International Journal of Scientific Reports
Kumar S et al. Int J Sci Rep. 2023 Oct;9(10):314-325
http://www.sci-rep.com
pISSN 2454-2156 | eISSN 2454-2164

DOI: https://dx.doi.org/10.18203/issn.2454-2156.IntJSciRep20232910

Original Research Article

Chain ratio and product estimators for population mode using two-phase sampling scheme

Sanjay Kumar*, Akhilesh Yadav

Department of Statistics, Central University of Rajasthan, Bandarsindri, Kishangarh, Ajmer, Rajasthan, India

Received: 10 July 2023
Revised: 20 August 2023
Accepted: 21 August 2023

***Correspondence:**
Sanjay Kumar,
E-mail: sanjay.kumar@curaj.ac.in

ABSTRACT

Background: For skewed datasets, the mode is utilized as a more appropriate measure of location. We formed chain ratio and product estimators for the population mode using two types of auxiliary information under the two-phase sampling scheme.

Methods: Expressions for biases and mean square errors for the formed estimators up to the first order of approximation are obtained. The confidence intervals of the estimators are obtained and the aspects related to fixed cost and fixed variance are also studied. A simulation study is performed to support the theoretical outcomes. A real dataset is also provided which was collected from the department of agriculture, United States.

Results: The simulation results for the fixed first-phase sample size 3450 and the second-phase sample size 90 show that the mean square error is 0.035, the bias is 0.012, the confidence interval is 2.67-3.15, and the cost of the survey under the fixed variance is ₹ 194.67 of the proposed estimator ($\tilde{T}_R$), which is lower than 0.039, 0.055, 2.70-3.20, and ₹ 247.37 of the ratio estimator ($\tilde{t}_r$) and 0.078, 0.041, 2.57-3.31, and ₹ 268.73 of the naive estimator ($\tilde{y}$).

Conclusions: The simulation results show that the proposed estimator ($\tilde{T}_R$) performs better than the ratio estimator ($\tilde{t}_r$) and the naive estimator ($\tilde{y}$).

Keywords: Mode estimation, Auxiliary information, Cost aspects, Mean square error, Confidence interval, Simulation study

INTRODUCTION

In sample surveys, surveyors often find that they are working with data such as income, abortions, drugs, AIDS, etc., that have skewed distributions. It is more appropriate to consider mode as a measure of location than the mean or the median in forecasting ready-made products, such as garments, shoes, etc., in manufacturing. The work presented here is intended to be helpful for social scientists, psychologists, demographers, business, and economics, where the mode is regularly used in practice. Y is a study variable with the population mean $\overline{Y} = N^{-1}\sum_{i=1}^{N} Y_i$, population mode $\tilde{Y}$, population median M_y, probability density function $f_Y(y)$, and cumulative distribution functions $F_Y(y)$, X is an auxiliary variable with the population mean $\overline{X} = N^{-1}\sum_{i=1}^{N} X_i$, population mode $\tilde{X}$, population median M_x, probability density function $f_X(x)$, and cumulative distribution function $F_X(x)$, and Z is an additional auxiliary variable with the population mean $\overline{Z} = N^{-1}\sum_{i=1}^{N} Z_i$, population mode $\tilde{Z}$, population median M_z, probability density function $f_Z(z)$, and cumulative distribution function $F_Z(z)$. ρ_{yx}, ρ_{xz}, and ρ_{yz} are the correlation coefficients among the study variable Y and the auxiliary variable X, and the additional auxiliary variable Z. Chand et al and Guha et al introduced chain estimators for the population mean and population total in the case when the population mean $\overline{X}$ of the auxiliary variable X, is not available, but the population mean $\overline{Z}$ or attribute of the additional auxiliary variable Z, closely related to X, is available, which may be cheaper and less correlated to the study variable Y.[1,4]

Table 1: Matrices of proportions.

Auxiliary variable X\ Study variable Y	**Study variable $Y \leq$ Population median M_y**	**Study variable $Y >$ Population median M_y**	**Total**
Auxiliary variable $X \leq$ Population median M_x	${}^{y}_{x}p_{11}$	${}^{y}_{x}p_{21}$	${}^{y}_{x}p_{.1}$
Auxiliary variable $X >$ Population median M_x	${}^{y}_{x}p_{12}$	${}^{y}_{x}p_{22}$	${}^{y}_{x}p_{.1}$
Total	${}^{y}_{x}p_{1.}$	${}^{y}_{x}p_{2.}$	1
Auxiliary variable Z\ Study variable Y	Study variable $Y \leq$ Population median M_y	Study variable $Y >$ Population median M_y	Total
Auxiliary variable $Z \leq$ Population median M_z	${}^{y}_{z}p_{11}$	${}^{y}_{z}p_{21}$	${}^{y}_{z}p_{.1}$
Auxiliary variable $Z >$ Population median M_z	${}^{y}_{z}p_{12}$	${}^{y}_{z}p_{22}$	${}^{y}_{z}p_{.1}$
Total	${}^{y}_{z}p_{1.}$	${}^{y}_{z}p_{2.}$	1
Auxiliary variable X\ Auxiliary variable Z	Auxiliary variable $Z \leq$ Population median M_z	Auxiliary variable $Z >$ Population median M_z	Total
Auxiliary variable $X \leq$ Population median M_x	${}^{z}_{x}p_{11}$	${}^{z}_{x}p_{21}$	${}^{z}_{x}p_{.1}$
Auxiliary variable $X >$ Population median M_x	${}^{z}_{x}p_{12}$	${}^{z}_{x}p_{22}$	${}^{z}_{x}p_{.1}$
Total	${}^{z}_{x}p_{1.}$	${}^{z}_{x}p_{2.}$	1

Similarly, for estimating the population mode, there may be a case when the population mode $\tilde{X}$ of the auxiliary variable X, is not available, but the population mode $\tilde{Z}$ of the additional auxiliary variable Z, closely related to X, is available, which may be cheaper and less correlated to the study variable Y. In such a situation, the unknown $\tilde{X}$ can be estimated using a two-phase sampling scheme invented by Neyman.[3] The collection of information on X is cheaper, so, a large preliminary sample of size n′ selected from the population using simple random sampling without replacement, is considered for collecting information on X and Z for estimating $\tilde{X}$ as $\hat{\tilde{X}} = \frac{\tilde{x}'}{\tilde{z}'}\tilde{Z}$, where $\tilde{x}'$ is the first-phase sample mode of X, and $\tilde{z}'$ is the first-phase sample mode of Z. A sub-sample of size n selected from the first-phase sample using simple random sampling without replacement is further used for noting both the variables Y and X. y_i is the second-phase sample measurements on Y with sample mode $\tilde{y}$, and sample median $\widehat{M}_y$. x_i is the second-phase sample measurements on X with sample mode $\tilde{x}$, and sample median $\widehat{M}_x$. We define:

$$\bar{y} = n^{-1}\sum_{i=1}^{n} y_i,\ \bar{x} = n^{-1}\sum_{i=1}^{n} x_i,\ \bar{z} = n^{-1}\sum_{i=1}^{n} z_i,$$

$$\bar{x}' = n'^{-1}\sum_{i=1}^{n'} x_i,\ \bar{z}' = n'^{-1}\sum_{i=1}^{n'} z_i,$$

$$s_y^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i-\bar{y})^2,\ s_x^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

$$s_z^2 = (n-1)^{-1}\sum_{i=1}^{n}(z_i-\bar{z})^2,\ S_y^2 = (N-1)^{-1}\sum_{i=1}^{N}(Y_i-\bar{Y})^2,$$

$$S_x^2 = (N-1)^{-1}\sum_{i=1}^{N}(X_i-\bar{X})^2,\ S_z^2 = (N-1)^{-1}\sum_{i=1}^{N}(Z_i-\bar{Z})^2,$$

$$S_{yx} = (N-1)^{-1}\sum_{i=1}^{N}(Y_i-\bar{Y})(X_i-\bar{X})$$

$$S_{zx} = (N-1)^{-1}\sum_{i=1}^{N}(Z_i-\bar{Z})(X_i-\bar{X})$$

$$S_{yz} = (N-1)^{-1}\sum_{i=1}^{N}(Y_i-\bar{Y})(Z_i-\bar{Z}),$$

$$Q_y(p) = \mathrm{Inf}(p \leq F(y), y \in \mathbb{R}),\ Q_x(p) = \mathrm{Inf}(p \leq F(x), x \in \mathbb{R}),$$

$$Q_z(p) = \mathrm{Inf}(p \leq F(z), z \in \mathbb{R})$$

Where $\mathbb{R}$ is the real number, and p is the p^{th} percentile. List the Y values of the second-phase sample units in ascending order as $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$. Suppose $p = \frac{I_0}{n}$ be the proportion of $y\ (\leq M_y)$ values $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$, where both the values p, and M_y are not known and I_0 is an integer such that $y_{I_0} \leq M_y \leq y_{I_0+1}$. Hence, M_y is approximately the sample p^{th} quantile $\widehat{Q}_y(p)$. Here, both p, and M_y can be estimated from the sample. p is estimated by $\hat{p}$ and hence $\widehat{Q}_y(\hat{p})$ is the estimator of $Q_y(p)$. $\widehat{M}_y$ can be regarded as the special estimator $\widehat{Q}_y(\hat{p})$ such that $\widehat{M}_y = \widehat{Q}_y(0.5)$. Kuk et al studied matrices of proportions ${}^{y}_{x}p_{ij}$, ${}^{y}_{z}p_{ij}$, and ${}^{z}_{x}p_{ij}$, which are given in (Table 1).[5]

With $N \to \infty$ under a super-population model, the distributions of the trivariate variables Y, X, and Z become continuous distributions with marginal densities $f_Y(y)$, $f_X(x)$, and $f_Z(z)$. Gross showed that the sample median $\widehat{M}_y$ is consistent and asymptotically normal with mean M_y and asymptotic variance.[3]

$$V(\widehat{M}_y) = \left(\frac{1-f}{4n}\right)\left(f_Y(M_y)\right)^{-2},$$

where $f = \frac{n}{N}$. When the distribution is moderately asymmetric, Doodson showed an empirical (also known as Karl Pearson) relationship between mean, median, and mode as Mode ≅ 3 × Median − 2 × Mean.[2]

That is, $\tilde{Y} \cong 3M_y - 2\bar{Y}, \tilde{X} \cong 3M_x - 2\bar{X}$, and $\tilde{Z} \cong 3M_z - 2\bar{Z}$. The naive estimators for $\tilde{Y}, \tilde{X}$, and $\tilde{Z}$ based on the second-phase sample are given by $\tilde{y} \cong 3\widehat{M}_y - 2\bar{y}$, $\tilde{x} \cong 3\widehat{M}_x - 2\bar{x}$, and $\tilde{z} \cong 3\widehat{M}_z - 2\bar{z}$. Based on the first-phase sample, we defined estimators for $\tilde{X}$ and $\tilde{Z}$ as $\tilde{x}' \cong 3\widehat{M}'_x - 2\bar{x}'$ and $\tilde{z}' \cong 3\widehat{M}'_z - 2\bar{z}'$, where $\widehat{M}'_x$ and $\widehat{M}'_z$ are the sample medians based on the first-phase sample of the size n'.

Table 2: Descriptive parameters of the generated dataset.

Variables	Mean	Median	Mode	Minimum	Maximum	First quartile	Third quartile	Standard Deviation
Study variable *Y*	3.17	3.08	2.89	0.34	8.62	2.42	3.80	1.05
Auxiliary variable *X*	3.18	3.08	2.88	0.46	8.77	2.40	3.82	1.08
Auxiliary variable *Z*	3.19	3.07	2.83	0.59	8.88	2.39	3.84	1.12

We define indicator functions I_{y_i} for Y, I_{x_i} for X and I_{z_i} for Z, such that

$$I_{y_i} = \begin{cases} 1, \text{if } Y_i \le M_y \\ 0, \text{otherwise} \end{cases}, I_{x_i} = \begin{cases} 1, \text{if } X_i \le M_x \\ 0, \text{otherwise} \end{cases}, \text{and } I_{z_i} = \begin{cases} 1, \text{if } Z_i \le M_z \\ 0, \text{otherwise} \end{cases}.$$

The variance and covariance of naive estimators $\tilde{y}, \tilde{x}$, and $\tilde{z}$ may be approximated as:

$$V(\tilde{y}) \cong \left(\frac{1-f}{n}\right)\tilde{V}_y^2 = \left(\frac{1-f}{n}\right)\left(\frac{9}{4}\left(f_Y(M_y)\right)^{-2} + 4S_y^2 + 12S_{yM_y}\left(f_Y(M_y)\right)^{-1}\right)$$

$$V(\tilde{x}) \cong \left(\frac{1-f}{n}\right)\tilde{V}_x^2 = \left(\frac{1-f}{n}\right)\left(\frac{9}{4}\left(f_X(M_x)\right)^{-2} + 4S_x^2 + 12S_{xM_x}\left(f_X(M_x)\right)^{-1}\right),$$

$$V(\tilde{z}) \cong \left(\frac{1-f}{n}\right)\tilde{V}_z^2 = \left(\frac{1-f}{n}\right)\left(\frac{9}{4}\left(f_Z(M_z)\right)^{-2} + 4S_z^2 + 12S_zM_z\left(f_Z(M_z)\right)^{-1}\right),$$

$$Cov(\tilde{y},\tilde{x}) \cong \left(\frac{1-f}{n}\right)\tilde{V}_{yx} = \left(\frac{1-f}{n}\right)\left(9\left(f_X(M_x)\,f_Y(M_y)\right)^{-1}(P_{11} - 0.25) + 6S_{yM_x}\left(f_X(M_x)\right)^{-1} + 6S_{xM_y}\left(f_Y(M_y)\right)^{-1} + 4S_{yx}\right)$$

$$Cov(\tilde{y},\tilde{z}) \cong \left(\frac{1-f}{n}\right)\tilde{V}_{yz} = \left(\frac{1-f}{n}\right)\left(9\left(f_Z(M_z)f_Y(M_y)\right)^{-1}(Q_{11} - 0.25) + 6S_{yM_z}\left(f_Z(M_z)\right)^{-1} + 6S_{zM_y}\left(f_Y(M_y)\right)^{-1} + 4S_{yz}\right),$$

$$Cov(\tilde{z},\tilde{x}) \cong \left(\frac{1-f}{n}\right)\tilde{V}_{zx} = \left(\frac{1-f}{n}\right)\left(9\left(f_X(M_x)\,f_Z(M_z)\right)^{-1}(P_{11} - 0.25) + 6S_{zM_x}\left(f_X(M_x)\right)^{-1} + 6S_{xM_z}\left(f_Z(M_z)\right)^{-1} + 4S_{zx}\right),$$

$$S_{yM_x} = (N-1)^{-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(I_{x_i} - 0.5),\ S_{xM_y} = (N-1)^{-1}\sum_{i=1}^{N}(X_i - \bar{X})(I_{y_i} - 0.5),\ S_{yM_y} = (N-1)^{-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(I_{y_i} - 0.5),\ S_{yM_z} = (N-1)^{-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(I_{z_i} - 0.5),\ S_{zM_y} = (N-1)^{-1}\sum_{i=1}^{N}(Z_i - \bar{Z})(I_{y_i} - 0.5),\ S_{zM_z} = (N-1)^{-1}\sum_{i=1}^{N}(Z_i - \bar{Z})(I_{z_i} - 0.5),\ S_{zM_x} = (N-1)^{-1}\sum_{i=1}^{N}(Z_i - \bar{Z})(I_{x_i} - 0.5),\ S_{xM_z} = (N-1)^{-1}\sum_{i=1}^{N}(X_i - \bar{X})(I_{z_i} - 0.5), \text{ and } S_{xM_x} = (N-1)^{-1}\sum_{i=1}^{N}(X_i - \bar{X})\,(I_{x_i} - 0.5).$$

Lamichhane et al proposed a naive ratio estimator of $\tilde{Y}$ using the two-phase sampling scheme as;[7]

$$\tilde{t}_r = \frac{\tilde{y}}{\tilde{x}}\tilde{x}'$$

Similarly, a naive product estimator can be defined as

$$\tilde{t}_p = \frac{\tilde{y}}{\tilde{x}'}\tilde{x},$$

Where the expressions for biases and mean square errors of the estimators $\tilde{t}_r$ and $\tilde{t}_p$ are given as:

$$B(\tilde{t}_r) \cong \tilde{Y}\left(\frac{1}{n} - \frac{1}{n'}\right)\left(\frac{\tilde{V}_x^2}{\tilde{X}^2} - \frac{\tilde{V}_{yx}}{\tilde{X}\tilde{Y}}\right)$$

$$MSE(\tilde{t}_r) \cong \left(\frac{1}{n'} - \frac{1}{N}\right)\tilde{V}_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\left(\tilde{V}_y^2 + R_1^2\tilde{V}_x^2 - 2R_1\tilde{V}_{yx}\right)$$

$$B(\tilde{t}_p) \cong \left(\frac{1}{n} - \frac{1}{n'}\right)\frac{\tilde{V}_{yx}}{\tilde{X}},$$

$$MSE(\tilde{t}_p) \cong \left(\frac{1}{n'} - \frac{1}{N}\right)\tilde{V}_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\left(\tilde{V}_y^2 + R_1^2\tilde{V}_x^2 + 2R_1\tilde{V}_{yx}\right)$$

Where $R_1 = \frac{\tilde{Y}}{\tilde{X}}$ is the ratio of two population modes of Y and X. Kumar et al studied naive ratio and product estimators under the two-phase sampling scheme for estimating the population mode using the information on a single auxiliary variable.[6]

METHODS

Proposed estimators and their properties

Following Lamichhane and Singh[7] and Chand[1], we have suggested chain ratio and product estimators of the population mode $\tilde{Y}$ of the study variable Y using two auxiliary variables X and Z where X is considered as the main auxiliary variable and Z, which is closer to X, as the additional auxiliary variable.

$$\tilde{T}_R = \tilde{y}\left(\frac{\tilde{x}'}{\tilde{x}}\right)\left(\frac{\tilde{z}}{\tilde{z}'}\right)$$

$$\tilde{T}_P = \tilde{y}\left(\frac{\tilde{x}}{\tilde{x}'}\right)\left(\frac{\tilde{z}'}{\tilde{z}}\right)$$

To obtain the properties of the suggested estimators, we defined

$$\tilde{y} = \tilde{Y}(1+\epsilon_0), \tilde{x} = \tilde{X}(1+\epsilon_1), \tilde{x}' = \tilde{X}(1+\epsilon_2), \tilde{z}' = \tilde{Z}(1+\epsilon_3),$$

Where $E(\epsilon_0) = E(\epsilon_1) = E(\epsilon_2) = E(\epsilon_3) = 0$

One can only see that:

$$E(\epsilon_0^2) = \left(\frac{1}{n}-\frac{1}{N}\right)\frac{\tilde{V}_y^2}{\tilde{Y}^2}, E(\epsilon_1^2) = \left(\frac{1}{n}-\frac{1}{N}\right)\frac{\tilde{V}_x^2}{\tilde{X}^2}, E(\epsilon_2^2) = \left(\frac{1}{n'}-\frac{1}{N}\right)\frac{\tilde{V}_x^2}{\tilde{X}^2}, E(\epsilon_3^2) = \left(\frac{1}{n'}-\frac{1}{N}\right)\frac{\tilde{V}_z^2}{\tilde{Z}^2},$$
$$E(\epsilon_0\epsilon_1) = \left(\frac{1}{n}-\frac{1}{N}\right)\frac{\tilde{V}_{yx}}{\tilde{Y}\tilde{X}}, E(\epsilon_0\epsilon_2) = \left(\frac{1}{n'}-\frac{1}{N}\right)\frac{\tilde{V}_{yx}}{\tilde{Y}\tilde{X}},$$
$$E(\epsilon_0\epsilon_3) = \left(\frac{1}{n'}-\frac{1}{N}\right)\frac{\tilde{V}_{yz}}{\tilde{Y}\tilde{Z}},$$
$$E(\epsilon_1\epsilon_2) = \left(\frac{1}{n'}-\frac{1}{N}\right)\frac{\tilde{V}_x^2}{\tilde{X}^2}, (\epsilon_1\epsilon_3) = \left(\frac{1}{n'}-\frac{1}{N}\right)\frac{\tilde{V}_{xz}}{\tilde{X}\tilde{Z}}, \text{ and}$$
$$E(\epsilon_2\epsilon_3) = \left(\frac{1}{n'}-\frac{1}{N}\right)\frac{\tilde{V}_{xz}}{\tilde{X}\tilde{Z}},$$

The expressions for biases and mean square errors of the proposed estimators $\tilde{T}_R$ and $\tilde{T}_P$ are given as

$$B(\tilde{T}_R) \cong \tilde{Y}\left(\left(\frac{1}{n}-\frac{1}{n'}\right)\left(\frac{\tilde{V}_x^2}{\tilde{X}^2}-\frac{\tilde{V}_{yx}}{\tilde{X}\tilde{Y}}\right)+\left(\frac{1}{n'}-\frac{1}{N}\right)\left(\frac{\tilde{V}_z^2}{\tilde{Z}^2}-\frac{\tilde{V}_{yz}}{\tilde{Y}\tilde{Z}}\right)\right)$$
$$= B(\tilde{t}_r) + \tilde{Y}\left(\frac{1}{n'}-\frac{1}{N}\right)\left(\frac{\tilde{V}_z^2}{\tilde{Z}^2}-\frac{\tilde{V}_{yz}}{\tilde{Y}\tilde{Z}}\right)$$

$$MSE(\tilde{T}_R) \cong \left(\frac{1}{n'}-\frac{1}{N}\right)\tilde{V}_y^2 + \left(\frac{1}{n}-\frac{1}{n'}\right)\left(\tilde{V}_y^2 + R_1^2\tilde{V}_x^2 - 2R_1\tilde{V}_{yx}\right) + \left(\frac{1}{n'}-\frac{1}{N}\right)R_2\left(R_2\tilde{V}_z^2 - 2\tilde{V}_{yz}\right)$$

$$\cong MSE(\tilde{t}_r) + \left(\frac{1}{n'}-\frac{1}{N}\right)R_2\left(R_2\tilde{V}_z^2 - 2\tilde{V}_{yz}\right)$$

$$B(\tilde{T}_P) \cong \left(\frac{1}{n}-\frac{1}{n'}\right)\frac{\tilde{V}_{yx}}{\tilde{X}} + \left(\frac{1}{n'}-\frac{1}{N}\right)\frac{\tilde{V}_{yz}}{\tilde{Z}} = B(\tilde{t}_P) + \left(\frac{1}{n'}-\frac{1}{N}\right)\frac{\tilde{V}_{yz}}{\tilde{Z}}$$

$$MSE(\tilde{T}_P) \cong \left(\frac{1}{n'}-\frac{1}{N}\right)\tilde{V}_y^2 + \left(\frac{1}{n}-\frac{1}{n'}\right)\left(\tilde{V}_y^2 + R_1^2\tilde{V}_x^2 + 2R_1\tilde{V}_{yx}\right) + \left(\frac{1}{n'}-\frac{1}{N}\right)R_2\left(2\tilde{V}_{yz} + R_2\tilde{V}_z^2\right)$$

$$\cong \mathrm{MSE}(\tilde{t}_p) + \left(\frac{1}{n'}-\frac{1}{N}\right)R_2\left(2\tilde{V}_{yz} + R_2\tilde{V}_z^2\right),$$

Where $R_2 = \frac{\tilde{Y}}{\tilde{Z}}$, is the ratio of two population modes of Y and Z.

Comparison of estimators

In this section, conditions for which the proposed estimators have lower mean square errors than the other relevant estimators are obtained.

$$MSE(\tilde{T}_R) \le V(\tilde{y}), K \ge \frac{1}{2}\left(1+\frac{f_1R_2}{R_1^2\tilde{V}_x^2}\left(R_2\,\tilde{V}_z^2 - 2\tilde{V}_{yz}\right)\right),$$

$$MSE(\tilde{T}_R) \le MSE(\tilde{t}_r) \text{ if } \tilde{V}_{yz} \ge \frac{1}{2}\left(R_2\tilde{V}_z^2\right),$$
$$MSE(\tilde{T}_P) \le V(\tilde{y}) \text{ if } K \le -\frac{1}{2}\left(1+\frac{f_1R_2}{R_1^2\tilde{V}_x^2}\left(R_2\,\tilde{V}_z^2 + 2\tilde{V}_{yz}\right)\right) \textit{and } MSE(\tilde{T}_P) \le MSE(\tilde{t}_p) \text{ if } \tilde{V}_{yz} \le -\frac{1}{2}R_2\tilde{V}_z^2$$

Where $= \rho_{\tilde{y}\tilde{x}}\frac{C_{\tilde{y}}}{C_{\tilde{x}}}$, $f_1 = \frac{\left(\frac{1}{n'}-\frac{1}{N}\right)}{\left(\frac{1}{n}-\frac{1}{n'}\right)}$, $\tilde{V}_{yx}$, and $\tilde{V}_{yz}$ are covariance terms.

A simulation study with a generated dataset

In this section, we have performed a simulation study with one generated dataset (considered as a population). We generated an artificial dataset by assuming the size $N = 5000$ of independent Gamma variables $Z_i \sim G(N, 8.00, 2.50)$. We estimated the main auxiliary variable $X_i = 0.15 + 0.90Z_i + 0.20e_1$, where $e_1 \sim N(0,1)$. Finally, we estimated the study variable using a linear relation $Y_i = 0.15 + 0.90X_i + 0.20e_2$ where $e_2 \sim N(0,1)$. The fitted distributions of the study and the auxiliary variables are given in Figure 1 (a) for the generated dataset, and we obtained the parametric estimates of Gamma distribution for the study and the auxiliary variables. We also calculated various descriptive parameters of the study and the auxiliary variables in Table 2. The correlation coefficients among the study and the auxiliary variables are $\rho_{yx} = 0.99$, $\rho_{yz} = 0.96$, and $\rho_{xz} = 0.98$ for the dataset, which are quite good for our study. The value of ${}_{x}^{y}p_{11} = 0.470$, ${}_{z}^{y}p_{11} = 0.462$, and ${}_{z}^{x}p_{11} = 0.472$ for the data set are also observed. The correlation coefficients among sample modes are $\rho_{\tilde{y}\tilde{x}} = 0.76$, $\rho_{\tilde{y}\tilde{z}} = 0.54$, and $\rho_{\tilde{x}\tilde{z}} = 0.74$ for the dataset. The relative efficiencies (REs) of the estimators with respect to the naive estimator of the population mode are calculated as:

$$RE(\tilde{y}) = \frac{V(\tilde{y})}{V(\tilde{y})}\times 100\%, RE(\tilde{t}_r) = \frac{V(\tilde{y})}{MSE(\tilde{t}_r)}\times 100\%, \text{ and}$$
$$RE(\tilde{T}_R) = \frac{V(\tilde{y})}{MSE(\tilde{T}_R)}\times 100\%$$

In our case, the subsequent sampling scheme is simple random sampling without replacement, so the possible number of samples is N_{C_n}, which is too large. So, we selected $M = 10{,}000$ samples randomly, each of the size n. We computed simulated mean square errors, simulated biases (Bs), simulated relative efficiencies (REs), and ratios (R) of approximate expressions of mean square errors to the simulated mean square errors in (Table 3) for different sizes of the second-phase sample $n = 90, 180, 360, 540, 900$, and 1800 at the fixed size $n' =$

3450 of the first-phase sample. We have shown simulated mean square errors and biases through graphical representations in (Figure 1).

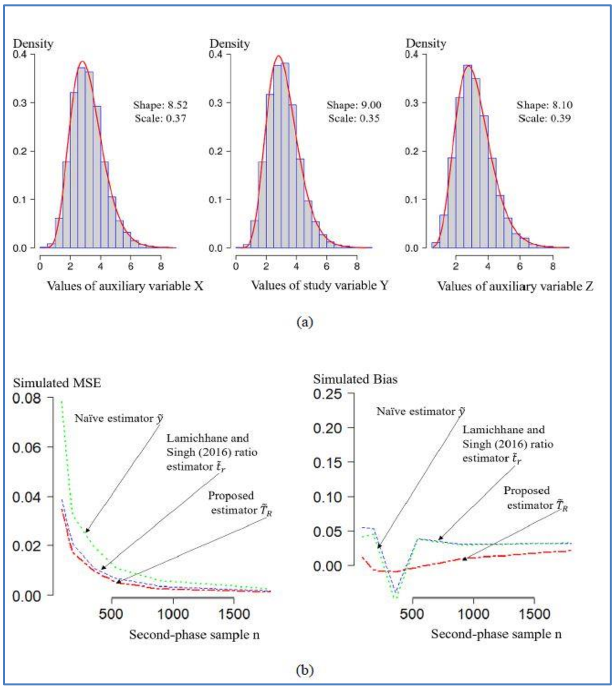

Figure 1: (a) Parameter values for the generated dataset, (b) simulated values of mean square errors and biases of the estimators for generated dataset.

The simulated mean square errors of the estimators of the population mode are given as:

$$MSE(\tilde{y}) = \frac{1}{M}\sum_{k=1}^{M}\left(\tilde{y}_{|k} - \tilde{Y}\right)^2, \mathrm{MSE}(\tilde{t}_r) = \frac{1}{M}\sum_{k=1}^{M}\left(\tilde{t}_{r|k} - \tilde{Y}\right)^2$$

$$MSE\left(\tilde{T}_R\right) = \frac{1}{M}\sum_{k=1}^{M}\left(\tilde{T}_{R|k} - \tilde{Y}\right)^2.$$

The simulated relative efficiencies of the estimators with respect to the naive estimator of the population mode are given as:

$$RE(\tilde{y}) = \frac{\sum_{k=1}^{M}(\tilde{y}_{|k}-\tilde{Y})^2}{\sum_{k=1}^{M}(\tilde{y}_{|k}-\tilde{Y})^2} \times 100\%\ , (\tilde{t}_r) = \frac{\sum_{k=1}^{M}(\tilde{y}_{|k}-\tilde{Y})^2}{\sum_{k=1}^{M}(\tilde{t}_{r|k}-\tilde{Y})^2} \times 100\%\ RE\left(\tilde{T}_R\right) = \frac{\sum_{k=1}^{M}(\tilde{y}_{|k}-\tilde{Y})^2}{\sum_{k=1}^{M}(\tilde{T}_{R|k}-\tilde{Y})^2} \times 100\%.$$

The simulated biases of the estimators $\tilde{T}_R$, $\tilde{t}_r$, and $\tilde{y}$ are given as:

$$B(\tilde{y}) = \frac{1}{M}\sum_{k=1}^{M}\left(\tilde{y}_{|k} - \tilde{Y}\right), B(\tilde{t}_r) = \frac{1}{M}\sum_{k=1}^{M}\left(\tilde{t}_{r|k} - \tilde{Y}\right), \text{ and } B\left(\tilde{T}_R\right) = \frac{1}{M}\sum_{k=1}^{M}\left(\tilde{T}_{R|k} - \tilde{Y}\right)$$

For the investigations of how far approximate variances are from the simulated mean square errors, we computed the three ratios (R) for generated data set given as:

$$\mathrm{R}(\tilde{y}) = \frac{\mathrm{V}(\tilde{y})}{\frac{1}{M}\sum_{i=1}^{M}(\tilde{y}_{|k}-\tilde{Y})^2}\ \mathrm{R}(\tilde{t}_r) = \frac{\mathrm{MSE}(\tilde{t}_r)}{\frac{1}{M}\sum_{i=1}^{M}(\tilde{t}_{r|k}-\tilde{Y})^2}, \mathrm{R}\left(\tilde{T}_R\right) = \frac{\mathrm{MSE}(\tilde{T}_R)}{\frac{1}{M}\sum_{i=1}^{M}(\tilde{T}_{R|k}-\tilde{Y})^2}.$$

RESULTS

From (Table 3, Figure 1b) we observe that the values of mean square errors of the suggested ratio estimator are less than Lamichhane et al ratio estimator and naive estimator.[7] Also, the mean square errors decrease when the second-phase sample size increases. The values of biases of the estimators are very low and close to zero. The ratios of the exact mean square errors to the simulated mean square errors are close to one, indicating that simulated mean square errors are approximately close to exact mean square errors. It means that exact mean square errors can also be used as simulated mean square errors. In (Table 3) we also computed the exact values of mean square errors and exact biases. The values of mean square errors of the suggested chain ratio estimator are less than that of other than Lamichhane and Singh[7] ratio estimator and naive estimator.

An application

To validate the theoretical conclusions with a real dataset, we have taken a dataset from the Department of Agriculture, United States.[9] This dataset was collected during the years 2003-2008. This dataset represents the price (US $)/centum weight (Cwt) of sweet corn in the year 2005 as the study variable Y, the price (US $)/Cwt of sweet corn in the year 2004 as the main auxiliary variable X, and price (US $)/Cwt of sweet corn in the year 2003 as the additional auxiliary variable Z. The correlation coefficients among the study and the auxiliary variables are $\rho_{yx} = 0.95$, $\rho_{yz} = 0.92$, and $\rho_{xz} = 0.93$ for the data set, which are acceptable for our study. We also observe that the value of ${}_{x}^{y}p_{11} = 0.48$, ${}_{z}^{y}p_{11} = 0.41$, and ${}_{z}^{x}p_{11} = 0.41$. The correlation coefficients among sample modes are $\rho_{\tilde{y}\tilde{x}} = 0.87, \rho_{\tilde{y}\tilde{z}} = 0.51$, and $\rho_{\tilde{x}\tilde{z}} = 0.55$ for the dataset. We calculated various descriptive parameters of the study and the auxiliary variables listed in (Table 4). We fitted an exponential distribution, a Gamma distribution, and a Weibull distribution to each of the variables used in this study, which are given in (Figure 2a) shows that the Gamma distribution gives the best fit for the dataset, and

we approximately get that $Y_i{\sim}G(14.56, 1.71)$, $X_i{\sim}G(10.38, 2.21)$, and $Z_i{\sim}G(10.39, 2.14)$.

Table 3: Simulated and exact values of the estimators for different values of n at the fixed first-phase sample size $n' = 3450$ for the generated dataset.

Simulated values of the estimators						
Parameters	**Second-phase sample size**					
	90	**180**	**360**	**540**	**900**	**1800**
Relative efficiency of proposed estimator $\tilde{T}_R$	225	189	204	218	215	192
Mean square error of proposed estimator $\tilde{T}_R$	0.035	0.017	0.010	0.005	0.003	0.001
Bias of proposed estimator $\tilde{T}_R$	0.012	-0.007	-0.010	-0.003	0.009	0.021
Ratios of the exact mean square error to the simulated mean square error of proposed estimator $\tilde{T}_R$	1.09	1.09	0.95	1.18	1.25	1.02
Relative efficiency of Lamichhane and Singh[7] estimator $\tilde{t}_r$	203	158	178	162	160	139
Mean square error of Lamichhane and Singh[7] estimator $\tilde{t}_r$	0.039	0.021	0.011	0.007	0.004	0.002
Bias of Lamichhane and Singh[7] estimator $\tilde{t}_r$	0.055	0.053	-0.040	0.039	0.031	0.031
Ratios of the exact mean square error to the simulated mean square error of Lamichhane and Singh[7] estimator $\tilde{t}_r$	0.98	0.92	0.85	0.91	0.99	0.85
Relative efficiency of naive estimator $\tilde{y}$	100	100	100	100	100	100
Variance of naive estimator $\tilde{y}$	0.078	0.032	0.019	0.011	0.006	0.003
Bias of naive estimator $\tilde{y}$	0.041	0.045	-0.053	0.037	0.028	0.032
Ratios of the exact variance to the simulated variance of naive estimator $\tilde{y}$	0.95	1.13	0.91	1.06	1.11	0.97
Exact values of the estimators	**Second-phase sample size**					
	90	**180**	**360**	**540**	**900**	**1800**
Relative efficiency of proposed estimator $\tilde{T}_R$	197	197	196	195	192	183
Mean square error of proposed estimator $\tilde{T}_R$	0.038	0.019	0.009	0.006	0.003	0.001
Bias of proposed estimator $\tilde{T}_R$	0.007	0.004	0.002	0.001	0.001	0.000
Relative efficiency of Lamichhane and Singh[7] estimator $\tilde{t}_r$	196	195	191	188	181	159
Mean square error of Lamichhane and Singh[7] estimator $\tilde{t}_r$	0.038	0.019	0.009	0.006	0.003	0.002
Bias of Lamichhane and Singh[7] estimator $\tilde{t}_r$	0.007	0.004	0.002	0.001	0.001	0.000
Relative efficiency of naive estimator $\tilde{y}$	100	100	100	100	100	100
Variance of naive estimator $\tilde{y}$	0.075	0.037	0.018	0.011	0.006	0.002
Bias of naive estimator $\tilde{y}$	0	0	0	0	0	0

Table 4: Various descriptive parameters for the real dataset.

Variables	Mean	Median	Mode	Minimum	Maximum	First quartile	Third quartile	Standard Deviation
Study variable *Y*	24.88	22.20	16.83	13.70	41.50	20.10	28.00	6.98
Auxiliary Variable *X*	22.91	20.80	16.57	11.70	42.00	18.40	27.00	7.62
Auxiliary variable *Z*	22.28	20.60	17.23	9.10	42.00	17.00	25.00	7.27

With the help of this real dataset, we have carried out a simulation study using R software. In our case, the sampling scheme used is simple random sampling without replacement, so the possible number of samples is N_{C_n}, which is too large. So, we selected $M = 10{,}000$ samples randomly, each of the size n. In Table 5, we computed simulated mean square errors using Eq. (34), simulated biases using Eq. (31), and simulated relative efficiencies using Eq. (35) of approximate expressions of mean square errors to the simulated mean square errors for different sizes of the second-phase sample and at the fixed size of the first-phase sample. We have shown simulated mean square errors and biases through graphical representations in Figure 2 (b), we note that the mean square errors and biases of the proposed estimator $\tilde{T}_R$ are lower than those of the relevant estimators $\tilde{t}_r$ and $\tilde{y}$. Hence, the proposed chain ratio estimator is more efficient than the than Lamichhane and Singh[7] ratio estimator and naive estimator.

Table 5: Simulated values of the estimators for different values of n at the fixed first-phase sample size $n' = 20$ for the real dataset.

Second-phase sample size n	Naive estimator $\tilde{y}$			Lamichhane and Singh[7] estimator $\tilde{t}_r$			Proposed estimator $\tilde{T}_R$		
	Relative efficiency	Mean square error	Bias	Relative efficiency	Mean square error	Bias	Relative efficiency	Mean square error	Bias
5	100	40.03	3.31	128	31.20	2.67	135	29.72	2.47
10	100	13.78	2.92	155	8.88	1.98	165	8.34	1.85
14	100	7.07	2.22	131	5.40	1.94	166	4.25	1.63

Confidence interval

The $100\,(1-\alpha)\%$ confidence intervals based on simulated estimates of the estimators $\tilde{T}_R$, $\tilde{t}_r$, and $\tilde{y}$ are given by:

$$\tilde{T}_R \pm t_{(n-1)}(\alpha/2)\left(\mathrm{MSE}(\tilde{T}_R)\right)^{1/2},\ \tilde{t}_r \pm t_{(n-1)}(\alpha/2)(\mathrm{MSE}(\tilde{t}_r))^{1/2}$$

$$\tilde{y} \pm t_{(n-1)}(\alpha/2)\left(\mathrm{V}(\tilde{y})\right)^{1/2}$$

Where $t_{(n-1)}(\alpha/2)$ is the value of the t-variate at $(n-1)$ degrees of freedom for a 95% level of confidence coefficient. We calculated 95% simulated confidence intervals of the estimated value for different values of $n = 90, 180, 360, 540, 900,$ and 1800 at the fixed size $n' = 3450$ for the generated dataset, and for $n = 5, 10,$ and 14 at the fixed size $n' = 20$ for the real dataset. For the generated dataset, the simulated as well as exact confidence intervals, percent coverage of the estimates, simulated estimates, and quartiles of the estimators $\tilde{T}_R, \tilde{t}_r$, and $\tilde{y}$ are calculated and given in (Table 6) and simulated values of confidence interval is presented graphically in (Figure 3a). For the real dataset, the simulated confidence intervals, percent coverage of the estimates, simulated estimates, and quartiles of the estimators $\tilde{T}_R, \tilde{t}_r$, and $\tilde{y}$ are calculated in (Table 6) and graphically presented in (Figure 3b). From (Table 6, Figures 3a and b), we observe that the proposed chain ratio estimator has a shorter confidence interval and more percent coverage than the than Lamichhane et al ratio estimator and naive estimator of population mode.[7] If we increase the sample size, the confidence intervals of the estimates become shorter.

Study to determine of n' and n for fixed cost $C \le C_0$

In practical applications, the cost aspect should also be taken into account. So, we define C_0 to be the total cost, i.e., fixed of the survey apart from overhead cost. The expected total cost of the survey, apart from the overhead cost, is given by a cost function:

$$C = (C_1' + C_2')n' + nC_1,$$

Where C_1' = The cost per unit of identifying and observing the main auxiliary variable x at the first-phase, C_2' = The cost per unit of identifying and observing additional auxiliary variable z at the first-phase, and C_1 = The cost per unit of mailing the questionnaire/visiting the unit in the second-phase. The expression for $MSE(T_i)$, $i = 1,2,3$ can be written as follows:

$$MSE(T_i) = \frac{V_{0i}}{n} + \frac{V_{1i}}{n'} + \text{ independent terms from } n' \text{ and } n;\ i = 1, 2, 3,$$

Where $T_1 = \tilde{y}, T_2 = \tilde{t}_r, T_3 = \tilde{T}_R$ and V_{0i},V_{1i} are the coefficient of the terms of $\frac{1}{n}$ and $\frac{1}{n'}$ respectively in the expression of $MSE(T_i), i = 1,2,3$. We consider ψ to be the function as follows:

$$\psi = MSE(T_i) + \lambda_i\left((C_1' + C_2')n' + nC_1\right)$$

Where λ_i is a Lagrange's multiplier. Differentiating ψ with respect to n' and n and equating them to zero, we obtained

$$n' = \sqrt{\frac{V_{1i}}{\lambda_i(C_1'+C_2')}},$$

$$n = \sqrt{\frac{V_{0i}}{\lambda_i C_1}}$$

We know that $n' > n$ so, we have $C_1' + C_2' < \frac{V_{1i}C_1}{V_{0i}}$.

Substituting the values of n' and n from above equations then we have,

$$\sqrt{\lambda_i} = \frac{1}{C_0}\left(\sqrt{(C_1' + C_2')V_{1i}} + \sqrt{C_1 V_{0i}}\right)$$

It has been observed that the determinant of the matrix of the second-order derivative of ψ with respect to n' and n is negative for the optimum values of n' and n, which shows that the solution for n' and n given by above equations for $C \le C_0$ minimizes $\mathrm{MSE}(T_i)$. The minimum value of MSE (T_i) for the optimum value of n' and n are given by:

$$MSE(T_i) = \frac{1}{C_0}\left(\sqrt{(C_1' + C_2')V_{1i}} + \sqrt{C_1 V_{0i}}\right)^2 - \frac{1}{N}\left(\tilde{V}_y^2 + R_2^2\tilde{V}_z^2 - 2R_2\tilde{V}_{yz}\right)$$

Study to determine of n' and n for a fixed variance $V = V_0$

We define V_0 be the variance of the estimator $T_{(i)}, i = 1,2,3$ in advance, then we have:

$$V_0 = \frac{V_{0i}}{n} + \frac{V_{1i}}{n'} - \frac{1}{N}\left(\tilde{V}_y^2 + R_2^2\tilde{V}_z^2 - 2R_2\tilde{V}_{yz}\right)$$

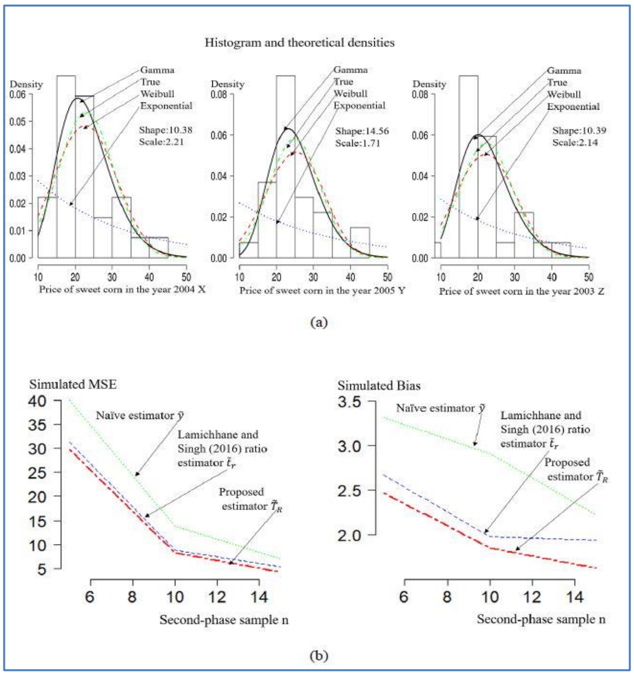

Figure 2: (a) Parameter values of the variables for the real dataset (b) Simulated values of mean square errors and biases of the estimators for different values of n at the fixed $n' = 20$ for the real dataset.

The total cost, apart from the overhead cost, is minimized by obtaining the optimum values of n' and n for specified precision $V = V_0$. For this purpose, we defined a function ϕ which is given as follows:

$$\phi = (C_1' + C_2')n' + C_1 n + \mu_i(MSE(T_i) - V_0),$$

Where $i = 1,2,3$, and μ_i is a Lagrange's multiplier. After differentiating ϕ with respect to n' and n and equating them to zero, we get,

$$n' = \sqrt{\frac{V_{1i}\mu_i}{(C_1' + C_2')}},$$

$$n = \sqrt{\frac{V_{0i}\mu_i}{C_1}}$$

Substituting the values of n' and n from Eq. (48) and (49), we get

$$\sqrt{\mu_i} = \frac{1}{V_0 + \frac{1}{N}(\tilde{V}_y^2 + R_2^2\tilde{V}_z^2 - 2R_2\tilde{V}_{yz})}\left(\sqrt{V_{0i}C_1} + \sqrt{V_{1i}(C_1' + C_2')}\right)$$

It has also been seen that the determinant of the matrix of a second-order derivative of ϕ with respect to n' and n is negative for the optimum values of n' and n, which shows that the solution for n', n given by Eq. (48) and (49). Putting the values of $\sqrt{\mu_i}$ from Eq. (50) in Eq. (48) and (49), we can obtain the value of n' and n for which the estimator MSE$(T_i), i = 1,2,3$ attains the variance V_0 with the expected cost given by

$$C(t_i) = \frac{1}{V_0 + \frac{1}{N}(\tilde{V}_y^2 + R_2^2\tilde{V}_z^2 - 2R_2\tilde{V}_{yz})}\left(\sqrt{V_{0i}C_1} + \sqrt{V_{1i}(C_1' + C_2')}\right)^2$$

From (Table 7), for both the datasets, we observe that for the fixed cost, the suggested chain ratio estimator $\tilde{T}_R$ shows the least mean square error in comparison to the than Lamichhane and Singh[7] ratio estimator $\tilde{t}_r$ and naive estimator $\tilde{y}$. Also, for the specified variance, $\tilde{T}_R$ has the lowest cost in comparison to the cost of the other estimators $\tilde{t}_r$ and $\tilde{y}$.

DISCUSSION

We formed chain ratio and product estimators for the population mode using two types of auxiliary information under the two-phase sampling scheme. We supported theoretical outcome through a simulation study. On the basis of simulated results, from (Table 3, Figure 1b), we observe that the values of mean square errors of the suggested ratio estimator are less than Lamichhane et al ratio estimator and naive estimator.[7] Also, the mean square errors decrease when the second-phase sample size increases.

The values of biases of the estimators are very low and close to zero. The ratios of the exact mean square errors to the simulated mean square errors are close to one, indicating that simulated mean square errors are approximately close to exact mean square errors. It means that exact mean square errors can also be used as simulated mean square errors. In (Table 3), we also computed the exact values of mean square errors and exact biases. The values of mean square errors of the suggested chain ratio estimator are less than that of other than Lamichhane et al ratio estimator and naive estimator.[7]

In Table 5, we computed simulated mean square errors using Eq. (34), simulated biases using Eq. (31), and simulated relative efficiencies using Eq. (35) of approximate expressions of mean square errors to the simulated mean square errors for different sizes of the second-phase sample $n = 5, 10,$ and 14 at the fixed size $n' = 20$ of the first-phase sample. We have shown simulated mean square errors and biases through graphical representations in (Figure 2b). From (Table 5, Figure 2b), we note that the mean square errors and biases of the proposed estimator $\widetilde{T}_R$ are lower than those of the relevant estimators $\tilde{t}_r$ and $\tilde{y}$.

Hence, the proposed chain ratio estimator is more efficient than the than Lamichhane and Singh[7] ratio estimator and naive estimator. From (Table 6, Figures 3 a and b) we observe that the proposed chain ratio estimator has a shorter confidence interval and more percent coverage than the than Lamichhane et al ratio estimator and naive estimator of population mode.[7]

Table 6: Simulated and exact confidence interval and its estimates of the estimators at different sizes of the second-phase sample at the fixed first-phase sample sizes for the generated dataset and the real dataset.

Simulated results for the generated dataset								
Estimators	**Lower Limit**	**Upper Limit**	**Coverage percent**	**Simulated estimates**	**Standard deviation**	**Lower Quartile**	**Me dian**	**Upper Quartile**
$\widetilde{Y} = 2.89, n' = 3450$								
Second-phase sample size n=90								
Proposed estimator $\widetilde{T}_R$	2.67	3.15	95.45	2.91	0.19	2.78	2.90	3.02
Lamichhane and Singh[7] $\tilde{t}_r$	2.70	3.20	94.45	2.95	0.19	2.82	2.94	3.07
Naive estimator $\tilde{y}$	2.57	3.31	81.84	2.94	0.28	2.76	2.94	3.12
Second-phase sample size n =180								
Proposed estimator $\widetilde{T}_R$	2.72	3.06	90.80	2.89	0.13	2.80	2.88	2.97
Lamichhane and Singh[7] $\tilde{t}_r$	2.76	3.14	92.54	2.95	0.13	2.86	2.94	3.04
Naive estimator $\tilde{y}$	2.71	3.18	82.42	2.94	0.17	2.83	2.95	3.05
Second-phase sample size n=360								
Proposed estimator $\widetilde{T}_R$	2.76	3.02	91.14	2.89	0.10	2.82	2.88	2.95
Lamichhane and Singh[7] $\tilde{t}_r$	2.72	3.00	93.70	2.86	0.10	2.79	2.85	2.92
Naive estimator $\tilde{y}$	2.66	3.03	83.91	2.84	0.13	2.76	2.85	2.93
Second-phase sample size n=540								
Proposed estimator $\widetilde{T}_R$	2.80	2.99	91.47	2.89	0.07	2.85	2.89	2.94
Lamichhane and Singh[7] $\tilde{t}_r$	2.83	3.04	95.00	2.94	0.07	2.89	2.93	2.98
Naive estimator $\tilde{y}$	2.80	3.07	84.79	2.93	0.10	2.87	2.93	3.00
Second-phase sample size n=900								
Proposed estimator $\widetilde{T}_R$	2.84	2.97	94.20	2.91	0.05	2.87	2.90	2.94
Lamichhane and Singh[7] $\tilde{t}_r$	2.85	3.00	95.13	2.93	0.05	2.89	2.93	2.96
Naive estimator $\tilde{y}$	2.83	3.02	85.01	2.93	0.07	2.88	2.92	2.97
Second-phase sample size n=1800								
Proposed estimator $\widetilde{T}_R$	2.87	2.97	96.35	2.92	0.02	2.90	2.92	2.94
Lamichhane and Singh[7] $\tilde{t}_r$	2.87	2.99	97.35	2.93	0.02	2.91	2.93	2.95
Naive estimator $\tilde{y}$	2.86	3.00	90.99	2.93	0.03	2.90	2.93	2.95

Exact results for the generated dataset				
Estimator	**Lower limit**	**Upper limit**	**Estimated value**	**U-L**
Second-phase sample size n=90				
Proposed estimator $\widetilde{T}_R$	2.53	3.17	2.85	0.64
Lamichhane and Singh[7] $\tilde{t}_r$	2.50	3.15	2.83	0.65
Naive estimator $\tilde{y}$	2.35	3.26	2.81	0.91

Continued.

Simulated results for the generated dataset				
Estimator	**Lower Limit**	**Upper Limit**	**Estimated value**	**U-L**
Second-phase sample size n=180				
Proposed estimator $\tilde{T}_R$	2.72	3.17	2.95	0.45
Lamichhane and Singh[7] $\tilde{t}_r$	2.74	3.20	2.97	0.46
Naive estimator $\tilde{y}$	2.69	3.33	3.01	0.64
Second-phase sample size n=360				
Proposed estimator $\tilde{T}_R$	2.77	3.08	2.93	0.31
Lamichhane and Singh[7] $\tilde{t}_r$	2.78	3.10	2.94	0.32
Naive estimator $\tilde{y}$	2.66	3.09	2.88	0.43
Second-phase sample size n=540				
Proposed estimator $\tilde{T}_R$	2.73	2.98	2.85	0.25
Lamichhane and Singh[7] $\tilde{t}_r$	2.74	3.00	2.87	0.26
Naive estimator $\tilde{y}$	2.70	3.05	2.87	0.35
Second-phase sample size n=900				
Proposed estimator $\tilde{T}_R$	2.79	2.98	2.89	0.19
Lamichhane and Singh[7] $\tilde{t}_r$	2.77	2.97	2.87	0.20
Naive estimator $\tilde{y}$	2.71	2.97	2.84	0.26
Second-phase sample size n=1800				
Proposed estimator $\tilde{T}_R$	2.81	2.93	2.87	0.12
Lamichhane and Singh[7] $\tilde{t}_r$	2.83	2.96	2.90	0.13
Naive estimator $\tilde{y}$	2.77	2.94	2.85	0.17

Second-phase sample size n=5								
	Lower Limit	**Upper Limit**	**Coverage percent**	**Simulated estimates**	**Standard deviation**	**Lower quartile**	**Median**	**Upper quartile**
$\tilde{Y} = 16.83, n' = 20$								
Second-phase sample size n=5								
Proposed estimator $\tilde{T}_R$	12.86	23.89	98.65	18.38	3.62	16.24	17.87	19.70
Lamichhane and Singh[7] $\tilde{t}_r$	12.27	27.91	98.55	20.09	3.96	17.75	19.53	21.53
Naive estimator $\tilde{y}$	10.93	31.26	96.15	21.09	4.98	17.80	21.10	22.78
Second-phase sample size n=10								
Proposed estimator $\tilde{T}_R$	14.70	22.69	98.10	18.70	2.21	17.31	18.40	19.78
Lamichhane and Singh[7] $\tilde{t}_r$	14.66	22.97	97.95	18.82	2.23	17.43	18.52	19.91
Naive estimator $\tilde{y}$	14.07	25.44	98.90	19.75	2.30	18.10	19.72	21.28
Second-phase sample size n=14								
Proposed estimator $\tilde{T}_R$	15.40	21.55	99.50	18.48	1.25	17.73	18.41	19.22
Lamichhane and Singh[7] $\tilde{t}_r$	15.23	22.33	99.50	18.78	1.27	18.02	18.71	19.53
Naive estimator $\tilde{y}$	15.03	23.09	99.60	19.06	1.46	17.91	19.02	20.07

Table 7: REs in percent of the estimators with respect to $\tilde{y}$ for the fixed cost $C \le C_0$ and expected cost of the different estimators for a specified variance $V = V_0$ for the generated and the real datasets.

Estimators	For generated dataset					
	Fixed cost C_0 = Rs. 100.00			**Fixed Variance $V_0 = 0.05$**		
	Cost C_1 = Rs. 2.00, cost C_1' = Rs. 0.10, and cost C_2' = Rs. 0.15			**Cost C_1 = Rs. 2.00, cost C_1' = Rs. 0.10, and cost C_2' = Rs. 0.15**		
	First-phase sample size	**Second-phase sample size**	**Relative efficiency (Mean SE)**	**First-phase sample size**	**Second-phase sample size**	**Cost**
Proposed estimator $\tilde{T}_R$	66	42	139 (0.098)	128	81	194.67

Continued.

Estimators	For generated dataset					
	Fixed cost $C_0 = \text{Rs.}\,100.00$			Fixed Variance $V_0 = 0.05$		
	Cost $C_1 = \text{Rs.}\,2.00$, cost $C_1' = \text{Rs.}\,0.10$, and cost $C_2' = \text{Rs.}\,0.15$			Cost $C_1 = \text{Rs.}\,2.00$, cost $C_1' =$ Rs. 0.10, and cost $C_2' = \text{Rs.}\,0.15$		
	First-phase sample size	**Second-phase sample size**	**Relative efficiency (Mean SE)**	**First-phase sample size**	**Second-phase sample size**	**Cost**
Lamichhane and Singh[7] $\tilde{t}_r$	104	37	109 (0.125)	257	92	247.37
Naive estimator $\tilde{y}$	~	50	100 (0.135)	~	134	268.73
Estimators	For real dataset					
	Fixed cost $C_0 = Rs.\,100.00$			Fixed Variance $V_0 = 19.50$		
	Cost $C_1 = \text{Rs.}\,20.00$, cost $C_1' =$ Rs. 0.75, and cost $C_2' = \text{Rs.}\,0.78$			Cost $C_1 = \text{Rs.}\,20.00$, cost $C_1' =$ Rs. 0.75, and cost $C_2' = \text{Rs.}\,0.78$		
	First-phase sample size	Second-phase sample size	Relative efficiency (Mean square error)	First-phase sample size	Second-phase sample size	Cost
Proposed estimator $\tilde{T}_R$	19	4	225 (19.63)	21	4	109.28
Lamichhane and Singh[7] $\tilde{t}_r$	20	3	170 (26.07)	22	4	112.89
Naive estimator $\tilde{y}$	~	5	100 (44.26)	~	9	188.52

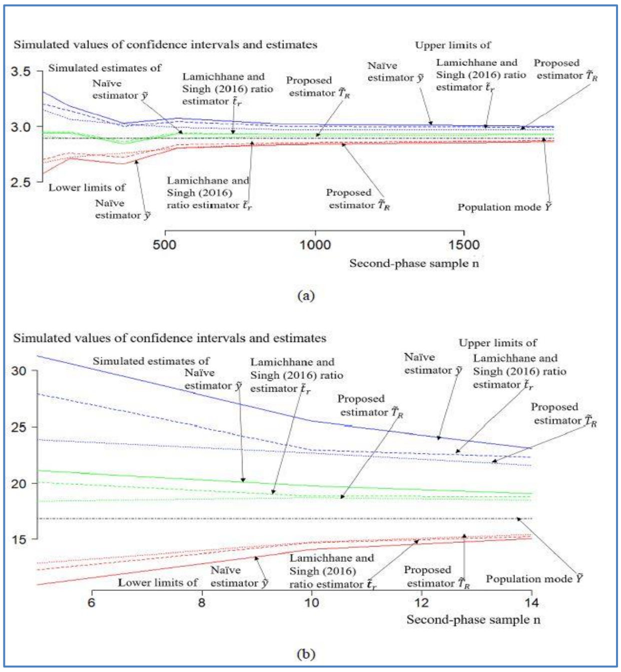

Figure 3: (a) Simulated values of the confidence interval and the estimates for generated dataset (b) Simulated values of the confidence interval and the estimates for real dataset.

If we increase the sample size, the confidence intervals of the estimates become shorter. From (Table 7), for both the datasets, we observe that for the fixed cost, the suggested chain ratio estimator $\tilde{T}_R$ shows the least mean square error in comparison to the than Lamichhane and Singh[7] ratio estimator $\tilde{t}_r$ and naive estimator $\tilde{y}$. Also, for the specified variance, $\tilde{T}_R$ has the lowest cost in comparison to the cost of the other estimators $\tilde{t}_r$ and $\tilde{y}$.

CONCLUSION

Using the information on two auxiliary variables, we have suggested chain ratio and product estimators for estimating the population mode. From the numerical outcomes through simulation studies with a generated and a real dataset, we found that the introduced chain ratio estimator has a minimum mean square error, shorter confidence interval, and a higher percentage of estimates coverage than Lamichhane and Singh[7] ratio estimator and naive estimator. After increasing the information related to all the variables used, it is also found that mean square errors and biases of the estimators decrease, confidence intervals become shorter, and covering percentages of the estimates become larger. Similar results can be obtained for the introduced chain product estimator for negatively correlated datasets. So we highly recommend preferring these suggested chain ratio and product estimators over than Lamichhane et al ratio estimator and naive estimator of the population mode.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Chand L. Some ratio type estimators based on two or more auxiliary variables. Available at: https://www.semanticscholar.org/. Accessed on February 2023.

2. Doodson AT. Relation of the mode, median and mean in frequency curves. Biometrika. 1917;11(4):425-9.
3. Gross S. Proceedings of the Section on Survey Research Methods Median estimation in sample surveys. Alexandria: American Statistical Association publication. 2012.
4. Guha, S, Chandra, H. Improved chain-ratio type estimator for population total in double sampling. Mathemat Popul Studies. 2020;27(4):216-31.
5. Kuk AY, Mak TK. Median estimation in the presence of auxiliary information. J Royal Stat Soc. 1989;51(2): 261-9.
6. Kumar S, Tiwari, N. Generalized naive ratio and product based estimators for estimating population mode in simple random sampling. Int J Stat Economics. 2019;20(1):57-79.
7. Lamichhane R, Singh S. Estimation of mode using two-phase sampling. Commu Stat Simul Comput. 2016:45(7):2586-97.
8. Neyman J. Contribution to the theory of sampling human populations. J Am Stat Assoc. 1938:33(201): 101-16.
9. Dataset from the Department of Agriculture, United States. Available at: www.nass.usda.gov. Accessed on 01 July 2022.

Cite this article as: Kumar S, Yadav A. Chain ratio and product estimators for population mode using two-phase sampling scheme. Int J Sci Rep 2023;9(10):314-25.