Journal of Statistical Research *https://doi.org/10.3329/jsr.v59i1.83686*
2025, Vol. 59, No. 1, pp. 81-98 *ISSN 0256 - 422 X*

INVESTIGATING THE ROLE OF DUAL USE OF AN AUXILIARY VARIABLE: A DIFFERENCE-CUM-EXPONENTIAL ESTIMATOR

SARHAD ULLAH KHAN
National College of Business Administration & Economics, Lahore, Pakistan
Email: 2213128@ncbae.edu.pk

MUHAMMAD HANIF*
National College of Business Administration & Economics, Lahore, Pakistan
Email: drhanif@ncbae.edu.pk

ZAHOOR AHMAD
University of Sargodha, Sargodha, Pakistan
Email: zahoor.ahmad@uos.edu.pk

KALIM ULLAH
Foundation University Medical College, Foundation University, Islamabad, Pakistan
Email: kalimullah@stat.qau.edu.pk

SUMMARY

The estimation of finite population mean is always of interest for different sampling techniques and it is the basic measure to find from sample to estimate one the most applicable central tendency. In literature, under simple random sampling without replacement people used auxiliary variable, its rank or empirical distribution function in different estimation approaches such as regression, ratio, exponential or combination of these to improve the efficiency of the estimator. In literature, either rank or empirical distribution function have been used while constructing the estimator because both cannot be used due to the fact that empirical distribution function of a variable is based on its rank, therefore, both are perfectly correlated. In this paper, our argument is that the dual use is not effective rather an additional independent auxiliary variable may be effective for efficiency improvement. To investigate this, we proposed difference-cum-exponential estimator using two auxiliary variables and also the dual use of one of the auxiliary variable in the form of its empirical distribution function. We also deduced some special cases of the proposed estimator. These special cases will help us to investigate the argument. The mean square errors of the proposed estimator and its special cases are derived. The proposed estimator, its special cases and potential existing estimators are compared using empirical study based on real life population for numerical investigation of the argument. The simulation study is also conducted for symmetric and skewed populations to asses the sampling stability of the competitive estimators using empirical mean square error and also it will help to further investigate the argument.

Keywords and phrases: finite population mean, difference-cum-exponential estimator, empirical distribution function, auxiliary variable.

---

* Corresponding author

1 Introduction

Sampling theory is a fundamental aspect of statistical research that enables researchers to draw inferences about population parameters from data collected from a finite population. The incorporation of auxiliary variables have been widely recognized as a valuable technique to improve the precision of estimators. Traditional auxiliary variables are typically used to reduce sampling errors and enhance the efficiency of estimators. However, recent research has explored alternative approaches, such as the utilization of rank or the empirical distribution function as a dual auxiliary variable, to further enhance the estimation process.

The empirical distribution function (EDF) is a non-parametric estimator of the cumulative distribution function, based on the observed values of the auxiliary variable in the sample. It has been applied in various statistical methods, such as hypothesis testing, goodness-of-fit tests and bootstrap procedures. The empirical distribution function is particularly useful when the underlying distribution of the auxiliary variable is unknown or difficult to model accurately. Several related papers have been investigated those used the empirical distribution function as a dual auxiliary variable in the context of survey sampling and finite population studies. Mak and Kuk (1993) proposed estimators for the population parameter using auxiliary variable and demonstrating its effectiveness in reducing bias and improving efficiency. Pandey et al. (2021) extended the application of auxiliary variable to address the issue of non-response in survey sampling, it has potential in handling missing data and improving estimations in the presence of non-response. Zaman and Kadilar (2021) introduced regression-type estimators for the finite population mean, leveraging the information embedded in empirical distribution function to enhance estimation precision. Singh and Solanki (2013) were among the early researchers to explore the utilization of the rank as a dual auxiliary variable in ratio type estimators. They proposed a novel rank-based ratio estimator for estimating the finite population mean. The rank-based ratio estimator is more precised in comparison to traditional ratio estimators using conventional auxiliary variables. Kadilar and Cingi (2006) explored the dual use of the rank of the auxiliary variable and proposed a hybrid estimator that combined ratio and regression estimators. Moreover, Haq et al. (2017) introduced a rank-based calibration estimator, leveraging the rank of the auxiliary variable to adjust estimations for finite population mean. Hussain et al. (2022) proposed a difference-cum-exponential estimator using empirical distribution function, which achieved more efficiency compared to conventional estimators.

The above researches claim that the dual use of auxiliary variables in improving the efficiency, but we argue otherwise. Our argument is based on the fact that the dual use of auxiliary variables may be ineffective due to redundancy, increased dependency, and added complexity. Highly correlated auxiliary variables often provide overlapping information, limiting efficiency gains. In contrast, a well-chosen independent auxiliary variable may offer greater efficiency without unnecessary complexity. Therefore, we proposed difference-cum-exponential estimator that utilizes two auxiliary variables X and Z and the dual use of X only in the form of its empirical distribution function. We will deduce the special cases from this estimator (i) using on X, (ii) using X with its empirical distribution function and (iii) using X and Z. The pairwise comparison between these estimators and with existing estimators will help us to investigate our argument properly.

After preliminary section, in section two we provide latest simple and shrinkage estimators available in literature along with their mean square errors. In section three, we proposed shrinkage difference-cum-exponential type estimator using empirical distribution function as dual of auxiliary variable and derived its minimum mean square error (MSE) by minimizing the optimizing constants. The special cases of the estimator are also discussed in this section. Simulation study in given in section 5, application of the proposed estimator to real-life data sets is in section six, and finally section seven concludes the work and discusses potential future research directions.

2 Preliminaries

In a sample survey we decide on certain properties that we attempt to measure and record for every unit that comes into the sample. These properties of the units are referred to as characteristics or, more simply, as items. The values obtained for any specific item y in the N units that comprise the population are denoted by $y_1,\, y_2,\, \dots,\, y_N$. The corresponding values for the units in the sample are denoted by $y_1,\, y_2,\, \dots,\, y_n$. Suppose the characteristic y is study variable and x and z are two auxiliary variables and it is assumed that both auxiliary variables have high correlation with study variable. Let the values of x for N units in the population are $x_1,\, x_2, \dots, x_N$ and the corresponding values for the units in the sample are denoted by $x_1,\, x_2, \dots, x_n$. Moreover, let $f_{x_1}, f_{x_2}, f_{x_3}, \dots, f_{x_N}$ denote the values of empirical distribution function of auxiliary variable Similarly the values of x for N units in the population are $z_1,\, z_2, \dots, z_N$ and the corresponding values for the units in the sample are denoted by $z_1,\, z_2,\, \dots,\, z_n$. The population and sample characteristics for study and auxiliary variables are listed in Table A1 and given in Appendix .

In order to derive MSE of estimators, we define the following errors terms. Let

$$\epsilon_0 = \bar{y} - \bar{Y},\ \epsilon_1 = \bar{x} - \bar{X},\ \epsilon_3 = \bar{f}_x - \bar{F}_x \text{ and } \epsilon_2 = \bar{z} - \bar{Z}. \tag{2.1}$$

Suppose $\bar{Y}$, $\bar{X}$ and $\bar{Z}$ are population mean of Y, X and Z respectively, C_y, C_x and C_z are population coefficient of variations of Y, X and Z respectively, ρ_{yx}, ρ_{yz} and ρ_{xz} are population coefficient of correlations Y and X, Y and Z, and X and Z respectively, and $\tau = (N-n)/Nn$ is finite population correction factor.

Then under simple random sampling without replacement, we have

$$\begin{cases} E(\epsilon_0) = E(\epsilon_1) = E(\epsilon_2) = E(\epsilon_3) = 0, \\ E(\epsilon_0^2) = \tau \bar{Y}^2 C_y^2,\ E(\epsilon_1^2) = \tau \bar{X}^2 C_x^2,\ E(\epsilon_3^2) = \tau \bar{F}_x^2 C_{F_x}^2,\ E(\epsilon_2^2) = \tau \bar{Z}^2 C_z^2 \\ E(\epsilon_0 \epsilon_1) = \tau \bar{Y} \bar{X} C_y C_x \rho_{yx},\ E(\epsilon_0 \epsilon_3) = \tau \bar{Y} \bar{F}_x C_y C_{F_x} \rho_{yF_x}, \\ E(\epsilon_0 \epsilon_2) = \tau \bar{Y} \bar{Z} C_y C_z \rho_{yz},\ E(\epsilon_1 \epsilon_3) = \tau \bar{X} \bar{F}_x C_x C_{F_x} \rho_{xF_x}, \\ E(\epsilon_1 \epsilon_2) = \tau \bar{X} \bar{Z} C_x C_z \rho_{xz}, E(\epsilon_2 \epsilon_3) = \tau \bar{F}_x \bar{Z} C_{F_x} C_z \rho_{F_x z}. \end{cases} \tag{2.2}$$

3 Some Known Estimators

In this section, we have provided the some available estimators already used for estimating the finite population mean. The variance and MSEs of these estimators are given up-to first order approximation.

The mean per unit estimator based on simple random sampling without replacement is

$$t_1 = \bar{y} = \sum_{i=1}^{n} y_i/n \text{ with } MSE(t_1) = \tau S_y^2 = \tau \bar{Y}^2 C_y^2. \tag{3.1}$$

Cochran (1953) suggested the following regression estimator for population mean $\bar{Y}$ using an auxiliary variable X

$$t_2 = \bar{y} + \beta_{yx}(\bar{X} - \bar{x}) \text{ with } MSE(t_2) = \tau \bar{Y}^2 C_y^2 (1 - \rho_{yx}^2). \tag{3.2}$$

The difference estimator, suggested by Hansen et al. (1942) and independently Tripathi (1970) is

$$t_3 = \bar{y} - d(\bar{x} - \bar{X}), \tag{3.3}$$

where d is optimizing constant and its optimum value which minimizes MSE is β_{yx}. The minimum MSE of t_3 is same as for t_2.

Searls (1964) presented the following modified version of mean per unit estimator and named as shrinkage estimator:

$$t_4 = k\bar{y} \tag{3.4}$$

where k is a constant, which is determined by minimizing mean square error of t_4. The optimum value of k is $(1 + \tau C_y^2)^{-1}$ and then the minimum MSE of t_4 is

$$MSE(t_4) = \frac{\bar{Y}^2 MSE(t_1)}{\bar{Y}^2 + MSE(t_1)}.$$

Riaz et al. (2014) proposed the following estimator

$$t_5 = d_1\bar{y} + d_2(\bar{X} - \bar{x}), \tag{3.5}$$

where d_1 and d_2 are optimizing constants and the optimum values of d_1 and d_2 which minimizes the MSE are $d_1 = \bar{Y}^2(\bar{Y}^2 + MSE(t_3))^{-1}$ and $d_2 = \beta_{yx} d_1$ respectively. Then the minimum MSE is

$$MSE(t_5) = \frac{\bar{Y}^2 MSE(t_3)}{\bar{Y}^2 + MSE(t_3)}.$$

Yaqoob, et al. (2017) proposed the following estimator using rank of auxiliary variable along with auxiliary variable itself and called it dual use of auxiliary variable

$$t_6 = \bar{y} + \pi_1(\bar{X} - \bar{x}) + \pi_2(\bar{R}_x - \bar{r}_x), \tag{3.6}$$

where π_1 and π_2 are optimizing constants and the optimum values of π_1 and π_2 are

$$\pi_1 = \frac{\bar{Y}}{\bar{X}} \times \frac{\beta_{yx} - \beta_{yR_x}\beta_{xR_x}}{1-\rho^2_{xR_x}}, \quad \pi_2 = \frac{\bar{Y}}{\bar{R}_x} \times \frac{\beta_{yR_x} - \beta_{yR_x}\beta_{xR_x}}{1-\rho^2_{xR_x}}.$$

or we may write as

$$\pi_1 = (-1)^{1+1}\frac{\bar{Y}}{\bar{X}}\frac{C_y}{C_x}\frac{|\mathbf{R}_{yx}|_{yxR_x}}{|\mathbf{R}|_{xR_x}}, \; \pi_2 = (-1)^{2+1}\frac{\bar{Y}}{\bar{R}_x}\frac{C_y}{C_{R_x}}\frac{|\mathbf{R}_{yR_x}|_{yxR_x}}{|\mathbf{R}|_{xR_x}},$$

where

$$|\mathbf{R}_{xR_x}| = \begin{vmatrix} 1 & \rho_{xR_x} \\ \rho_{xR_x} & 1 \end{vmatrix}, \; |\mathbf{R}_{yx}|_{yxR_x} = \begin{vmatrix} \rho_{yx} & \rho_{xR_x} \\ \rho_{yR_x} & 1 \end{vmatrix}, |\mathbf{R}_{yR_x}|_{yxR_x} = \begin{vmatrix} \rho_{yx} & 1 \\ \rho_{yR_x} & \rho_{xR_x} \end{vmatrix}.$$

Then the minimum MSE of t_6 is

$$MSE(t_6) = \tau\bar{Y}^2 C_y^2(1-\rho^2_{y.xR_x}), \text{ where } \rho^2_{y.xr_x} = \frac{\rho^2_{yx} + \rho^2_{yR_x} - 2\rho_{yx}\rho_{yR_x}\rho_{xR_x}}{1-\rho^2_{xR_x}}.$$

Hussain et al. (2022) also proposed the following class of regression-cum-exponential estimator using empirical distribution function as a dual use of auxiliary variable as:

$$t_7 = \left(\lambda_1\hat{Y} + \lambda_2(\bar{X} - \hat{\bar{X}}) + \lambda_3(\bar{F} - \hat{\bar{F}})\right)\exp\left(\frac{a(\bar{X} - \hat{\bar{X}})}{a(\bar{X} + \hat{\bar{X}}) + 2b}\right), \tag{3.7}$$

where λ_1, λ_2, and λ_3 are the optimizing constants which minimize the MSE of the proposed class. Moreover, a and b are generalising constants to produced various members of the suggested class. Hussain et al. (2022) produced ten members of the class (named here $t_{71}, t_{72}, \ldots, t_{80}$) for different values of a and b, e.g. population correlation coefficient, moments ratio, coefficient of variation, population total or any suitable numeric value. The optimum values λ_1, λ_2, and λ_3 and expression of minimum MSE for (3.7) given by Hussain et al. (2022) are complex in writing, therefore, following Ahmad(2008), we reproduced their expressions as:

$$\begin{aligned}
\lambda_1 &= \left[1 + \tau C_y^2 \frac{|\mathbf{R}|_{yxF_x}}{|\mathbf{R}|_{xF_x}}\right]^{-1}, \\
\lambda_2 &= \lambda_1\left[(-1)^{1+1}\frac{\bar{Y}}{\bar{X}}\frac{C_y}{C_x}\frac{|\mathbf{R}_{yx}|_{yxF_x}}{|\mathbf{R}|_{xF_x}} - K\right], \\
\lambda_3 &= \lambda_1\left[(-1)^{2+1}\frac{\bar{Y}}{\bar{F}_x}\frac{C_y}{C_{F_x}}\frac{|\mathbf{R}_{yF_x}|_{yxF_x}}{|\mathbf{R}|_{xF_x}}\right], \text{and} \\
MSE(t_7) &= \frac{\tau\bar{Y}^2 C_y^2\left(1-\rho^2_{y.xF_x}\right)}{1+\tau C_y^2\left(1-\rho^2_{y.xF_x}\right)} = \left[\tau\bar{Y}^2 C_y^2\frac{|\mathbf{R}|_{yxF_x}}{|\mathbf{R}|_{xF_x}}\right]\left[1+\tau C_y^2\frac{|\mathbf{R}|_{yxF_x}}{|\mathbf{R}|_{xF_x}}\right]^{-1},
\end{aligned} \tag{3.8}$$

where $K = a/[2(a\bar{X} + b)]$. We can see that for any value of a and b the theoretical MSE is identical and it is evident from the empirical study given by Hussain et al. (2022) that MSEs of all members are same for a given population. The utility of these members is based on the availability of population parameters used while producing the members. Therefore, there is need to conduct simulation study to find empirical standard error to know the sampling stability of all members. In simulation study the performance will vary and one can know the best member of the class and then a specific member can be suggested for practical use. We have included all members in our simulation study.

4 Proposed Difference-cum-Exponential Type Estimator

In order to develop an estimator for study variable Y, we propose the following difference-cum-exponential type estimator using two auxiliary variables X and Z and dual use of auxiliary variable X in the form of empirical distribution function F_x. Later on we will define special cases of this estimator with and without F_x to investigate the role of dual use of auxiliary variable in improving the efficiency of estimator and similarly with and with out Z for the role of additional auxiliary variable

$$t_{xzF} = \left[\gamma_1 \ \bar{y} + \pi_1(\bar{X} - \bar{x}) + \pi_2(\bar{Z} - \bar{z})\right] \exp\left(\pi_3 \frac{\bar{F}_x - \bar{f}_x}{\bar{F}_x + \bar{f}_x}\right), \tag{4.1}$$

where γ_1, π_1, π_2 and π_3 are optimizing constants to be obtained by minimizing MSE.

We can write (4.1) as

$$t_{xzF} = \left[\gamma_1 \left\{\bar{y} + \pi_1^*(\bar{X} - \bar{x}) + \pi_2^*(\bar{Z} - \bar{z})\right\}\right] \exp\left(\pi_3 \frac{\bar{F}_x - \bar{f}_x}{\bar{F}_x + \bar{f}_x}\right),$$

where $\pi_1^* = \pi_1/\gamma_1$ and $\pi_2^* = \pi_2/\gamma_1$.

Let

$$\hat{\bar{Y}}^* = \left[\bar{y} + \pi_1^*(\bar{X} - \bar{x}) + \pi_2^*(\bar{Z} - \bar{z})\right] exp\left(\pi_3 \frac{\bar{F}_x - \bar{f}_x}{\bar{F}_x + \bar{f}_x}\right). \tag{4.2}$$

Therefore,

$$\hat{\bar{Y}}_{RcE} = \gamma_1 \hat{\bar{Y}}^* \tag{4.3}$$

In order to find MSE of $\hat{\bar{Y}}_{Reg}$, first we need to find $MSE(\hat{\bar{Y}}^*)$. Therefore, considering (4.2) and using (2.1), we can write

$$\hat{\bar{Y}}^* = \left[\bar{Y} + \epsilon_0 - \pi_1^*\epsilon_1 - \pi_2^*\epsilon_2\right](1 - \pi_3 \frac{1}{2\bar{F}_x}\epsilon_3),$$

As the $\hat{\bar{Y}}^*$ is biased, hence to the first order of approximation, the MSE can be written as:

$$MSE(\hat{\bar{Y}}^*) = E(\hat{\bar{Y}}^* - \bar{Y})^2 = E(\epsilon_0 - \pi_1^*\epsilon_1 - \pi_2^*\epsilon_2 - \pi_3 \frac{\bar{Y}}{2\bar{F}_x}\epsilon_3)^2. \tag{4.4}$$

For optimum values of π_1^*, π_2^* and π_3, we need to partially differentiate (4.4) with respect to π_1^*, π_2^* and π_3 then we will have the following three normal equations.

$$E(\epsilon_0\epsilon_1) - \pi_1^* E(\epsilon_1^2 - \pi_2^* E(\epsilon_1\epsilon_2) - \pi_3 \frac{\bar{Y}}{2\bar{F}_x} E(\epsilon_1\epsilon_3) = 0 \tag{4.5}$$

$$E(\epsilon_0\epsilon_2) - \pi_1^* E(\epsilon_1\epsilon_2) - \pi_2^* E(\epsilon_2^2) - \pi_3 \frac{\bar{Y}}{2\bar{F}_x} E(\epsilon_2\epsilon_3) = 0 \tag{4.6}$$

$$\frac{\bar{Y}}{2\bar{F}_x}\left(E(\epsilon_0\epsilon_3) - \pi_1^* E(\epsilon_1\epsilon_3) - \pi_2^* E(\epsilon_2\epsilon_3) - \pi_3 \frac{\bar{Y}}{2\bar{F}_x} E(\epsilon_3^2)\right) = 0. \tag{4.7}$$

Solving the above normal equations and using Result 3(ii) of Ahmad (2008), we can write the optimum values of π_1^*, π_2^* and π_3 as:

$$\pi_{1_{opt}}^* = \frac{(-1)^{1+1}\,\bar{Y}\,C_y\,|\mathbf{R}_{yx}|_{yxzF_x}}{\bar{X}\,C_x\,|\mathbf{R}|_{xzF_x}}, \pi_{2_{opt}}^* = \frac{(-1)^{2+1}\bar{Y}\,|\mathbf{R}_{yz}|_{yxzF_x}}{\bar{Z}\,|\mathbf{R}|_{xzF_x}} \text{ and } \pi_{3opt} = \frac{2\,(-1)^{3+1}\,C_y|\mathbf{R}_{yF_x}|_{yxzF_x}}{C_{F_x}\,|\mathbf{R}|_{xzF_x}},$$

where

$$|\mathbf{R}_{xzF_x}| = \begin{vmatrix} 1 & \rho_{xz} & \rho_{xF_x} \\ \rho_{xz} & 1 & \rho_{zF_x} \\ \rho_{xF_x} & \rho_{zF_x} & 1 \end{vmatrix}, |\mathbf{R}_{yx}|_{yxzF_x} = \begin{vmatrix} \rho_{yx} & \rho_{xz} & \rho_{xF_x} \\ \rho_{yz} & 1 & \rho_{zF_x} \\ \rho_{yF_x} & \rho_{zF_x} & 1 \end{vmatrix},$$

$$|\mathbf{R}_{yz}|_{yxzF_x} = \begin{vmatrix} \rho_{yx} & 1 & \rho_{xF_x} \\ \rho_{yz} & \rho_{xz} & \rho_{zF_x} \\ \rho_{yF_x} & \rho_{xF_x} & 1 \end{vmatrix}, |\mathbf{R}_{yF_x}|_{yxF_x} = \begin{vmatrix} \rho_{yx} & 1 & \rho_{xz} \\ \rho_{yz} & \rho_{xz} & 1 \\ \rho_{yF_x} & \rho_{xF_x} & \rho_{zF_x} \end{vmatrix}.$$

Now using normal equations (4.5), we can write (4.4) as

$$MSE(\hat{\bar{Y}}^*) = E(\epsilon_0(\epsilon_0 - \pi_1^*\epsilon_1 - \pi_2^*\epsilon_2 - \pi_3\frac{\bar{Y}}{2\bar{F}_x}\epsilon_3)) \tag{4.8}$$

$$\begin{aligned}
MSE(\hat{\bar{Y}}^*) &= E(\epsilon_0^2) - \pi_1^* E(\epsilon_0\epsilon_1) - \pi_2^*(\epsilon_0\epsilon_2) - \pi_3\frac{\bar{Y}}{2\bar{F}_x}E(\epsilon_0\epsilon_3) \\
&= \tau\bar{Y}^2C_y^2 - \pi_1^*\tau\bar{Y}\bar{X}C_yC_x\rho_{yx} - \pi_2^*\tau\bar{Y}\bar{Z}C_yC_z\rho_{yz} - \pi_3\frac{\bar{Y}}{2\bar{F}_x}\tau\bar{Y}\bar{F}_xC_yC_{F_x}\rho_{yF_x} \text{ (using Eq. (2.2))} \\
&= \tau\bar{Y}^2C_y^2 - (-1)^{1+1}\frac{\bar{Y}}{\bar{X}}\frac{C_y}{C_x}\frac{|\mathbf{R}_{yx}|_{yxzF_x}}{|\mathbf{R}|_{xzF_x}}\tau\bar{Y}\bar{X}C_yC_x\rho_{yx} - (-1)^{2+1}\frac{\bar{Y}}{\bar{Z}}\frac{C_y}{C_z}\frac{|\mathbf{R}_{yz}|_{yxzF_x}}{|\mathbf{R}|_{xzF_x}}\tau\bar{Y}\bar{Z}C_yC_z\rho_{yz} \\
&- (-1)^{3+1}\frac{2C_y}{C_{F_x}}\frac{|\mathbf{R}_{yz}|_{yxzF_x}}{|\mathbf{R}|_{xzF_x}}\frac{\bar{Y}}{2\bar{F}_x}\tau\bar{Y}\bar{F}_xC_yC_{F_x}\rho_{yz} \text{ (using Eqs. (3.7)-(3.9))} \\
&= \tau\bar{Y}^2C_y^2 - \tau\bar{Y}^2C_y^2\rho_{yx}\frac{|\mathbf{R}_{yx}|_{yxzF_x}}{|\mathbf{R}|_{xzF_x}} + \tau\bar{Y}^2C_y^2\rho_{yz}\frac{|\mathbf{R}_{yz}|_{yxzF_x}}{|\mathbf{R}|_{xzF_x}} - \tau\bar{Y}^2C_y^2\rho_{yF_x}\frac{|\mathbf{R}_{yF_x}|_{yxzF_x}}{|\mathbf{R}|_{xzF_x}} \\
&= \frac{\tau\bar{Y}^2C_y^2}{|\mathbf{R}|_{xzF_x}}\left(|\mathbf{R}|_{xzF_x} - \rho_{yx}|\mathbf{R}_{yx}|_{yxzF_x} + \rho_{yz}|\mathbf{R}_{yz}|_{yxzF_x} - \rho_{yF_x}|\mathbf{R}_{yF_x}|_{yxzF_x}\right) \\
&= \tau\bar{Y}^2C_y^2\frac{|\mathbf{R}|_{yxzF_x}}{|\mathbf{R}|_{xzF_x}} = \tau\bar{Y}^2C_y^2(1 - R_{yxzF_x}^T\mathbf{R}_{xzF_x}^{-1}R_{yxzF_x}),
\end{aligned}$$

where

$$R_{yxF_x} = \begin{bmatrix} \rho_{yx} \\ \rho_{yz} \\ \rho_{yF_x} \end{bmatrix} \text{ and } |\mathbf{R}|_{yxF_x} = \begin{vmatrix} 1 & \rho_{yx} & \rho_{yz} & \rho_{yF_x} \\ \rho_{yx} & 1 & \rho_{xz} & \rho_{xF_x} \\ \rho_{yz} & \rho_{xz} & 1 & \rho_{zF_x} \\ \rho_{yF_x} & \rho_{xF_x} & \rho_{zF_x} & 1 \end{vmatrix}.$$

Now using Result 2 of Ahmad (2008), we have

$$MSE(\hat{\hat{Y}}^*) = \tau \bar{Y}^2 C_y^2 \left(1 - \rho^2_{y.xzF_x}\right).$$

From Theorem 2.1 of Ahmad and Hanif (2016), we can write

$$\gamma_{1_{opt}} = \frac{\bar{Y}^2}{\bar{Y}^2 + MSE(\hat{\hat{Y}}^*)} = \left[1 + \tau C_y^2 \left(1 - \rho^2_{y.xzF_x}\right)\right]^{-1} = \left[1 + \tau C_y^2 \frac{|\mathbf{R}|_{yxzF_x}}{|\mathbf{R}|_{xzF_x}}\right]^{-1}.$$

Now $\pi_{1_{opt}} = \pi^*_{1_{opt}} \gamma_{1_{opt}}$, $\pi_{2_{opt}} = \pi^*_{2_{opt}} \gamma_{1_{opt}}$ and $\pi_{3_{opt}} = \pi^*_{3_{opt}} \gamma_{1_{opt}}$. Again using Theorem 2.1 of Ahmad and Hanif (2016), we can write

$$\begin{aligned} MSE_{opt}(t_{xzF}) &= \frac{\bar{Y}^2 MSE(\hat{\hat{Y}}^*)}{\bar{Y}^2 + MSE(\hat{\hat{Y}}^*)} = \frac{\tau \bar{Y}^2 C_y^2 \left(1 - \rho^2_{y.xzF_x}\right)}{1 + \tau C_y^2 \left(1 - \rho^2_{y.xzF_x}\right)} \\ &= \left[\tau \bar{Y}^2 C_y^2 \frac{|\mathbf{R}|_{yxzF_x}}{|\mathbf{R}|_{xzF_x}}\right] \left[1 + \tau C_y^2 \frac{|\mathbf{R}|_{yxzF_x}}{|\mathbf{R}|_{xzF_x}}\right]^{-1}. \end{aligned} \tag{4.9}$$

4.1 Special cases

In this section, we define three different estimators based on the available auxiliary information. Each estimator incorporates different combinations of the auxiliary variable X, its rank R_x, and its empirical distribution function F_x to investigate the role of dual use of auxiliary variable.

4.1.1 Estimator Using X

$$t_x = \left[\gamma_1 \, \bar{y} + \pi_1(\bar{x} - \bar{X})\right], \tag{4.10}$$

The $\pi_{1_{opt}} = \pi^*_{1_{opt}} \gamma_{1_{opt}}$ where,

$$\pi^*_{1_{opt}} = (-1)^{1+1} \frac{\bar{Y}}{\bar{X}} \frac{C_y}{C_x} \frac{|\mathbf{R}_{yx}|_{yx}}{|\mathbf{R}|_x} \text{ and } \gamma_{1_{opt}} = \left[1 + \tau C_y^2 \frac{|\mathbf{R}|_{yx}}{|\mathbf{R}|_x}\right]^{-1}.$$

The corresponding MSE is given by

$$MSE_{opt}(t_x) = \left[\tau \bar{Y}^2 C_y^2 \frac{|\mathbf{R}|_{yx}}{|\mathbf{R}|_x}\right] \left[1 + \tau C_y^2 \frac{|\mathbf{R}|_{yx}}{|\mathbf{R}|_x}\right]^{-1}. \tag{4.11}$$

4.1.2 Estimator Using X and Z

$$t_{xz} = \left[\gamma_1 \; \bar{y} + \pi_1(\bar{X} - \bar{x}) + \pi_3(\bar{Z} - \bar{z})\right], \tag{4.12}$$

The $\pi_{1_{opt}} = \pi^*_{1_{opt}} \gamma_{1_{opt}}, \quad \pi_{3_{opt}} = \pi^*_{3_{opt}} \gamma_{1_{opt}}$, where the optimum values of π_1 and π_3 are given by

$$\pi^*_{1_{opt}} = (-1)^{1+1} \frac{\bar{Y}}{\bar{X}} \frac{C_y}{C_x} \frac{|\mathbf{R}_{yx}|_{yxZ}}{|\mathbf{R}|_{xz}} \text{ and } \pi^*_{3_{opt}} = (-1)^{3+1} \frac{\bar{Y}}{\bar{Z}} \frac{C_y}{C_Z} \frac{|\mathbf{R}_{yz}|_{yxz}}{|\mathbf{R}|_{xz}}. \tag{4.13}$$

an the optimum value of γ_1 is

$$\gamma_{1_{opt}} = \left[1 + \tau C_y^2 \frac{|\mathbf{R}|_{yxZ}}{|\mathbf{R}|_{xZ}}\right]^{-1}.$$

The corresponding MSE is given by

$$MSE_{opt}(t_{xz}) = \left[\tau \bar{Y}^2 C_y^2 \frac{|\mathbf{R}|_{yxz}}{|\mathbf{R}|_{xz}}\right] \left[1 + \tau C_y^2 \frac{|\mathbf{R}|_{yxz}}{|\mathbf{R}|_{xz}}\right]^{-1}. \tag{4.14}$$

4.1.3 Estimator Using X and EDF F_x

$$t_{xF} = \left[\gamma_1 \; \bar{y} + \pi_1(\bar{X} - \bar{x})\right] \exp\left(\pi_4 \frac{\bar{F}_x - \bar{f}_x}{\bar{F}_x + \bar{f}_x}\right), \tag{4.15}$$

The $\pi_{1_{opt}} = \pi^*_{1_{opt}} \gamma_{1_{opt}}, \quad \pi_{4_{opt}} = \pi^*_{4_{opt}} \gamma_{1_{opt}}$, where the optimum values of π_1 and π_4 are given by

$$\pi^*_{1_{opt}} = (-1)^{1+1} \frac{\bar{Y}}{\bar{X}} \frac{C_y}{C_x} \frac{|\mathbf{R}_{yx}|_{yxF_x}}{|\mathbf{R}|_{xF_x}} \text{ and } \pi^*_{4_{opt}} = (-1)^{4+1} \frac{2C_y}{C_{F_x}} \frac{|\mathbf{R}_{yF_x}|_{yxF_x}}{|\mathbf{R}|_{xF_x}}. \tag{4.16}$$

an the optimum value of γ_1 is

$$\gamma_{1_{opt}} = \left[1 + \tau C_y^2 \frac{|\mathbf{R}|_{yxF_x}}{|\mathbf{R}|_{xF_x}}\right]^{-1}.$$

The corresponding MSE is given by

$$MSE_{opt}(t_{xF}) = \left[\tau \bar{Y}^2 C_y^2 \frac{|\mathbf{R}|_{yxF_x}}{|\mathbf{R}|_{xF_x}}\right] \left[1 + \tau C_y^2 \frac{|\mathbf{R}|_{yxF_x}}{|\mathbf{R}|_{xF_x}}\right]^{-1}. \tag{4.17}$$

5 Simulation Study

We have conducted a simulation study to compare the performance of the proposed estimator, its special cases and estimators given in Section 2 to investigate the role of dual use of auxiliary variable$X$ and independent auxiliary variable Z, on the basis of their empirical MSE. For this purpose, we used an arbitrary population of size N (= 10000). As we are dealing with two-auxiliary variables, the true values of two auxiliary variables X and Z are generated, where $X \sim N(10, 2)$

and $Z \sim lognorm(1, 0.5)$. The values of empirical distribution function of X are obtained using R function $ecfd()$ and denoted by F_x. The population values of the study variable are generated using the model $Y = 10 + 2X + 2.5Z + 2.75F_x + \epsilon$, where $\epsilon \sim N(0, 1)$. The coefficients show the direction and strength of relationship of auxiliary variables with the study variable.

The simulation study is based on $S = 10000$ samples. The empirical MSE of proposed estimator, t_1 to t_6 and all members of t_7 are computed for three sample sizes 5%, 10% and 20% of N. The empirical MSE of an estimator t is computed by

$$MSE(t) = \frac{1}{S}\sum_{i=1}^{S}(t_i - T)^2, \text{ where } T = \frac{1}{S}\sum_{i=1}^{S} t_i.$$

The percent relative efficiency (PRE) of an estimator t_i with respect to t_1 can be obtained by

$$PRE = \frac{MSE(t_i)}{MSE(t_1)} \times 100.$$

Using above simulation setting, the empirical MSEs and PREs of proposed estimator, t_1 to t_6 and all members of t_7 are computed using symmetric and skewed Y. For skewed Y, we simulated the auxiliary variable using Lognormal distribution as $X \sim lognormal(1, 0.75)$ and $Z \sim lognormal(1, 1)$. The histograms of Y for symmetric and populations are given in the following figure.

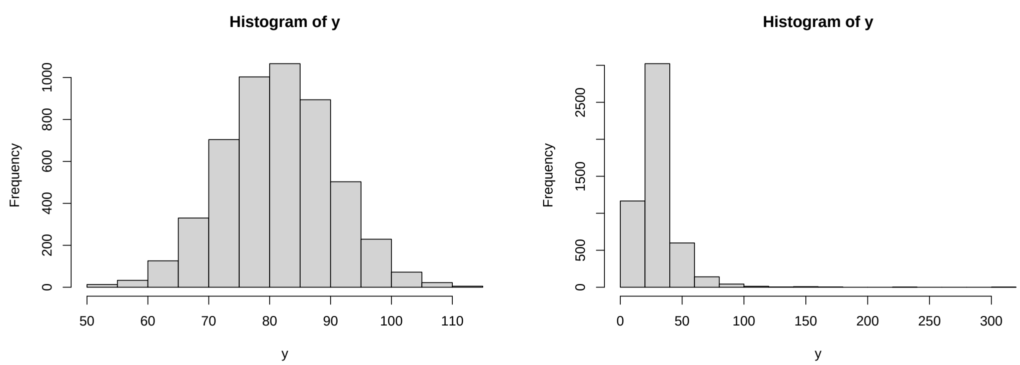

Figure 1: Histograms of Y for Symmetric and Skewed Populations

The population characteristics in terms of mean, coefficient of variation and coefficient of correlations of Y, X, Z and F_x are shown in table below. The results related to empirical MSEs and PREs are given in the following tables.

Table 1: Parameters of Symmetric and Skewed populations

Population	$\bar{Y}$	$\bar{X}$	$\bar{Z}$	$\bar{F}_x$	C_y	C_x	C_z	C_{F_x}	ρ_{yx}	ρ_{yz}	ρ_{yFx}	ρ_{xz}	ρ_{xF_x}	ρ_{zF_x}
Symmetric	81.3	10	20	0.5	0.11	0.20	0.15	0.58	0.53	0.83	0.52	-0.011	0.977	0.014
Skewed	30.1	3.6	4.6	0.5	0.59	0.86	1.42	0.58	0.36	0.92	0.30	-0.022	0.814	-0.029

Table 2: Empirical MSE of each estimator for symmetric and skewed populations

Estimator	**Symmetric**			**Skewed**		
	$n = 20\%$	$n = 10\%$	$n = 5\%$	$n = 20\%$	$n = 10\%$	$n = 5\%$
t_1	0.07	0.124	0.215	0.19	0.33	0.59
t_2	0.06	0.104	0.174	0.17	0.29	0.50
t_3	0.06	0.104	0.174	0.17	0.29	0.50
t_4	0.07	0.124	0.216	0.19	0.33	0.59
t_5	0.06	0.104	0.174	0.17	0.29	0.50
t_6	0.66	1.073	1.268	1.78	2.30	2.83
t_{71}	0.08	0.151	0.272	0.20	0.36	0.65
t_{72}	0.07	0.132	0.232	0.17	0.29	0.50
t_{73}	0.08	0.152	0.275	0.22	0.41	0.73
t_{74}	0.06	0.110	0.186	0.17	0.29	0.50
t_{75}	0.08	0.148	0.266	0.21	0.38	0.68
t_{76}	0.07	0.133	0.236	0.21	0.38	0.68
t_{77}	0.08	0.149	0.269	0.19	0.33	0.59
t_{78}	0.08	0.151	0.273	0.22	0.41	0.74
t_{79}	0.08	0.151	0.273	0.22	0.41	0.74
t_{80}	0.06	0.105	0.175	0.17	0.29	0.50
t_{xzF}	0.03	0.055	0.064	0.09	0.11	0.13
t_{xz}	0.03	0.054	0.062	0.09	0.11	0.13
t_{xF}	0.06	0.104	0.174	0.17	0.29	0.50
t_x	0.06	0.104	0.174	0.17	0.29	0.50

Table 3: Empirical PRE of each estimator for symmetric and skewed populations

Estimator	**Symmetric**			**Skewed**		
	$n = 20\%$	$n = 10\%$	$n = 5\%$	$n = 20\%$	$n = 10\%$	$n = 5\%$
t_1	100.00	100.00	100.00	100.00	100.00	100.00
t_2	114.60	119.13	124.11	110.51	113.63	118.21
t_3	114.60	119.13	124.11	110.51	113.63	118.21
t_4	99.96	99.93	99.91	99.78	99.70	99.67
t_5	114.55	119.05	124.01	110.29	113.31	117.81
t_6	10.16	11.55	16.99	10.43	14.16	20.74
t_{71}	84.57	82.20	79.09	91.53	89.31	90.75
t_{72}	94.72	94.28	93.02	110.06	112.98	117.61
t_{73}	83.94	81.46	78.27	83.72	80.02	80.19
t_{74}	109.44	112.68	115.67	110.13	113.08	117.67
t_{75}	86.03	83.90	81.01	87.95	84.99	85.82
t_{76}	93.62	92.95	91.47	88.62	85.79	86.72
t_{77}	85.42	83.18	80.19	97.95	97.27	99.93
t_{78}	84.45	82.06	78.93	83.52	79.80	79.94
t_{79}	84.45	82.06	78.93	83.52	79.80	79.94
t_{80}	114.24	118.46	123.24	110.13	113.08	117.28
t_{xzF}	197.68	224.39	337.12	206.15	289.27	443.46
t_{xz}	199.12	227.79	346.05	206.60	290.35	447.62
t_{xF}	114.36	118.71	123.57	110.24	113.26	117.51
t_x	114.55	119.05	124.01	110.29	113.31	117.81

The Table 2 contains the MSEs of proposed and the estimators discussed in Section 2. As mentioned above that Hussain, et al. (2022) deduced ten members of their class using different combinations of a and b. The MSEs of all these ten estimators are also given in this table. The percent relative efficiencies based on the MSEs considering the mean per unit estimator as base estimator are given in Table 3. From this table, we can see that the members t_{74} and t_{80} are better then the base estimator the rest are performing bad then even mean per unit estimator. The performance of t_{80} is same as t_2,t_3,t_5, t_{xF} and t_x. This means that the use of dual auxiliary variable as empirical distribution function is useless. Moreover, it is noticed that using rank of auxiliary variable as dual use even worsen the performance of an estimator in terms of efficiency e.g. t_6 that is almost 70% bad in efficiency then men per unit estimator. The PREs of t_{xF} and t_x also show that there is no role of empirical distribution function in improving the efficiency. The similar argument can be made from PREs of t_{xzF} and t_{xz}. However, an independent auxilairy variable can improve effiecieny as evedent from the PREs of t_{xz} and t_x or t_{xzF} and t_{xF}. Based on this simulation study, we do not recommend dual use of auxiliary variable either in the form of rank or empirical distribution function, however, new auxiliary variable(s) can be used to improve the performance of estimator provided the new variable(s) are significantly correlated with study variable. Below we provided PRE based on theoretical MSEs calculated form real data.

6 Real Life Applications

In this section, we have used five real populations data sets to numerically compare the MSE of suggested estimator to the estimators discussed in Section 2. The detail of populations, variable used and the parameters used for calculation of MSEs are given in Tables A2 and A3 of Appendix. Below we provided the MSEs and percentage relative efficiency(PRE) of each estimator.

Table 4: MSE and PRE of each estimator

	MSE					**PRE**				
Est.	**Pop.-I**	**Pop.-II**	**Pop.-III**	**Pop.-IV**	**Pop.-V**	**Pop.-I**	**Pop.-II**	**Pop.-III**	**Pop.-IV**	**Pop.-V**
t_1	2829.13	8707.99	5537.26	13112.87	6945.41	100.00	100.00	100.00	100.00	100.00
t_2	2318.34	4379.75	43.03	4732.20	816.23	122.03	198.82	12867.16	277.10	850.91
t_3	2318.34	4379.75	43.03	4732.20	816.23	122.03	198.82	12867.16	277.10	850.91
t_4	2818.35	8674.92	5512.10	13024.88	6902.64	100.38	100.38	100.46	100.68	100.62
t_5	2311.09	4371.37	43.03	4720.69	815.64	122.41	199.21	12867.62	277.77	851.53
t_6	1239.59	2770.81	43.03	2484.74	759.86	228.23	314.28	12867.18	527.74	914.04
t_7	1237.27	2768.11	43.03	2481.56	759.35	228.66	314.58	12867.64	528.41	914.65
t_{xzF}	1232.79	2648.93	41.62	267.78	57.96	229.49	328.74	13303.70	4896.92	11982.85
t_{xz}	3287.36	15073.26	631.52	268.20	58.32	86.06	57.77	876.81	4889.20	11909.88
t_{xF}	1159.56	3374.50	60.88	2709.41	602.11	243.98	258.05	9095.69	483.98	1153.51
t_x	2311.09	4371.37	43.03	4720.69	815.64	122.41	199.21	12867.62	277.77	851.53

For the first two populations, the efficiency values of t_{xzF}, t_{xF}, and t_x are relatively close to each other, suggesting that the inclusion of Z and F does not significantly alter the estimation performance. However, t_{xzF} does show a moderate improvement over the others, indicating that the combination of X, Z, and F is beneficial, albeit not overwhelmingly so in these cases. Interestingly, t_{xz} has the lowest efficiency among these estimators in the first two populations, which suggests that excluding F leads to a loss in precision when Z alone is not highly correlated with Y.

A drastic shift occurs in the last three populations, particularly the third, fourth, and fifth, where Z is highly correlated with Y. The efficiency of t_{xzF} jumps significantly (13303.70, 4896.92, and 11982.85), making it the most efficient estimator. Similarly, t_{xz} also experiences a significant increase (876.81, 4889.20, and 11909.88), indicating that Z is now a dominant auxiliary variable. However, the difference between t_{xzF} and t_{xz} becomes relatively small in the last two populations, suggesting that the inclusion of F does not provide much additional gain when Z is already highly correlated with Y.

The estimator t_{xF}, which uses X and F but not Z, performs well in the first three populations but fails to maintain high efficiency in the last two (9095.69 in the third but only 483.98 and 1153.51 in the fourth and fifth). This confirms that F is unable to compensate for the lack of Z when Z has a strong association with Y. Similarly, the estimator t_x, which relies solely on X, shows a significant drop in efficiency in the last two populations (277.77 and 851.53), reinforcing the idea that relying on X alone is insufficient, particularly when other strong auxiliary variables are available.

Overall, these results suggest that Z plays a crucial role in improving efficiency, especially in populations where it has a strong correlation with Y. While F provides some benefit in certain cases, its contribution diminishes when Z is already an effective predictor. The estimator t_{xzF} remains the most efficient overall, but in cases where Z is highly correlated with Y, t_{xz} performs nearly as well, demonstrating that F becomes redundant in such scenarios. The simulation study confirms that dual use of auxiliary variable is not useful in improving the efficiency but here in improves the efficiency where Z is weakly correlated with Y and one possible reason may be that the theoretical expressions of MSEs contains the coefficient of multiple determination and it always increases by adding new variable irrespective of the fact that the auxiliary variable is significantly affecting the study variable or not.

7 Conclusions

The findings from both the real population data analysis and the simulation study provide valuable insights into the efficiency of different estimators. The simulation study demonstrated that incorporating the empirical distribution function (EDF) or rank of an auxiliary variable as a dual-use technique does not enhance estimator performance. In some cases, such as t_6, the inclusion of rank even led to a significant drop in efficiency—nearly 70% lower than the mean per unit estimator. Similarly, the PREs of t_{xF} and t_x confirmed that EDF does not contribute to efficiency improvement. However, the study highlighted that adding an independent auxiliary variable can significantly improve estimator efficiency, as observed in the comparisons of t_{xz} with t_x and t_{xzF} with t_{xF}.

The real population data analysis further supports these conclusions. Across the five studied populations, it was observed that when the variable Z was highly correlated with the study variable Y (as in the last two populations), estimators incorporating Z demonstrated better performance. In contrast, when the empirical distribution function was included, particularly in cases where Z was more correlated with Y, the efficiency of the estimators decreased. This confirms that EDF fails to provide any substantial benefit, especially when a more strongly correlated auxiliary variable is available.

Overall, based on both studies, we conclude that the dual use of auxiliary variables either through ranks or empirical distribution functions does not improve efficiency and may even be detrimental. Instead, the inclusion of additional auxiliary variables, provided they have a strong correlation with the study variable, can lead to meaningful efficiency gains. Thus, we recommend focusing on selecting independent and highly correlated auxiliary variables rather than using ranks or EDF transformations. We also recommend simulation study along with empirical study because theoretical results do no guarantee the sampling stability of the estimator as we seen in our case.

Acknowledgements

We are thankful to the reviewers for their constructive comments and suggestions which have certainly improved the presentation and quality of paper. The research data associated with the paper

is available with author Zahoor Ahmad, and data can be provided by requesting on email.

References

Abdollahi Nanvapisheh, A., MirMostafaee, S., and Altun, E. (2019), "A new two-parameter distribution: properties and applications," *Journal of Mathematical Modeling*, 7, 35–48.

Akaike, H. (1974), "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, 19, 716–723.

Al-Ta'ani, O. and Gharaibeh, M. M. (2023), "Ola distribution: A new one parameter model with applications to engineering and Covid-19 data," *Applied Mathematics and Information Sciences*, 17, 243–252.

Bittmann, F. (2021), *Bootstrapping: an integrated approach with Python and Stata*, Walter de Gruyter GmbH & Co KG.

Brent, R. P. (1973), *Algorithms for minimization without derivatives*, Prentice-Hall.

Canty, A. and Ripley, B. (2017), "Package 'boot'," *Bootstrap Functions. CRAN R Proj.*

Davison, A. C. and Hinkley, D. V. (1997), *Bootstrap methods and their application*, no. 1, Cambridge University Press.

Elechi, O., Okereke, E. W., Chukwudi, I. H., Chizoba, K. L., and Wale, O. T. (2022), "Iwueze's distribution and its application," *Journal of Applied Mathematics and Physics*, 10, 3783–3803.

Gharaibeh, M. (2021), "Gharaibeh distribution and its applications," *Journal of Statistics Applications and Probability*, 10, 441–452.

Ghitany, M. E., Atieh, B., and Nadarajah, S. (2008), "Lindley distribution and its application," *Mathematics and Computers in Simulation*, 78, 493–506.

Henningsen, A. and Toomet, O. (2011), "maxLik: A package for maximum likelihood estimation in R," *Computational Statistics*, 26, 443–458.

Kostyshak, S. (2024), "Package bootstrap. Functions for the book 'An introduction to the bootstrap'," `https://cran.r-project.org/web/packages/bootstrap`,

Kwan, K., Breault, G., Umbenhauer, E., McMahon, F., and Duggan, D. (1976), "Kinetics of indomethacin absorption, elimination, and enterohepatic circulation in man," *Journal of Pharmacokinetics and Biopharmaceutics*, 4, 255–280.

Mbegbu, J. I. and Echebiri, U. V. (2022), "Juchez probability distribution: Properties and applications," *Asian Journal of Probability and Statistics*, 20, 56–71.

McAtamney, M. (2019), “Industrial service design: an examination of Chinese choice preferences for shipping services,” Ph.D. thesis, University of Otago.

Nassar, M. and Nada, N. (2011), “The beta generalized Pareto distribution,” *Journal of Statistics: Advances in Theory and Applications*, 6, 1–17.

Olufemi-Ojo, O. B., Onyeagu, S. I., and Obiora Ilouno, H. O. (2024), “On the application of two-parameter Shanker distribution,” *International Journal of Innovative Science and Research Technology*, 9, 807–820.

Onyekwere, C. K. and Obulezi, O. J. (2022), “Chris-Jerry distribution and its applications,” *Asian Journal of Probability and Statistics*, 20, 16–30.

Pawitan, Y. (2001), *In all likelihood: statistical modelling and inference using likelihood*, Oxford University Press.

R Core Team (2023), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, `https://www.R-project.org/`.

Rama, S. (2016), “Aradhana distribution and its applications,” *International Journal of Statistics and Applications*, 6, 23–34.

Saleh, A. M. E. (2006), *Theory of preliminary test and Stein-type estimation with applications*, John Wiley & Sons.

Schwarz, G. (1978), “Estimating the dimension of a model,” *The Annals of Statistics*, 6, 461–464.

Severini, T. A. (2000), *Likelihood methods in statistics*, Oxford University Press.

Shanker, R. (2015a), “Akash distribution and its applications,” *International Journal of Probability and Statistics*, 4, 65–75.

— (2015b), “Shanker distribution and its applications,” *International Journal of Statistics and Applications*, 5, 338–348.

— (2016a), “Garima distribution and its application to model behavioral science data,” *Biometrics & Biostatistics International Journal*, 4, 1–9.

— (2016b), “Sujatha distribution and its applications,” *Statistics in Transition. New Series*, 17, 391–410.

— (2017a), “Rama distribution and its application,” *International Journal of Statistics and Applications*, 7, 26–35.

— (2017b), “Rani distribution and its application,” *Biometrics and Biostatistics International Journal*, 6, 1–10.

— (2023a), "Komal distribution with properties and application in survival analysis," *Biometrics and Biostatistics International Journal*, 12, 40–44.

— (2023b), "Pratibha distribution with properties and application," *Biometrics and Biostatistics International Journal*, 13, 136–142.

Shanker, R., Sharma, S., and Shanker, R. (2013), "A two-parameter Lindley distribution for modeling waiting and survival times data," *Applied Mathematics*, 4, 363–368.

Shanker, R. and Shukla, K. (2017), "Ishita distribution and its applications," *Biometrics and Biostatistics International Journal*, 5, 1–9.

Shanker, R., Shukla, K., Ranjan, A., and Shanker, R. (2021), "Adya distribution with properties and application," *Biometrics and Biostatistics International Journal*, 10, 81–88.

Shanker, R., Shukla, K. K., Shanker, R., and Leonida, T. A. (2017), "A three-parameter Lindley distribution," *American Journal of Mathematics and Statistics*, 7, 15–26.

Wilks, S. S. (1938), "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, 9, 60–62.

Received: January 13, 2025

Accepted: June 22, 2025

A Appendix

Table A1: Population and sample characteristics for study and auxiliary variables, where Y, X, and Z are the population of currently married, 15–49 years old women, and 18 years old and above individuals, respectively

Characteristic	Population	Sample
Mean	$\bar{Y} = \sum_{i=1}^{N} y_i/N, \bar{X} = \sum_{i=1}^{N} x_i/N, \bar{F}_x = \sum_{i=1}^{N} f_{x_i}/N, \bar{Z} = \sum_{i=1}^{N} z_i/N$	$\bar{y} = \sum_{i=1}^{n} y_i/n, \bar{x} = \sum_{i=1}^{n} x_i/n, \bar{f}_x = \sum_{i=1}^{n} f_{x_i}/n, \bar{z} = \sum_{i=1}^{n} z_i/n$
Variance	$S_y^2 = \sum_{i=1}^{N}(y_i - \bar{Y})^2/N, S_x^2 = \sum_{i=1}^{N}(x_i - \bar{X})^2/N$ $S_{F_x}^2 = \sum_{i=1}^{N}(r_{x_i} - \bar{F}_x)^2/N, S_z^2 = \sum_{i=1}^{N}(z_i - \bar{Z})^2/N$	$s_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/n, s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$ $s_{F_x}^2 = \sum_{i=1}^{n}(r_{x_i} - \bar{f}_x)^2/n, s_z^2 = \sum_{i=1}^{n}(z_i - \bar{z})^2/n$
Covariance	$S_{xy} = \sum_{i=1}^{N}(y_i - \bar{Y})(x_i - \bar{X})/N, S_{yF_x} = \sum_{i=1}^{N}(y_i - \bar{Y})(r_{x_i} - \bar{F}_x)/N$ $S_{yz} = \sum_{i=1}^{N}(y_i - \bar{Y})(z_i - \bar{Z})/N, S_{xF_x} = \sum_{i=1}^{N}(x_i - \bar{X})(r_{x_i} - \bar{F}_x)/N$ $S_{xz} = \sum_{i=1}^{N}(x_i - \bar{X})(z_i - \bar{Z})/N, S_{zF_x} = \sum_{i=1}^{N}(z_i - \bar{Z})(r_{x_i} - \bar{F}_x)/N$	$s_{xy} = \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})/n, s_{yF_x} = \sum_{i=1}^{n}(y_i - \bar{y})(r_{x_i} - \bar{f}_x)/n$ $s_{yz} = \sum_{i=1}^{n}(y_i - \bar{y})(z_i - \bar{z})/n, s_{xF_x} = \sum_{i=1}^{n}(x_i - \bar{x})(r_{x_i} - \bar{f}_x)/n$ $s_{xz} = \sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})/n, s_{zF_x} = \sum_{i=1}^{n}(z_i - \bar{z})(r_{x_i} - \bar{f}_x)/n$
Correlation Coefficient	$\rho_{xy} = S_{yx}/S_y S_x, \rho_{yF_x} = S_{yF_x}/S_y S_{F_x}, \rho_{yz} = S_{yz}/S_y S_z$ $\rho_{xF_x} = S_{xF_x}/S_x S_{F_x}, \rho_{xz} = S_{xz}/S_x S_z, \rho_{zF_x} = S_{zF_x}/S_z S_{F_x}$	$r_{yx} = s_{yx}/s_y s_x, r_{yF_x} = s_{yF_x}/s_y s_{F_x}, r_{yz} = s_{yz}/s_y s_z$ $r_{xF_x} = s_{xF_x}/s_x s_{F_x}, r_{xz} = s_{xz}/s_x s_z, r_{zF_x} = s_{zF_x}/s_z s_{F_x}$

Table A2: Detail of Populations

Sr. No.	Source of Populations
1	Population census report of Jhang district (1998), Pakistan
2	Population census report of Faisalabad district (1998), Pakistan
3	Population census report of Gujrat district (1998), Pakistan
4	Population census report of Kasur (1998), Pakistan
5	Population census report of Sialkot district (1998), Pakistan

Table A3: Parameters of populations

District	N	n	$\bar{Y}$	C_y	$\bar{X}$	$\bar{Z}$	$\bar{F}_x$	C_x	C_z	C_{F_x}	ρ_{yx}	ρ_{yz}	ρ_{yFx}	ρ_{xz}	ρ_{xF_x}	ρ_{zF_x}
Jhang	368	184	0860.11	0.59	3159.24	1311.44	184.50	0.77	0.81	0.57	0.48	0.42	0.73	0.33	0.65	0.49
Faisalabad	283	142	1511.26	0.52	6173.16	2457.68	142.00	1.02	0.60	0.58	0.50	0.71	0.82	0.44	0.49	0.75
Gujrat	204	102	1101.28	0.48	4326.28	1703.48	102.50	1.88	0.49	0.58	0.49	0.99	0.89	0.50	0.31	0.89
Kasur	181	091	1393.20	0.55	4984.15	2114.84	091.00	0.55	0.64	0.58	0.99	0.80	0.91	0.80	0.92	0.72
Sailkot	269	135	1058.74	0.65	3812.47	1684.70	135.00	0.65	0.68	0.58	0.99	0.94	0.82	0.94	0.82	0.75