EURASIAN JOURNAL OF ECONOMICS AND STATISTICS
Vol. 2, No. 1, 2025, pp. 49-58 • ISSN: 2584-2684

URL: http://www.pbjournals.com/sjss
Publisher of Open Access Journals

An Improved Difference and Regression Type Estimator of Finite Population Mean

Abhijeet Vishwakarma[1], S.A.H. Rizvi[2] and Nazia Naqvi[3]
[1]*University of Lucknow, India (abhijeetvishwakarma12jan@gmail.com)*
[2]*Shia P.G. College, India (rizvi_athar@hotmail.com)*
[3]*Shia P.G. College, India (naqvi.nazia8@gmail.com)*

To Cite this Article
Abhijeet Vishwakarma, S.A.H. Rizvi & Nazia Naqvi (2025). An Improved Difference and Regression Type Estimator of Finite Population Mean. *Eurasian Journal of Economics and Statistics,* 2: 1, pp. 49-58.

Abstract: For estimating finite population mean, an efficient difference and regression type estimator using prior information in the form of population mean and coefficient of variation of auxiliary variable is proposed, its bias and mean square error are found, estimator based on estimated optimum values is obtained and its comparative study with the usual linear regression estimator is made. A simulated empirical study has also been conducted to support the theoretical findings.

Keywords: Difference and Regression type estimator, Auxiliary information, Bias, Mean Square Error (M.S.E).

1.1. Introduction

Let the study variable under consideration and the auxiliary variable be $\boldsymbol{y}$ and $\boldsymbol{x}$ respectively. For population values $\boldsymbol{Y_1}, \boldsymbol{Y_2}, \dots, \boldsymbol{Y_N}$ on the study variable $\boldsymbol{y}$ and population values $\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_N}$ on auxiliary variable $\boldsymbol{x}$;

let,

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i \,,$$

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$S_y{}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2,$$

Received : 16 May 2025 • **Revised :** 14 June 2025 • **Accepted :** 18 June 2025 • **Published :** 30 June 2025

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

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

$$\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^r (X_i - \bar{X}\,)^s,$$

$$C_y = \frac{S_y}{\bar{Y}} = \frac{\sqrt{\mu_{20}}}{\bar{Y}},\ C_x = \frac{S_x}{\bar{X}} = \frac{\sqrt{\mu_{02}}}{\bar{X}},\ \beta = \frac{S_{yx}}{S_x{}^2} = \rho\,\frac{S_y}{S_x}$$

$$\rho = \frac{\mu_{11}}{\sqrt{\mu_{20}\,\mu_{02}}} = \frac{S_{yx}}{S_x S_y} \text{ and}$$

$$\beta_2(x) = \frac{\mu_{04}}{\mu_{02}{}^2}$$

Let n pairs $(y_i, x_i), i = 1,2, \ldots, n$ be observations on the variable (y, x) taken under simple random sampling without replacement scheme from a population of size N.

Further, let, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\,,$

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

$$s_{yx} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})\,(x_i - \bar{x}), \text{ and } b = \frac{s_{yx}}{s_x{}^2}$$

A few sampling methods which depend on the information on auxiliary variable x have been commonly used for the estimation of population mean $\bar{Y}$ with increased efficiency. When the population mean $\bar{X}$ of auxiliary variable x is known, estimators such as ratio, product and the linear regression estimators are

$$\bar{y}_r = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right) \tag{1.1.1}$$

$$\bar{y}_p = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right) \tag{1.1.2}$$

$$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x}) \tag{1.1.3}$$

Using known information in form of the population mean $\bar{X}$ and coefficient of variation C_x of auxiliary variable x, proposed estimator of population mean $\bar{Y}$ is

$$\bar{y}_k = \bar{y} + \frac{s_{yx}}{\bar{x}^2 C_x^2}(\bar{X} - \bar{x}) + k\left[\bar{y}\frac{s_x^2}{\bar{x}^2 C_x^2} - \left\{\bar{y} + \frac{s_{yx}}{\bar{x}^2 C_x^2}(\bar{X} - \bar{x})\right\}\right] \quad (1.1.4)$$

1.2. BIAS AND MSE OF $\bar{y}_k$

For simplicity, we have assumed that the population size *N* is large enough as compared to the sample size *n*, so that the finite population correction terms may be ignored.

Let,

$$\bar{y} = \bar{Y} + e_0, \bar{x} = \bar{X} + e_1, s_x^2 = S_x^2 + e_2 \text{ and } s_{yx} = S_{yx} + e_3$$

So that,

$$E(e_0) = E(e_1) = E(e_2) = E(e_3) = 0 \quad (1.2.1)$$

$$E(e_0^2) = \frac{S_y^2}{n} = \frac{\mu_{20}}{n}, \quad E(e_1^2) = \frac{S_x^2}{n} = \frac{\mu_{02}}{n},$$

$$E(e_2^2) = \left\{\frac{\beta_2(x)-1}{n}\right\}S_x^4 = \frac{\mu_{02}^2}{n}\{\beta_2(x) - 1\},$$

$$E(e_0e_1) = \frac{S_{yx}}{n} = \frac{\mu_{11}}{n}, \quad E(e_0e_2) = \frac{\mu_{12}}{n}, \quad E(e_1e_2) = \frac{\mu_{03}}{n}, \quad E(e_1e_3) = \frac{\mu_{12}}{n} \quad (1.2.2)$$

From (1.1.4), we have

$$\bar{y}_k = (\bar{Y} + e_0) - e_1\left(\frac{S_{yx}+e_3}{\bar{X}^2 C_x^2}\right) + k\left[\frac{(\bar{Y}+e_0)(S_x^2+e_2)}{\bar{X}^2 C_x^2} - \left\{(\bar{Y} + e_0) - \frac{e_1(S_{yx}+e_3)}{\bar{X}^2 C_x^2}\right\}\right]$$

$$\bar{y}_k - \bar{Y} = e_0 - \frac{S_{yx}\, e_1}{\bar{X}^2 C_x^2} - \frac{e_1\, e_3}{\bar{X}^2 C_x^2} + k\left[\frac{\bar{Y} S_x^2}{\bar{X}^2 C_x^2} + \frac{\bar{Y}\, e_2}{\bar{X}^2 C_x^2} + \frac{S_x^2\, e_0}{\bar{X}^2 C_x^2} + \frac{e_0\, e_2}{\bar{X}^2 C_x^2} - \bar{Y} - e_0 + \frac{S_{yx}\, e_1}{\bar{X}^2 C_x^2} + \frac{e_1\, e_3}{\bar{X}^2 C_x^2}\right]$$

$$= e_0 - \frac{S_{yx}\, e_1}{S_x^2} - \frac{e_1\, e_3}{S_x^2} + k\left[\bar{Y} + \frac{\bar{Y}\, e_2}{S_x^2} + e_0 + \frac{e_0\, e_2}{S_x^2} - \bar{Y} - e_0 + \frac{S_{yx}\, e_1}{S_x^2} + \frac{e_1\, e_3}{S_x^2}\right] \quad (1.2.3)$$

Taking the expectation on both side of *equation* (3.2.3), we obtain bias of $\bar{y}_k$ to the order of $O\left(\frac{1}{n}\right)$ to be

$$Bias(\bar{y}_k) = \frac{1}{n}\left[\frac{\mu_{12}}{\mu_{02}}\{2k-1\}\right] = \frac{1}{n}\left[\frac{\mu_{12}}{S_x^2}\{2k-1\}\right] \quad (1.2.4)$$

Squaring both the sides of (1.2.3), and ignoring the terms involving e_i's having degree higher than two, we have MSE of $\bar{y}_k$ to the order $O\left(\frac{1}{n}\right)$ to be

$$E(\bar{y}_k - \bar{y})^2 = E\left[e_0 - \frac{S_{yx}\,e_1}{S_x^2} - \frac{e_1\,e_3}{S_x^2} + k\left\{\bar{Y} + \frac{\bar{Y}\,e_2}{S_x^2} + e_0 + \frac{e_0\,e_2}{S_x^2} - \bar{Y} - e_0 + \frac{S_{yx}\,e_1}{S_x^2} + \frac{e_1\,e_3}{S_x^2}\right\}\right]^2$$

$$E(\bar{y}_k - \bar{y})^2 = E\left[e_0^2 + \frac{S_{yx}^2\,e_1^2}{S_x^4} - \frac{2S_{yx}\,e_0 e_1}{S_x^2} + k^2\left\{\frac{\bar{Y}^2\,e_2^2}{S_x^4} + \frac{S_{yx}^2\,e_1^2}{S_x^4} + \frac{2\bar{Y}S_{yx}\,e_1\,e_2}{S_x^4}\right\} + 2k\left\{\frac{\bar{Y}\,e_0\,e_2}{S_x^2} + \frac{S_{yx}\,e_0\,e_1}{S_x^2} - \frac{\bar{Y}\,S_{yx}\,e_1\,e_2}{S_x^4} - \frac{S_{yx}^2\,e_1^2}{S_x^4}\right\}\right]$$

$$\text{MSE}(\bar{y}_k) = \frac{1}{n}\left[\mu_{02} - \frac{\mu_{11}^2}{\mu_{02}} + k^2\left\{\bar{Y}^2\,\{\beta_2(x)-1\} + \frac{\mu_{11}^2}{\mu_{02}} + \frac{2\bar{Y}\mu_{11}\mu_{03}}{\mu_{02}^2}\right\} + 2\bar{Y}k\left\{\frac{\mu_{12}}{\mu_{02}} - \frac{\mu_{11}\mu_{03}}{\mu_{02}^2}\right\}\right] \quad (1.2.5)$$

Minimizing $\text{MSE}(\bar{y}_k)$ with respect to k and equate to zero, we get:

$$2k\left\{\bar{Y}^2\,\{\beta_2(x)-1\} + \frac{\mu_{11}^2}{\mu_{02}} + \frac{2\bar{Y}\mu_{11}\mu_{03}}{\mu_{02}^2}\right\} + 2\bar{Y}k\left\{\frac{\mu_{12}}{\mu_{02}} - \frac{\mu_{11}\mu_{03}}{\mu_{02}^2}\right\} = 0$$

which is minimum for,

$$k = \frac{\bar{Y}(\mu_{11}\mu_{03} - \mu_{12}\mu_{02})}{\{\bar{Y}^2\mu_{02}^2(\beta_2(x)-1) + \mu_{11}^2\mu_{02} + 2\bar{Y}\mu_{11}\mu_{03}\}} = k_{opt}(say) \quad (1.2.6)$$

And the minimum square error of $\bar{y}_k$ is:

$$\text{MSE}(\bar{y}_k)_{min} = \frac{1}{n}\left(\mu_{20} - \frac{\mu_{11}^2}{\mu_{02}}\right) - \frac{1}{n}\left[\frac{\bar{Y}^2(\mu_{11}\mu_{03} - \mu_{12}\mu_{02})^2}{\mu_{02}^2\{\bar{Y}^2\mu_{02}^2(\beta_2(x)-1) + \mu_{11}^2\mu_{02} + 2\bar{Y}\mu_{11}\mu_{03}\}}\right] \quad (1.2.7)$$

1.3. Estimator Based on the Estimated Optimum $\hat{k}$

Generally, the guessed value of k may not be known, hence the alternative is to replace it by its estimate based on sample observations. Replacing $\bar{Y}, \mu_{11}, \mu_{03}, \mu_{12}$ and μ_{04} involved in optimum k by their unbiased or consistent estimators given by $\bar{y}, \hat{\mu}_{11}, \hat{\mu}_{03}, \hat{\mu}_{12}$ and $\hat{\mu}_{04}$ respectively, we get the estimated optimum $\hat{k}$ as:

$$\hat{k} = \frac{\bar{y}(\hat{\mu}_{11}\,\hat{\mu}_{03} - \hat{\mu}_{12}\,\mu_{02})}{\left\{\bar{y}^2 \mu_{02}{}^2\left(\frac{\hat{\mu}_{04}}{\mu_{02}{}^2} - 1\right) + \hat{\mu}_{11}{}^2 \mu_{02} + 2\bar{y}\hat{\mu}_{11}\hat{\mu}_{03}\right\}} \tag{1.3.1}$$

and so the resulting estimator of the population mean $\bar{Y}$ based on estimated optimum $\hat{k}$ is

$$\bar{y}_{k_{opt}} = \bar{y} + \frac{s_{yx}}{\bar{X}^2\, C_x{}^2}(\bar{X} - \bar{x}) + \hat{k}\left[\frac{\bar{y}\, s_x{}^2}{\bar{X}^2\, C_x{}^2} - \left\{\bar{y} + \frac{s_{yx}(\bar{X} - \bar{x})}{\bar{X}^2\, C_x{}^2}\right\}\right] \tag{1.3.2}$$

Let $\hat{\mu}_{03} = \mu_{03} + e_4$, $\hat{\mu}_{12} = \mu_{12} + e_5$ and $\hat{\mu}_{04} = \mu_{04} + e_6$, so that from (1.3.1), we get:

$$\hat{k} = \frac{(\bar{Y}+e_0)\{(\mu_{11}+e_3)(\mu_{03}+e_4) - (\mu_{12}+e_5)\mu_{02}\}}{\left[(\bar{Y}+e_0)^2 \mu_{02}{}^2\left\{\frac{(\mu_{04}+e_6)}{\mu_{02}{}^2} - 1\right\} + (\mu_{11}+e_3)^2 \mu_{02} + 2(\bar{Y}+e_0)(\mu_{11}+e_3)(\mu_{03}+e_4)\right]}$$

$$= \frac{(\bar{Y}+e_0)(\mu_{11}\ \mu_{03} + \mu_{11}e_4 + \mu_{03}e_3 - \mu_{12}\mu_{02} - \mu_{02}e_5)}{\left\{\begin{array}{c}(\bar{Y}^2 + e_0{}^2 + 2\bar{Y}e_0)(\mu_{04} + e_6 - \mu_{02}{}^2) + (\mu_{11}{}^2 + e_3{}^2 + 2\mu_{11}e_3)\mu_{02} \\ +2(\bar{Y}+e_0)(\mu_{11}\mu_{03} + \mu_{11}e_4 + \mu_{03}e_3 + e_3e_4)\end{array}\right\}}$$

$$= \frac{\begin{array}{c}\bar{Y}(\mu_{11}\mu_{03} + \mu_{11}e_4 + \mu_{03}e_3 - \mu_{12}\mu_{02} - \mu_{02}e_5) + (\mu_{11}\mu_{03}e_0 + \mu_{11}e_4e_0 + \mu_{03}e_3e_0 \\ e_0e_3e_4 - \mu_{12}\mu_{02}e_0 - \mu_{02}e_0e_5)\end{array}}{\left\{\begin{array}{c}\bar{Y}^2(\mu_{04} - \mu_{02}{}^2 + e_6) + \mu_{04}e_0{}^2 + e_0{}^2e_6 - \mu_{02}{}^2e_0{}^2 + 2\bar{Y}\mu_{04}e_0 + 2\bar{Y}e_0e_6 - 2\bar{Y}\mu_{02}{}^2e_0 \\ +\mu_{11}{}^2\mu_{02} + \mu_{02}e_3{}^2 + 2\mu_{11}e_3\mu_{02} + 2\bar{Y}(\mu_{11}\mu_{03} + \mu_{11}e_4 + \mu_{03}e_3 + e_3e_4) \\ +2(\mu_{11}\mu_{03}e_0 + \mu_{11}e_4e_0 + \mu_{03}e_3e_0 + e_0e_3e_4)\end{array}\right\}}$$

$$= \frac{\left\{ \begin{matrix} \bar{Y}(\mu_{11}\mu_{03} - \mu_{12}\mu_{02}) + \bar{Y}(\mu_{11}e_4 + \mu_{03}e_3 - \mu_{02}e_5) + e_0(\mu_{11}\mu_{03} - \mu_{12}\mu_{02}) \\ +\bar{Y}e_3e_4 + \mu_{11}e_4e_0 + \mu_{03}e_3e_0 - \mu_{02}e_0e_5 + e_0e_3e_4 \end{matrix} \right\}}{\left\{ \begin{matrix} \bar{Y}^2(\mu_{04} - {\mu_{02}}^2) + {\mu_{11}}^2\mu_{02} + 2\bar{Y}\mu_{11}\mu_{03} + \bar{Y}^2e_6 + 2\bar{Y}^2e_0(\mu_{04} - {\mu_{02}}^2) + 2\mu_{11}e_3\mu_{02} + 2\bar{Y}\mu_{11}e_4 \\ +2\bar{Y}\mu_{03}e_3 + 2\mu_{11}e_0\mu_{03} + {e_0}^2(\mu_{04} - {\mu_{02}}^2) + \mu_{02}{e_3}^2 + 2\bar{Y}e_0e_6 + 2\bar{Y}e_3e_4 + 2\mu_{11}e_0e_4 \\ +2\mu_{03}e_0e_3 + {e_0}^2e_6 + 2e_0e_3e_4 \end{matrix} \right\}}$$

$$= \frac{\bar{Y}(\mu_{11}\mu_{03} - \mu_{12}\mu_{02}) \left[1 + \frac{\begin{matrix} \bar{Y}(\mu_{11}e_4 + \mu_{03}e_3 - \mu_{02}e_5) + e_0(\mu_{11}\mu_{03} - \mu_{12}\mu_{02}) \\ +\bar{Y}e_3e_4 + \mu_{11}e_4e_0 + \mu_{03}e_3e_0 - \mu_{02}e_0e_5 + e_0e_3e_4 \end{matrix}}{\bar{Y}(\mu_{11}\mu_{03} - \mu_{12}\mu_{02})} \right]}{\left\{ \begin{matrix} \bar{Y}^2(\mu_{04} - {\mu_{02}}^2) \\ +{\mu_{11}}^2\mu_{02} + 2\bar{Y}\mu_{11}\mu_{03} \end{matrix} \right\} \left[1 + \frac{\left\{ \begin{matrix} \bar{Y}^2e_6 + 2\bar{Y}^2e_0(\mu_{04} - {\mu_{02}}^2) + 2\mu_{11}e_3\mu_{02} + 2\bar{Y}\mu_{11}e_4 + 2\bar{Y}\mu_{03}e_3 \\ +2\mu_{11}e_0\mu_{03} + {e_0}^2(\mu_{04} - {\mu_{02}}^2) + \mu_{02}{e_3}^2 + 2\bar{Y}e_0e_6 \\ +2\bar{Y}e_3e_4 + 2\mu_{11}e_0e_4 + 2\mu_{03}e_0e_3 + {e_0}^2e_6 + 2e_0e_3e_4 \end{matrix} \right\}}{\{\bar{Y}^2(\mu_{04} - {\mu_{02}}^2) + {\mu_{11}}^2\mu_{02} + 2\bar{Y}\mu_{11}\mu_{03}\}} \right]} \quad (1.3.3)$$

$$= \frac{\bar{Y}(\mu_{11}\mu_{03} - \mu_{12}\mu_{02})}{\{\bar{Y}^2(\mu_{04} - {\mu_{02}}^2) + {\mu_{11}}^2\mu_{02} + 2\bar{Y}\mu_{11}\mu_{03}\}}(1 + \Delta_1)(1 + \Delta_2)^{-1} \quad (1.3.4)$$

Where,

$$\Delta_1 = \frac{\bar{Y}(\mu_{11}e_4 + \mu_{03}e_3 - \mu_{02}e_5) + e_0(\mu_{11}\mu_{03} - \mu_{12}\mu_{02}) + \bar{Y}e_3e_4 + \mu_{11}e_4e_0 + \mu_{03}e_3e_0 - \mu_{02}e_0e_5 + e_0e_3e_4}{\bar{Y}(\mu_{11}\mu_{03} - \mu_{12}\mu_{02})}$$

and

$$\Delta_2 = \frac{\left\{ \begin{matrix} \bar{Y}^2e_6 + 2\bar{Y}^2e_0(\mu_{04} - {\mu_{02}}^2) + 2\mu_{11}e_3\mu_{02} + 2\bar{Y}\mu_{11}e_4 + 2\bar{Y}\mu_{03}e_3 + 2\mu_{11}e_0\mu_{03} + {e_0}^2(\mu_{04} - {\mu_{02}}^2) + \\ \mu_{02}{e_3}^2 + 2\bar{Y}e_0e_6 + 2\bar{Y}e_3e_4 + 2\mu_{11}e_0e_4 + 2\mu_{03}e_0e_3 + {e_0}^2e_6 + 2e_0e_3e_4 \end{matrix} \right\}}{\{\bar{Y}^2(\mu_{04} - {\mu_{02}}^2) + {\mu_{11}}^2\mu_{02} + 2\bar{Y}\mu_{11}\mu_{03}\}} \quad (1.3.5)$$

From (1.3.2), we have

$$\left(\bar{y}_{k_{opt}} - \bar{Y}\right) = e_0 - \frac{S_{yx}e_1}{{S_x}^2} - \frac{e_1e_3}{{S_x}^2} + \hat{k}\left[\frac{\bar{Y}e_2}{{S_x}^2} + \frac{e_0e_2}{{S_x}^2} + \frac{S_{yx}}{{S_x}^2}e_1 + \frac{e_1e_3}{{S_x}^2}\right] \quad (1.3.6)$$

Substituting $\hat{k}$ from (1.3.4) in (1.3.3), squaring both the sides and taking expectation, taking terms of $e_i{}'s$ upto the second degree, we obtain

$$E\left(\bar{y}_{k_{opt}}-\bar{Y}\right)^2 = E\left[e_0 - \frac{S_{yx}e_1}{S_x{}^2} - \frac{e_1 e_3}{S_x{}^2} + \left\{\frac{\bar{Y}(\mu_{11}\mu_{03}-\mu_{12}\mu_{02})(1+\Delta_1)(1+\Delta_2)^{-1}}{\bar{Y}^2\mu_{02}{}^2\{\beta_2(x)-1\}+\mu_{11}{}^2\mu_{02}+2\bar{Y}\mu_{11}\mu_{03}}\right\}\left\{\frac{\bar{Y}e_2}{S_x{}^2}+\frac{e_0 e_2}{S_x{}^2}+\frac{S_{yx}}{S_x{}^2}e_1+\frac{e_1 e_3}{S_x{}^2}\right\}\right]^2$$

Keeping the terms of $e_i{}'s$ upto the second degree, we get the MSE of $\bar{y}_{k_{opt}}$ as

$$MSE\left(\bar{y}_{k_{opt}}\right) = \frac{1}{n}\left(\mu_{20}-\frac{\mu_{11}{}^2}{\mu_{02}}\right) - \frac{1}{n}\left[\frac{\bar{Y}^2(\mu_{11}\mu_{03}-\mu_{12}\mu_{02})^2}{\mu_{02}{}^2\{\bar{Y}^2\mu_{02}\{\beta_2(x)-1\}+\mu_{11}{}^2\mu_{02}+2\bar{Y}\mu_{11}\mu_{03}\}}\right]$$

$$= MSE(\bar{y}_{lr}) - \frac{1}{n}\left[\frac{\bar{Y}^2(\mu_{11}\mu_{03}-\mu_{12}\mu_{02})^2}{\mu_{02}{}^2\{\bar{Y}^2\mu_{02}\{\beta_2(x)-1\}+\mu_{11}{}^2\mu_{02}+2\bar{Y}\mu_{11}\mu_{03}\}}\right] \quad (1.3.7)$$

which shows that MSE of estimator $\bar{y}_{k_{opt}}$ in (1.3.7) based on estimated optimum values, to the terms of order $O\left(\frac{1}{n}\right)$ is the same as that of minimum MSE of the proposed estimator $\bar{y}_k$ in (1.2.7), i.e.,

$$MSE(\bar{y}_k)_{min} = MSE(\bar{y}_{k_{opt}})$$

1.4. Result

The conventional linear regression estimator $\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x})$ of finite population mean $\bar{Y}$.

has its MSE

$$MSE(\bar{y}_{lr}) = \frac{1}{n}\left(\mu_{20} - \frac{\mu_{11}{}^2}{\mu_{02}}\right)$$

The proposed estimators $\bar{y}_k$ and $\bar{y}_{k_{opt}}$have their MSE given by (1.3.7) as:

$$MSE(\bar{y}_k)_{min} = MSE\left(\bar{y}_{k_{opt}}\right) = MSE(\bar{y}_{lr}) - \frac{1}{n}\left[\frac{\bar{Y}^2(\mu_{11}\mu_{03}-\mu_{12}\mu_{02})^2}{\mu_{02}{}^2\{\bar{Y}^2\mu_{02}\{\beta_2(x)-1\}+\mu_{11}{}^2\mu_{02}+2\bar{Y}\mu_{11}\mu_{03}\}}\right]$$

It may be noted here that,

$(\mu_{11}\mu_{03} - \mu_{12}\mu_{02})^2 > 0$ and $(\mu_{04} - \mu_{02}{}^2) = \mu_{02}{}^2\{\beta_2(x) - 1\} > 0$

So that for positively skewed distribution

$$MSE\left(\bar{y}_{k_{opt}}\right) < MSE(\bar{y}_{lr})$$

Which shows that proposed estimators $\bar{y}_k$ and $\bar{y}_{k_{opt}}$ are more efficient than the convectional linear regression estimator $\bar{y}_{lr}$.

NUMERICAL ILLUSTRATION

To evaluate the efficiency of the proposed estimator relative to the conventional linear regression estimator, a simulation study was conducted under a positively skewed population distribution. The goal was to compute the Mean Square Error using the theoretical expressions and assess the Percent Relative Efficiency (PRE). Here we have considered:

- **Population Size (N)** = 1000
- **Sample Size (n)** = 100
- **Number of Simulations** = 5
- **Distribution** = The study variable y and auxiliary variable x were generated from a **bivariate log-normal distribution**, inducing a strong **positive skew** and correlation of approximately 0.85.

For each simulation, the population moments and the sample estimates have been calculated and then *MSE* of the Linear Regression Estimator and Proposed Estimator is calculated. The calculations give the following results:

Table 1: Population Moments

Simulation	$\bar{Y}$	$\bar{X}$	μ_{20}	μ_{11}	μ_{03}	μ_{12}	$\beta_2(x)$	ρ
1	1.105	1.087	1.217	0.983	2.841	1.912	3.892	0.847
2	1.098	1.082	1.208	0.974	2.815	1.894	3.876	0.843
3	1.102	1.085	1.213	0.980	2.832	1.905	3.885	0.845
4	1.108	1.091	1.223	0.989	2.856	1.925	3.901	0.849
5	1.100	1.083	1.210	0.977	2.823	1.899	3.880	0.844

Table 2: Sample Estimates

Simulation	$\bar{y}$	$\bar{x}$	$s_y^{\ 2}$	$s_x^{\ 2}$	s_{yx}	b	$\bar{y}_{lr}$	$\bar{y}_k$
1	1.092	1.075	0.012	0.011	0.009	0.818	1.099	1.097
2	1.086	1.070	0.012	0.011	0.009	0.821	1.093	1.091
3	1.090	1.073	0.012	0.011	0.009	0.819	1.097	1.095
4	1.096	1.079	0.013	0.011	0.010	0.815	1.103	1.101
5	1.088	1.071	0.012	0.011	0.009	0.820	1.095	1.093

Table 3: Efficiency Comparison

Simulation	$MSE(\bar{y}_{lr})$	$MSE(\bar{y}_{k})$	PRE (%)
1	0.0087	0.0039	223.1
2	0.0083	0.0037	224.3
3	0.0085	0.0038	223.7
4	0.0090	0.0040	225.0
5	0.0084	0.0037	227.0
Mean	**0.0086**	**0.0038**	**224.6**

1.6. CONCLUSION

1. The proposed estimator consistently yields lower MSE than the conventional linear regression estimator.
2. The PRE values across simulations lie between **223.1% and 227%**, with an average of **224.6%**.
3. This confirms that the proposed estimator is significantly more efficient than the conventional linear regression estimator when the population is positively skewed.
4. This simulation validates the theoretical result and supports the improved efficiency of the proposed difference and regression type estimator in practical sampling scenarios.

References

Cochran, W. G. (1977) – 'Sampling Techniques', 3rd Edition, John Wiley and Sons, New York.

Das, A.K., & Tripathi, T.P. (1980) – Sampling strategies for population mean when the coefficient of variation of an auxiliary character is known; Sankhya, C, 42, 76-86.

Dhansawad, N. (2022). Efficient Modified Estimators for the Population Mean Using Auxiliary Variables, Naresuan University Journal: Science and Technology, Bangkok, Thailand 30(4): 84-93.

Gupta and Singh (1998) – Estimation of population mean using coefficient of variation and auxiliary information, REBRAPE (Brazilian Journal of Probability and Statistics), 12, 35-40.

Kataria, P., & Singh, S. (1989) – On the estimation of mean when population variance is known; Jour. Ind. Soc. Agri. Statist., Vol. XLI, No. 2, 173-175.

Khan, R.A. (1968) – A note on estimation of the mean of a normal distribution with known coefficient of variation. J. Amer. Statist. Assoc. 63, 1039-41.

M. S. Ahmed (2004) – Some estimators for a finite population mean under two-stage sampling using multivariate auxiliary information, Applied Mathematics and Computation, 153, 505-511.

Rizvi, S.A.H. and Singh R. Karan (1997). A Generalized class of Estimators of Finite Population Mean Using Known Coefficient of Variation, Proc. Nat. Acad. Sci. India, 67 (A), IV, pp. 395-404, ISSN No. 0369-8203, India.

Rizvi, S.A.H. and Singh R. Karan (1997). On Estimation of Population Mean Using Known Coefficient of Variation, Microelectron. Reliab., Vol. 37, No. 5, pp. 841-843, 1997, Elsevier Science Ltd., ISSN No. 0026-2714, Great Britain.

Singh H.P. and Vishwakarma G.K. (2008). Some families of estimators of variance of stratified random sample mean using auxiliary information. *Journal of Statistical Theory and Practice,* vol. 2 , No. 1, pp. 21-43.