Estimating peppermint oil yields with auxiliary variable information

S.K. Yadav*

Department of Statistics,
Babasaheb Bhimrao Ambedkar University,
Lucknow-226025, UP, India
Email: drskystats@gmail.com
*Corresponding author

Dinesh K. Sharma and Kate Brown

Department of Business, Management and Accounting,
University of Maryland Eastern Shore,
Princess Anne, MD 21853, USA
Email: profdksharma@gmail.com
Email: kbrown@umes.edu

Abstract: In this article, we propose an improved method for estimation of the population mean using an auxiliary variable and apply it to the peppermint oil yield for a block level in the Barabanki District of Uttar Pradesh State in India. We consider a new family of estimators for the population mean, using the area of the peppermint field as the auxiliary variable. We study the sampling properties of the proposed family, through the bias and the mean squared error (MSE) to the first order approximation. We compare the suggested estimators with competing estimators theoretically and verify the conditions under which they outperform the competing estimators with actual data collected from the Siddhaur Block of the Barabanki District.

Keywords: study variable; auxiliary variable; regression-cum-ratio estimator; bias; mean squared error; MSE; percentage relative efficiency; PRE.

Reference to this paper should be made as follows: Yadav, S.K., Sharma, D.K. and Brown, K. (2022) 'Estimating peppermint oil yields with auxiliary variable information', *Int. J. Operational Research*, Vol. 44, No. 1, pp.122–139.

Biographical notes: S.K. Yadav is a faculty in the Department of Statistics at Babasaheb Bhimrao Ambedkar University Lucknow, UP, India. He earned his MSc and PhD in Statistics from the Lucknow University and qualified the National Eligibility Test. He has published 38 papers in national and international journals of repute indexed in Scopus/Web of Science and two books from an international publisher. He is a referee for 20 reputed international journals. He has been awarded Young Scientist Award in 2016 for the contribution in the field survey sampling by Venus International Research Foundation, Chennai, India and Best Paper Award in 2018 by MTMI, USA. He has presented papers in more than 20 national and international conferences and also delivered 13 invited talks in several conferences.

Dinesh K. Sharma is a Professor of Quantitative Methods and Computer Applications in the Department of Business, Management and Accounting at the University of Maryland Eastern Shore. He earned his MS in Mathematics, MS in Computer Science, PhD in Operations Research, and a second PhD in Management. He has published over 95 refereed journal articles and 115 conference proceedings and has also won 12 best paper awards. Some of his work has appeared in the *International Journal of Computers and Applications*, *International Journal of Computer Mathematics*, *Computers and Mathematics with Applications*, *Journal of Applied Mathematics and Computing*, *International Journal of Production Economics*, and *International Journal of Logistics*, to name a few. He is the Editor-in-Chief of the *Journal of Global Information Technology* and *Review of Business and Technology Research*.

Kate Brown is currently an Associate Professor with teaching specialties in Finance and Statistics. She received her Bachelor of Arts, Master of Business Administration, and PhD from the University of Connecticut. Prior to her appointment at UMES, she spent 17 years in New Zealand, where she was the Chair of the Department of Finance and Quantitative Analysis at the University of Otago in Dunedin. Her experiences in New Zealand included a five-year appointment on the New Zealand Commerce Commission. She has authored numerous publications in refereed journals, including one best paper awards. Her research interests include dividend policy, entrepreneurship, financial market analysis, mergers, and bankruptcy prediction. She developed the Finance Concentration for the Business Administration Major, and has introduced three new finance courses since her appointment in 2005.

1 Introduction

Describing the parameters of a large population generally requires sampling which introduces differences from the true parameter values. Additional information collected at various stages of sampling including planning, data collection, and estimation can often be used to improve the estimation of the parameters under investigation. When this information is collected about an auxiliary variable that has a high degree of association, either positively and negatively correlated with the primary variable, it can be used to improve the efficacy of the estimation.

Following Watson (1937), researchers have worked to find more efficient unbiased estimators of the population mean, including ratio and product types. In both the ratio and the product methods, the line of regression should pass through the origin. Ratio estimators have been studied and improved with auxiliary information since Cochran (1940), including Hartley and Ross (1954), Robson (1957), Goodman and Hartley (1958) and Murthy (1964). Hutchison (1971) compared some ratio estimators of the population mean using simulated data obtained through the Monte Carlo technique and Robinson (1987) suggested the conditioning of ratio estimates under a simple random sampling scheme.

Prasad (1989) proposed some improved ratio type estimators for finite populations; and Rao (1991) suggested specific methods for the improvements in ratio and regression type estimators. Upadhyaya and Singh (1999) used a transformed auxiliary variable, and Singh and Tailor (2003) used known information on the correlation coefficient between the study variable and auxiliary information for an improved estimation. More recently, Gupta and Shabbir (2008), Koyuncu and Kadilar (2009) and Al-Omari et al. (2009) suggested new efficient ratio type estimators using known parameters of an auxiliary variable under simple random sampling and rank set sampling techniques. Shabbir and Gupta (2011) and Singh and Solanki (2012) proposed improved ratio type estimators of the population mean of the primary variable using auxiliary information in quantitative and qualitative forms under simple random sampling and stratified random sampling schemes.

Yadav and Kadilar (2013a, 2013b), Sharma and Singh (2013) and Subramani and Kumarapandiyan (2012a, 2012b, 2013) suggested improved ratio and product type estimators of the population mean of the main variable using known information on the parameters of an auxiliary variable while Yadav and Mishra (2015), Yadav et al. (2016), Subramani (2016) and Abid et al. (2016) proposed enhanced ratio type estimators of the population mean of the study variable using known information on the median of study variable and some conventional and non-conventional parameters of auxiliary variable. Pal and Singh (2016) and Khan and Al-Hossain (2016) suggested elevated estimators using known parameters of a secondary variable.

Yadav and Pandey (2017) and Yadav et al. (2017) suggested some improved estimators using auxiliary information. Gupta and Yadav (2017, 2018), Subramani and Master Ajith (2017) and Subzar et al. (2017) suggested ratio type estimators of the population mean of the main variable using known traditional and non-traditional parameters of a supplementary variable while Srija and Subramani (2018), Ijaz and Ali (2018), Yadav et al. (2018), Singh et al. (2018) and Zatezalo et al. (2018) proposed improved ratio and ratio-cum-regression type estimators of the population mean of the study variable using known conventional and non-conventional location parameters. Jaroengeratikun and Lawson (2019), Yadav et al. (2019) and Zaman (2019) used known information on traditional and non-traditional parameters of an auxiliary variable for enhanced estimation of the population mean.

We examine the sampling properties of a set of competing estimators, and observe that it is possible to improve estimators based on moving the sampling distributions closer to the parameter under consideration. The objective of this study is to search for the most efficient estimators for the population mean in the class of the selected recently proposed estimators. Motivated by Subzar et al. (2018b) and Jerajuddin and Kishun (2016), we consider an enhanced estimator of the population mean of peppermint oil yields (primary variable), using the area of the field as the secondary variable. Here the peppermint yield and the area of the fields are highly positively correlated; therefore, we are taking into account only ratio type estimators as candidates for improved estimators. The known information on the secondary variable is being used to enhance the efficiency of the estimators under consideration.

The rest of the paper is organised as follows: following the introduction, a review of the competing estimators is presented in Section 2. The proposed estimators are presented in Section 3. The efficiency comparison is given in Section 4. An empirical study of the proposed approach is presented in Section 5. Results and discussion are presented in Section 6, and a conclusion is given in Section 7.

2 Literature review of competing estimators

In this section, different families of estimators for improved estimation of the population mean of the main variable using known auxiliary variables are presented chronologically along with their biases and the MSEs up to an approximation of order one.

2.1 Kadilar and Cingi (2004) estimators

Kadilar and Cingi (2004) utilised known information on a secondary variable and proposed the following estimators as,

$$t_{a1} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}} \bar{X}$$

$$t_{a2} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + C_x)} (\bar{X} + C_x)$$

$$t_{a3} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + \beta_2)} (\bar{X} + \beta_2)$$

$$t_{a4} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_2 + C_x)} (\bar{X}\beta_2 + C_x)$$

$$t_{a5} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}C_x + \beta_2)} (\bar{X}C_x + \beta_2)$$

The biases and the MSEs of Kadilar and Cingi (2004) estimators are respectively given by,

$$B(t_{ai}) = \frac{1-f}{n} \frac{S_x^2}{\bar{Y}} R_{ai}^2, \; i = 1, 2, \ldots, 5$$

$$MSE(t_{ai}) = \frac{1-f}{n} \left[R_{ai}^2 S_x^2 + S_y^2 (1 - \rho^2) \right] \tag{1}$$

where $R_{a1} = \frac{\bar{Y}}{\bar{X}}, R_{a2} = \frac{\bar{Y}}{(\bar{X} + C_x)}, R_{a3} = \frac{\bar{Y}}{(\bar{X} + \beta_2)}, R_{a4} = \frac{\bar{Y}\beta_2}{(\bar{X}\beta_2 + C_x)}, R_{a5} = \frac{\bar{Y}C_x}{(\bar{X}C_x + \beta_2)}$.

2.2 Kadilar and Cingi (2006) estimators

Kadilar and Cingi (2006) made use of known parameters of the secondary variable and suggested the following estimators as,

$$t_{b1} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + \rho)} (\bar{X} + \rho)$$

$$t_{b2} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}C_x + \rho)} (\bar{X}C_x + \rho)$$

$$t_{b3}=\frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\rho+C_x)}(\bar{X}\rho+C_x)$$

$$t_{b4}=\frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\beta_2+\rho)}(\bar{X}\beta_2+\rho)$$

$$t_{b5}=\frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\rho+\beta_2)}(\bar{X}\rho+\beta_2)$$

The biases and the MSEs of Kadilar and Cingi (2006) estimators are respectively given by,

$$B(t_{bi})=\frac{1-f}{n}\frac{S_x^2}{\bar{Y}}R_{bi}^2,\ i=1,2,\ldots,5$$

$$MSE(t_{bi})=\frac{1-f}{n}\left[R_{bi}^2S_x^2+S_y^2\left(1-\rho^2\right)\right] \quad (2)$$

where $R_{b1}=\frac{\bar{Y}}{\bar{X}+\rho}$, $R_{b2}=\frac{\bar{Y}C_x}{(\bar{X}C_x+\rho)}$, $R_{b3}=\frac{\bar{Y}\rho}{(\bar{X}\rho+C_x)}$, $R_{b4}=\frac{\bar{Y}\beta_2}{(\bar{X}\beta_2+\rho)}$, $R_{b5}=\frac{\bar{Y}\rho}{(\bar{X}\rho+\beta_2)}$.

2.3 Subramani and Kumarapandiyan (2012a, 2012b) estimators

Subramani and Kumarapandiyan (2012a, 2012b) suggested the estimators given below, using auxiliary information in the form of its various parameters as,

$$t_{c1}=\frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}\beta_1+M_d)}(\bar{X}\beta_1+M_d)$$

$$t_{c2}=\frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}+M_d)}(\bar{X}+M_d)$$

$$t_{c3}=\frac{\bar{y}+b(\bar{X}-\bar{x})}{(\bar{x}C_x+M_d)}(\bar{X}C_x+M_d)$$

The biases and the MSEs of Subramani and Kumarapandiyan (2012a, 2012b) class are respectively given by,

$$B(t_{ci})=\frac{1-f}{n}\frac{S_x^2}{\bar{Y}}R_{ci}^2,\ i=1,2,3$$

$$MSE(t_{ci})=\frac{1-f}{n}\left[R_{ci}^2S_x^2+S_y^2\left(1-\rho^2\right)\right] \quad (3)$$

where $R_{c1} = \dfrac{\bar{Y}\beta_1}{\bar{X}\beta_1 + M_d}$, $R_{c2} = \dfrac{\bar{Y}\rho}{(\bar{X} + M_d)}$, $R_{c3} = \dfrac{\bar{Y}C_x}{(\bar{X}C_x + M_d)}$.

2.4 Abid et al. (2016) estimators

Abid et al. (2016) used the known correlation coefficient, the coefficient of variation and the decile mean (DM) of the secondary variable and suggested the following estimators as,

$$t_{d1} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + DM)}(\bar{X} + DM)$$

$$t_{d2} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}C_x + DM)}(\bar{X}C_x + DM)$$

$$t_{d3} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\rho + DM)}(\bar{X}\rho + DM)$$

The biases and the MSEs of Abid et al. (2016) estimators are respectively given by,

$$B(t_{di}) = \frac{1-f}{n}\frac{S_x^2}{\bar{Y}}R_{di}^2,\ i = 1, 2, 3$$

$$MSE(t_{di}) = \frac{1-f}{n}\left[R_{di}^2 S_x^2 + S_y^2\left(1-\rho^2\right)\right] \tag{4}$$

where $R_{d1} = \dfrac{\bar{Y}}{\bar{X} + DM}$, $R_{d2} = \dfrac{\bar{Y}C_x}{(\bar{X}C_x + DM)}$, $R_{d3} = \dfrac{\bar{Y}\rho}{(\bar{X}\rho + DM)}$.

2.5 Subzar et al. (2017) estimators

Subzar et al. (2017) utilised information on deciles and the quartile deviations of the secondary variable and suggested the estimators given below as,

$$t_{ei} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}QD + D_j)}(\bar{X}QD + D_j),\ i = 1, 2, \ldots, 9 \text{ and } j = 1, 2, \ldots, 9$$

$$t_{ei} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}D_j + QD)}(\bar{X}D_j + QD),\ i = 10, 11, \ldots, 18 \text{ and } j = 1, 2, \ldots, 9$$

The biases and the MSEs of Subzar et al. (2017) estimators are respectively given by,

$$B(t_{ei}) = \frac{1-f}{n}\frac{S_x^2}{\bar{Y}}R_{ei}^2,\ i = 1, 2, \ldots, 18$$

$$MSE(t_{ei}) = \frac{1-f}{n}\left[R_{ei}^2 S_x^2 + S_y^2\left(1-\rho^2\right)\right] \tag{5}$$

where

$$R_{ei} = \frac{\bar{Y}QD}{\bar{X}QD + D_j}, i = 1, 2, \ldots, 9 \text{ and } j = 1, 2, \ldots, 9$$

and

$$R_{ei} = \frac{\bar{Y}D_j}{\bar{X}D_j + QD}, i = 10, 11, \ldots, 18 \text{ and } j = 1, 2, \ldots, 9$$

2.6 Subzar et al. (2018a) estimators

Subzar et al. (2018a) employed information on known non-traditional parameters of a secondary variable and suggested the following estimators as,

$$t_{f1} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_1 + TM)}(\bar{X}\beta_1 + TM)$$

$$t_{f2} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_1 + MR)}(\bar{X}\beta_1 + MR)$$

$$t_{f3} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_1 + HL)}(\bar{X}\beta_1 + HL)$$

$$t_{f4} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_2 + TM)}(\bar{X}\beta_2 + TM)$$

$$t_{f5} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_2 + MR)}(\bar{X}\beta_2 + MR)$$

$$t_{f6} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_2 + HL)}(\bar{X}\beta_2 + HL)$$

The biases and the MSEs of the Subzar et al. (2018a) estimators are respectively given by,

$$B(t_{fi}) = \frac{1-f}{n}\frac{S_x^2}{\bar{Y}}R_{fi}^2, i = 1, 2, \ldots, 6$$

$$MSE(t_{fi}) = \frac{1-f}{n}\left[R_{fi}^2 S_x^2 + S_y^2(1-\rho^2)\right] \tag{6}$$

where $R_{f1} = \frac{\bar{Y}\beta_1}{\bar{X}\beta_1 + TM}$, $R_{f2} = \frac{\bar{Y}\beta_1}{\bar{X}\beta_1 + MR}$, $R_{f3} = \frac{\bar{Y}\beta_1}{\bar{X}\beta_1 + HL}$, $R_{f4} = \frac{\bar{Y}\beta_2}{\bar{X}\beta_2 + TM}$, $R_{f5} = \frac{\bar{Y}\beta_2}{\bar{X}\beta_2 + MR}$, $R_{f6} = \frac{\bar{Y}\beta_2}{\bar{X}\beta_2 + HL}$.

2.7 Subzar et al. (2018b) estimators

Subzar et al. (2018b) suggested the following estimators utilising information on some known traditional and non-traditional parameters of the secondary variable as,

$$t_{gi} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + \psi_j)}(\bar{X} + \psi_j), i = 1, 2, \ldots, 6 \text{ and } j = 1, 2, \ldots, 6$$

$$t_{gi} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\rho + \psi_j)}(\bar{X}\rho + \psi_j), i = 7, 8, \ldots, 12 \text{ and } j = 1, 2, \ldots, 6$$

$$t_{gi} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}C_x + \psi_j)}(\bar{X}C_x + \psi_j), i = 13, 14, \ldots, 18 \text{ and } j = 1, 2, \ldots, 6$$

The biases and the MSEs of the Subzar et al. (2018b) estimators are respectively given by,

$$B(t_{gi}) = \frac{1-f}{n}\frac{S_x^2}{\bar{Y}}R_{gi}^2, i = 1, 2, \ldots, 18$$

$$MSE(t_{gi}) = \frac{1-f}{n}\left[R_{gi}^2 S_x^2 + S_y^2(1-\rho^2)\right] \tag{7}$$

where

$$R_{gi} = \frac{\bar{Y}}{\bar{X} + \psi_j}, i = 1, 2, \ldots, 6 \text{ and } j = 1, 2, \ldots, 6$$

$$R_{gi} = \frac{\bar{Y}\rho}{\bar{X}\rho + \psi_j}, i = 7, 8, \ldots, 12 \text{ and } j = 1, 2, \ldots, 6$$

$$R_{gi} = \frac{\bar{Y}C_x}{\bar{X}C_x + \psi_j}, i = 13, 14, \ldots, 18 \text{ and } j = 1, 2, \ldots, 6$$

and $\psi_1 = (M_d + G)$, $\psi_2 = (M_d + G)$, $\psi_3 = (M_d + S_{pw})$, $\psi_4 = (QD + G)$, $\psi_5 = (QD + G)$, $\psi_6 = (QD + S_{pw})$.

3 Proposed estimators

We are searching for the most efficient estimators for the mean of the primary variable under consideration. As suggested by the list of recent research looking for improved estimators using auxiliary information, researchers consider improvement to be possible. Motivated by Subzar et al. (2018b) and Jerajuddin and Kishun (2016), we suggest the following estimators of population mean using known non-conventional parameters of the auxiliary variable as,

$$t_{pi} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + \varphi_j)}(\bar{X} + \varphi_j), i = 1, 2, \ldots, 8, \text{ and } j = 1, 2, \ldots, 8$$

$$t_{pi} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\rho + \varphi_j)}(\bar{X}\rho + \varphi_j), i = 9, 10, \ldots, 16 \text{ and } j = 1, 2, \ldots, 8$$

$$t_{pi} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}C_x + \varphi_j)}(\bar{X}C_x + \varphi_j), i = 17, 18, \ldots, 24 \text{ and } j = 1, 2, \ldots, 8$$

where

$$R_{pi} = \frac{\bar{Y}}{\bar{X} + \varphi_j}, i = 1, 2, \ldots, 8 \text{ and } j = 1, 2, \ldots, 8$$

$$R_{pi} = \frac{\bar{Y}\rho}{\bar{X}\rho + \varphi_j}, i = 9, 10, \ldots, 16 \text{ and } j = 1, 2, \ldots, 8$$

$$R_{pi} = \frac{\bar{Y}C_x}{\bar{X}C_x + \varphi_j}, i = 17, 18, \ldots, 24 \text{ and } j = 1, 2, \ldots, 8$$

and $\varphi_1 = (QD + n)$, $\varphi_2 = (DM + n)$, $\varphi_3 = (TM + n)$, $\varphi_4 = (MR + n)$, $\varphi_5 = (HL + n)$, $\varphi_6 = (G + n)$, $\varphi_7 = (D + n)$, $\varphi_8 = (S_{pw} + n)$.

3.1 Bias and MSE of the proposed class

In this section, the large sampling properties including bias and the MSE of the proposed estimators using the Taylor series method are derived up to the approximation of order one and are given by,

$$h(\bar{x}, \bar{y}) \cong h(\bar{X}, \bar{Y}) + \frac{\partial(c, d)}{\partial c}|_{\bar{X},\bar{Y}} (\bar{x} - \bar{X}) + \frac{\partial(c, d)}{\partial d}|_{\bar{X},\bar{Y}} (\bar{y} - \bar{Y}) \tag{8}$$

where $h(\bar{x}, \bar{y}) = \hat{R}_{pj}$ and $h(\bar{X}, \bar{Y}) = R$ with $R = \bar{Y} / \bar{X}$.

Applying the method in equation (8) to the proposed estimators as Wolter (1985) has discussed earlier, the MSE of suggested class is obtained as,

$$\begin{aligned} \hat{R}_{pj} - R &\cong \frac{\partial\left(\left(\bar{y} + b(\bar{X} - \bar{x})\right)/(\varphi_j\bar{x} + \phi_j)\right)}{\partial\bar{x}}|_{\bar{X},\bar{Y}} (\bar{x} - \bar{X}) \\ &+ \frac{\partial\left(\left(\bar{y} + b(\bar{X} - \bar{x})\right)/(\varphi_j\bar{x} + \phi_j)\right)}{\partial\bar{y}}|_{\bar{X},\bar{Y}} (\bar{y} - \bar{Y}) \\ &\cong -\left(\frac{\bar{y}}{(\bar{x}\varphi_j + \phi_j)^2} + \frac{b(\bar{X}\varphi_j + \phi_j)}{(\bar{x}\varphi_j + \phi_j)^2}\right)|_{\bar{X},\bar{Y}} (\bar{x} - \bar{X}) + \frac{1}{(\bar{x}\varphi_j + \phi_j)}|_{\bar{X},\bar{Y}} (\bar{y} - \bar{Y}) \\ &\cong -\left(\frac{\bar{y} + b(\bar{X}\varphi_j + \phi_j)}{(\bar{x}\varphi_j + \phi_j)^2}\right)|_{\bar{X},\bar{Y}} (\bar{x} - \bar{X}) + \frac{1}{(\bar{x}\varphi_j + \phi_j)}|_{\bar{X},\bar{Y}} (\bar{y} - \bar{Y}) \end{aligned} \tag{9}$$

Squaring on both sides of equation (9), we have,

$$E\left(\hat{R}_{pj}-R\right)^2 \cong \left(\frac{\left(\bar{Y}+B\left(\bar{X}\varphi_j+\phi_j\right)\right)^2}{\left(\bar{X}\varphi_j+\phi_j\right)^4}\right)V(\bar{x})$$
$$-2\left(\frac{\bar{Y}+B\left(\bar{X}\varphi_j+\phi_j\right)}{\left(\bar{X}\varphi_j+\phi_j\right)^3}\right)Cov(\bar{x},\bar{y})+\left(\frac{1}{\left(\bar{X}\varphi_j+\phi_j\right)^2}\right)V(\bar{y})$$
$$\cong \frac{1}{\left(\bar{X}\varphi_j+\phi_j\right)^2}\left[\left(\frac{\left(\bar{Y}+B\left(\bar{X}\varphi_j+\phi_j\right)\right)^2}{\left(\bar{X}\varphi_j+\phi_j\right)^2}\right)V(\bar{x})\right.$$
$$\left.-2\left(\frac{\bar{Y}+B\left(\bar{X}\varphi_j+\phi_j\right)}{\left(\bar{X}\varphi_j+\phi_j\right)}\right)Cov(\bar{x},\bar{y})+V(\bar{y})\right]$$

where

$$B=\frac{S_{yx}}{S_x^2}=\frac{\rho S_x S_y}{S_x^2}=\frac{\rho S_y}{S_x} \tag{10}$$

and φ_j and ϕ_j are the parameters of the auxiliary variable.

It is worth noting that the difference $[E(b)-B]$ is omitted because it is assumed that the line of regression is passing through the origin as an unbiased ratio estimator. Sukhatme (1954) noted that the condition for the bias to be zero is meaningful since the regression estimates, by both the weighted least square method and ordinary least square method which is homoscedastic, are both unbiased for B. Thus, the design-based ratio estimators are unbiased when the line of regression passes through the origin with homoscedastic regression coefficients.

Thus, the MSE of the suggested estimators is,

$$MSE\left(t_{pj}\right)=\left(\bar{X}\varphi_j+\phi_j\right)^2 E\left(\hat{R}_{pj}-R\right)^2$$
$$\cong\left[\left(\frac{\left(\bar{Y}+B\left(\bar{X}\varphi_j+\phi_j\right)\right)^2}{\left(\bar{X}\varphi_j+\phi_j\right)^2}\right)V(\bar{x})-2\left(\frac{\bar{Y}+B\left(\bar{X}\varphi_j+\phi_j\right)}{\left(\bar{X}\varphi_j+\phi_j\right)}\right)Cov(\bar{x},\bar{y})+V(\bar{y})\right]$$
$$\cong\left[\left(\frac{\bar{Y}^2+2B\left(\bar{X}\varphi_j+\phi_j\right)\bar{Y}+B^2\left(\bar{X}\varphi_j+\phi_j\right)^2}{\left(\bar{X}\varphi_j+\phi_j\right)^2}\right)V(\bar{x})\right.$$
$$\left.-\left(\frac{2\bar{Y}+2B\left(\bar{X}\varphi_j+\phi_j\right)}{\left(\bar{X}\varphi_j+\phi_j\right)}\right)Cov(\bar{x},\bar{y})+V(\bar{y})\right]$$
$$\cong\frac{1-f}{n}\left[\left(\frac{\bar{Y}^2}{\left(\bar{X}\varphi_j+\phi_j\right)^2}+\frac{+2B\bar{Y}}{\left(\bar{X}\varphi_j+\phi_j\right)}+B^2\right)S_x^2-\left(\frac{2\bar{Y}}{\left(\bar{X}\varphi_j+\phi_j\right)}+2B\right)S_{yx}+S_y^2\right]$$
$$\cong\frac{1-f}{n}\left[\left(R_{pj}^2+2BR_{pj}+B^2\right)S_x^2-2\left(R_{pj}+B\right)S_{yx}+S_y^2\right]$$

Putting the value of B from equation (10) in above equation and simplifying, we get,

$$MSE\left(t_{pj}\right) \cong \frac{1-f}{n}\left[(R_{pj}^2 S_x^2 - \left(1-\rho^2\right) S_y^2\right], \; j = 1, 2, \ldots, 24 \tag{11}$$

Similarly the expression for the bias of the proposed estimators is given by,

$$B\left(t_{pj}\right) \cong \frac{1-f}{n} \frac{S_x^2}{\bar{Y}} R_{pj}^2, \; j = 1, 2, \ldots, 24 \tag{12}$$

4 Efficiency comparison

The various competing classes of estimators are theoretically compared with the suggested class of estimators in this section. The theoretical conditions of efficiencies under which the suggested class is better than the competing one are also mentioned.

4.1 Comparison with Kadilar and Cingi (2004) estimators

The suggested family performs better than Kadilar and Cingi (2004) family if,

$$MSE\left(t_{ai}\right) - MSE\left(t_{pj}\right) > 0, i = 1, 2, \ldots, 5, \; j = 1, 2, \ldots, 24 \; or \; R_{ai}^2 - R_{pj}^2 > 0$$

4.2 Comparison with Kadilar and Cingi (2006) estimators

The proposed family is more efficient than the Kadilar and Cingi (2006) family if,

$$MSE\left(t_{bi}\right) - MSE\left(t_{pj}\right) > 0, i = 1, 2, \ldots, 5, \; j = 1, 2, \ldots, 24 \text{ or } R_{bi}^2 - R_{pj}^2 > 0$$

4.3 Comparison with Subramani and Kumarapandiyan (2012a, 2012b) estimators

The suggested family performs better than the Subramani and Kumarapandiyan (2012a, 2012b) family if,

$$MSE\left(t_{ci}\right) - MSE\left(t_{pj}\right) > 0, i = 1, 2, 3, \; j = 1, 2, \ldots, 24 \text{ or } R_{ci}^2 - R_{pj}^2 > 0$$

4.4 Comparison with Abid et al. (2016) estimators

The suggested family performs better than the Abid et al. (2016) family if,

$$MSE\left(t_{di}\right) - MSE\left(t_{pj}\right) > 0, i = 1, 2, 3, \; j = 1, 2, \ldots, 24 \text{ or } R_{di}^2 - R_{pj}^2 > 0$$

4.5 Comparison with Subzar et al. (2017) estimators

The suggested family is better than the Subzar et al. (2017) family if,

$$MSE\left(t_{ei}\right) - MSE\left(t_{pj}\right) > 0, i = 1, 2, \ldots, 18, \; j = 1, 2, \ldots, 24 \text{ or } R_{ei}^2 - R_{pj}^2 > 0$$

4.6 Comparison with Subzar et al. (2018a) estimators

The suggested family performs better than the Subzar et al. (2018) family if,

$$MSE\left(t_{fi}\right)-MSE\left(t_{pj}\right)>0,\ i=1,2,\ldots,9,\ j=1,2,\ldots,24 \text{ or } R_{fi}^{2}-R_{pj}^{2}>0$$

4.7 Comparison with Subzar et al. (2018b) estimators

The suggested family is better than the Subzar et al. (2018) family if,

$$MSE\left(t_{gi}\right)-MSE\left(t_{pj}\right)>0,\ i=1,2,\ldots,18,\ j=1,2,\ldots,24 \text{ or } R_{gi}^{2}-R_{pj}^{2}>0$$

The above efficiency conditions are verified through the numerical example in the following section.

5 Empirical study

This section presents a case example to demonstrate the usefulness of the proposed approach. In this study, we have used actual peppermint oil yields along with the area of the associated fields collected from Siddhaur Block of Barabanki District at Uttar Pradesh State in India. The parameters of the population under investigation are given in Table 1. The dependent variable and the auxiliary variable are as follows:

Y the production (yield) of peppermint oil in kilograms

X the area of the field in Bigha (1 Bigha = 0.2529 hectare).

Table 1 Parameters of the population under consideration

$N = 150$, $n = 40$, $\gamma = 0.018333$, $\bar{Y} = 79.58$, $\bar{X} = 6.5833$, $\rho = 0.9369$, $C_y = 0.781333$, $C_x = 0.661726$, $S_y^2 = 3{,}866.165$, $S_x^2 = 18.97791$, $\beta_1 = 1.4984$, $\beta_2 = 5.408$, $Q_1 = 4$, $M_d = 5$, $Q_3 = 10$, $D_1 = 2$, $D_2 = 3$, $D_3 = 4$, $D_4 = 5$, $D_5 = 5$, $D_6 = 6$, $D_7 = 8$, $D_8 = 10$, $D_9 = 13$, $QD = 3$, $DM = 6.22$, $TM = 6$, $MR = 11$, $HL = 7$, $G = 8.2298$, $D = 9.2542$, $S_{pw} = 9.3707$

The MSEs of the proposed and the competing families of estimators for the above population are presented in Tables 2a and 2b.

Table 2 MSEs of competing and proposed class of estimators

a

Estimators	*MSE*	*Estimators*	*MSE*	*Estimators*	*MSE*
t_{a1}	59.5772	t_{d2}	17.3618	t_{e13}	51.4388
t_{a2}	50.7146	t_{d3}	21.3321	t_{e14}	51.4388
t_{a3}	24.0605	t_{e1}	50.6571	t_{e15}	52.6530
t_{a4}	57.7388	t_{e2}	47.0528	t_{e16}	54.2450
t_{a5}	18.8559	t_{e3}	43.8941	t_{e17}	55.2422
t_{b1}	47.7047	t_{e4}	41.1105	t_{e18}	56.1919
t_{b2}	43.1799	t_{e5}	41.1105	t_{f1}	28.3932
t_{b3}	50.1977	t_{e6}	38.6448	t_{f2}	20.1009

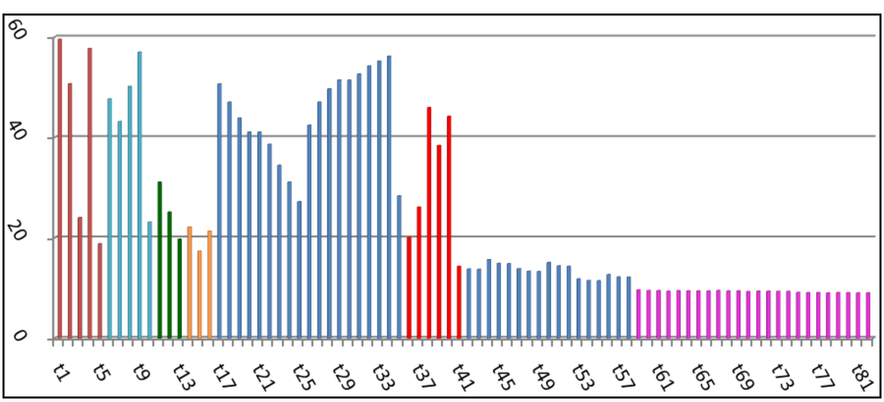
Table 2 MSEs of competing and proposed class of estimators (continued)

		a			
Estimators	*MSE*	*Estimators*	*MSE*	*Estimators*	*MSE*
t_{b4}	57.0050	t_{e7}	34.4891	t_{f3}	26.1312
t_{b5}	23.1617	t_{e8}	31.1429	t_{f4}	45.9701
t_{c1}	31.1269	t_{e9}	27.2260	t_{f5}	38.4091
t_{c2}	25.1589	t_{e10}	42.4592	t_{f6}	44.2426
t_{c3}	19.7578	t_{e11}	47.0528		
t_{d1}	22.1784	t_{e12}	49.7098		

		b			
Estimators	*MSE*	*Estimators*	*MSE*	*Estimators*	*MSE*
t_{g1}	14.3495	t_{g15}	11.4875	t_{p11}	9.4465
t_{g2}	13.8112	t_{g16}	12.7080	t_{p12}	*9.3277*
t_{g3}	13.7549	t_{g17}	12.2333	t_{p13}	9.4200
t_{g4}	15.6806	t_{g18}	12.1847	t_{p14}	9.3895
t_{g5}	14.9459	t_{p1}	9.6328	t_{p15}	9.3655
t_{g6}	14.8698	t_{p2}	9.5269	t_{p16}	9.3629
t_{g7}	13.8723	t_{p3}	9.5335	t_{p17}	9.1668
t_{g8}	13.3699	t_{p4}	*9.4011*	t_{p18}	9.1138
t_{g9}	13.3174	t_{p5}	9.5040	t_{p19}	9.1171
t_{g10}	15.1212	t_{p6}	9.4700	t_{p20}	*9.0514*
t_{g11}	14.4307	t_{p7}	9.4433	t_{p21}	9.1024
t_{g12}	14.3594	t_{p8}	9.4404	t_{p22}	9.0855
t_{g13}	11.8561	t_{p9}	9.5357	t_{p23}	9.0723
t_{g14}	11.5221	t_{p10}	9.4405	t_{p24}	9.0708

The MSEs of the various proposed (*p*) and competing (*a*–*g*) families of estimators are shown in Figure 1. The MSEs of the suggested family is shown in pink.

Figure 1 MSE of various suggested and competing families of estimators (see online version for colours)



The percentage relative efficiency (PRE) of the proposed class over the competing classes of estimators have been shown in Figures 2–4 respectively. The PREs have been calculated for the three types of proposed estimators having the smallest value in each type which were highlighted in Table 2b in ital.

Figure 2 PRE of type-1 proposed estimator with least MSE over competing estimators (see online version for colours)

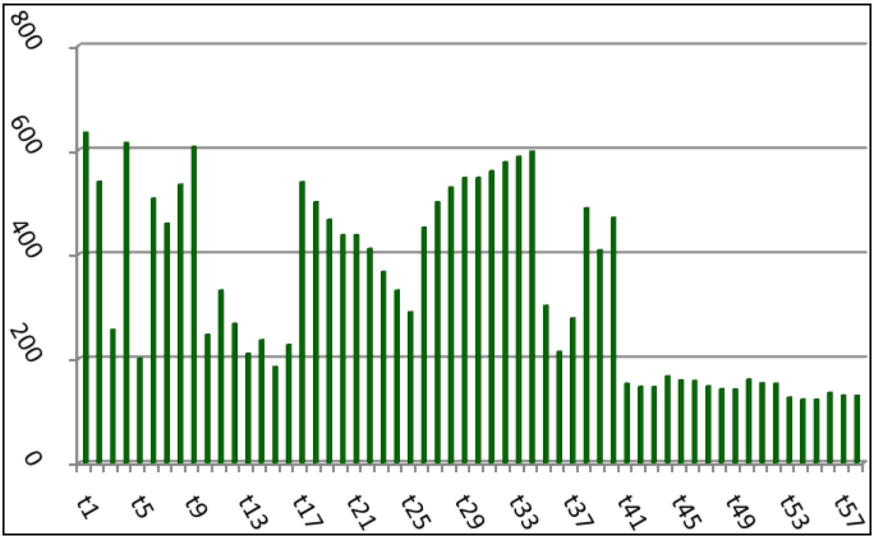


Figure 3 PRE of type-2 proposed estimator with least MSE over competing estimators (see online version for colours)

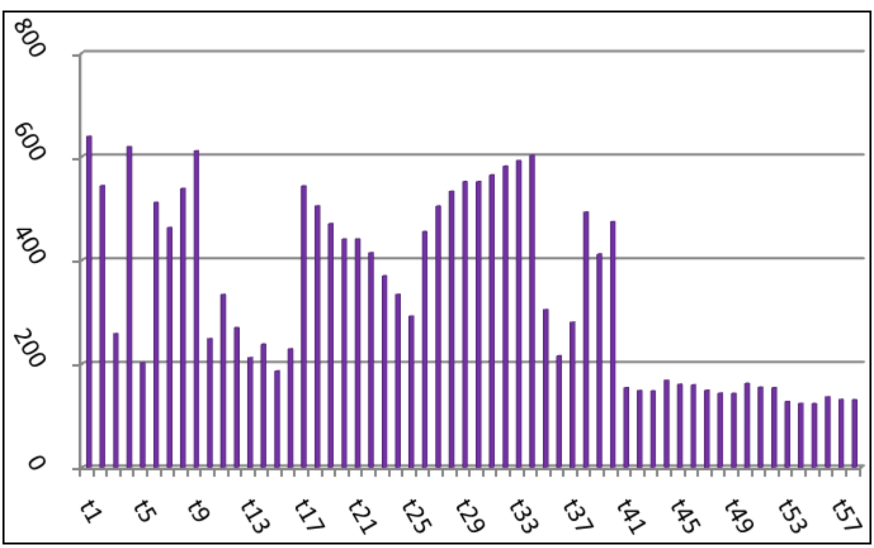
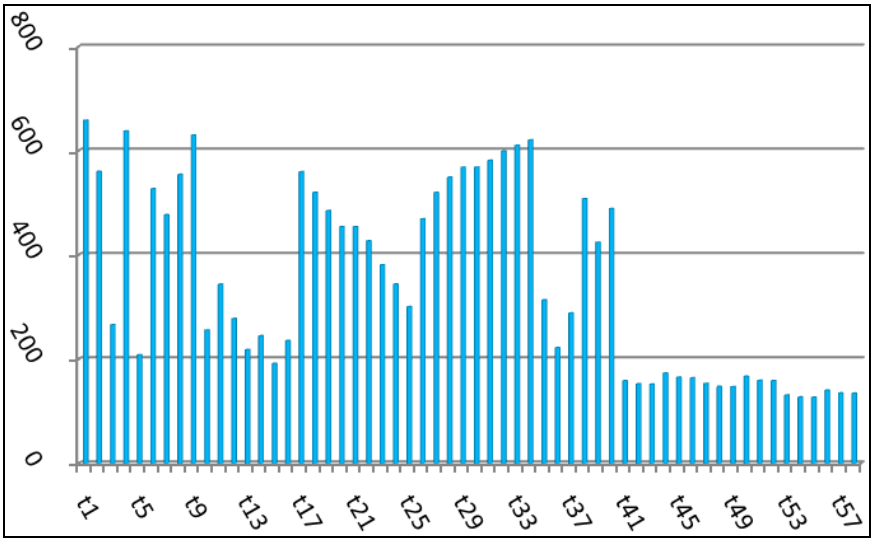


Figure 4 PRE of type-3 proposed estimator with least MSE over competing estimators (see online version for colours)



6 Results and discussion

From Tables 2a and 2b, it is readily evident that the mean squared errors of the competing classes of the population mean range from 23.1617 to 59.5772 while the MSEs of proposed estimators range from 9.0514 to 9.6328. Further, it is worth noting that the 24 suggested estimators have been divided into three types of sub-classes, each sub-class having eight estimators. The first type of sub-class used non-traditional known parameters of the auxiliary variable along with its population mean. The second type of sub-class used correlation coefficients along with the population mean of the secondary variable and its non-traditional parameters. The third type of sub-class used information on the coefficient of variation of a supplementary variable along with its population mean and non-traditional parameters. From Table 2b, it is evident that the suggested family has lesser MSE than competing families of estimators under a simple random sampling scheme. Our results indicate that the smallest values of MSEs of the sub-classes are 9.4011, 9.3277 and 9.0514 respectively. Thus, the proposed estimators are the most efficient among the class of competing estimators of the population mean of the primary variable. Hence, the suggested estimator will provide a more efficient estimate of average yield of peppermint than any other competing estimator, which will be beneficial in policy decisions.

7 Conclusions

In the present study, we suggested an improved family of estimators for the population mean of peppermint oil yields using the area of the field as the auxiliary variable. We derived the expressions for the bias and MSEs of the suggested family up to an approximation of order one using Taylor series expansion. The suggested family is compared with competing families of estimators both theoretically and empirically. It was demonstrated through a numerical study that the suggested family is more efficient than the competing classes by Kadilar and Cingi (2004, 2006), Subramani and Kumarpandiyan (2012a, 2012b), Abid et al. (2016) and Subzar et al. (2017, 2018a, 2018b) respectively as the proposed class has smaller mean squared errors in comparison to all these competing classes under a simple random sampling scheme. Our results show that the proposed estimators are the most efficient to estimate the average yield of peppermint oil and can be applied to determine how much area to plough to meet the requirement for that yield. The proposed estimators may be used for enhanced estimation of the average yield of any crop under a simple random sampling scheme. In future studies, as we search for better estimators for the parameters under consideration, it may be possible to get more efficient estimators using different auxiliary variable information.

Acknowledgements

The authors are thankful to unknown referees and the chief editor for minutely observing the things which improved the earlier draft. The first author Dr. S.K. Yadav (PI) is thankful to Council of Science and Technology, Uttar Pradesh, Lucknow for financial assistance in carrying out this research under the Project No. CST/4052.

References

Abid, M., Abbas, N. Sherwani, R.A.K. and Nazir, H.Z. (2016) 'Improved ratio estimators for the population mean using non-conventional measure of dispersion', *Pakistan Journal of Statistics and Operations Research*, Vol. 12, No. 2, pp.353–367.

Al-Omari, A.I., Jemain, A.A. and Ibrahim, K. (2009) 'New ratio estimators of the mean using simple random sampling and ranked set sampling methods', *Revista Investigación Operacional*, Vol. 30, No. 2, pp.97–108.

Cochran, W.G. (1940) 'The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce', *The Journal of Agric. Science*, Vol. 30, No. 2, pp.262–275.

Goodman, L.A. and Hartley, H.O. (1958) 'The precision of unbiased ratio-type estimators', *Journal of the American Statistical Association*, Vol. 53, No. 282, pp.491–508.

Gupta, R.K. and Yadav, S.K. (2017) 'New efficient estimators of population mean using non-traditional measures of dispersion', *Open Journal of Statistics*, Vol. 7, pp.394–404.

Gupta, R.K. and Yadav, S.K. (2018) 'Improved estimation of population mean using information on size of the sample', *American Journal of Mathematics and Statistics*, Vol. 8, No. 2, pp.27–35.

Gupta, S. and Shabbir, J. (2008) 'On the improvement in estimating the population mean in simple random sampling', *Journal of Applied Statistics*, Vol. 35, No. 5, pp.559–566.

Hartley, H.O. and Ross, A. (1954) 'Unbiased ratio estimators', *Nature*, August, Vol. 7, No. 174, pp.270–271.

Hutchison, M.C. (1971) 'A Monte Carlo comparison of some ratio estimators', *Biometrika*, Vol. 58, No. 2, pp.313–321.

Ijaz, M. and Ali, H. (2018) 'Some improved ratio estimators for estimating mean of finite population', *Research & Reviews: Journal of Statistics and Mathematical Sciences*, Vol. 4, No. 2, pp.18–23.

Jaroengeratikun, U. and Lawson, N. (2019) 'A combined family of ratio estimators for population mean using an auxiliary variable in simple random sampling', *Journal of Mathematics and Fundamental Sciences*, Vol. 51, No. 1, pp.1–12.

Jerajuddin, M. and Kishun, J. (2016) 'Modified ratio estimators for population mean using size of the sample, selected from population', *IJSRSET*, Vol. 2, No. 2, pp.10–16.

Kadilar, C. and Cingi, H. (2004) 'Ratio estimators in simple random sampling', *Applied Mathematics and Computation*, Vol. 151, pp.893–902.

Kadilar, C. and Cingi, H. (2006) 'An improvement in estimating the population mean by using the correlation coefficient', *Hacettepe Journal of Mathematics and Statistics*, Vol. 35, No. 1, pp.103–109.

Khan, M. and Al-Hossain, A.Y. (2016) 'A note on a difference type estimator for population mean under two phase sampling design', *SpringerPlus*, Vol. 5, No. 723, pp.1–7.

Koyuncu, N. and Kadilar, C. (2009) 'Efficient estimators for the population mean', *Hacettepe Journal of Mathematics and Statistics*, Vol. 38, No. 2, pp.217–225.

Murthy, M.N. (1964) 'Product method of estimation', *Sankhya A*, Vol. 26, pp.69–74.

Pal, S.K. and Singh, H.P. (2016) 'Finite population mean estimation through a two parameter ratio estimator using auxiliary information in presence of non-response', *JAMSI*, Vol. 12, No. 2, pp.5–39.

Prasad, B. (1989) 'Some improved ratio type estimators of population mean and ratio in finite population sample surveys', *Communications in Statistics-Theory and Methods*, Vol. 18, No. 3, pp.379–392.

Rao, T.J. (1991) 'On certain methods of improving ratio and regression estimators', *Communications in Statistics-Theory and Methods*, Vol. 20, No. 10, pp.3325–3340.

Robinson, J. (1987) 'Conditioning ratio estimates under simple random sampling', *Journal of the American Statistical Association*, Vol. 82, No. 399, pp.826–831.

Robson, D.S. (1957) 'Application of multivariate polykays to the theory of unbiased ratio-type estimation', *Journal of American Statistical Association*, Vol. 59, pp.1225–1226.

Shabbir, J. and Gupta, S. (2011) 'On estimating finite population mean in simple and stratified sampling', *Communications in Statistics – Theory and Methods*, Vol. 40, No. 2, pp.199–212.

Sharma, P. and Singh, R. (2013) 'Improved estimators for simple random sampling and stratified random sampling under second order of approximation', *Statistics in Transition – New Series*, Vol. 14, No. 3, pp.379–390.

Singh, H.P and Solanki, R.S. (2012) 'Improved estimation of population mean in simple random sampling using information on auxiliary attribute', *Applied Mathematics and Computation*, Vol. 218, No. 15, pp.7798–7812.

Singh, H.P. and Tailor, R. (2003) 'Use of known correlation coefficient in estimating the finite population means', *Stat Transit*, Vol. 6, No. 4, pp.555–560.

Singh, L., Yadav, S.K. and Mishra, S.S. (2018) 'A new efficient estimator for the population mean', *International journal of Agricultural and Statistical Sciences*, Vol. 14, No. 2, pp.671–677.

Srija and Subramani, J. (2018) 'Modified ratio cum product estimators for finite population mean with known median and mean', *SSRG International Journal of Economics Management Studies*, Special Issue ICRTECITA, April, pp.1–9.

Subramani, J. (2016) 'A new median based ratio estimator for estimation of the finite population mean', *Statistics in Transition New Series*, Vol. 17, No. 4, pp.1–14.

Subramani, J. and Kumarapandiyan, G. (2012a) 'Estimation of population mean using co-efficient of variation and median of an auxiliary variable', *International Journal of Probability and Statistics*, Vol. 1, No. 4, pp.111–118.

Subramani, J. and Kumarapandiyan, G. (2012b) 'Estimation of population mean using known median and co-efficient of skewness', *American Journal of Mathematics and Statistics*, Vol. 2, No. 5, pp.101–107.

Subramani, J. and Kumarapandiyan, G. (2013) 'A new modified ratio estimator of population mean when median of the auxiliary variable is known', *Pakistan Journal of Statistics and Operation Research*, Vol. 9, No. 2, pp.137–145.

Subramani, J. and Master Ajith, S. (2017) 'Improved ratio cum product estimators for finite population mean with known quartiles and their functions', *SM J. Biometrics Biostat*., Vol. 2, No. 1, p.1008.

Subzar, M., Maqbool, S., Raja, T.A. and Abid, M. (2018a) 'Ratio estimators for estimating population mean in simple random sampling using auxiliary information', *Applied Mathematics and Information Sciences Letters*, Vol. 6, No. 3, pp.123–130.

Subzar, M., Maqbool, S., Raja, T.A. and Shabeer, M. (2017) 'A new ratio estimators for estimation of population mean using conventional location parameters', *World Applied Sciences Journal*, Vol. 35, No. 3, pp.377–384.

Subzar, M., Maqbool, S., Raja, T.A., Pal, S.K. and Sharma, P. (2018b) 'Efficient estimators of population mean using auxiliary information under simple random sampling', *Statistics in Transition New Series*, Vol. 19, No. 2, pp.219–238.

Sukhatme, P.V. (1954) *Sampling Theory of Surveys with Applications*, The Lowa State College Press.

Upadhyaya, L.N. and Singh, H.P. (1999) 'Use of transformed auxiliary variable in estimating the finite population mean', *Biometrical Journal*, Vol. 41, No. 5, pp.627–636.

Watson, D.J. (1937) 'The estimation of leaf area in field crops', *The Journal of Agricultural Science*, Vol. 27, No. 3, pp.474–483.

Wolter, K.M. (1985) *Introduction to Variance Estimation*, Springer-Verlag.

Yadav, S.K. and Kadilar, C. (2013a) 'Improved class of ratio and product estimators', *Applied Mathematics and Computation*, Vol. 219, pp.10726–10731.

Yadav, S.K. and Kadilar, C. (2013b) 'Efficient family of exponential estimator for population mean', *Hacettepe Journal of Mathematics and Statistics*, Vol. 42, No. 6, pp.671–677.

Yadav, S.K. and Mishra, S.S. (2015) 'Developing improved predictive estimator for finite population mean using auxiliary information', *Statistika*, Vol. 95, No. 1, pp.76–85.

Yadav, S.K. and Pandey, H. (2017) 'A new difference type median based estimator of the finite population mean', *International journal of Agricultural and Statistical Sciences*, Vol. 13, No. 1, pp.289–295.

Yadav, S.K., Gupta, S., Mishra, S.S. and Shukla, A.K. (2016) 'Modified ratio and product estimators for estimating population mean in two-phase sampling', *American Journal of Operational Research*, Vol. 6, No. 3, pp.61–68.

Yadav, S.K., Sharma, D.K. and Mishra, S.S. (2019) 'New modified ratio type estimator of population mean using known median of study variable', *International Journal of Operational Research*, Vol. 35, No. 2, 10.1504/IJOR.2021.10016339.

Yadav, S.K., Sharma, D.K., Mishra, S.S. and Shukla, A.K. (2018) 'Use of auxiliary variables in searching efficient estimator of population mean', *International Journal of Multivariate Data Analysis*, Vol. 1, No. 3, pp.230–244.

Yadav, S.K., Singh, L., Mishra, S.S., Mishra, P.P. and Kumar, S. (2017) 'A median based regression type estimator of the finite population mean', *International Journal of Agricultural and Statistical Sciences*, Vol. 13, No. 1, pp.265–271.

Zaman, T. (2019) 'Improvement in estimating the population mean in simple random sampling using coefficient of skewness of auxiliary attribute', *Journal of Natural and Applied Sciences*, Vol. 23, No. 1, pp.98–102.

Zatezalo, T., Gupta, S., Yadav, S.K. and Shabbir, J. (2018) 'Assessing the adequacy of first order approximations in ratio type estimators', *Journal of Interdisciplinary Mathematics*, Vol. 21, No. 6, pp.1395–1411.