http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

OPEN ACCESS Information Technology Journal

ISSN 1812-5638
DOI: 10.3923/itj.2021.15.20

Research Article

An Innovative Class of Chain Ratio-Type Estimators Using Auxiliary Information in Sample Survey

[1]Khalid Ul Islam Rather and [2]Tanveer Ahmad Tarray

[1]Division of Statistics and Computer Science, SKUAST, 180009, Jammu and Kashmir
[2]Department of Mathematical Sciences, Islamic University of Science and Technology, 192122, Jammu and Kashmir

Abstract

In survey sampling, the use of auxiliary information at the estimation stage has been discussed by various statisticians to improve the efficiency of their formulated estimators for estimating the population mean. The crux of this study was to propose some modified and improved compromised ratio type estimators to estimate the population mean of the study variate using the information on different parameters. The properties of the suggested estimation procedures have been examined. It has been shown that the proposed class of estimators is better than the usual unbiased estimator, ratio estimator under some realistic conditions. The study has been performed to show that the proposed class of estimators gives better results in comparison to some of the existing estimators. Suitable recommendations are made to the survey practitioners.

Key words: Study variate, auxiliary variate, simple random sampling, ratio estimator, chain ratio-type estimator, correlation coefficient, skewness

Citation: Rather, K.U.I. and T.A. Tarray, 2021. An innovative class of chain ratio-type estimators using auxiliary information in sample survey. Inform. Technol. J., 20: 15-20.

Corresponding Author: Tanveer Ahmad Tarray, Department of Mathematical Sciences, Islamic University of Science and Technology, 192122, Jammu and Kashmir

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Out of many ratios, product and regression methods of estimation are good examples in this context. The use of auxiliary information dates back to the year (1934) when Neyman used it for the stratification of the finite population. Using auxiliary information in estimation procedure and envisaged ratio method of estimation to provide a more efficient estimator of the population mean or total compared to the simple mean per unit estimator under certain conditions when the correlation between the study variable and the auxiliary variable is positive (high). On the other hand, if the correlation is negative (high), the use of the production method of estimation suggested by Singh *et al.*[1] .

Consider a finite population $U = \{U_1, U_2, \cdots, U_N\}$ of size N.

Let y and x be study variate and auxiliary variate taking values y_i and x_i respectively of the ith unit U_i ($i = 1,2,\cdots, N$). The parameter of attention is the population mean $\bar{Y}$ of the study variate y. The precision of estimators of the population mean $\bar{Y}$ can be increased by utilizing advanced information about a suitable auxiliary variable x correlated with y. Out of many ratios, product, difference and regression methods of estimation are good examples in this context. Let $\bar{X}$ be the known population mean of the auxiliary variate and $\bar{y}$, $\bar{x}$ respectively denote the sample means of the variates y and x based on a simple random sample of size n drawn without replacement from the population U. Also let:

$$S_y^2 = (N-1)^{-1}\Sigma_{(j-1)}^{N}(y_i - \bar{Y})^2$$

and:

$$S_x^2 = (N-1)^{-1}\,\Sigma_{(j-1)}^{N}(x_i - \bar{X})^2$$

with:

$$f = \frac{n}{N},\ R = \frac{\bar{Y}}{\bar{X}}$$

$$S_{xy}^2 = (N-1)^{-1}\,\Sigma_{(j-1)}^{N}(x_i - \bar{X})(y_i - \bar{Y})$$

and:

$$C_y^2 = \frac{S_y^2}{(\bar{Y})^2},\ C_x^2 = \frac{S_x^2}{(\bar{X})^2}$$

having $\rho = S_{xy}/(S_xS_y)$ denotes the correlation coefficient between the study variate y and the auxiliary variate x.

The classical ratio estimator for the population mean of the study variate y is defined by:

$$\bar{y}R = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right) \tag{1}$$

where, it is assumed that the population mean of the auxiliary variate x is known.

Replacing by in (1), suggested the chain ratio-type estimator for the population mean $\bar{Y}$:

$$\bar{y}R = \overline{yR}\left(\frac{\bar{X}}{\bar{x}}\right) = \bar{y}\left(\frac{\bar{X}^2}{\bar{x}^2}\right) \tag{2}$$

Singh *et al.*[2] have shown that the chain ratio-type estimator is more efficient than the usual ratio estimator if:

$$k\left(=\rho\frac{C_y}{C_x}\right) > \frac{3}{2}$$

It is observed in the sampling theory literature that in recent years considerable attempts have been made by various authors to increase the precision of the conventional estimators by using suitable transformation on the auxiliary variate x, for instance[2-5].

Using the transformation:

$$z_i = \gamma x_i + \mu,\ i = 1,2,\cdots,N \tag{3}$$

On the auxiliary variate x, a ratio-type estimator for population mean $\bar{y}$ is defined by:

$$t_R = \bar{y}\left(\frac{\bar{Z}}{\bar{z}}\right) = \bar{y}\left(\frac{\bar{Z}}{\bar{z}}\right) = \bar{y}\left(\frac{\gamma\bar{X}+\mu}{\gamma\bar{x}+\mu}\right) \tag{4}$$

Where:

$$\bar{Z} = (2\gamma\bar{X}+\mu),\ \bar{z} = (2\gamma\bar{x}+\mu)$$

and:

$$\gamma = \frac{\eta}{n},\ \eta(\neq 0)$$

μ are either real number or function of the known parameters of the auxiliary variate x as such as the standard deviation (S_x), coefficient of variation (C_x), skewness ($\beta_{1(x)}$),

Table 1: Some members of the estimator

Estimator	γ	μ
$t_{R(1)} = \bar{y}\left(\frac{2\bar{X}+S_x}{2\bar{x}+S_x}\right)$	2	S_x
$t_{R(2)} = \bar{y}\left(\frac{2\bar{X}+X}{2\bar{x}+X}\right)$	2	X

X(= N$\bar{X}$) is the population total of the auxiliary variate x

Table 2: Some members of the estimator

Estimator	γ	μ
$t_{CT(1)} = \bar{y}\left(\frac{2\bar{X}+S_x}{2\bar{x}+S_x}\right)^2$	2	S_x
$t_{CT(2)} = \bar{y}\left(\frac{2\bar{X}+X}{2\bar{x}+X}\right)^2$	2	X

kurtosis ($\beta_{2(x)}$), $\Delta = \beta_{2(x)}-\beta_{1(x)}-1$ and the correlation coefficient (ρ) of the population. Some members of t_R are listed in Table 1.

It is well known under Simple Random Sampling Without Replacement (SRSWOR) that the variance/Mean Squared Error (MSE) of the sample mean $\bar{y}$ is given by:

$$\text{var}(\bar{y}) = \text{MSE}(\bar{Y}) = \frac{(1-f)}{n}S_y^2 = \frac{(1-f)}{n}\bar{y}^2C_y^2 \quad (5)$$

To the first degree of approximation, the mean squared error of the estimators $\bar{y}_R$, $\bar{y}_{CT}$ and t_R are respectively given by:

$$\text{MSE}\ (\bar{y}_R) = \frac{(1-f)}{n}\ (S_y^2 + R^2S_x^2 - 2R\rho S_xS_y) \quad (6)$$

$$\text{MSE}\ (\bar{y}_{CT}) = \frac{(1-f)}{n}\ (S_y^2 + 4R^2S_x^2 - 4R\rho S_xS_y) \quad (7)$$

$$\text{MSE}\ (t_R) = \frac{(1-f)}{n}\ (S_y^2 + R^{*2}\ S_x^2 - 4R^*\rho S_xS_y) \quad (8)$$

where, $R^* = \bar{y}/(\gamma\ \bar{x}+\mu)$.

The plan of this article is as follows. The first section is the introduction. The class of chain ratio-type estimators t_{CT} has been proposed with its bias and MSE expressions. The authors presented the efficiency comparisons of the proposed estimator t_{CT} with estimators $\bar{y}$, $\bar{y}_R$, $\bar{y}_{CT}$ and t_R. An empirical study is carried out to judge the merits of the suggested class of chain ratio-type estimators t_{CT} over other estimators.

Class of chain ratio-type estimators: When $\bar{y}$ in (4) is replaced with t_T, the class of chain ratio-type estimator is obtained as:

$$t_{CT} = \frac{t_R\ \ (\gamma\bar{X}+\mu)}{\gamma\bar{x}+\mu} \quad (9)$$

We can re-write (9) using (4) as:

$$t_{CT} = \frac{\bar{y}\ (\gamma\bar{X}+\mu)^2}{(\gamma\bar{x}+\mu)^2} \quad (10)$$

Which may be further generalized as:

$$t_{CT(C)} = \bar{y}\left(\frac{\gamma\bar{X}+\mu}{\gamma\bar{x}+\mu}\right)^c \quad (11)$$

where, α is a suitably chosen constant. The optimum value of c is k{1+μ/($\gamma\bar{x}$)}. Call this generalized form of the estimator the repeated substitution estimator. Some members of the estimator t_{CT} are listed in given Table 2.

The MSE of the class of chain ratio-type estimator can be obtained using the Taylor series expansion. In general, this expansion for a function of p variables can be given as:

$$L\ (\bar{x}_1,\ \bar{x}_2,...,\bar{x}_p) = L(\bar{X}_1,\ \bar{X}_2,...,\bar{X}_p) + \sum_{J-1}^{p} b_j\ (\bar{x}_j - \bar{X}_j) + T_p(\bar{X}_p,\ \alpha)$$

Where:

$$b_j = \frac{\partial L\ (\alpha_1,\ \alpha_2,...,\alpha_p)}{\partial\alpha_j}$$

And:

$$T_p\ (\bar{X}_p,\ a) = \sum_{j=1}^{p}\ \sum_{i=1}^{p}\frac{1}{2!}\frac{\partial^2 L\ (\bar{X}_1,\ \bar{X}_2,...,\bar{X}_p)}{\partial\bar{X}_i\partial\bar{X}_j}(\bar{x}_j - \bar{X}_j)\ (\bar{x}_i - \bar{X}_j) + O_p$$

Where denotes the remainder of the Taylor series expansion having terms of degree higher than two. Omitted the term Rp ($\bar{x}_p$, α), we obtain the Taylor series method for two variables as:

$$L\ (\bar{x},\ \bar{y})\text{-}L(\bar{X},\ \bar{Y}) \cong \frac{\partial L(c,d)}{\partial c}\Big|_{(\bar{X},\ \bar{Y})}\ (\bar{x},\ \bar{X}) + \frac{\partial L(c,d)}{\partial d}\Big|_{(\bar{X},\ \bar{Y})}\ (\bar{y},\ \bar{Y}) \quad (12)$$

See previous studies suggested by Singh *et al.*[2] in this area of research. Here:

$$L(\bar{x}, \bar{y}) = \frac{\bar{y}}{(\gamma\bar{x}+\mu)^2} = \hat{T} \text{ and } h(\bar{X}, \bar{Y}) = \frac{\bar{Y}}{(\gamma\bar{X}+\mu)^2} = T_c \quad (13)$$

From (12) and (13) we can write:

$$\hat{T}_c - T_c \cong -\frac{2\bar{Y}\eta(\gamma\bar{X}+\mu)^2}{(\gamma\bar{X}+\mu)^4}(\bar{x}, \bar{X}) + \frac{1}{(\gamma\bar{X}+\mu)^2}(\bar{y}-\bar{Y}) \quad (14)$$

Squaring both sides of (14) we have:

$$(\hat{T}_c - T_c)^2 \cong -\frac{4\bar{Y}^2\gamma^2(\gamma\bar{X}+\mu)^2}{(\gamma\bar{X}+\mu)^8}(\bar{x}, \bar{X})^2 - \frac{4\bar{Y}\gamma(\gamma\bar{X}+\mu)}{(\gamma\bar{X}+\mu)^2}(\bar{x}, \bar{X})(\bar{y}-\bar{Y}) + \frac{1}{(\gamma\bar{X}+\mu)^4}(\bar{y}-\bar{Y})^2$$

or:

$$(\hat{T}_c - T_c)^2 \cong -\frac{1}{(\gamma\bar{X}+\mu)^4}\{4\gamma^2 R^{*2} V(\bar{x}) - 4\gamma R^*(\bar{x}, \bar{X})(\bar{y}-\bar{Y}) + (\bar{y}-\bar{Y})^2\} \quad (15)$$

Taking expectation of both sides of (15) we get:

$$E(\hat{T}_c - T_c)^2 \cong -\frac{1}{(\gamma\bar{X}+\mu)^4}\{4\gamma^2 R^{*2} V(\bar{x}) - 4\gamma R^* \operatorname{cov}(\bar{x}, \bar{y}) + V(\bar{y})\}$$

$$E(\hat{T}_c - T_c)^2 \cong -\frac{(1-f)}{\gamma(\gamma\bar{X}+\mu)^4}\{S_y^2 - 4\gamma^2 R^{*2} S_x^2 + 4\gamma R^* \rho S_x S_y\}$$

Thus, the MSE of to the first degree of approximation is given by:

$$MSE(t_{CT}) \cong -\frac{(1-f)}{n}\{S_y^2 - 4\gamma^2 R^{*2} S_x^2 + 4\gamma R^* \rho S_x S_y\} \quad (16)$$

The bias of t_{CT} is given by:

$$B(t_{CT}) = E(t_{CT} - \bar{Y}) \quad (17)$$

$$= E\{\hat{T}_c(\gamma\bar{X}+\mu)^2 - T_c(\gamma\bar{X}+\mu)^2\}$$

$$= (\gamma\bar{X}+\mu)^2 E(\hat{T}_c - T_c)$$

$$= (\gamma\bar{X}+\mu)^2 B(\hat{T}_c)$$

where, $B(_{tCT})$ stands for the bias of $\hat{T}_c$. To the first degree of approximation, the bias of $\hat{T}_c$ is obtained as follows:

$$\begin{aligned}\hat{T}_c - T_c &= L(\bar{x}, \bar{y}) - L(\bar{X}, \bar{Y}) \\ &= \Bigg[\frac{\partial L(c,d)}{\partial c}\Big|_{(\bar{X}, \bar{Y})}(\bar{x}, \bar{X}) + \frac{\partial L(c,d)}{\partial d}\Big|_{(\bar{X}, \bar{Y})}(\bar{y}, \bar{Y}) \\ &+ \frac{1}{2!}\Bigg\{\frac{\partial^2 L(c,d)}{\partial c^2}\Big|_{(\bar{X}, \bar{Y})}(\bar{x}, \bar{X})^2 + 2\frac{\partial^2 L(c,d)}{\partial d\partial c}\Big|_{(\bar{X}, \bar{Y})}(\bar{x}, \bar{X})(\bar{y}, \bar{Y}) \\ &+ \frac{\partial^2 L(c,d)}{\partial d^2}\Big|_{(\bar{X}, \bar{Y})}(\bar{y}, \bar{Y})^2\Bigg\}\Bigg]\end{aligned} \quad (18)$$

We note that:

$$\frac{\partial L(c,d)}{\partial c}\Big|_{(\bar{X}, \bar{Y})} = \frac{-2\gamma\bar{Y}}{(\gamma\bar{X}+\mu)^3} \quad (19)$$

$$\frac{\partial L(c,d)}{\partial c}\Big|_{(\bar{X}, \bar{Y})} = \frac{1}{(\gamma\bar{X}+\mu)^2} \quad (20)$$

$$\frac{\partial^2 L(c,d)}{\partial c^2}\Big|_{(\bar{X}, \bar{Y})} = \frac{-6\gamma^2\bar{Y}}{(\gamma\bar{X}+\mu)^4} \quad (21)$$

$$\frac{\partial^2 L(c,d)}{\partial c^2}\Big|_{(\bar{X}, \bar{Y})} = 0 \quad (22)$$

$$\frac{\partial L(c,d)}{\partial d\partial c}\Big|_{(\bar{X}, \bar{Y})} = \frac{-2\gamma}{(\gamma\bar{X}+\mu)^3} \quad (23)$$

Putting (19), (23) in (18) we have:

$$\begin{aligned}\hat{T}_c - T_c &= \frac{2\gamma\bar{Y}}{(\gamma\bar{X}+\mu)^3}(\bar{x}, \bar{X})^2 + \frac{1}{(\gamma\bar{X}+\mu)^2}(\bar{y}-\bar{Y}) \\ &+ \frac{1}{2}\left\{\frac{6\gamma^2\bar{Y}}{(\gamma\bar{X}+\mu)^2}\times(\bar{x}, \bar{X})^2 - \frac{4\gamma}{(\gamma\bar{X}+\mu)^3}(\bar{x}, \bar{X})(\bar{y}-\bar{Y})\right\}\end{aligned} \quad (24)$$

Taking expectation of both sides of (24) we get the bias of $\hat{R}_c$ to the first degree of approximation as:

$$B(\hat{T}_C) = \frac{(1-f)}{n}\frac{\gamma}{(\gamma\bar{X}+\mu)^3}(3\gamma R^* S_x^2 + 2\rho S_x S_y) \quad (25)$$

Substitution of (25) in (17) yields the bias of to the first degree of approximation as:

$$B(t_{CR}) = \frac{(1-f)}{n}\frac{\gamma}{(\gamma\bar{X}+\mu)}(3\gamma R^* S_x^2 + 2\rho S_x S_y) \quad (26)$$

The bias of t_{CR} will vanish either (i) the sample size n is large enough, or (ii) $3/2\,\{1+\mu/(\gamma\bar{x})\}^{-1}$.

The bias term is omitted for the first degree of approximation in the Taylor series expansion while obtaining the MSE of the proposed estimator T_{CT}.

Efficiency comparisons: From (5), (6), (7), (8) and (16) we have:

- $$\text{MSE}(\bar{y}) - \text{MSE}(t_{CT}) = \frac{4(1-f)}{n}\gamma R^* S_x^2(\beta - \eta R^*)$$

$$= \frac{4(1-f)}{n}\gamma R R^* S_x^2 (k-\theta) > 0 \text{ if } \theta < k \quad (27)$$

Where:

$$\theta = \frac{\gamma\bar{X}}{(\gamma\bar{X}+\mu)}$$

and k = β/R:

- $$\text{MSE}(\bar{y}_R) - \text{MSE}(t_{CT}) = \frac{(1-f)}{n}(1-2\theta)(1-2\theta-2k)R^2S_x^2 > 0 \text{ if}$$

$$\left.\begin{array}{l}\text{either } \frac{1}{2} < \theta < \frac{1}{2}\ (2k-1) \\ \text{or } \frac{1}{2}\ (2k-1) < \theta < \frac{1}{2}\end{array}\right\}$$

Or equivalently:

$$\min\left\{\frac{1}{2},(2k-1)\right\} < \theta < \max\left\{\frac{1}{2},(2k-1)\right\} \quad (28)$$

- $$\text{MSE}(\bar{y}_{CT}) - \text{MSE}(t_{CT}) = \frac{4(1-f)}{n}R^2S_x^2(1-\theta)(1-\theta-k) > 0 \text{ if}$$

$$\left.\begin{array}{l}\text{either}(k-1) < \theta < 1 \\ \text{or } 1 < \theta < (k-1)\end{array}\right\}$$

Or equivalently:

$$\min\{1,(k-1)\} < \theta < \max\{1,(k-1)\} \quad (29)$$

$$\text{MSE}(t_R) - \text{MSE}(t_{CT}) = \frac{(1-f)}{n}\theta R^2S_x^2(2k-3\theta) > 0 \text{ if } \frac{3}{2} < \frac{k}{0} \quad (30)$$

Empirical study: In this section, we have compared the performance of the estimators and with the following data set.

Table 3: PRE of $t_{R(i)}$ and $t_{CT(i)}$ (i = 1,2) with respect to $\bar{y}$

Estimator PRE (.,$\bar{y}$) $t_{R(i)}$ (i = 1,2)	Estimator PRE (.,$\bar{y}$) $t_{CT(i)}$ (i = 1,2)
$t_{R(1)}$ 171.93	$t_{CT(1)}$ 266.16
$t_{R(2)}$ 125.123	$t_{CT(2)}$ 298.40

Population I: The data consists of 106 villages in the Marmara region of Turkey in 1999.

The variates are defined as:

- y = The level of apple production
- x = The number of apple trees (1 unit = 100 trees)

The required values of the parameters are:

$$N = 106,\ n = 20,\ \bar{x} = 243.76,\ S_x = 491.89,\ S_y = 491.89,$$
$$\rho = 0.82,\ \beta_{2(x)} = 25.71,\ K = 1.70$$

We have computed the Percent Relative Efficiency (PRE) of the estimators $t_{R(i)}$ and $t_{CT(i)}$, (i = 1-21) to the usual unbiased estimator $\bar{y}$ by using the PRE($t_{R(i)}$, $\bar{y}$) and PRE($t_{R(i)}$, $\bar{y}$), we have:

$$\text{PRE}(t_{R(i)}\bar{y}) = \left\{1 + T_{(i)}^2\left(\frac{S_x^2}{S_y^2}\right)4\gamma_i^2 - 2T_i\left(\rho\frac{S_x^2}{S_y^2}\right)2\gamma_i\right\}^{-1}\times 100,\ (i = 1\text{-}21) \quad (31)$$

$$\text{PRE}(t_{R(i)}\bar{y}) = \left\{1 + 4T_{(i)}^2\left(\frac{S_x^2}{S_y^2}\right)4\gamma_i^2 - 2T_i\left(\rho\frac{S_x^2}{S_y^2}\right)2\gamma_i\right\}^{-1}\times 100,\ (i = 1\text{-}21) \quad (32)$$

where, γ_i = 2 for i = 1-6:

$$R_{(1)} = \left(\frac{\bar{Y}}{\bar{x}+S_x}\right)$$

$$R_{(2)} = \left(\frac{\bar{Y}}{\bar{x}+X}\right)$$

Percent relative efficiency regarding the estimator to has been mentioned in Table 3.

CONCLUSION

The performance of an estimator is generally judged based on relative efficiency. The proposed estimator utilizes the information on the relationship between auxiliary and study variables more efficiently as compared to existing estimators. It is further observed that the proposed estimator results in a high gain in efficiency between study and auxiliary variables.

In other words, the proposed estimator continues to be superior to the existing estimator.

SIGNIFICANCE STATEMENT

This study discovers the major enabling technologies and key applications domains that can be beneficial for researchers engaged in this field. This study will help the researcher to uncover the critical areas of survey sampling that many researchers were not able to explore. Thus, a new theory on survey sampling may be arrived at.

REFERENCES

1. Kadilar, C. and H. Cingi, 2003. Ratio estimators in stratified random sampling. Biomet. J., 45: 218-225.
2. Singh, H.P., R. Tailor, R. Tailor and M.S. Kakran, 2004. An improved estimator of population means using power transformation. J. Ind. Soc. Agri. Statist., 58: 223-230.
3. Tailor, R., S. Chouhan, R. Tailor and N. Garg, 2012. A ratio-cum-product estimator of population mean in stratified random sampling using two auxiliary variables. Statistica, 72: 287-297.
4. Singh, H.P., N. Sharma and T.A. Tarray, 2015. An efficient class of two-phase exponential ratio and product-type estimators for estimating the finite population mean. J. Statist. Appl. Prob. Let., 21: 71-88.
5. Singh, H.P., M.K. Srivastava, N. Srivastava, T.A. Tarray, V. Singh and S. Dixit, 2015. Chain regression-type estimator using multiple auxiliary information in successive sampling. Hacettepe J. Math. Stat., 44: 1247-1256.