Pak.j.stat.oper.res. Vol.18 No. 4 2022 pp 985-994 DOI: http://dx.doi.org/10.18187/pjsor.v18i4.3955

Pakistan Journal of Statistics and Operation Research

Modified Regression Estimators for Improving Mean Estimation -Poisson Regression Approach

Zakir Hussain Wani[1*], S.E.H. Rizvi[2], Manish Sharma[3], M. Iqbal Jeelani[4], Saqib Mushtaq[5]

* Corresponding Author

1. Division of Statistics and Computer Science, Main Campus SKUAST-J, Chatha Jammu-180009, India, wanizakir927@gmail.com
2. Division of Statistics and Computer Science, Main Campus SKUAST-J, Chatha Jammu-180009, India, sehrizvistat@gmail.com
3. Division of Statistics and Computer Science, Main Campus SKUAST-J, Chatha Jammu-180009, India, manshstat@gmail
4. Division of Statistics and Computer Science, Main Campus SKUAST-J, Chatha Jammu-180009, India, jeelani.miqbal@gmail.com
5. Department of Mathematics, Main Campus University of Kashmir Srinagar-190006-India, iamsaqib01@gmail.com

Abstract

In this article, a class of Poisson-regression based estimators has been proposed for estimating the finite population mean in simple random sampling without replacement *(SRSWOR)*. The Poisson-regression model is the most common method used to model count responses in many studies. The expression for bias and mean square error (*MSE)* of proposed class of estimators are obtained up to first order of approximation. The proposed estimators have been compared theoretically with the existing estimators, and the condition under which the proposed class of estimators perform better than existing estimators have been obtained. Two real data sets are considered to assess the performance of the proposed estimators. Numerical findings confirms that the proposed estimators dominate over the existing estimators such as Koc (2021) and Usman et al. (2021) in terms of mean squared error.

Key Words: Ratio estimator; Poisson regression; Mean Square error; Bias; Efficiency; Auxiliary variable

Mathematical Subject Classification: 62D05

1. Introduction

Ratio type estimators take advantage of the correlation between the auxiliary variable g and the study variable y. When data on the auxiliary variable is available, the ratio estimator is a good choice for estimating the population mean. In sampling theory, population information of the auxiliary variable, such as coefficient of variation or kurtosis, is frequently employed to improve the efficiency of the estimation for a population mean for the ratio estimator. However, the outlier problem, which occurs when data has extreme values, reduces efficiency because traditional estimators are sensitive to extreme values. In order to handle this problem, Kadilar, Candan, and Cingi (2007) adapted Huber-M estimate related to ratio estimator presented in Kadilar and Cingi (2004). They obtained the MSE equations in order to decrease the effect of outlier problem. Oral and Kadilar (2011a, 2011b) considered maximum likelihood approach and they modified maximum likelihood estimators into Kadilar–Cingi estimators. Zaman and Bulut (2019) introduced new robust ratio estimators based on the estimators given in Kadilar, Candan, and Cingi (2007). Zaman (2019) provided some combining robust estimators for the population mean using the ratio estimators presented in Zaman and Bulut (2019). Ali et al. (2021) developed a class of robust-regression type estimators, by utilizing Zaman and Bulut (2019), in case of sensitive research under simple random sampling scheme. Zaman and Bulut (2020) suggested new regression-type estimators by using robust regression estimates and robust covariance matrices in the stratified random sampling.

Furthermore, it is problematic of using a linear regression model when the mean is large enough for the count data, this is the situation when Poisson distribution converges to normal distribution. The regression model is the most widely used method for modelling. The linear model links the estimated value with supplementary variables, negative prediction values are possible. The validity of hypothesis testing in linear regression is also dependent on the supposition of variables constant variance. These suppositions are erroneous for count data. As a result, in the applied sciences, modelling count data, the used technique is the Poisson regression. Koc (2021) and Usman et al. (2021) used the Poisson regression method for improving the population mean of the study variable. In this article, we attempted to propose estimators for estimating the population mean of the study variable Y using information on the auxiliary variable G using Poisson regression model. The mathematical properties of proposed estimators, such as bias and mean square error were examined using large sample approximation. The proposed estimators have been shown to outperform all other estimators tested in the literature. Numerical illustrations have also been done in support of current investigation.

In Poisson regression, the study variable $Y_i(y_i = 0,1,2,\dots)$is the number of events that occur at a particular period, with a Poisson distribution given by

$$P(y_i,\theta_i) = \frac{e^{-\theta_i}\lambda_i^{y_i}}{y_i!}, \theta > 0 \tag{1}$$

The mean and variance are same in Poisson distribution and is given by

$E(Y_i) = Var(Y_i) = \theta_i$

The $L(\theta, y)$ i.e., log-likelihood of $P(y_i,\theta_i)$ is defined as;

$$L(\theta, y) = \sum_{i=1}^{n}\left(y_i ln(\theta_i) - \theta_i - ln(y_i!)\right) \tag{2}$$

Let G be the matrix of order $n \times (k+1)$ of the auxiliary variable. Then, the association between Y_i and ith row of matrix, g_i associating through $d(\theta_i), is$

$$ln(\theta_i) = \xi_i = g_i^T B \tag{3}$$

Where, $B = B_0, B_1, \dots, B_k$ are the regression parameters.

Differentiating (3) with respect to B yields $\breve{B}$ which represents the maximum likelihood estimators of B.

$$\sum_{i=1}^{n}(y_i exp(g_i^T B) g_i) \tag{4}$$

For solving such k equations, we use iterative methods such as Newton-Raphson algorithms and Fisher Scoring algorithms (see Cameron and Trivedi (1998), Montgomery et al. (2006), and Koc (2021).

2. Estimators from the literature

This section gives a brief introduction of some well-known estimators/ classes of estimators from the literature.

1. Kadilar and Cingi (2004) propose the following estimators for the estimation of population mean $\bar{Y}$ in simple random sampling.

$$\bar{y}_{KC1} = \frac{\bar{y} + b(\bar{G} - \bar{g})}{\bar{g}}\bar{G} \tag{5}$$

$$\bar{y}_{KC2} = \frac{\bar{y} + b(\bar{G} - \bar{g})}{(\bar{g} + C_g)}(\bar{G} + C_g) \tag{6}$$

$$\bar{y}_{KC3} = \frac{\bar{y} + b(\bar{G} - \bar{x})}{(\bar{g} + \beta_2(g))}(\bar{G} + \beta_2(g)) \tag{7}$$

$$\bar{y}_{KC4} = \frac{\bar{y} + b(\bar{G} - \bar{g})}{(\bar{g}\beta_2(g) + C_g)}(\bar{G}\beta_2(g) + C_g) \tag{8}$$

$$\bar{y}_{KC5} = \frac{\bar{y} + b(\bar{G} - \bar{g})}{\left(\bar{g}C_g + \beta_2(g)\right)}\left(\bar{G}C_g + \beta_2(g)\right) \quad (9)$$

Here, C_g and $\beta_2(g)$ are the population coefficient of variation and the population coefficient of kurtosis respectively of the auxiliary variable. $\bar{y}$ and $\bar{g}$ are the sample means of the study and auxiliary variable respectively and it is assumed that the population mean $\bar{G}$ of the auxiliary variable G is known. Here $b = \frac{s_{gy}}{s_g^2}$ is obtained by the least square. Where s_g^2 and s_y^2 are the sample variances of auxiliary and study variable respectively and s_{gy} is the sample covariance between the auxiliary and study variable.

The *MSE* of the estimators 5-9 can be found using a first degree of approximation of the Taylor series expansion and is as follows;

$$\mathrm{MSE}(\bar{y}_{KCi}) = \lambda\left[R_{KCi}^2 S_g^2 + 2BR_{KCi}S_g^2 + B^2S_g^2 - 2R_{KCi}S_{gy} - 2BS_{gy} + S_y^2\right] \quad (10)$$

Where i=1, 2, …, 5, $B = \frac{S_{gy}}{S_g^2}$ is obtained by least square method $f = \frac{n}{N}$; n is the sample size and N is the population size and $\lambda = \frac{1}{n} - \frac{1}{N}$

Where:

$$R_{KC1} = R = \frac{\bar{Y}}{\bar{G}}, \quad R_{KC2} = \frac{\bar{Y}}{\bar{G} + C_g}, \quad R_{KC3} = \frac{\bar{Y}}{\bar{G} + \beta_{2(g)}}$$

$$R_{KC4} = \frac{\bar{Y}\beta_{2(g)}}{\bar{G}\beta_{2(g)} + C_g}, \quad R_{KC5} = \frac{\bar{Y}C_g}{\bar{G}C_g + \beta_{2(g)}}$$

2. Motivated from Kadilar and Cingi (2004), Koc (2021) proposed a new way to improve the estimators by adding Poisson regression-based ratio estimators as follows;

$$\bar{y}_1 = \frac{\bar{y} + b_P(\bar{G} - \bar{g})}{\bar{g}}\bar{G} \quad (11)$$

$$\bar{y}_2 = \frac{\bar{y} + b_P(\bar{G} - \bar{g})}{\left(\bar{g} + C_g\right)}\left(\bar{G} + C_g\right) \quad (12)$$

$$\bar{y}_3 = \frac{\bar{y} + b_P(\bar{G} - \bar{x})}{\left(\bar{g} + \beta_2(g)\right)}\left(\bar{G} + \beta_2(g)\right) \quad (13)$$

$$\bar{y}_4 = \frac{\bar{y} + b_P(\bar{G} - \bar{g})}{\left(\bar{g}\beta_2(g) + C_g\right)}\left(\bar{G}\beta_2(g) + C_g\right) \quad (14)$$

$$\bar{y}_5 = \frac{\bar{y} + b_P(\bar{G} - \bar{g})}{\left(\bar{g}C_g + \beta_2(g)\right)}\left(\bar{G}C_g + \beta_2(g)\right) \quad (15)$$

The mean square error of $\bar{y}_i, i = 1,2,..,5$ is given as;

$$\mathrm{MSE}(\bar{y}_i) = \lambda\left[R_{KCi}^2 S_g^2 + 2B_P R_{KCi}S_g^2 + B_P^2 S_g^2 - 2R_{KCi}S_{gy} - 2B_P S_{gy} + S_y^2\right], i = 1, 2, \dots, 5 \quad (16)$$

The value of B_p can be obtained by using Poisson regression model.

3. Usman et al. (2021) proposed a regression estimator using Poisson regression model and is given as

$$\bar{y}_U = \bar{y} + b_P(\bar{G} - \bar{g}) \quad (17)$$

The mean square of $\bar{y}_U$ is given as;

$$\mathrm{MSE}(\bar{y}_U) = \lambda\left[S_y^2 - 2B_P S_{gy} + B_P^2 S_g^2\right] \quad (18)$$

3. Proposed Estimators

In this section, a new efficient class of ratio-cum-regression type estimator for estimation the population mean under *SRSWOR* based on Poisson regression method is proposed. Some members of the family of estimators are written in Table 1. Expressions for bias and *MSE* are obtained up to first degree of approximation.

$$Z_{Pi} = \bar{y}\left\{\left(\frac{\alpha_i\bar{G} + \beta_i}{\alpha_i\bar{g} + \beta_i}\right)^{\frac{\alpha_i\bar{G}}{\alpha_i\bar{G} + \beta_i}} + b_P\left(1 - \frac{\bar{g}}{\bar{G}}\right)\right\} \quad (19)$$

Where, $\alpha_i \neq 0$ and β_i may be any constant or functions of some known parameters of auxiliary variable G, which are determined such that the *MSE* of $Z_{Pi}, i = 1, 2, \dots, 15$ is minimum.

Table1: Some members of the proposed family of estimators using different values of α and β

S.No.	Values		Estimators
	α_i	β_i	
1	$\alpha_1 = 1$	$\beta_1 = S_g$	$Z_{P1} = \bar{y}\left\{\left(\frac{\alpha_1\bar{G}+\beta_1}{\alpha_1\bar{g}+\beta_1}\right)^{\frac{\alpha_1\bar{G}}{\alpha_1\bar{G}+\beta_1}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
2	$\alpha_2 = C_g$	$\beta_2 = S_g$	$Z_{P2} = \bar{y}\left\{\left(\frac{\alpha_2\bar{G}+\beta_2}{\alpha_2\bar{g}+\beta_2}\right)^{\frac{\alpha_2\bar{G}}{\alpha_2\bar{G}+\beta_2}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
3	$\alpha_3 = \rho_{gy}$	$\beta_3 = S_g$	$Z_{P3} = \bar{y}\left\{\left(\frac{\alpha_3\bar{G}+\beta_3}{\alpha_3\bar{g}+\beta_3}\right)^{\frac{\alpha_3\bar{G}}{\alpha_3\bar{G}+\beta_3}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
4	$\alpha_4 = 1$	$\beta_4 = fS_g$	$Z_{P4} = \bar{y}\left\{\left(\frac{\alpha_4\bar{G}+\beta_4}{\alpha_4\bar{g}+\beta_4}\right)^{\frac{\alpha_4\bar{G}}{\alpha_4\bar{G}+\beta_4}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
5	$\alpha_5 = \rho_{gy}$	$\beta_5 = fS_g$	$Z_{P5} = \bar{y}\left\{\left(\frac{\alpha_5\bar{G}+\beta_5}{\alpha_5\bar{g}+\beta_5}\right)^{\frac{\alpha_5\bar{G}}{\alpha_5\bar{G}+\beta_5}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
6	$\alpha_6 = \rho_{gy}C_g$	$\beta_6 = fS_g$	$Z_{P6} = \bar{y}\left\{\left(\frac{\alpha_6\bar{G}+\beta_6}{\alpha_6\bar{g}+\beta_6}\right)^{\frac{\alpha_6\bar{G}}{\alpha_6\bar{G}+\beta_6}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
7	$\alpha_7 = 1$	$\beta_7 = f(1-f)S_g$	$Z_{P7} = \bar{y}\left\{\left(\frac{\alpha_7\bar{G}+\beta_7}{\alpha_7\bar{g}+\beta_7}\right)^{\frac{\alpha_7\bar{G}}{\alpha_7\bar{G}+\beta_7}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
8	$\alpha_8 = 1$	$\beta_8 = 2f(1-f)S_g$	$Z_{P8} = \bar{y}\left\{\left(\frac{\alpha_8\bar{G}+\beta_8}{\alpha_8\bar{g}+\beta_8}\right)^{\frac{\alpha_8\bar{G}}{\alpha_8\bar{G}+\beta_8}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
9	$\alpha_9 = 1$	$\beta_9 = \left(\frac{f}{1-f}\right)S_g$	$Z_{P9} = \bar{y}\left\{\left(\frac{\alpha_9\bar{G}+\beta_9}{\alpha_9\bar{g}+\beta_9}\right)^{\frac{\alpha_9\bar{G}}{\alpha_9\bar{G}+\beta_9}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
10	$\alpha_{10} = 1$	$\beta_{10} = \left(\frac{f}{1+f}\right)S_g$	$Z_{P10} = \bar{y}\left\{\left(\frac{\alpha_{10}\bar{G}+\beta_{10}}{\alpha_{10}\bar{g}+\beta_{10}}\right)^{\frac{\alpha_{10}\bar{G}}{\alpha_{10}\bar{G}+\beta_{10}}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
11	$\alpha_{11} = 1$	$\beta_{11} = \left(\frac{f}{1+2f}\right)S_g$	$Z_{P11} = \bar{y}\left\{\left(\frac{\alpha_{11}\bar{G}+\beta_{11}}{\alpha_{11}\bar{g}+\beta_{11}}\right)^{\frac{\alpha_{11}\bar{G}}{\alpha_{11}\bar{G}+\beta_{11}}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
12	$\alpha_{12} = 1$	$\beta_{12} = \left(\frac{2f}{1+2f}\right)S_g$	$Z_{P12} = \bar{y}\left\{\left(\frac{\alpha_{12}\bar{G}+\beta_{12}}{\alpha_{12}\bar{g}+\beta_{12}}\right)^{\frac{\alpha_{12}\bar{G}}{\alpha_{12}\bar{G}+\beta_{12}}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
13	$\alpha_{13} = C_g$	$\beta_{13} = \left(\frac{2f}{1-f}\right)S_g$	$Z_{P13} = \bar{y}\left\{\left(\frac{\alpha_{13}\bar{G}+\beta_{13}}{\alpha_{13}\bar{g}+\beta_{13}}\right)^{\frac{\alpha_{13}\bar{G}}{\alpha_{13}\bar{G}+\beta_{13}}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
14	$\alpha_{14} = \rho_{gy}$	$\beta_{14} = \left(\frac{2f}{1-f}\right)S_g$	$Z_{P14} = \bar{y}\left\{\left(\frac{\alpha_{14}\bar{G}+\beta_{14}}{\alpha_{14}\bar{g}+\beta_{14}}\right)^{\frac{\alpha_{14}\bar{G}}{\alpha_{14}\bar{G}+\beta_{14}}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$
10	$\alpha_{15} = 1$	$\beta_{15} = \left(\frac{1-2f}{1+2f}\right)S_g$	$Z_{P15} = \bar{y}\left\{\left(\frac{\alpha_{15}\bar{G}+\beta_{15}}{\alpha_{15}\bar{g}+\beta_{15}}\right)^{\frac{\alpha_{15}\bar{G}}{\alpha_{15}\bar{G}+\beta_{15}}} + b_P\left(1-\frac{\bar{g}}{\bar{G}}\right)\right\}$

The Bias of the proposed estimator $Z_{Pi}, i = 1, 2, \ldots, 15$ up to first order of approximation is given as,

$$Bias(Z_{Pi}) = \lambda \bar{Y}\left[\frac{\varphi_i^3(\varphi_i + 1)}{2} C_g^2 - (\varphi_i^2 + B_P)C_{gy}\right] \quad (20)$$

*The MSE*of the proposed estimator $Z_{Pi}, i = 1, 2, \ldots, 15$ up to first order of approximation is given as

$$MSE(Z_{Pi}) = \lambda\left[S_y^2 + R^2(\varphi_i^2 + B_P)^2 S_g^2 - 2R(\varphi_i^2 + B_P)S_{gy}\right] \quad (21)$$

Where:

$$\varphi_i = \frac{\alpha_i \bar{G}}{\alpha_i \bar{G} + \beta_i} \quad , \mathrm{i} = 1, 2, 3, , 15$$

$$\varphi_1 = \frac{\bar{G}}{\bar{G} + S_g}, \quad \varphi_2 = \frac{C_g \bar{G}}{C_g \bar{G} + S_g}, \quad \varphi_3 = \frac{\rho_{gy}\bar{G}}{\rho_{gy}\bar{G} + S_g}, \quad \varphi_4 = \frac{\bar{G}}{\bar{G} + f S_g},$$

$$\varphi_5 = \frac{\rho_{gy}\bar{G}}{\rho_{gy}\bar{G} + f S_g}, \quad \varphi_6 = \frac{\rho_{gy} C_g \bar{G}}{\rho_{gy} C_g \bar{G} + f S_g}, \quad \varphi_7 = \frac{\bar{G}}{\bar{G} + f(1-f)S_g}, \quad \varphi_8 = \frac{\bar{G}}{\bar{G} + 2f(1-f)S_g},$$

$$\varphi_9 = \frac{\bar{G}}{\bar{G} + \left(\frac{f}{1-f}\right)S_g}, \quad \varphi_{10} = \frac{\bar{G}}{\bar{G} + \left(\frac{f}{1-f}\right)S_g}, \quad \varphi_{11} = \frac{\bar{G}}{\bar{G} + \left(\frac{f}{1+2f}\right)S_g}, \quad \varphi_{12} = \frac{\bar{G}}{\bar{G} + \left(\frac{2f}{1+2f}\right)S_g},$$

$$\varphi_{13} = \frac{C_g \bar{G}}{C_g \bar{G} + \left(\frac{2f}{1-f}\right)S_g}, \quad \varphi_{14} = \frac{\rho_{gy}\bar{G}}{\rho_{gy}\bar{G} + \left(\frac{2f}{1-f}\right)S_g}, \quad \varphi_{15} = \frac{\bar{G}}{\bar{G} + \left(\frac{1-2f}{1+2f}\right)S_g}$$

For the detailed derivation of the expression of Bias and *MSE* of the proposed estimator $Z_{Pi}, i = 1, 2, \ldots, 15$ one can see the Appendix-A.

4. Efficiency Comparison:

In this section, we present the comparison of the proposed estimators $(Z_{Pi}, \mathrm{i} = 1,2, \ldots, 15)$with Koc (2021) $(\bar{y}_\mathrm{i}, \mathrm{i} = 1, 2, \ldots, 5)$ and Usman et al. (2021)$(\bar{y}_\mathrm{U})$

4.1. Comparing the MSE of proposed estimator $Z_{Pi}, \mathbf{i = 1, 2, \ldots, 15}$with Koc (2021) $\bar{y}_\mathbf{i}, \mathbf{i = 1, 2, \ldots, 5}$

$Z_{Pi}, \mathrm{i} = 1,2, \ldots, 15$ Perform better than $\bar{y}_\mathrm{i}, \mathrm{i} = 1, 2, \ldots, 5$ if

$$MSE(\bar{y}_\mathrm{i}) > MSE(Z_{Pi})$$

$$\lambda\left[R_{KCi}^2 S_g^2 + 2B_P R_{KCi} S_g^2 + B_P^2 S_g^2 - 2R_{KCi} S_{gy} - 2B_P S_{gy} + S_y^2\right] > \lambda\left[S_y^2 + R^2(\varphi_i^2 + B_P)^2 S_g^2 - 2R(\varphi_i^2 + B_P)S_{gy}\right]$$

$$\{R_{KCi}^2 - R^2(\varphi_i^2 + B_P)^2 + 2B_P R_{KCi} + B_P^2\}S_g^2 - 2\{R_{KCi} - R(\varphi_i^2 + B_P) + B_P\}S_{gy} > 0$$

4.1. Comparing the MSE of proposed estimator $Z_{Pi}, \mathbf{i = 1, 2, \ldots, 15}$with Usman et al. (2021) $\bar{y}_\mathbf{U}$

$Z_{Pi}, \mathrm{i} = 1,2, \ldots, 15$ Perform better than $\bar{y}_\mathrm{U}$ if

$$MSE(\bar{y}_\mathrm{U}) > MSE(Z_{Pi})$$

$$\lambda\left[S_y^2 - 2B_P S_{gy} + B_P^2 S_g^2\right] > \lambda\left[S_y^2 + R^2(\varphi_i^2 + B_P)^2 S_g^2 - 2R(\varphi_i^2 + B_P)S_{gy}\right]$$

$$\{B_P^2 - R^2(\varphi_i^2 + B_P)^2\}S_g^2 + 2\{R(\varphi_i^2 + B_P) - B_P\}S_{gy} > 0$$

5. Empirical Study:

To examine the performance of the proposed class of estimator $Z_{Pi}, i = 1, 2, \ldots, 15$ over the other well-known estimators, two real data sets have been considered. The descriptions of the populations along with the values of various parameters are listed in Table 2 and Table 3.

Population: 1

We consider the dataset collected between 2006 and 2010 from the Afyon Respiratory Disease Hospital and the Afyon Environmental Department Air Pollution Unit, which was used by Koc (2021). The number of patients admitted to the hospital on a weekly basis was taken as the dependent variable Y, and PM10 was taken as the explanatory variable G.

Table: 2The descriptive statistics of population 1 are

N =213	$n = 40$	$\bar{Y} = 4.676056338$	$\bar{G} = 116.1915$
C_y =0.633555159	$C_g = 0.439846$	$\beta_{2(g)} = -0.6908$	$S_y = 2.962539616$
S_g =51.10636	$S_{gy} = 53.51706$	$\rho_{gy} = 0.353$	$B_l = 0.02049002$
B_P =0.004			

Population: 2

We consider the dataset obtained from TUIK of 81 provinces in 2019 used by Koc (2021). The number of people who died due to traffic accidents was taken as a dependent variable Y, and the number of motor vehicles was taken as explanatory variable G.

Table: 3 The descriptive statistics of population 2 are

$N = 81$	$n = 20$	$\bar{Y} = 82.40740741$	$\bar{G} = 274308$
$C_y = 1.025504149$	$C_g = 1.90331$	$\beta_{2(g)} = 35.77509$	$S_y = 84.50913823$
$S_g = 522093$	$S_{gy} = 38124520.6$	$\rho_{gy} = 0.864077723$	$B_l = 0.02049002$
$B_P = 6.131E-07$			

Table: 4 Mean square error of the existing and proposed estimators for population 1 and population 2

Estimators	**Population 1**	**Population 2**
$\bar{y}_1$	0.18587	334.73926
$\bar{y}_2$	0.18549	334.73236
$\bar{y}_3$	0.18648	334.60966
$\bar{y}_4$	0.18643	334.73906
$\bar{y}_5$	0.18728	334.67116
$\bar{y}_U$	0.17037	267.16300
Z_{P1}	0.15599	179.62946
Z_{P2}	0.161537	111.18094
Z_{P3}	0.164036	193.63367
Z_{P4}	0.166353	68.14308
Z_{P5}	0.157918	70.09819
Z_{P6}	0.156482	86.12029
Z_{P7}	0.1679	74.05904
Z_{P8}	0.162312	82.08443
Z_{P9}	0.164673	75.05373
Z_{P10}	0.167644	71.68617
Z_{P11}	0.168662	79.98185
Z_{P12}	0.163265	75.44276
Z_{P13}	0.156066	77.54750
Z_{P14}	0.15685	150.51009
Z_{P15}	0.159036	76.67033

Table: 5The PRE of the proposed estimators with the existing estimators for population 1

Estimators	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	$\bar{y}_5$	$\bar{y}_U$
Z_{P1}	119.15562	118.91125	119.54637	119.51546	120.05664	109.21617
Z_{P2}	115.06181	114.82584	115.43913	115.40929	115.93188	105.46385
Z_{P3}	113.30824	113.07587	113.67982	113.65044	114.16506	103.85656
Z_{P4}	111.73480	111.50566	112.10122	112.07224	112.57972	102.41437
Z_{P5}	117.69937	117.45799	118.08535	118.05482	118.58939	107.88140
Z_{P6}	118.78249	118.53890	119.17202	119.14122	119.68070	108.87417
Z_{P7}	110.70330	110.47627	111.06634	111.03763	111.54042	101.46891
Z_{P8}	114.51595	114.28111	114.89149	114.86179	115.38190	104.96353
Z_{P9}	112.87475	112.64326	113.24490	113.21563	113.72828	103.45922
Z_{P10}	110.87500	110.64762	111.23859	111.20984	111.71341	101.62628
Z_{P11}	110.20446	109.97846	110.56586	110.53728	111.03780	101.01168
Z_{P12}	113.84959	113.61611	114.22294	114.19342	114.71050	104.35275
Z_{P13}	119.09454	118.85030	119.48509	119.45420	119.99510	109.16019
Z_{P14}	118.50229	118.25927	118.89090	118.86017	119.39838	108.61734
Z_{P15}	116.87050	116.63082	117.25376	117.22345	117.75425	107.12167

Table: 6 The PRE of the proposed estimators with the existing estimators for population 2

Estimators	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	$\bar{y}_5$	$\bar{y}_U$
Z_{P1}	186.34986	186.34603	186.27772	186.34976	186.31196	148.73006
Z_{P2}	301.07612	301.06991	300.95956	301.07594	301.01487	240.29569
Z_{P3}	172.87244	172.86888	172.80552	172.87234	172.83728	137.97342
Z_{P4}	491.23001	491.21988	491.03983	491.22972	491.13007	392.06183
Z_{P5}	477.52910	477.51926	477.34423	477.52882	477.43196	381.12682
Z_{P6}	388.68803	388.68002	388.53755	388.68780	388.60896	310.22074
Z_{P7}	451.98973	451.98042	451.81475	451.98947	451.89778	360.74327
Z_{P8}	407.79872	407.79032	407.64084	407.79849	407.71576	325.47342
Z_{P9}	445.99949	445.99030	445.82683	445.99924	445.90876	355.96233
Z_{P10}	466.95095	466.94133	466.77018	466.95069	466.85596	372.68416
Z_{P11}	418.51902	418.51040	418.35699	418.51878	418.43388	334.02954
Z_{P12}	443.69964	443.69050	443.52787	443.69939	443.60938	354.12676
Z_{P13}	431.65706	431.64816	431.48994	431.65681	431.56924	344.51530
Z_{P14}	222.40320	222.39862	222.31710	222.40307	222.35796	177.50505
Z_{P15}	436.59556	436.58656	436.42653	436.59531	436.50674	348.45683

Table 4 shows that our proposed estimators Z_{Pi} , $i = 1,2,\dots,15$ have minimum MSE than Koc (2021) $\bar{y}_i$ $i = 1,2,\dots,5$ and Usman et al. (2021)$\bar{y}_U$ for both population 1 and population 2.

Table 5 presents the efficiency comparison based on percent relative efficiency of our proposed estimators Z_{Pi} , $i = 1,2,\dots,15$ with $\bar{y}_i$ $i = 1,2,\dots,5$ and $\bar{y}_U$ using population 1. It can be seen that our proposed estimators are more efficient than $\bar{y}_i$ $i = 1,2,\dots,5$ and $\bar{y}_U$. Among all the proposed estimators we can also see from table 5 that the proposed estimator Z_{P1} is more efficient using population 1.

Table 6 presents the efficiency comparison based on percent relative efficiency of our proposed estimators Z_{Pi} , $i = 1,2,\dots,15$ with $\bar{y}_i$ $i = 1,2,\dots,5$ and $\bar{y}_U$ using population 2. It can be seen that our proposed estimators are more efficient than $\bar{y}_i$ $i = 1,2,\dots,5$ and $\bar{y}_U$. Among all the proposed estimators we can also see from Table 6 that the proposed estimator Z_{P4} is more efficient using population 2.

Also, from Table 5 and Table 6 we see that our proposed estimators Z_{Pi} , $i = 1,2,\dots,15$ perform much better in population 2 than population 1.

6. Conclusion:

In this paper, we have proposed a class of estimators Z_{Pi}, $i = 1,2,\ldots,15$ by using simple random sampling for estimation of population mean based on the Poisson regression model. The proposed estimator was found to generate more efficient results than estimators suggested by Koc (2021) $\bar{y}_i$ $i = 1, 2, \ldots, 5$and Usman (2021)$\bar{y}_U$. A comparative study of the proposed estimators with the existing estimators has also been presented. To evaluate the benefits of the proposed estimator over others, a real data set was investigated in support of the current study, and it is clear that the proposed class of estimators is more competent as compared to existing estimators in terms of percent relative efficiency, as all values exceed 100. As a result, for count data analysis, we strongly indorse choosing proposed class of estimators instead of existing estimator.

Acknowledgement:

The authors would like to thank the Chief editor and anonymous referees for their constructive suggestions which led to improvement over an earlier version of this manuscript.

References

1. Abid, M., Abbas, N., Nazir, H. Z. and Lin, Z. (2016). Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters, *Revista Colombiana de Estadistica*, 39(1), 63–79.
2. Ali, N., Ahmad, I., Hanif, M.and Shahzad, U. (2021). Robust-regression type estimators for improving mean estimation of sensitive variables by using auxiliary information, *Communications in Statistics-theory and Methods*, 50(4), 979–992.
3. Cameron, A. C. and Trivedi, P. K. (1998). Regression Analysis of Count Data, *Cambridge University Press, Cambridge*.
4. Kadilar, C., Candan, M. and Cingi, H. (2007). Ratio estimators using robust regression, *Hacettepe Journal of Mathematics and Statistics*, 36(2), 181–188.
5. Kadilar,C. and Cingi, H. (2004). Ratio estimators in simple random sampling, *Applied Mathematics and Computation*, 151(3), 893–902.
6. Koc, H. (2021). Ratio-type estimators for improving mean estimation using Poisson regression method, *Communications in Statistics -theory and Methods*, 50(20), 4685–4691.
7. Koyuncu, N. (2012). Efficient estimators of population mean using auxiliary attributes, *Applied Mathematics and Computation*, 218(22), 10900–10905.
8. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2006). Introduction to Linear Regression Analysis, *John Wiley and Sons, Hoboken, 4th edition*.
9. Oral, E. and Kadilar, C. (2011a). Improved ratio estimators via modified maximum likelihood, *Pakistan Journal of Statistics*, 27(3), 269–282.
10.Oral, E. and Kadilar, C. (2011b). Robust ratio-type estimators in simple random sampling, *Journal of the Korean Surgical Society*, 40(4),457–467.
11. Shahzad, U., Perri, P. F. and Hanif, M. (2019). A new class of ratio type estimators for improving mean estimation of non-sensitive and sensitive variables by using supplementary information, *Communications in Statistics - Simulation and Computation*, 48(9), 2566–2585.
12. Shahzad, U., Shahzadi, S., Afshan, N., Al-Noor, N.H., Alilah, D. A., Hanif, M. and Anas, M.M. (2021). Poisson Regression-Based Mean Estimator, *Mathematical Problems in Engineering*, https://doi.org/10.1155/2021/9769029.
13. Sisodia, B. V. S. and Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable, *Journal of the Indian Society of Agricultural Statistics*, 33(2), 13–18.
14. Upadhyaya, L. N. and Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean, *Biometrical Journal*, 41(5), 627–636.
15. Upadhyaya, L. N. Singh, H. P. and Vos, J. W. E. (1985). On the estimation of population means and ratios using supplementary information, *Statistical Neerlandica*, 39(3),309–318.
16. Zaman, T. and Bulut, H. (2019). Modified ratio estimators using robust regression methods, *Communications in Statistics- Theory and Methods*, 48(8), 2039–2048.
17. Zaman, T. and Bulut, H. (2020). Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling, *Communications in Statistics- theory and Methods*, 49(14), 3407–3420.
18. Zaman, T. (2019). Improvement of modified ratio estimators using robust regression method ratio estimators using robust regression methods, *Applied Mathematics and Computation*, 348, 627–631.

Appendix-A: A detailed derivation of Bias and MSE for the proposed class of estimators in the present study.

To obtain the bias and MSE of the proposed estimator $Z_{Pi}, i = 1, 2, \ldots, 15$, we write

$$\eta_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}} \ , \qquad and \quad \eta_1 = \frac{\bar{g} - \bar{G}}{\bar{G}}$$

Such that $E(\eta_0) = E(\eta_1) = 0$ and to first degree of approximation

$$E(\eta_o^2) = \lambda C_y^2 \ , \ E(\eta_1^2) = \lambda C_g^2 \ , \ E(\eta_0\eta_1) = \lambda C_{gy}$$

Now expressing the (19) in terms of η's, we have

$$Z_{Pi} = \bar{Y}(1+\eta_0)\left\{\left(\frac{\alpha_i\bar{G} + \beta_i}{\alpha_i\bar{G}(1+\eta_1) + \beta_i}\right)^{\frac{\alpha_i\bar{G}}{\alpha_i\bar{G}+\beta_i}} + b_P\left(1 - \frac{\bar{G}(1+\eta_1)}{\bar{G}}\right)\right\} \tag{22}$$

Now expanding the right-hand side of $Z_{Pi}, i = 1, 2, \ldots, 15$ at (22) and neglecting terms of η's having power greater than two we have

$$Z_{Pi} = \bar{Y}(1+\eta_0)\left\{\left(\frac{1}{1 + \frac{\alpha_i\bar{G}}{\alpha_i\bar{G}+\beta_i}\eta_1}\right)^{\frac{\alpha_i\bar{G}}{\alpha_i\bar{G}+\beta}} - b_P\eta_1\right\}$$

$$Z_{Pi} = \bar{Y}(1+\eta_0)[(1+\varphi_i\eta_1)^{-\varphi_i} - b_P\eta_1]$$

Where:

$$\varphi_i = \frac{\alpha_i\bar{G}}{\alpha_i\bar{G} + \beta_i} \ , \text{i} = 1, 2, 3, , 15$$

$$\varphi_1 = \frac{\bar{G}}{\bar{G} + S_g} \ , \qquad \varphi_2 = \frac{C_g\bar{G}}{C_g\bar{G} + S_g} \ , \qquad \varphi_3 = \frac{\rho_{gy}\bar{G}}{\rho_{gy}\bar{G} + S_g} \ , \qquad \varphi_4 = \frac{\bar{G}}{\bar{G} + fS_g} \ ,$$

$$\varphi_5 = \frac{\rho_{gy}\bar{G}}{\rho_{gy}\bar{G} + fS_g}, \qquad \varphi_6 = \frac{\rho_{gy}C_g\bar{G}}{\rho_{gy}C_g\bar{G} + fS_g} \ , \qquad \varphi_7 = \frac{\bar{G}}{\bar{G} + f(1-f)S_g} \ , \quad \varphi_8 = \frac{\bar{G}}{\bar{G} + 2f(1-f)S_g} \ ,$$

$$\varphi_9 = \frac{\bar{G}}{\bar{G} + \left(\frac{f}{1-f}\right)S_g}, \qquad \varphi_{10} = \frac{\bar{G}}{\bar{G} + \left(\frac{f}{1-f}\right)S_g}, \qquad \varphi_{11} = \frac{\bar{G}}{\bar{G} + \left(\frac{f}{1+2f}\right)S_g}, \qquad \varphi_{12} = \frac{\bar{G}}{\bar{G} + \left(\frac{2f}{1+2f}\right)S_g},$$

$$\varphi_{13} = \frac{C_g\bar{G}}{C_g\bar{G} + \left(\frac{2f}{1-f}\right)S_g}, \qquad \varphi_{14} = \frac{\rho_{gy}\bar{G}}{\rho_{gy}\bar{G} + \left(\frac{2f}{1-f}\right)S_g}, \qquad \varphi_{15} = \frac{\bar{G}}{\bar{G} + \left(\frac{1-2f}{1+2f}\right)S_g}$$

$$Z_{Pi} = \bar{Y}(1+\eta_0)\left[1 - \varphi_i^2\eta_1 + \frac{\varphi_i^3(\varphi_i+1)}{2}\eta_1^2 - b_P\eta_1\right]$$

$$Z_{Pi} = \bar{Y}\left[1 - \varphi_i^2\eta_1 + \frac{\varphi_i^3(\varphi_i+1)}{2}\eta_1^2 - b_P\eta_1 + \eta_0 - \varphi_i^2\eta_0\eta_1 - b_P\eta_0\eta_1\right]$$

$$Z_{Pi} - \bar{Y} = \bar{Y}\left[\eta_0 - \varphi_i^2\eta_1 + \frac{\varphi_i^3(\varphi_i+1)}{2}\eta_1^2 - b_P\eta_1 - \varphi_i^2\eta_0\eta_1 - b_P\eta_0\eta_1\right] \tag{23}$$

Taking expectation of both sides of (23) we get the bias of $Z_{Pi}, i = 1, 2, \ldots, 15$up to first order of approximation as

$$E(Z_{Pi}-\bar{Y})=\bar{Y}\left[E(\eta_0)-\varphi_i^2E(\eta_1)+\frac{\varphi_i^3(\varphi_i+1)}{2}E(\eta_1^2)-b_PE(\eta_1)-\varphi_i^2E(\eta_0\eta_1)-b_PE(\eta_0\eta_1)\right]$$

$$E(Z_{Pi}-\bar{Y})=\bar{Y}\left[E(\eta_0)-\varphi_i^2E(\eta_1)+\frac{\varphi_i^3(\varphi_i+1)}{2}E(\eta_1^2)-b_PE(\eta_1)-\varphi_i^2E(\eta_0\eta_1)-b_PE(\eta_0\eta_1)\right]$$

$$E(Z_{Pi}-\bar{Y})=\bar{Y}\left[\frac{\varphi_i^3(\varphi_i+1)}{2}E(\eta_1^2)-(\varphi_i^2+B_P)E(\eta_0\eta_1)\right]$$

$$Bias(Z_{Pi})=\lambda\bar{Y}\left[\frac{\varphi_i^3(\varphi_i+1)}{2}C_g^2-(\varphi_i^2+B_P)C_{gy}\right] \quad (24)$$

Squaring both sides (23) and neglecting terms of η'shaving power greater than two, we have

$$(Z_{Pi}-\bar{Y})^2=\bar{Y}^2\left[\eta_0-\varphi_i^2\eta_1+\frac{\varphi_i^3(\varphi_i+1)}{2}\eta_1^2-b_P\eta_1-\varphi_i^2\eta_0\eta_1-b_P\eta_0\eta_1\right]^2$$

$$(Z_{Pi}-\bar{Y})^2=\bar{Y}^2\left[\eta_0^2+\left(\varphi_i^4+b_p^2+2b_p\varphi_i^3\right)\eta_0^2-2\left(\varphi_i^2+b_p\right)\eta_0\eta_1\right] \quad (25)$$

Taking expectation of both sides (25), we get the MSE of the proposed class of estimator $Z_{Pi}, i=1,2,\ldots,15$ to the first degree of approximation.

$$E(Z_{Pi}-\bar{Y})^2=MSE(Z_{Pi})=\bar{Y}^2\left[E(\eta_0^2)+(\varphi_i^2+B_P)^2E(\eta_1^2)-2\left(\varphi_i^2+b_p\right)E(\eta_0\eta_1)\right]$$

$$MSE(Z_{Pi})=\lambda\left[S_y^2+R^2(\varphi_i^2+B_P)^2S_g^2-2R(\varphi_i^2+B_P)S_{gy}\right] \quad (26)$$