

On the Restricted Poisson Ridge Regression Estimator

Enas Gawdat Yehia

Department of Statistics, Faculty of Commerce, Al-Azhar University, Tafahna Al-Ashraf, Egypt

Email address:

enasyehia538.el@azhar.edu.eg

To cite this article:

Enas Gawdat Yehia. On the Restricted Poisson Ridge Regression Estimator. *Science Journal of Applied Mathematics and Statistics*. Vol. 9, No. 4, 2021, pp. 106-112. doi: 10.11648/j.sjams.20210904.12

Received: August 2, 2021; **Accepted:** August 19, 2021; **Published:** August 26, 2021

Abstract: For modeling count data, the Poisson regression model is widely used in which the response variable takes non-negative integer values. However, the presence of strong correlation between the explanatory variables causes the problem of multicollinearity. Due to multicollinearity, the variance of the maximum likelihood estimator (MLE) will be inflated causing the parameters estimation to become unstable. Multicollinearity can be tackled by using biased estimators such as the ridge estimator in order to minimize the estimated variance of the regression coefficients. An alternative approach is to specify exact linear restrictions on the parameters in addition to regression model. In this paper, the restricted Poisson ridge regression estimator (RPRRE) is suggested to handle multicollinearity in Poisson regression model with exact linear restrictions on the parameters. In addition, the conditions of superiority of the suggested estimator in comparison to some existing estimators are discussed based on the mean squared error (MSE) matrix criterion. Moreover, a simulation study and a real data application are provided to illustrate the theoretical results. The results indicate that the suggested estimator, RPRRE, outperforms the other existing estimators in terms of scalar mean squared error (SMSE). Therefore, it is recommended to use the RPRRE for the Poisson regression model when the problem of multicollinearity is present.

Keywords: Poisson Regression, Multicollinearity, Ridge Regression Estimator, Restricted Maximum Likelihood Estimator, Restricted Ridge Regression Estimator

1. Introduction

In a Poisson regression, the maximum likelihood estimator (MLE) can be affected by multicollinearity problem in which inflates the variance of the estimates, and hence, the estimate of the parameters will be unstable (Kibria et al [14]). To deal with multicollinearity, many biased estimators have been suggested in linear regression model such as, the ridge regression estimator (RRE) by Hoerl and Kennard [10], the Liu regression estimator (LRE) by Liu [15], the Liu-type regression estimator (LTRE) by Liu [16], the two-parameter estimator (TPE) by Özkale and Kaçiranlar [24], and the modified Jackknifed ridge regression estimator (MJRRE) by Batah et al [7]. Further, in Poisson regression model, Månsson and Shukur [20] suggested the Poisson ridge regression estimator (PRRE), Månsson et al [19] suggested the Poisson Liu regression estimator (PLRE), Türkan and Özel [30] suggested the modified Jackknifed Poisson ridge regression estimator (MJPRRE), Asar and Genç [6] suggested the two-parameter Poisson estimator (TPPE), and

Alkhateeb and Algamal [3] suggested the Jackknifed Poisson Liu-type regression estimator (JPLTRE). For more details on biased estimators, one can refer to Muniz and Kibria [22], Kibria et al [13], Inan and Erdogan [11], Asar [4], Kibria et al [14], Şiray et al [29], Alanaz and Algamal [1], Algamal [2], and Qasim et al [25].

Another technique for dealing with multicollinearity is to include exact linear restrictions for the parameters in addition to regression model. Duffy and Santner [9] presented the restricted maximum likelihood estimator (RMLE) in logistic regression model. Then, by combining the RMLE and the biased estimators, many restricted biased estimators were presented in the context of logistic regression model such as, the restricted logistic ridge regression estimator (RLRRE) by Saleh and Kibria [28], the restricted logistic Liu regression estimator (RLLRE) by Şiray et al [29], and the restricted logistic Liu-type regression estimator (RLLTRE) by Asar et al [5]. Recently, Månsson and Kibria [17] introduced both the unrestricted and restricted Poisson Liu regression estimators for the Poisson regression model.

This paper aims to suggest a new restricted estimator

named as, restricted Poisson ridge regression estimator (RPRRE) to handle the problem of multicollinearity in Poisson regression model by combining the RMLE and the PRRE, and compare the estimators considered in this paper with the suggested estimator through a simulation study and a real data application.

The rest of this paper is planned as follows. In Section 2, the Poisson regression model specification and estimation are given. In Section 3, the restricted Poisson ridge regression estimator is suggested and its statistical properties are explained. In Section 4, mean squared error (MSE) matrix comparisons of the suggested estimator, RPRRE, and some existing estimators are derived. In Section 5, a simulation study is conducted to examine the performance of the suggested estimator according to the scalar mean squared error (SMSE) criterion. In Section 6, a real data is analyzed to clarify some of theoretical results. Finally, the conclusion is provided in Section 7.

2. Poisson Regression Model Specification and Estimation

For analyzing count data, the Poisson regression model is the most widely employed. It assumes that the mean and the variance are equal which is known as equidispersion. Let $Y_i, i = 1, 2, \dots, n$ is the count random variable that has a Poisson distribution with probability mass function as follows:

$$P(Y_i = y_i; \mu_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, y_i = 0, 1, 2, \dots \quad (1)$$

where $\mu_i = \exp(\mathbf{x}_i' \beta)$ is the mean of the response variable Y_i , $\mathbf{x}_i = (\mathbf{x}_{i0}, \dots, \mathbf{x}_{ip})'$ is i^{th} row of X , which is an $n \times (p+1)$ matrix of p explanatory variables, and $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ is a $(p+1) \times 1$ vector of regression coefficients.

Using the maximum likelihood estimation method, one can estimate the Poisson regression coefficients by differentiating the log likelihood function with regard to β and solving them to zero.

From (1), the likelihood function is given by

$$L = \prod_{i=1}^n \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, \quad (2)$$

Then, the log likelihood function is given as follows:

$$\begin{aligned} \ell(\beta) &= \sum_{i=1}^n y_i \ln(\mu_i) - \mu_i - \ln(y_i!) \\ &= \sum_{i=1}^n y_i \mathbf{x}_i' \beta - \exp(\mathbf{x}_i' \beta) - \ln(y_i!). \end{aligned} \quad (3)$$

From (3), the first partial derivative with regard to β is given as follows:

$$\begin{aligned} \frac{\partial \ell(\beta)}{\partial \beta} &= \sum_{i=1}^n (y_i - \mu_i) \mathbf{x}_i \\ &= \sum_{i=1}^n (y_i - \exp(\mathbf{x}_i' \beta)) \mathbf{x}_i, \end{aligned} \quad (4)$$

Since Equation (4) is nonlinear in β , therefore, the iterative weighted least squares (IWLS) algorithm can be used to get

an appropriate solution. Then, the MLE of β can be obtained by

$$\hat{\beta}_{\text{MLE}} = G^{-1} X' \hat{W} \hat{u}, \quad (5)$$

where $G = X' \hat{W} X$, $\hat{W} = \text{diag}[\hat{\mu}_i]$, and \hat{u} is a column vector with i^{th} element equals $\ln(\hat{\mu}_i) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}$.

The statistical properties of MLE are

$$E(\hat{\beta}_{\text{MLE}}) = \beta, \quad (6)$$

$$\text{COV}(\hat{\beta}_{\text{MLE}}) = D(\hat{\beta}_{\text{MLE}}) = G^{-1}, \quad (7)$$

Consequently, the MSE and SMSE of $\hat{\beta}_{\text{MLE}}$ are given by

$$\begin{aligned} \text{MSE}(\hat{\beta}_{\text{MLE}}) &= D(\hat{\beta}_{\text{MLE}}) + B(\hat{\beta}_{\text{MLE}}) B'(\hat{\beta}_{\text{MLE}}) \\ &= G^{-1}, \end{aligned} \quad (8)$$

where $B(\hat{\beta}_{\text{MLE}})$ is the bias vector of $\hat{\beta}_{\text{MLE}}$.

$$\text{SMSE}(\hat{\beta}_{\text{MLE}}) = \text{tr}[\text{MSE}(\hat{\beta}_{\text{MLE}})] \quad (9)$$

$$= \text{tr}(G^{-1}) = \sum_{j=1}^p \frac{1}{\lambda_j},$$

where λ_j is the j^{th} eigenvalue of the matrix, G .

Since multicollinearity between the explanatory variables leads to inaccurate parameters estimation, and large variance for the estimates (Kibria et al [14]). Therefore, many other estimators have been suggested for Poisson regression to tackle this problem. One of the most important estimation approaches is to use biased estimators, among those, the Poisson ridge regression estimator (PRRE) by Månsson and Shukur [20].

The PRRE is defined as follows:

$$\begin{aligned} \hat{\beta}_{\text{PRRE}} &= (X' \hat{W} X + kI)^{-1} X' \hat{W} X \hat{\beta}_{\text{MLE}} \\ &= (G + kI)^{-1} G \hat{\beta}_{\text{MLE}} \\ &= Z_k \hat{\beta}_{\text{MLE}}, \end{aligned} \quad (10)$$

where $Z_k = (G + kI)^{-1} G$, and k is the biasing parameter known as the ridge parameter, $k \geq 0$.

The statistical properties of $\hat{\beta}_{\text{PRRE}}$ are given by

$$\begin{aligned} E(\hat{\beta}_{\text{PRRE}}) &= E(Z_k \hat{\beta}_{\text{MLE}}) \\ &= Z_k \beta, \end{aligned} \quad (11)$$

$$\begin{aligned} \text{COV}(\hat{\beta}_{\text{PRRE}}) &= D(\hat{\beta}_{\text{PRRE}}) \\ &= \text{COV}(Z_k \hat{\beta}_{\text{MLE}}) \\ &= Z_k G^{-1} Z_k' \\ &= (G + kI)^{-1} G (G + kI)^{-1}, \end{aligned} \quad (12)$$

Then, the bias vector, MSE, and SMSE of $\hat{\beta}_{\text{PRRE}}$ are given respectively by

$$B(\hat{\beta}_{\text{PRRE}}) = E(\hat{\beta}_{\text{PRRE}}) - \beta \quad (13)$$

$$\begin{aligned}
&= (Z_k - I) \beta = \delta_1, \text{ (say)} \\
\text{MSE}(\hat{\beta}_{\text{PRRE}}) &= D(\hat{\beta}_{\text{PRRE}}) + B(\hat{\beta}_{\text{PRRE}}) B'(\hat{\beta}_{\text{PRRE}}) \quad (14) \\
&= Z_k G^{-1} Z_k' + \delta_1 \delta_1' \\
&= (G + kI)^{-1} G (G + kI)^{-1} + \delta_1 \delta_1', \\
\text{SMSE}(\hat{\beta}_{\text{PRRE}}) &= \text{tr}[\text{MSE}(\hat{\beta}_{\text{PRRE}})] \quad (15) \\
&= \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^p \frac{\alpha_j}{(\lambda_j + k)^2},
\end{aligned}$$

where α_j is the j^{th} element of $Q'\beta$, and Q is an orthogonal matrix defined so that

$$Q'GQ = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_p); \lambda_1 \geq \dots \geq \lambda_p > 0.$$

An alternative technique to improve estimation due to multicollinearity is to include prior information for the parameters as in form of exact linear restrictions in addition to regression model. The resulting estimator is called a restricted estimator.

According to (1), assume that the following exact linear restriction is considered for the parameter vector, β

$$H\beta = h, \quad (16)$$

where H is a $q \times (p+1)$ known matrix, and h is a $q \times 1$ vector of known constants.

Then, the restricted maximum likelihood estimator (RMLE) according to Duffy and Santner [9] is as follows:

$$\hat{\beta}_{\text{RMLE}} = \hat{\beta}_{\text{MLE}} + G^{-1}H'(HG^{-1}H')^{-1}(h - H\hat{\beta}_{\text{MLE}}). \quad (17)$$

The statistical properties of $\hat{\beta}_{\text{RMLE}}$ are

$$E(\hat{\beta}_{\text{RMLE}}) = \beta + G^{-1}H'(HG^{-1}H')^{-1}(h - H\beta), \quad (18)$$

$$\begin{aligned}
\text{COV}(\hat{\beta}_{\text{RMLE}}) &= D(\hat{\beta}_{\text{RMLE}}) \quad (19) \\
&= G^{-1} - G^{-1}H'(HG^{-1}H')^{-1}HG^{-1} \\
&= A, \text{ (say)}
\end{aligned}$$

Thus, the bias vector, MSE, and SMSE of $\hat{\beta}_{\text{RMLE}}$ are given respectively by

$$\begin{aligned}
B(\hat{\beta}_{\text{RMLE}}) &= E(\hat{\beta}_{\text{RMLE}}) - \beta \quad (20) \\
&= G^{-1}H'(HG^{-1}H')^{-1}(h - H\beta) \\
&= \delta_2, \text{ (say)}
\end{aligned}$$

$$\begin{aligned}
\text{MSE}(\hat{\beta}_{\text{RMLE}}) &= D(\hat{\beta}_{\text{RMLE}}) + B(\hat{\beta}_{\text{RMLE}}) B'(\hat{\beta}_{\text{RMLE}}) \quad (21) \\
&= G^{-1} - G^{-1}H'(HG^{-1}H')^{-1}HG^{-1} + \delta_2 \delta_2' \\
&= A + \delta_2 \delta_2',
\end{aligned}$$

Following Şiray et al [29] and Månsson et al [18], the SMSE of $\hat{\beta}_{\text{RMLE}}$ is as follows:

$$\begin{aligned}
\text{SMSE}(\hat{\beta}_{\text{RMLE}}) &= \text{tr}(A + \delta_2 \delta_2') \quad (22) \\
&= \sum_{j=1}^p a_{jj} + \delta_j^2,
\end{aligned}$$

where a_{jj} is the j^{th} diagonal element of the matrix $Q'AQ$, and δ_j is the j^{th} element of the vector $\delta_2'Q$.

3. The Restricted Poisson Ridge Regression Estimator

In this section, following (10) and (17), the restricted Poisson ridge regression estimator (RPRRE) is suggested and is defined by the following form:

$$\begin{aligned}
\hat{\beta}_{\text{RPRRE}} &= (X'\hat{W}X + kI)^{-1}X'\hat{W}X\hat{\beta}_{\text{RMLE}} \quad (23) \\
&= (G + kI)^{-1}G\hat{\beta}_{\text{RMLE}} \\
&= Z_k\hat{\beta}_{\text{RMLE}},
\end{aligned}$$

where $Z_k = (G + kI)^{-1}G, k \geq 0$.

The statistical properties of $\hat{\beta}_{\text{RPRRE}}$ are given by

$$\begin{aligned}
E(\hat{\beta}_{\text{RPRRE}}) &= E(Z_k\hat{\beta}_{\text{RMLE}}) \quad (24) \\
&= Z_k\beta,
\end{aligned}$$

$$\begin{aligned}
\text{COV}(\hat{\beta}_{\text{RPRRE}}) &= D(\hat{\beta}_{\text{RPRRE}}) \quad (25) \\
&= \text{COV}(Z_k\hat{\beta}_{\text{RMLE}}) \\
&= Z_kAZ_k',
\end{aligned}$$

$$\begin{aligned}
B(\hat{\beta}_{\text{RPRRE}}) &= E(\hat{\beta}_{\text{RPRRE}}) - \beta \quad (26) \\
&= (Z_k - I)\beta = \delta_1,
\end{aligned}$$

Consequently, the MSE of $\hat{\beta}_{\text{RPRRE}}$ is given as

$$\begin{aligned}
\text{MSE}(\hat{\beta}_{\text{RPRRE}}) &= D(\hat{\beta}_{\text{RPRRE}}) + B(\hat{\beta}_{\text{RPRRE}}) B'(\hat{\beta}_{\text{RPRRE}}) \quad (27) \\
&= Z_kAZ_k' + \delta_1 \delta_1',
\end{aligned}$$

Following Najarian et al [23], the SMSE of $\hat{\beta}_{\text{RPRRE}}$ can be obtained by

$$\begin{aligned}
\text{SMSE}(\hat{\beta}_{\text{RPRRE}}) &= \text{tr}(Z_kAZ_k' + \delta_1 \delta_1') \quad (28) \\
&= \sum_{j=1}^p \frac{\lambda_j(\lambda_j + k - r_{jj})^2}{(\lambda_j + k)^4} + k^2 \sum_{j=1}^p \left[\frac{\alpha_j(\lambda_j + k - r_{jj})}{(\lambda_j + k)^2} \right]^2,
\end{aligned}$$

where $r_{jj} = \text{diag}(R)$, and $R = Q'H'(HG^{-1}H')^{-1}HQ$.

4. The Comparisons of the Estimators

The superiority of the suggested estimator, RPRRE, over the estimators MLE, PRRE, and RMLE is compared in this section using the MSE matrix criterion.

For two given estimators $\hat{\beta}_1$ and $\hat{\beta}_2$, in the sense of MSE matrix, the estimator $\hat{\beta}_2$ is superior to $\hat{\beta}_1$ if and only if

$$\Delta(\hat{\beta}_1, \hat{\beta}_2) = \text{MSE}(\hat{\beta}_1) - \text{MSE}(\hat{\beta}_2) \geq 0. \quad (29)$$

The following Lemmas will be used in the comparisons among the estimators:

Lemma 1: Assume that A is a $n \times n$ matrix, and B is a

$n \times n$ matrix such that $A > 0$ and $B \geq 0$. Then, $A + B > 0$. (Rao and Toutenburg [26]).

Lemma 2: Assume that M and N are two matrices where $M > 0$, and $N > 0$. Then, $M > N$ if and only if $\lambda_{\max}(NM^{-1}) < 1$. (Rao et al [27]).

4.1. RPRRE Versus MLE

From (8) and (27), the difference is computed as

$$\begin{aligned}\Delta_1 &= \text{MSE}(\hat{\beta}_{\text{MLE}}) - \text{MSE}(\hat{\beta}_{\text{RPRRE}}) \\ &= G^{-1} - [Z_k A Z'_k + \delta_1 \delta'_1] \\ &= M_1 - N_1,\end{aligned}\quad (30)$$

where $M_1 = G^{-1}$ and $N_1 = Z_k A Z'_k + \delta_1 \delta'_1$.

Theorem 4.1. *The RPRRE is superior to MLE if and only if the largest eigenvalue of $N_1 M_1^{-1}$, $\lambda_{\max}(N_1 M_1^{-1})$ be less than one.*

Proof. It is obvious to say that M_1 and $Z_k A Z'_k$ are positive definite matrices, and $\delta_1 \delta'_1$ is non-negative definite matrix. Also, N_1 is positive definite matrix using Lemma 1. Further, according to Lemma 2, if $\lambda_{\max}(N_1 M_1^{-1}) < 1$, then, $M_1 - N_1$ is a positive definite matrix where $\lambda_{\max}(N_1 M_1^{-1})$ is the largest eigenvalue of $N_1 M_1^{-1}$. Therefore, the RPRRE, is superior to MLE if and only if $\lambda_{\max}(N_1 M_1^{-1}) < 1$.

4.2. RPRRE Versus PRRE

From (14) and (27), the difference is computed as

$$\begin{aligned}\Delta_2 &= \text{MSE}(\hat{\beta}_{\text{PRRE}}) - \text{MSE}(\hat{\beta}_{\text{RPRRE}}) \\ &= Z_k G^{-1} Z'_k - Z_k A Z'_k \\ &= Z_k (G^{-1} - A) Z'_k \\ &= Z_k [G^{-1} H' (H G^{-1} H')^{-1} H G^{-1}] Z'_k \\ &= Z_k F Z'_k,\end{aligned}\quad (31)$$

where $F = G^{-1} H' (H G^{-1} H')^{-1} H G^{-1}$.

Theorem 4.2. *The RPRRE is always superior to PRRE.*

Proof. It is clear that $Z_k F Z'_k$ is positive definite. Hence, the RPRRE, is always superior to PRRE in the sense of MSE.

4.3. RPRRE Versus RMLE

From (21) and (27), the difference is computed as

$$\begin{aligned}\Delta_3 &= \text{MSE}(\hat{\beta}_{\text{RMLE}}) - \text{MSE}(\hat{\beta}_{\text{RPRRE}}) \\ &= A + \delta_2 \delta'_2 - [Z_k A Z'_k + \delta_1 \delta'_1] \\ &= G^{-1} - F + \delta_2 \delta'_2 - Z_k G^{-1} Z'_k + Z_k F Z'_k - \delta_1 \delta'_1 \\ &= (G^{-1} + Z_k F Z'_k + \delta_2 \delta'_2) - (F + Z_k G^{-1} Z'_k + \delta_1 \delta'_1) \\ &= M_2 - N_2,\end{aligned}\quad (32)$$

where $M_2 = G^{-1} + Z_k F Z'_k + \delta_2 \delta'_2$, and $N_2 = F + Z_k G^{-1} Z'_k + \delta_1 \delta'_1$.

Theorem 4.3. *The RPRRE is superior to RMLE if and only if the largest eigenvalue of $N_2 M_2^{-1}$, $\lambda_{\max}(N_2 M_2^{-1})$ be less*

than one.

Proof. Since $G^{-1} + Z_k F Z'_k$ and $F + Z_k G^{-1} Z'_k$ are positive definite, and $\delta_2 \delta'_2$ and $\delta_1 \delta'_1$ are non-negative definite matrices, then, N_2 is a positive definite matrix using Lemma 1. Further, it is obvious that $M_2 - N_2$ is positive definite matrix according to Lemma 2 if and only if $\lambda_{\max}(N_2 M_2^{-1}) < 1$, where $\lambda_{\max}(N_2 M_2^{-1})$ is the largest eigenvalue of $N_2 M_2^{-1}$. Therefore, the RPRRE, is superior to RMLE if and only if $\lambda_{\max}(N_2 M_2^{-1}) < 1$.

5. Simulation Study

In order to assess the performance of the suggested estimator, RPRRE over the MLE, PRRE, and RMLE by means of SMSE, a Monte Carlo simulation study is performed. The response variable of the Poisson regression model is generated by pseudo-random numbers from the Poisson distribution with $\mu_i = \exp(x'_i \beta)$, where $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ in which are chosen so that $\beta_0 = 0$, $\sum_{j=1}^p \beta_j^2 = 1$, $\beta_1 = \beta_2 = \dots = \beta_p$.

According to McDonald and Galarneau [21] and Kibria [12], the following formula is considered to generate the explanatory variables:

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}} w_{ij} + \rho w_{ip+1}, i = 1, 2, \dots, n, j = 1, 2, \dots, p, \quad (33)$$

where ρ^2 represents the coefficient of correlation between any two explanatory variables, and w_{ij} are independent pseudo-random numbers generated from the standard normal distribution. Four explanatory variables are generated by (33) with the values of ρ corresponding to 0.85, 0.95, and 0.99. The various values of sample size, n are used corresponding to $n = 20, 50$, and 100 . Further, for the biasing parameter k , a set of values has been chosen so that $k = 0.1, 0.3, 0.5, 0.7, 0.9$, and 1 . Moreover, the following restriction is considered:

$$H = [1, -2, 1, 0], h = [0].$$

The simulation is repeated 1000 times for each combination of n , ρ , and k and the simulated SMSE is computed by

$$\text{SMSE}(\hat{\beta}) = \frac{1}{1000} \sum_{r=1}^{1000} (\hat{\beta}_r - \beta)' (\hat{\beta}_r - \beta). \quad (34)$$

where $\hat{\beta}_r$ is any estimator used in the r^{th} replication.

The results of the simulation are given in Tables 1-3. According to Tables 1-3, it can be noticed that the estimated SMSE values of MLE, PRRE, RMLE, and RPRRE are increasing as the degree of correlation increases. In addition, with the increasing of sample size, the estimated SMSE values of all existing estimators are decreasing. Also, the increases in biasing parameter, k values decrease the estimated SMSE values of PRRE and RPRRE. In all cases, the MLE has the worst performance (having the largest SMSE value). Moreover, for all selected values of n, ρ , and k , the RPRRE has the best performance compared to MLE, PRRE, and RMLE since it has the least SMSE value. Therefore, the RPRRE can be used in practical applications to tackle multicollinearity in Poisson regression.

Table 1. The estimated SMSE values of the MLE, PRRE, RMLE, and RPRRE when $n = 20$.

ρ	Estimator	k					
		0.1	0.3	0.5	0.7	0.9	1
0.85	MLE	0.4711	0.4711	0.4711	0.4711	0.4711	0.4711
	PRRE	0.4561	0.4292	0.4058	0.3852	0.3669	0.3585
	RMLE	0.3875	0.3875	0.3875	0.3875	0.3875	0.3875
	RPRRE	0.3764	0.3563	0.3386	0.3229	0.3089	0.3024
0.95	MLE	1.2568	1.2568	1.2568	1.2568	1.2568	1.2568
	PRRE	1.1259	0.9324	0.7973	0.6980	0.6221	0.5904
	RMLE	0.9281	0.9281	0.9281	0.9281	0.9281	0.9281
	RPRRE	0.8406	0.7104	0.6185	0.5500	0.4969	0.4745
0.99	MLE	6.0979	6.0979	6.0979	6.0979	6.0979	6.0979
	PRRE	4.5889	3.1808	2.5208	2.1322	1.8696	1.7660
	RMLE	4.1389	4.1389	4.1389	4.1389	4.1389	4.1389
	RPRRE	3.3794	2.5836	2.1632	1.8928	1.6971	1.6167

Table 2. The estimated SMSE values of the MLE, PRRE, RMLE, and RPRRE when $n = 50$.

ρ	Estimator	k					
		0.1	0.3	0.5	0.7	0.9	1
0.85	MLE	0.0719	0.0719	0.0719	0.0719	0.0719	0.0719
	PRRE	0.0715	0.0709	0.0704	0.0698	0.0692	0.0689
	RMLE	0.0532	0.0532	0.0532	0.0532	0.0532	0.0532
	RPRRE	0.0529	0.0524	0.0519	0.0515	0.0510	0.0508
0.95	MLE	0.1831	0.1831	0.1831	0.1831	0.1831	0.1831
	PRRE	0.1812	0.1775	0.1739	0.1706	0.1673	0.1657
	RMLE	0.1251	0.1251	0.1251	0.1251	0.1251	0.1251
	RPRRE	0.1237	0.1210	0.1184	0.1159	0.1134	0.1124
0.99	MLE	0.8280	0.8280	0.8280	0.8280	0.8280	0.8280
	PRRE	0.7737	0.6841	0.6130	0.5549	0.5064	0.4849
	RMLE	0.7081	0.7081	0.7081	0.7081	0.7081	0.7081
	RPRRE	0.6559	0.5710	0.5048	0.4517	0.4079	0.3889

Table 3. The estimated SMSE values of the MLE, PRRE, RMLE, and RPRRE when $n = 100$.

ρ	Estimator	k					
		0.1	0.3	0.5	0.7	0.9	1
0.85	MLE	0.0357	0.0357	0.0357	0.0357	0.0357	0.0357
	PRRE	0.0355	0.0353	0.0350	0.0348	0.0345	0.0344
	RMLE	0.0262	0.0262	0.0262	0.0262	0.0262	0.0262
	RPRRE	0.0261	0.0259	0.0257	0.0255	0.0253	0.0252
0.95	MLE	0.0916	0.0916	0.0916	0.0916	0.0916	0.0916
	PRRE	0.0908	0.0893	0.0878	0.0864	0.0849	0.0842
	RMLE	0.0648	0.0648	0.0648	0.0648	0.0648	0.0648
	RPRRE	0.0643	0.0633	0.0623	0.0613	0.0604	0.0599
0.99	MLE	0.4180	0.4180	0.4180	0.4180	0.4180	0.4180
	PRRE	0.4110	0.3977	0.3851	0.3733	0.3621	0.3567
	RMLE	0.2843	0.2843	0.2843	0.2843	0.2843	0.2843
	RPRRE	0.2791	0.2692	0.2599	0.2512	0.2430	0.2392

6. Real Data Application

In this section, an application of real data of FIFA World Cup in 2018 due to <https://www.fifa.com> is considered. The data set involves 32 teams in which the response variable is defined as the number of won matches with 5 explanatory variables include the number of goals scored (X_1), the number of goals conceded (X_2), the number of clean sheets (X_3), the number of shoots (X_4), and the number of assists (X_5).

First, for checking the adequacy of fit the Poisson regression model to this data, the residual deviance test is used. The result of residual deviance is 11.121 with 27 degrees of freedom and the p – value is 0.9970. It is clear that data set is well fitted to Poisson regression model.

Also, the condition number (CN) is used to check for multicollinearity among the explanatory variables as follows:

$$CN = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}, \quad (35)$$

where λ_{max} and λ_{min} are the maximum and minimum eigenvalues of the matrix, G respectively. The eigenvalues of the matrix, G are equal to 424434.6152, 679.5312, 390.9653, 213.8038, and 28.3011 and the condition number is 122.4627 indicating that there is a high degree of multicollinearity in the data (Belsley et al [8]). The following correlation matrix shows the bivariate correlation between the explanatory variables as:

$$\begin{bmatrix} 1 & 0.34 & 0.63 & 0.81 & 0.78 \\ 0.34 & 1 & -0.27 & 0.20 & 0.18 \\ 0.63 & -0.27 & 1 & 0.59 & 0.57 \\ 0.81 & 0.20 & 0.59 & 1 & 0.99 \\ 0.78 & 0.18 & 0.57 & 0.99 & 1 \end{bmatrix},$$

which one can see that there are some high bivariate correlations, namely, 0.99, 0.81, and 0.78.

In addition, the restriction $H = [1, -2, 1, 0, 0]$ with $h = [0]$ is used. Moreover, the estimated SMSE values of the estimators MLE, PRRE, RMLE, and RPRRE are given in Table 4, and the coefficients of Poisson regression and corresponding standard errors (SE) of this estimators are given in Table 5 for different values of k .

From Table 4, it is obvious that the estimator, RPRRE, has the lowest SMSE value for all different values of k , while the largest is obtained by the MLE which suffers from multicollinearity.

Table 4. The estimated SMSE values of the estimators.

k	Estimator			
	MLE	PRRE	RMLE	RPRRE
0.1	0.0440	0.0438	0.0386	0.0367
0.3	0.0440	0.0433	0.0386	0.0362
0.5	0.0440	0.0428	0.0386	0.0358
0.7	0.0440	0.0424	0.0386	0.0354
0.9	0.0440	0.0419	0.0386	0.0349
1	0.0440	0.0417	0.0386	0.0347

Additionally, from Table 5, the standard errors of all coefficients decrease as k increases for all estimators and the RPRRE has the lowest values confirming its superiority over MLE, PRRE, and RMLE.

Table 5. The estimated coefficients, corresponding standard errors (SE) of the estimators.

Estimator	k	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$		$\hat{\beta}_4$		$\hat{\beta}_5$	
		Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
MLE		0.1405	0.0597	-0.1507	0.0820	0.1467	0.1762	0.0145	0.0336	-0.0167	0.0396
PRRE	0.1	0.1405	0.0597	-0.1507	0.0818	0.1464	0.1756	0.0145	0.0336	-0.0167	0.0396
	0.3	0.1405	0.0595	-0.1507	0.0815	0.1459	0.1743	0.0145	0.0335	-0.0167	0.0395
	0.5	0.1404	0.0593	-0.1506	0.0812	0.1454	0.1732	0.0145	0.0334	-0.0167	0.0394
	0.7	0.1404	0.0592	-0.1506	0.0809	0.1450	0.1720	0.0145	0.0333	-0.0167	0.0394
	0.9	0.1404	0.0591	-0.1505	0.0806	0.1445	0.1709	0.0146	0.0333	-0.0167	0.0393
RMLE	1	0.1404	0.0589	-0.1505	0.0805	0.1442	0.1703	0.0146	0.0332	-0.0167	0.0392
		0.0786	0.0573	0.0137	0.0684	-0.0511	0.1676	-0.0054	0.0332	0.0129	0.0388
RPRRE	0.1	0.0785	0.0571	0.0138	0.0681	-0.0509	0.1667	-0.0054	0.0331	0.0129	0.0387
	0.3	0.0784	0.0570	0.0139	0.0678	-0.0506	0.1658	-0.0055	0.0330	0.0129	0.0387
	0.5	0.0783	0.0568	0.0140	0.0674	-0.0502	0.1646	-0.0055	0.0329	0.0130	0.0386
	0.7	0.0782	0.0567	0.0141	0.0671	-0.0498	0.1635	-0.0055	0.0329	0.0130	0.0385
	0.9	0.0781	0.0565	0.0142	0.0667	-0.0495	0.1623	-0.0055	0.0328	0.0130	0.0384
	1	0.0780	0.0564	0.0141	0.0665	-0.0493	0.1618	-0.0055	0.0328	0.0130	0.0384

7. Conclusion

In this paper, the RPRRE was suggested for Poisson regression model with exact linear restrictions on the parameters to tackle the problem of multicollinearity. Further, based on the MSE matrix criterion, the conditions for superiority of the suggested estimator, RPRRE over the estimators MLE, PRRE, and RMLE are given. Moreover, a Monte Carlo simulation study and a real data application were conducted to evaluate the performance of the RPRRE with the MLE, PRRE, and RMLE according to SMSE criterion. The results indicate that the RPRRE is superior to the MLE, PRRE, and RMLE in the SMSE sense. So, the RPRRE is a better alternative to MLE, PRRE, and RMLE in Poisson regression when multicollinearity is present. Therefore, for future work, it is recommended to use the RPRRE in many applications to tackle the problem of multicollinearity.

References

- [1] Alanaz, M. M., & Algarnal, Z. Y. (2018). Proposed methods in estimating the ridge regression parameter in Poisson regression model. *Electronic Journal of Applied Statistical Analysis*, 11 (2): 506-515.
- [2] Algarnal, Z. Y. (2018). Biased estimators in Poisson regression model in the presence of multicollinearity: A subject review. *Al-Qadisiyah Journal for Administrative and Economic Sciences*, 20 (1): 36-43.
- [3] Alkhateeb, A. N., & Algarnal, Z. Y. (2020). Jackknifed Liu-type estimator in Poisson regression model. *Journal of the Iranian Statistical Society*, 19 (1): 21-37.
- [4] Asar, Y. (2015). Some new methods to solve multicollinearity in logistic regression. *Communications in Statistics-Simulation and Computation*, 46 (4): 2576-2586.
- [5] Asar, Y., Erişoglu, M., & Arashi, M. (2017). Developing a restricted two-parameter Liu-type estimator: A comparison of restricted estimators in the binary logistic regression model. *Communications in Statistics-Theory and Methods*, 46 (14): 6864-6873.
- [6] Asar, Y., & Genç, A. (2018). A new two-parameter estimator for the Poisson regression model. *Iranian Journal of Science and Technology, Transactions A: Science*, 42 (2): 793-803.
- [7] Batah, F. S. M., Ramanathan, T. V., & Gore, S. D. (2008). The efficiency of modified Jackknife and ridge type regression estimators: A comparison. *Surveys in Mathematics and its Applications*, 3: 111-122.
- [8] Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons, New York, Inc.
- [9] Duffy, D. E., & Santner, T. J. (1989). On the small sample properties of norm-restricted maximum likelihood estimators for logistic regression models. *Communications in Statistics-Theory and Methods*, 18 (3): 959-980.

- [10] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12 (1): 55-67.
- [11] Inan, D., & Erdogan, B. E. (2013). Liu-type logistic estimator. *Communications in Statistics-Simulation and Computation*, 42 (7): 1578-1586.
- [12] Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation*, 32 (2): 419-435.
- [13] Kibria, B. M. G., Månsson, K., & Shukur, G. (2012). Performance of some logistic ridge regression estimators. *Computational Economics*, 40 (4): 401-414.
- [14] Kibria, B. M. G., Månsson, K., & Shukur, G. (2015). A simulation study of some biasing parameters for the ridge type estimation of Poisson regression. *Communications in Statistics-Simulation and Computation*, 44 (4): 943-957.
- [15] Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics-Theory and Methods*, 22 (2): 393-402.
- [16] Liu, K. (2003). Using Liu-type estimator to combat collinearity. *Communications in Statistics-Theory and Methods*, 32 (5): 1009-1020.
- [17] Månsson, K., & Kibria, B. M. G. (2021). Estimating the unrestricted and restricted Liu estimators for the Poisson regression model: Method and application. *Computational Economics*, 58: 311-326.
- [18] Månsson, K., Kibria, B. M. G., & Shukur, G. (2016). A restricted Liu estimator for binary regression models and its application to an applied demand system. *Journal of Applied Statistics*, 43 (6): 1119-1127.
- [19] Månsson, K., Kibria, B. M. G., Sjölander, P., & Shukur, G. (2012). Improved Liu estimators for the Poisson regression model. *International Journal of Statistics and Probability*, 1 (1): 2-6.
- [20] Månsson, K., & Shukur, G. (2011). A Poisson ridge regression estimator. *Economic Modelling*, 28 (4): 1475-1481.
- [21] McDonald, G. C., & Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70 (350): 407-416.
- [22] Muniz, G., & Kibria, B. M. G. (2009). On some ridge regression estimators: An empirical comparisons. *Communications in Statistics-Simulation and Computation*, 38 (3): 621-630.
- [23] Najarian, S., Arashi, M., & Kibria, B. M. G. (2013). A simulation study on some restricted ridge regression estimators. *Communications in Statistics-Simulation and Computation*, 42 (4): 871-890.
- [24] Özkale, M. R., & Kaçiranlar, S. (2007). The restricted and unrestricted two-parameter estimators. *Communications in Statistics-Theory and Methods*, 36 (15): 2707-2725.
- [25] Qasim, M., Kibria, B. M. G., Månsson, K., & Sjölander, P. (2020). A new Poisson Liu regression estimator: Method and application. *Journal of Applied Statistics*, 47 (12): 2258-2271.
- [26] Rao, C. R., & Toutenburg, H. (1995). *Linear models: Least squares and alternatives*, 2nd edition, Springer-Verlag, New York, Inc.
- [27] Rao, C. R., Toutenburg, H., Shalabh, & Heumann, C. (2008). *Linear models and generalizations*, Springer, Berlin.
- [28] Saleh, A. M. E., & Kibria, B. M. G. (2013). Improved ridge regression estimators for the logistic regression model. *Computational Statistics*, 28 (6): 2519-2558.
- [29] Şiray, G. Ü., Toker, S., & Kaçiranlar, S. (2015). On the restricted Liu estimator in the logistic regression model. *Communications in Statistics-Simulation and Computation*, 44 (1): 217-232.
- [30] Türkan, S., & Özel, G. (2016). A new modified Jackknifed estimator for the Poisson regression model. *Journal of Applied Statistics*, 43 (10): 1892-1905.