters using auxiliary information from the larger first-stage sample to improve precision in the second-stage estimates. (Särndal *et al.*, 1992).

RESULTS AND DISCUSSION

Results

The study population consisted of 2,093 farmers distributed across 48 cooperative society clusters. Twenty-four clusters were selected for the sample. The population mean of the study variable was 19.114. The correlation coefficient between fertilizer application and crop yield was 0.9758, indicating a very strong positive association and supporting the use of fertilizer as auxiliary information. Some of the preliminary results obtained from the data analysis are as obtained in table 1 below

Table 1: Summary Statistics and Results of some Parameters

Parameters	Value	Parameter	Value
N	48	$\bar{Y}$	19.114
n	24	$\bar{X}$	3.033
M	2093	$\bar{x}_r$	3.045
$\bar{x}$	3.165	$\bar{M}$	43.602
$\bar{y}$	19.181	ρ	0.9758
$\bar{x}'$	3.117	$\bar{Y}_{r'}$	18.026
S_y^2	18.273	$\bar{Y}_r$	19.524
$\bar{y}'$	19.562	t	6.060
S_x^2	0.448	T	6.303
S_{xy}	2.795	Z_r	18.455
R	6.303		
f	0.500		

Table 2: Predictive Ratio and Classical Ratio Estimators and Summary Statistics

Estimator	Mean	MSE	Standard Error	CV(%)	RE(%)
Classical (R, d)	18.886465	0.156041	0.395021	2.091554	100.000
Predictive (d_{01})	19.016662	0.154235	0.392728	2.065180	101.170

The table 2 above indicates that the predictive estimator had a slightly lower MSE of 0.154235 than the classical estimator with MSE of 0.156041. Also the standard error of the predictive ratio estimator was 0.392728 which is slightly lower when compared with 0.395021 of the classical estimator. The coefficient of variation for the predictive ratio estimator was 2.065180% and again slightly lower when compared with 2.091554% of the classical ratio estimator. The resulting relative efficiency of 101.17% indicates that the predictive estimator has an improvement over the classical estimator under the study design.

Discussion

The results indicate that the predictive ratio estimator provided a modest improvement over the classical ratio estimator under the sampling conditions considered in this study. The predictive estimator produced a mean that was closer to the population mean, together with a lower mean squared error (MSE), standard error, and coefficient of variation. Taken together, these results indicate that the predictive ratio estimator achieved slightly greater precision in estimating the population mean than the classical ratio estimator. The relative efficiency of 101.17%, with the classical ratio estimator taken as the 100% reference, further confirms that the gain in efficiency was positive, although relatively small.

The observed result can be understood from the role of auxiliary information in ratio estimation. Ratio estimators are particularly useful when an auxiliary variable has a strong relationship with the study variable because information on the auxiliary variable can be used to improve the estimation of the study variable. The effectiveness of such estimators is therefore closely related to the extent to which the auxiliary variable provides useful information about the variable of interest (Cochran, 1977; Särndal et al., 1992). In the present study, fertilizer application and crop yield had a correlation coefficient of 0.9758, indicating a very strong positive association between the auxiliary and study variables. This strong association provides a statistical basis for the favourable performance observed for the predictive ratio estimator, since fertilizer application contains substantial information that can be utilized in estimating crop yield.

The finding is also relevant to the distinction between the classical and predictive ratio approaches. Both estimators make use of auxiliary information, but the predictive approach is formulated within the two-stage sampling framework to make use of information obtained through the sampling process for prediction. Swain et al., (2013) in their study also demonstrates that predictive ratio estimators can be evaluated directly against classical ratio estimators in terms of relative efficiency. Thus, the slightly higher efficiency observed in the present study is consistent with the broader development of predictive ratio estimation as an approach for improving estimation when useful auxiliary information is available.

The strength of the relationship between the auxiliary and study variables is particularly important in the present result. A strong auxiliary relationship generally provides more useful information for estimation than a weak relationship. Consequently, the high correlation of 0.9758 observed in this study represents a favourable condition for the use of ratio-

type estimation. Conversely, where the association between the auxiliary and study variables is considerably weaker, the auxiliary variable would provide less information for predicting the study variable, and the efficiency advantage associated with the predictive approach may be reduced. Previous work on ratio estimation has similarly demonstrated that the performance of estimators using auxiliary information is evaluated in relation to the correlation between the study and auxiliary variables and other characteristics of the sampling design (Al-Omari, 2012). Therefore, the 101.17% relative efficiency obtained in this study is interpreted as evidence of a small advantage under the particular relationship between fertilizer application and crop yield and the sampling configuration examined, rather than as evidence that the predictive estimator will always outperform the classical ratio estimator.

In double or two-stage sampling, auxiliary information is obtained through an initial sampling phase, while more detailed information on the study variable is collected from a subsequent sample. The relative sizes of these samples affect the amount of information available for prediction and the precision with which the study variable is ultimately estimated. Consequently, a change in the allocation between the first and second phases could alter the relative performance of competing estimators. A larger first-phase sample may provide more information on the auxiliary variable, while the size of the second-phase sample determines, in part, the precision with which the study variable is measured. The appropriate balance between the two phases is therefore relevant to the efficiency achieved by a predictive estimator. The theoretical importance of auxiliary information and sampling-phase structure is well established in model-assisted and two-phase survey sampling (Särndal et al., 1992).

The findings also support the practical use of double sampling when auxiliary information is not completely available at the required sampling-unit level. Rather than requiring detailed auxiliary information for every unit before the survey, an initial phase can be used to obtain the auxiliary information, after which the study variable is observed for a second-phase sample. This approach is consistent with established two-phase sampling procedures (Hidiroglou & Särndal, 1998) and provides a useful framework for agricultural surveys.

Conclusion

The study compared classical and predictive ratio estimators in a two-stage sampling design using double sampling, with fertilizer application as auxiliary information and rice crop yield as the study variable. The predictive ratio estimator produced a mean estimate of 19.016662, closer to the population mean of 19.114 than the classical estimate of 18.886465. It also recorded lower MSE of 0.154235, standard error of 0.392728, and coefficient of variation of 2.065180% than the classical ratio estimator. Its relative efficiency of 101.17% indicates improved performance relative to the classical estimator. The predictive ratio estimator is therefore recommended for two-stage agricultural surveys where strong relationship exists between the auxiliary and study variables, as it provide an improvement in estimates of population characteristics for the sample under study.

REFERENCES

- Al-Omari, A. I. (2012). Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling. *Statistics & Probability Letters*, 82(11), 1883–1890. <https://doi.org/10.1016/j.spl.2012.07.001>
- Arbia, G., Solano-Hermosilla, G., Micale, F., Nardelli, V., & Genovese, G. (2020). Post-sampling crowdsourced data to allow reliable statistical inference: The case of food price indices in Nigeria. *International Journal of Geo-Information*, 8(11), 498–505. <https://doi.org/10.1038/s41597-023-02211-1>
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley and Sons.
- Hidiroglou, M. A., & Särndal, C. E. (1998). Use of auxiliary information for two-phase sampling. *Survey Methodology*, 24(1), 11–20. [statcan.gc.ca](https://doi.org/10.2306/1875-9463.24.1.11)
- Khan, S., Farooq, M., Ahmed, S., SidAhmed, M. M., Idriss, M. E., Hussain, E. M. M., & Albadwi, F. I. A. (2025). On use of predictive technique for estimation of population mean using auxiliary information: Application on real data sets. *European Journal of Pure and Applied Mathematics*, 18(2), 6019–6035. <https://doi.org/10.29020/nybg.ejpam.v18i2.6019>
- Kumar, M., & Saini, R. (2020). Exponential estimators for finite population mean using auxiliary attributes under predictive approach. *Revista Investigación Operacional*, 41(3), 413–420.
- Lohr, S. L. (2021). *Sampling: Design and analysis* (3rd ed.). Chapman and Hall/CRC.
- Michael, C. U., & Mbanefo, S. M. (2022). Two-stage cluster sampling with unequal probability sampling in the first stage and ranked set sampling in the second stage. *Statistics in Transition New Series*, 23(3), 199–214. <https://doi.org/10.2478/stattrans-2022-0038>
- Ozturk, O. (2019). Two-stage cluster samples with ranked set sampling designs. *Annals of the Institute of Statistical Mathematics*, 71, 63–91. <https://doi.org/10.1007/s10463-017-0623-z>
- Panda, K. B. (2015). An efficient predictive approach to estimation in two-phase sampling. *IOSR Journal of Mathematics*, 11(3), 78–81.
- Saini, M. A. (2013). A class of predictive estimators in two-stage sampling when auxiliary character is estimated at SSU level. *International Journal of Pure and Applied Mathematics*, 85(2), 285–295. <https://doi.org/10.12732/ijpam.v85i2.8>
- Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. Springer.
- Swain, A. K. P. C., Mishra, S. S., and Pradhan, B. K. (2013). On some predictive ratio type estimators in two stage sampling. *Journal of Statistical Theory and Applications*, 12(1), 83–91. <https://doi.org/10.2991/jsta.2013.12.1.7>
- Yadav, S.K., Mishra, S.S., Kumar, S., Kumar, R., (2018). New Efficient Predictive Estimator of Population Mean using Median of study variable, *Jnanabha*, 48 (1), 17-25.
- Yadav, S. K., & Mishra, P. (2015). Upgraded ratio-cum-product type predictive estimators to estimate the population mean. *Statistika: Statistics and Economy Journal*, 95(1), 76–85. [csu.gov.cz](https://doi.org/10.2478/s11226-015-0000-0)
- Yadav, S. K., Shabbir, J., Pandey, S. K., & Gupta, S. (2018). Predictive estimation of finite population mean using coefficient of kurtosis and median of an auxiliary variable under simple random sampling scheme. *International Journal of Mathematical Archive*, 9(5), 137–144.

