



The r-k class estimator in generalized linear models applicable with simulation and empirical study using a Poisson and Gamma responses

Atif Abbasi^{*1,2} , M. Revan Özkale² 

¹The University of Azad Jammu and Kashmir Muzaffarabad, Department of Statistics, 13100, AJK, Pakistan

²Çukurova University, Faculty of Science and Letters, Department of Statistics, 01330, Adana, Turkey

Abstract

Multicollinearity is considered to be a significant problem in the estimation of parameters not only in general linear models, but also in generalized linear models (GLMs). Thus, in order to alleviate the serious effects of multicollinearity a new estimator is proposed by combining the ridge and PCR estimators in GLMs. This new estimator is called the r-k class estimator in GLMs. The various comparisons of the new estimator are made with already existing estimators in the literature, which are maximum likelihood (ML) estimator, ridge and PCR estimators, respectively. The comparisons are to be made in terms of scalar MSE criterion. So that, a numerical example and application through simulation are mentioned in the study for Poisson and Gamma response variables, respectively. On the basis of results it is found that, the proposed estimator outperforms all of its competitors comprehensively.

Mathematics Subject Classification (2020). 62J07, 62J12, 62F10

Keywords. maximum likelihood estimator, principal component regression, ridge estimator, r-k class estimator, reduction rate, Gamma, Poisson data

1. Introduction

Generalized linear models (GLMs) were first introduced by [16], which covered a broad class of models. These include normal linear model, binary logistic regression model, log-linear model, etc. GLMs spread the idea of well implicit linear regression models. The GLMs deal with discrete as well as continuous response variable without making any usual assumptions which are required in case of GLMs such as normality, constant variance of the response, etc. Hence, GLMs offer a combined illustration for a huge class of models for discrete and continuous response variables. Suppose that the sample observations for the response and explanatory variables are (x_i, y_i) , $i = 1, 2, \dots, n$. Let $X = (1, X_1, \dots, X_q)$ is an $n \times q$ matrix of explanatory variables having the observations $x_i^T = (1, x_{i1}, x_{i2}, \dots, x_{iq})$ with $i = 1, 2, \dots, n$, and $Y = (y_1, \dots, y_n)^T$ is an $n \times 1$ vector of observations on the response

*Corresponding Author.

Email addresses: atifbbs3@gmail.com (A. Abbasi), mrevan@cu.edu.tr (M.R. Özkale)

Received: 06.04.2020; Accepted: 31.12.2020

variable. Every element of Y belongs to the class of exponential family having the following probability density function,

$$f_{Y_i}(y_i, \theta_i, \phi) = \exp \left[\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right], \quad i = 1, 2, \dots, n \quad (1.1)$$

where ϕ is a dispersion parameter considered to be a fixed quantity and does not change as the values of y_i change and $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are known specified functions conforming to the specific type of the probability density function. A GLM is generally based on three components. The Y_i is a random component, the mean and variance of Y_i are defined as $E(Y_i) = \mu_i = b'(\theta_i) = \frac{\partial b(\theta_i)}{\partial \theta_i}$ and $Var(Y_i) = b''(\theta_i) = \frac{\partial^2 b(\theta_i)}{\partial \theta_i^2}$, where the mean μ_i is connected to the set of explanatory variables by a link function $g(\mu_i) = \eta_i$, where $\eta_i = x_i^T \beta$ is the systematic part of the linear predictors and the third component link function $g(\cdot)$ is a differentiable monotonic function, whereas β is a $(q+1) \times 1$ vector of the parameters $\beta = (\beta_0, \beta_1, \dots, \beta_q)^T$. In any statistical modeling, one of the major objectives is to find the better estimates of the parameters. Generally, the method of maximum likelihood (ML) is used to estimate the parameters in GLMs. Suppose one might want to estimate the parameter β by using the sample of n independent and random observations. Then by applying the method of ML, the log likelihood function is obtained from the probability density function given by Eq.(1.1). Then to estimate the parameters, partial derivatives of the log likelihood functions are taken, which are set equal to 0 and named as the score functions. Method of ML solves the nonlinear score equations iteratively, that is by the method of Fisher's scoring algorithm. A complete discussion of Fisher's scoring method and other iterative methods are established by [4] as follows.

$$\beta^{(t+1)} = \beta^{(t)} + \left[I(\beta^{(t)}) \right]^{-1} U(\beta^{(t)}).$$

This method of Fisher's scoring updates the values of parameters at each iterative step and the procedure continues until the desired convergence criterion is achieved and the resulting ML estimator is obtained as

$$\hat{\beta}^{(t+1)} = \left(X^T \hat{W}^{(t)} X \right)^{-1} X^T \hat{W}^{(t)} \hat{z}^{(t)},$$

where t denotes the iteration step, $W^{(t)} = \text{diag}(w_{ii})$ is an $n \times n$ diagonal weight matrix with weights $w_{ii} = \left(\left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 / \text{Var}(y_i) \right)$ and $z^{(t)}$ is the $n \times 1$ working response vector having the elements $z_i^{(t)} = x_i^T \beta^{(t)} + \left(y_i - \mu_i^{(t)} \right) \frac{\partial \eta_i^{(t)}}{\partial \mu_i^{(t)}}$, where $\mu_i^{(t)}$ and $\frac{\partial \eta_i^{(t)}}{\partial \mu_i^{(t)}}$ are assessed at $\hat{\beta}^{(t)}$. Under slight regularity conditions $\hat{\beta}^{(t+1)}$ approaches the ML estimator,

$$\hat{\beta}_{ML} = \left(X^T \hat{W}_{ML} X \right)^{-1} X^T \hat{W}_{ML} \hat{z},$$

which is normally distributed for large sample (asymptotically) with mean β and variance-covariance matrix $a(\phi) \left(X^T \hat{W}_{ML} X \right)^{-1}$.

Even though the method of ML is most commonly used for estimating the parameters in GLMs, this method has drawbacks in some particular circumstances. For example, the method of ML does not give the precise results when there is a problem of multicollinearity. The multicollinearity severely affects the ML estimator resulting in large variance-covariance matrix of the ML estimator $\hat{\beta}_{ML}$. As a result, the regression coefficients having large standard error and prediction would be inefficient. Since the estimation of parameters in GLMs is affected by the collinearity among explanatory variables, researchers move towards the alternative methods of ML estimator which provide better estimates in the presence of multicollinearity. An overwhelming number of alternative techniques have been proposed to address multicollinearity problem in GLMs. Smith and Marx [24] developed

a principal component regression (PCR) estimator to estimate the effects of acid rain. Nyquist [17] explored ML estimation of the GLMs by using a penalty function approach under linear restrictions on the parameters. Segerstedt [23] proposed a conventional ridge regression estimator in GLMs. Marx and Smith [14] explored PCR for GLMs. Kurtoğlu and Özkale [8] developed a first order approximated Liu estimator in GLMs by considering a response from the gamma distribution. Kurtoğlu and Özkale [9] established a restricted ridge estimator in GLMs. Kurtoğlu and Özkale [10] published an iterative restricted Liu estimator to combat multicollinearity in GLMs by considering response from Gamma and Poisson distributions. Özkale [18] developed a $r - d$ class estimator in GLMs by combining Liu and PCR estimators. Özkale and Nyquist [21] presented a stochastic restricted ridge estimator in GLMs under stochastic restrictions.

Many of the authors alleviate multicollinearity by applying some special cases of the GLMs, for example binary logistic regression, Poisson regression or by combining with other regression models. Månsson *et al.* [12] proposed a Liu estimator in logistic regression model. Inan [7] combined the Liu-type estimator and the PCR estimator in binary logistic regression. Özkale and Arıcan [20] explored the $r - k$ class estimator in binary logistic regression model which combines the ridge logistic and principal components logistic regression. Wu and Asar [28] derived on almost unbiased ridge logistic estimator for the binary logistic regression model. Varathan and Wijekoon [26] introduced a stochastic restricted logistic estimator. Månsson and Shukur [13] proposed a Poisson ridge regression estimator to overcome the problem of multicollinearity. Türkan and Özel [25] presented a new modified jackknifed estimator to handle the problem of multicollinearity for the Poisson regression model. Batah *et al.* [2] combined the unbiased ridge and PCR estimator to alleviate the collinearity problem in the linear regression model. Baye and Parker [3] jointly presented the ridge and PCR estimators against the multicollinearity and introduced the $r - k$ class estimator. However, research which combines both the PCR and ridge regression in GLMs has not been reported in the literature to the best of our knowledge. Therefore, this is an important gap in the literature and this study is an effort to fill this gap.

In this study, we propose the r - k class estimator by combining the PCR and ridge regression to overcome the severe effects of multicollinearity in GLMs. The paper is structured as follows: In Section 2, we propose the $r - k$ class estimator in GLMs. In Section 3, mean square error (MSE) of the first-order approximated version of the proposed estimator is given and its properties are discussed. In Section 4, estimation and selection of the biasing parameter k are presented. In Section 5, we illustrate the performance of the $r - k$ class estimator with a numerical example having a response variable which follows a Poisson distribution. In Section 6, simulation studies are conducted when the response variable follows Gamma and Poisson distributions, respectively. Finally, we conclude the study in Section 7.

2. The proposed r - k class estimator in GLMs

Here, we utilize from the singular value decomposition which was used by [24] for proposing the PCR in GLMs. The linear predictor $\eta = X\beta$ can be expressed as $\eta = XMM^T\beta = V\alpha$, where $V = XM$, $\alpha = M^T\beta$ and $M = [m_1, \dots, m_q]$ is a $q \times q$ orthogonal matrix through $V^T\hat{W}_{ML}V = \Lambda = \text{diag}(\lambda_j)$ is a $q \times q$ diagonal matrix of the eigenvalues of $X^T\hat{W}_{ML}X$ ($\lambda_1 = \lambda_{\max} \geq \lambda_2, \dots, \geq \lambda_q = \lambda_{\min}$) and m_j are the corresponding eigenvectors accompanied by their eigenvalues λ_j . The linear predictor for the i th component of the V matrix can be written as $\eta_i = x_i^T MM^T\beta = v_i^T\alpha$ where, $v_i^T = x_i^T M$ is the i th row vector of the V matrix. To overcome the problem of multicollinearity, sometimes it is worthwhile to use the reduced set of principal components (PCs). Therefore, V matrix and the α vector

can be divided into separate parts such as, $V = [V_r \ V_{q-r}]$ and $\alpha = [\alpha_r^T \ \alpha_{q-r}^T]$, where $V_r = XM_r$ ($r \leq q$) consisting of the PCs having large eigenvalues so that it will be retained in the model. The M and Λ can also similarly be partitioned into separate parts that are $M = [M_r \ M_{q-r}]$ and $\Lambda = \begin{bmatrix} \Lambda_r & 0 \\ 0 & \Lambda_{q-r} \end{bmatrix}$, where $\Lambda_r = V_r^T \hat{W}_{ML} V_r = M_r^T X^T \hat{W}_{ML} X M_r$ and $\Lambda_{q-r} = V_{q-r}^T \hat{W}_{ML} V_{q-r} = M_{q-r}^T X^T \hat{W}_{ML} X M_{q-r}$. Thus, the linear predictor for the reduced PCs can be defined as $\eta_{r,i} = v_{r,i}^T \alpha_r$, where $v_{r,i}^T$ is the i th row vector of the V_r matrix.

In order to obtain the r - k class estimator in GLMs, we maximize the following objective function.

$$F(\alpha_r, k) = \text{Max}_{\alpha_r} \left\{ \frac{2}{k} l(\alpha_r) - \alpha_r^T \alpha_r \right\},$$

where $l(\alpha_r) = \sum_{i=1}^n l_i$ is the log-likelihood function of the reduced model.

To find the parameter α_r , we differentiate the objective function $F(\alpha_r, k)$ with respect to parameter $\alpha_{r,j}$, which are the elements of α_r . Then, we have

$$\begin{aligned} \frac{\partial F(\alpha_r, k)}{\partial \alpha_{r,j}} &= \frac{2}{k} \sum_{i=1}^n \frac{\partial l_i}{\partial \alpha_{r,j}} - \frac{\partial(\alpha_r^T \alpha_r)}{\partial \alpha_{r,j}} \\ &= \frac{2}{k} \sum_{i=1}^n \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \alpha_{r,j}} - 2\alpha_{r,j} \\ &= \frac{2}{k} \frac{1}{a(\phi)} \sum_{i=1}^n \frac{y_i - \mu_i}{\text{Var}(y_i)} v_{r,ij} \frac{\partial \mu_i}{\partial \eta_{r,i}} - 2\alpha_{r,j}. \end{aligned} \tag{2.1}$$

In matrix notation Eq.(2.1) can be written as

$$\frac{\partial F(\alpha_r, k)}{\partial \alpha_{r,j}} = U(\alpha_r, k) = \frac{2}{k} \frac{1}{a(\phi)} V_r^T \hat{W}_{ML} D^{-1} (y - \mu) - 2\alpha_r,$$

where $D = \text{diag} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) = \text{diag} \left(\frac{1}{g'(\mu_i)} \right)$. Differentiate Eq.(2.1) w.r.t. $\alpha_{r,u}$ we get:

$$\frac{\partial^2 F(\alpha_r, k)}{\partial \alpha_{r,j} \partial \alpha_{r,u}} = \frac{-2}{k} \frac{1}{a(\phi)} \sum_{i=1}^n \left(\frac{y_i - \mu_i}{\text{Var}(y_i)} v_{r,ij} \frac{\partial \mu_i}{\partial \eta_{r,i}} \right) \left(\frac{y_i - \mu_i}{\text{Var}(y_i)} v_{r,iu} \frac{\partial \mu_i}{\partial \eta_{r,i}} \right) - 2\delta_{ju}, \tag{2.2}$$

where $\delta_{ju} = 1$ if $j = u$ and 0 otherwise. Taking the expectation of Eq.(2.2), we have the following form.

$$E \left[\frac{\partial^2 F(\alpha_r, k)}{\partial \alpha_{r,j} \partial \alpha_{r,u}} \right] = \frac{-2}{k} \frac{1}{a(\phi)} \sum_{i=1}^n \frac{v_{r,ij} v_{r,iu}}{\text{Var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 - 2. \tag{2.3}$$

Then by virtue of Eq.(2.3), we get:

$$Q(\alpha_r, k) = -E \left[\frac{\partial^2 F(\alpha_r, k)}{\partial \alpha_{r,j} \partial \alpha_{r,u}} \right] = 2 \left(\frac{1}{k} \frac{1}{a(\phi)} V_r^T \hat{W}_{ML} V_r + I_r \right).$$

By applying the Fisher's scoring method, we get the estimator as

$$\hat{\alpha}_{rk}^{(t+1)} = \hat{\alpha}_{rk}^{(t)} + \{ (Q(\alpha_r, k))^{-1} (U(\alpha_r, k))^{(t)} \}.$$

On making substitution, we have

$$\begin{aligned}
 \hat{\alpha}_{rk}^{(t+1)} &= \hat{\alpha}_{rk}^{(t)} + \frac{1}{2} \left(\frac{1}{k} \frac{1}{a(\phi)} V_r^T \hat{W}_{ML} V_r + I_r \right)^{-1} \left(\frac{2}{k} \frac{1}{a(\phi)} V_r^T \hat{W}_{ML} (D^{(t)})^{-1} \right. \\
 &\quad \left. \times (y - \mu^{(t)}) - 2\hat{\alpha}_{rk}^{(t)} \right) \\
 &= \hat{\alpha}_{rk}^{(t)} + \left(\frac{1}{k} V_r^T \hat{W}_{ML} V_r + I_r \right)^{-1} \left[\frac{1}{k} V_r^T \hat{W}_{ML} (D^{(t)})^{-1} (y - \mu^{(t)}) - \hat{\alpha}_{rk}^{(t)} \right] \\
 &= \left[I - \left(\frac{1}{k} V_r^T \hat{W}_{ML} V_r + I_r \right)^{-1} \right] \hat{\alpha}_{rk}^{(t)} + \left(\frac{1}{k} V_r^T \hat{W}_{ML} V_r + I_r \right)^{-1} \\
 &\quad \times \left(\frac{1}{k} V_r^T \hat{W}_{ML} (D^{(t)})^{-1} (y - \mu^{(t)}) \right).
 \end{aligned}$$

Using the fact $\left(\frac{1}{k} V_r^T \hat{W}_{ML} V_r + I_r \right)^{-1} \left(\frac{1}{k} V_r^T \hat{W}_{ML} V_r + I_r \right) = I$, which gives

$$I - \left(\frac{1}{k} V_r^T \hat{W}_{ML} V_r + I_r \right)^{-1} = \left(\frac{1}{k} V_r^T \hat{W}_{ML} V_r + I_r \right)^{-1} \frac{1}{k} V_r^T \hat{W}_{ML} V_r,$$

we obtain

$$\begin{aligned}
 \hat{\alpha}_{rk}^{(t+1)} &= \left(\frac{1}{k} V_r^T \hat{W}_{ML} V_r + I_r \right)^{-1} \frac{1}{k} V_r^T \hat{W}_{ML} V_r \hat{\alpha}_{rk}^{(t)} + \left(\frac{1}{k} V_r^T \hat{W}_{ML} V_r + I_r \right)^{-1} \\
 &\quad \times \frac{1}{k} V_r^T \hat{W}_{ML} (D^{(t)})^{-1} (y - \mu^{(t)}) \\
 &= \left(V_r^T \hat{W}_{ML} V_r + kI_r \right)^{-1} \left[V_r^T \hat{W}_{ML} V_r \hat{\alpha}_{rk}^{(t)} + V_r^T \hat{W}_{ML} (D^{(t)})^{-1} (y - \mu^{(t)}) \right] \\
 &= \left(V_r^T \hat{W}_{ML} V_r + kI_r \right)^{-1} V_r^T \hat{W}_{ML} \left[V_r \hat{\alpha}_{rk}^{(t)} + (D^{(t)})^{-1} (y - \mu^{(t)}) \right],
 \end{aligned}$$

where D and μ are evaluated at $\hat{\alpha}_{rk}^{(t)}$. Converting back to the original parameters, given as $\hat{\beta}_{rk}^{(t+1)} = M_r \hat{\alpha}_{rk}^{(t+1)}$, we obtain the $r - k$ class estimator in GLMs as

$$\begin{aligned}
 \hat{\beta}_{rk}^{(t+1)} &= M_r (M_r^T X^T \hat{W}_{ML} X M_r + kI_r)^{-1} M_r^T X^T \hat{W}_{ML} \\
 &\quad \times \left(X M_r M_r^T \hat{\beta}_{rk}^{(t)} + (D^{(t)})^{-1} (y - \mu^{(t)}) \right) \\
 &= M_r \left(M_r^T X^T \hat{W}_{ML} X M_r + kI_r \right)^{-1} M_r^T X^T \hat{W}_{ML} z_{rk}^{(t)}, \tag{2.4}
 \end{aligned}$$

where $z_{rk}^{(t)} = X M_r M_r^T \hat{\beta}_{rk}^{(t)} + (D^{(t)})^{-1} (y - \mu^{(t)})$. The iteration proceeds until the difference between successive values of the estimates is sufficiently small. The inequality $\|\hat{\beta}_{rk}^{(t+1)} - \hat{\beta}_{rk}^{(t)}\| \leq \delta$ can be selected as a convergence criterion where δ is some small number, say 1×10^{-6} . Let us denote the $r - k$ class estimator in GLMs as $\hat{\beta}_{rk} = M_r \left(M_r^T X^T \hat{W}_{ML} X M_r + kI_r \right)^{-1} M_r^T X^T \hat{W}_{ML} z_{rk}$.

When, $k = 0$, $\hat{\beta}_{rk}^{(t+1)}$ in Eq.(2.4) gives the iterative PCR estimator given by [24] as

$$\hat{\beta}_r^{(t+1)} = M_r \left(M_r^T X^T \hat{W}_{ML} X M_r \right)^{-1} M_r^T X^T \hat{W}_{ML} z_r^{(t)},$$

where $z_r^{(t)} = X M_r M_r^T \hat{\beta}_r^{(t)} + (D^{(t)})^{-1} (y - \mu^{(t)})$.

The first order approximated form of Eq.(2.4) is

$$\hat{\beta}_{rk}^{(1)} = M_r (M_r^T X^T \hat{W}^{(0)} X M_r + kI_r)^{-1} M_r^T X^T \hat{W}^{(0)} z_{rk}^{(0)}, \tag{2.5}$$

where $z_r^{(0)} = X M_r M_r^T \beta^{(0)} + (D^{(0)})^{-1}(y - \mu^{(0)})$, with $\beta^{(0)}$ as the initial value of the regression coefficients and $D^{(0)}$ and $\mu^{(0)}$ are evaluated at $\beta^{(0)\dagger}$. We call the estimator in Eq.(2.5) as the first order approximated $r - k$ class estimator in GLMs.

Eq.(2.5) implies that:

- (1) The first-order approximated ML estimator:

$$\hat{\beta}_{r=p,k=0}^{(1)} = \hat{\beta}^{(1)} = \left(X^T W^{(0)} X \right)^{-1} X^T W^{(0)} z^{(0)},$$

where $z^{(0)} = X \beta^{(0)} + (D^{(0)})^{-1} (y - \mu^{(0)})$.

- (2) The first-order approximated PCR estimator:

$$\hat{\beta}_{r,k=0}^{(1)} = \hat{\beta}_r^{(1)} = M_r \left(M_r^T X^T W^{(0)} X M_r \right)^{-1} M_r^T X^T W^{(0)} z_r^{(0)} = M_r M_r^T \hat{\beta}^{(1)}. \tag{2.6}$$

- (3) The first-order approximated ridge estimator:

$$\hat{\beta}_{r=p,k}^{(1)} = \hat{\beta}^{(1)}(k) = \left(X^T W^{(0)} X + k I_p \right)^{-1} X^T W^{(0)} z^{(0)}.$$

$\hat{\beta}_{rk}^{(1)}$ can also be written in terms of $\hat{\beta}^{(1)}$ defined by Eq.(2.6):

$$\hat{\beta}_{rk}^{(1)} = M_r \left(M_r^T X^T W^{(0)} X M_r + k I_r \right)^{-1} M_r^T X^T W^{(0)} X M_r M_r^T \hat{\beta}^{(1)}. \tag{2.7}$$

3. Mean square error of the first order approximated r-k class estimator

There are many criteria to assess the performance of a good estimator, among those the most widely used criterion is the MSE. In this section, we develop the MSE of proposed $r - k$ class estimator. As, we know from the literature that asymptotically (by [1])

$$E \left(\hat{\beta}^{(1)} \right) = \beta^{(0)}, \quad Var \left(\hat{\beta}^{(1)} \right) = a \left(\phi \right) \left(X^T W^{(0)} X \right)^{-1}.$$

From [24], we approximately write the expected value and variance of the first-order approximated PCR estimator as

$$E \left(\hat{\beta}_r^{(1)} \right) = M_r \alpha^{(0)}, \quad Var \left(\hat{\beta}_r^{(1)} \right) = a \left(\phi \right) \left(M_r \Lambda_r^{-1} M_r^T \right),$$

where $\alpha^{(0)} = M_r^T \beta^{(0)}$. Then, by Eq. (2.7), the asymptotic bias and variance of $\hat{\beta}_{rk}^{(1)}$ can be found. Since

$$E \left(\hat{\beta}_{rk}^{(1)} \right) = M_r \left(\Lambda_r^{(0)} + k I_r \right)^{-1} \Lambda_r^{(0)} M_r^T \beta^{(0)},$$

where $\Lambda_r^{(0)} = M_r^T X^T W^{(0)} X M_r$, we obtain

$$Bias \left(\hat{\beta}_{rk}^{(1)} \right) = E \left(\hat{\beta}_{rk}^{(1)} \right) - \beta^{(0)} = \left[M_r \left(\Lambda_r^{(0)} + k I_r \right)^{-1} \Lambda_r^{(0)} M_r^T - I_p \right] \beta^{(0)}.$$

Since $I_p = M_r M_r^T + M_{p-r} M_{p-r}^T$, on making substitution and simplifying we get:

$$Bias \left(\hat{\beta}_{rk}^{(1)} \right) = \left[-k M_r \left(\Lambda_r^{(0)} + k I_r \right)^{-1} M_r^T - M_{p-r} M_{p-r}^T \right] \beta^{(0)}.$$

The asymptotic variance of $\hat{\beta}_{rk}^{(1)}$ is given as

$$Var \left(\hat{\beta}_{rk}^{(1)} \right) = a \left(\phi \right) M_r \left(\Lambda_r^{(0)} + k I_r \right)^{-1} \Lambda_r^{(0)} \left(\Lambda_r^{(0)} + k I_r \right)^{-1} M_r^T.$$

Then by using

$$MSE \left(\hat{\beta}_{rk}^{(1)} \right) = Var \left(\hat{\beta}_{rk}^{(1)} \right) + [Bias \left(\hat{\beta}_{rk}^{(1)} \right)][Bias \left(\hat{\beta}_{rk}^{(1)} \right)]^T,$$

the asymptotic matrix mean square error (MMSE) of $\hat{\beta}_{rk}^{(1)}$ is defined as

[†]One can also use \hat{W}_{ML} instead of $\hat{W}^{(0)}$ which was used by [20]

$$\begin{aligned}
MSE(\hat{\beta}_{rk}^{(1)}) &= a(\phi) M_r (\Lambda_r^{(0)} + kI_r)^{-1} \Lambda_r^{(0)} (\Lambda_r^{(0)} + kI_r)^{-1} M_r^T \\
&\quad + [-kM_r (\Lambda_r^{(0)} + kI_r)^{-1} M_r^T - M_{p-r} M_{p-r}^T] \beta^{(0)} \\
&\quad \times (\beta^{(0)})^T [-kM_r (\Lambda_r^{(0)} + kI_r)^{-1} M_r^T - M_{p-r} M_{p-r}^T]. \quad (3.1)
\end{aligned}$$

As we know that the scalar MSE is obtained from $mse(\hat{\beta}_{rk}^{(1)}) = tr(MSE(\hat{\beta}_{rk}^{(1)}))$ and also $M_r^T M_{p-r} = 0$, by using these results and properties of trace, we get the scalar MSE as

$$mse(\hat{\beta}_{rk}^{(1)}) = \sum_{i=1}^r \frac{a(\phi) \lambda_i + k^2 (\alpha_i^{(0)})^2}{(\lambda_i + k)^2} + \sum_{i=r+1}^p (\alpha_i^{(0)})^2. \quad (3.2)$$

Here, we discuss some of the special cases of the $r - k$ class estimator. It can be seen that if we replace $r = p$ in Eqs.(3.1) and (3.2), the matrix MSE and scalar MSE of the first order approximated ridge estimator $\hat{\beta}^{(1)}(k)$ are obtained, respectively as

$$\begin{aligned}
MSE(\hat{\beta}^{(1)}(k)) &= a(\phi) M (\Lambda^{(0)} + kI_p)^{-1} \Lambda^{(0)} (\Lambda^{(0)} + kI_p)^{-1} M^T \\
&\quad + k^2 M (\Lambda^{(0)} + kI_p)^{-1} M^T \beta^{(0)} (\beta^{(0)})^T M (\Lambda^{(0)} + kI_p)^{-1} M^T, \quad (3.3)
\end{aligned}$$

and

$$mse(\hat{\beta}^{(1)}(k)) = \sum_{i=1}^p \frac{a(\phi) \lambda_i + k^2 (\alpha_i^{(0)})^2}{(\lambda_i + k)^2}. \quad (3.4)$$

Similarly, the matrix MSE and scalar MSE for the first order approximated PCR estimator $\hat{\beta}_r^{(1)}$ can be obtained by substituting $k = 0$ in Eqs.(3.1) and (3.2); that is:

$$MSE(\hat{\beta}_r^{(1)}) = a(\phi) M_r (\Lambda_r^{(0)})^{-1} M_r^T + M_{p-r} M_{p-r}^T \beta^{(0)} (\beta^{(0)})^T M_{p-r} M_{p-r}^T, \quad (3.5)$$

and

$$mse(\hat{\beta}_r^{(1)}) = \sum_{i=1}^r \frac{a(\phi)}{\lambda_i} + \sum_{i=r+1}^p (\alpha_i^{(0)})^2. \quad (3.6)$$

Whereas, the scalar MSE for the first order approximated ML estimator is given below.

$$mse(\hat{\beta}_{ML}^{(1)}) = \sum_{i=1}^p \frac{1}{\lambda_i}. \quad (3.7)$$

4. Estimation of the parameter k

In this section, we estimate the value of parameter k in the $r - k$ class estimator. A plausible value of k can be obtained from the scalar MSE of $\hat{\beta}_{rk}^{(1)}$. Thus, to find the value of k by differentiating Eq.(3.2) with respect to k for a specified r , we get:

$$\frac{dmse(\hat{\beta}_{rk}^{(1)})}{dk} = \sum_{i=1}^r \frac{-2a(\phi) \lambda_i + 2\lambda_i k (\alpha_i^{(0)})^2}{(\lambda_i + k)^3} = 0. \quad (4.1)$$

Considering the right hand side of Eq.(4.1) and solving it for k , we get the required value of k for $i = 1, 2, \dots, r$, as $k = \frac{a(\phi)}{(\alpha_i^{(0)})^2}$, which is minimum at that value of k .

It is to be noted that $\alpha_i = m_i^T \beta$ is not fixed, but k should be fixed, therefore to choose the value of k in $r - k$ class estimator we can adopt any of the procedure from the list given in [29]. Although an overwhelming number of choices are available in the literature for the selection of ridge parameter k , we might use the same procedure suggested by [5] and [6]; the value of k in $r - k$ class estimator can be found by using $k = \frac{1}{\max(m_i^T \beta^{(0)})^2}$ or by taking the harmonic mean of $\frac{a(\phi)}{(\alpha_i^{(0)})^2}$ for $i = 1, 2, \dots, r$, that is $k = \frac{ra(\phi)}{\beta^{(0)T} M_r M_r^T \beta^{(0)}}$.

5. Numerical example

In this section, we illustrate the application of ML, ridge, PCR and $r - k$ class estimators to a real life data set. We consider the Swedish football data set, which is obtained on request from [22]. There are 242 observations in the data set. This data set includes two response variables: the number of full-time home team goals (y_1) and the number of full-time away team goals (y_2) and six explanatory variables, which are pinnacle home win odds (x_1), pinnacle away win odds (x_2), maximum oddsportal maximum home win (x_3), oddsportal maximum away win (x_4), average oddsportal home win (x_5) and average oddsportal away win (x_6). Qasim *et al.* [22] showed that each of the response variable follows a Poisson distribution. Therefore, the log link function is to be considered and the regression model is $\log(\hat{\mu}_i) = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{\beta}_4 x_{4i} + \hat{\beta}_5 x_{5i} + \hat{\beta}_6 x_{6i}$. Our computations were carried out by using MATLAB programming language.

Before computations, we standardized the explanatory variables by unit length standardization to avoid any difficulty regarding comparisons of the regression estimates. As a result of standardization, the intercept term is also vanished. Therefore, the standardization is also helpful in reducing an ill-conditioned situation which may arise because of the intercept term and the standardizing also increases the speed of convergence. The standardization is also very useful in order to achieve the desired convergence rapidly. The number of full-time home team goals (y_1) is considered as a response variable.

We first obtained the ML estimator by an iterative procedure. We used the convergence criterion 1×10^{-6} as a rule of thumb, which is close enough to zero. The iteration carries on until the absolute value of the difference of parameter estimates smaller than 1×10^{-6} is achieved. As an initial estimate, the ordinary least square estimator (OLS) is used; that is $\hat{\beta}_{ols} = (X^T X)^{-1} X^T y$, having values:

$$\hat{\beta}_{ols} = [13.8992, 2.0119, -18.8602, -20.2691, 3.0540, 24.1594]^T.$$

It is observed that the iterative ML estimates are obtained at the 3rd iteration and given in Table 1.

After obtaining the ML estimator, the collinearity diagnostics are applied by following [19], who investigated the effects of centering and scaling the information matrix on the sensitivity of the collinearity diagnostics. She found that when the columns of $X^* = \hat{W}_{ML}^{1/2} X$ have been centered and scaled, where \hat{W}_{ML} is the weight matrix at final iteration of ML estimation, the collinearity in such a coordinate system will be meaningful. Therefore, after getting \hat{W}_{ML} , we centered and scaled the columns of X^* . Following [11] and [27], the condition number, $\varsigma_{\max} = \left(\frac{v_{\max}}{v_{\min}}\right)^{1/2}$, is computed as $k = (4623.714059)^{1/2} = 67.997897$, which is large enough to confirm that there exists a strong multicollinearity problem among the explanatory variables. Here, v_{\max} and v_{\min} denote respectively the maximum and minimum eigenvalues of $X^{*T} X^*$. Furthermore, the values of variance inflation factor (VIF) are obtained as 357.9890, 90.7073, 596.7983, 205.6173, 432.4859 and 211.3393, which are all greater than 10 and revealed that there exists a severe problem of multicollinearity among all the explanatory variables.

Table 1. Results of Swedish football data estimated coefficients and the scalar MSE values for k_1 and k_2

	$\hat{\beta}_{ML}$	$\hat{\beta}_r$	$\hat{\beta}(k_1)$	$\hat{\beta}_r(k_1)$	$\hat{\beta}(k_2)$	$\hat{\beta}_r(k_2)$
x_1	44.5967	-2.3055	-1.0360	-2.2614	39.3589	-2.3054
x_2	2.27943	0.7011	0.7788	0.7154	2.0959	0.7011
x_3	-60.9131	-2.2586	-3.3571	-2.2154	-52.2539	-2.2585
x_4	-11.6586	0.5974	-0.7965	0.6144	-11.0773	0.5974
x_5	9.3973	-2.3201	-2.5623	-2.2762	6.0965	-2.3199
x_6	12.3219	0.79224	2.1481	0.8042	11.8368	0.7923
SMSE	2975.6742	1546.0270	1299.9149	1545.9282	2336.2812	1546.0267

As it is well known that for selecting the number of PCs, numerous strategies have been proposed in the literature. Among those one of the most commonly used strategy is the percentage of total variation (PTV), which we used in this example and defined as

$$PTV = \frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^p \lambda_i} \times 100,$$

where r denotes the number of PCs, there is no rule existing in the literature for selecting the cut-off point for the value of PTV.

By using this method, the number of PCs retained in the model were $r = 2$, which explained almost 95% of the total variation. There is no hard and fast rule for selecting the cut-off point for PTV value so, it is arbitrarily chosen as 0.95[‡].

The value of k is found by using the formula $k_1 = \frac{r}{\hat{\beta}_{ols}^T M_r M_r^T \hat{\beta}_{ols}}$ = 0.0325 and another value of k is also chosen arbitrarily as $k_2 = 0.0001$. The results of the parameter estimates corresponding to iterative procedure are shown in Table 1 for both the values of k . Whereas, scalar MSE values of the estimators given in Table 1 are attained in their first-order approximated form, because of comparative purpose.

Table 1 shows the effects of multicollinearity on the ML estimates. The signs and magnitude of some of the ML estimates coefficients become unstable by making comparisons with ridge, PCR and $r - k$ class estimators. The signs of the coefficients of x_1 , x_4 and x_5 are changed from positive to negative and negative to positive when ridge, PCR and $r - k$ class estimators are used. It can be seen that both the k values and the estimator are affecting this change. By making comparisons of the estimators on the basis of scalar MSE values, it is clear that all of the biased estimators have smaller scalar MSE value than the ML estimator so, it is evident that all the biased estimators are superior to the ML estimator, particularly when there is a problem of multicollinearity. Moreover, it is perceived that the ridge estimator has smaller scalar MSE value as compared to all other estimators for k_1 . However, it is important to note that the value of k also has a significant impact on the performance of estimators. As, it could be shown that for k_2 , the proposed $r - k$ class estimator outperformed all its competitors in terms of the scalar MSE criterion, because it contained smallest scalar MSE value when compared to its counterparts.

Table 1 presents the performance of estimators only for particular values of k as discussed above. To make comparisons for other values of k , Figures 1 and 2 are provided. Figure 1 displays the performance of PCR and $r - k$ class estimators on the basis of scalar MSE for different values of k . It is examined that the $r - k$ class estimator outperformed the PCR estimator for small values of k . However, when the values of k are increased that is when k is greater than approximately 2.5, then the PCR estimator performed better

[‡]0.90 and 0.99 also gave the same number of PCs for this data set

than the $r - k$ class estimator. Similarly, Figure 2 expresses the performance of ridge estimator and the ML estimator on the basis of scalar MSE criterion for different values of k . It is to be noted that the scalar MSE values of ridge estimator decreased as k value shifted a bit far from 0 and then increased gradually. Nevertheless, regardless of the values of k it is clear that the ridge estimator dominated over the ML estimator comprehensively. Thus, the graphical results confirm the tabular results.

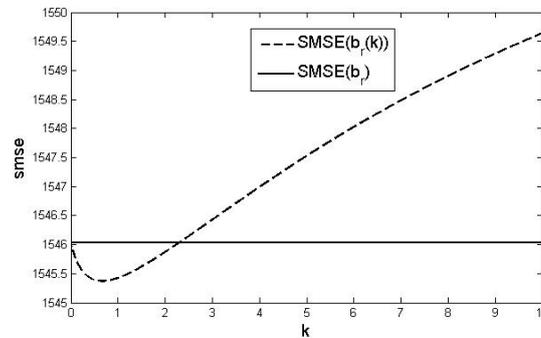


Figure 1. Graph of scalar MSE values against k for PCR and $r - k$ class estimators

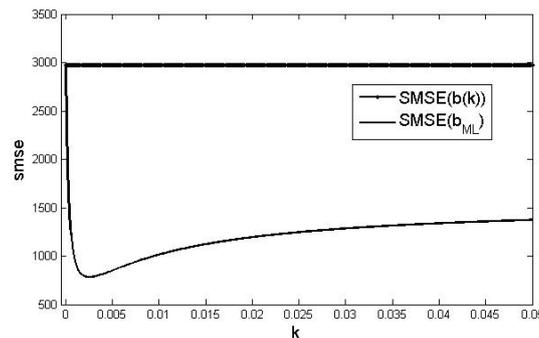


Figure 2. Graph of scalar MSE values against k for ML and ridge estimators

This numerical example indicates that there is always a prevailing value of k for which the proposed $r - k$ class estimator is superior to all of its counterparts and performs efficiently, particularly for small values of k in terms of scalar MSE criterion. However, for larger values of k results may be reverse.

6. Simulation studies

This section provides a brief overview about Monte Carlo simulation studies having response variables from Poisson and Gamma distributions respectively, for making comparisons of the proposed $r - k$ class estimator to its counterparts that are ML, ridge, and PCR estimators, respectively. The comparison is provided under different conditions such as various levels of multicollinearity, choices about the number of explanatory variables as well as for various sample sizes. The simulation results are done by using MATLAB programming language.

6.1. Simulation study with Poisson response

In this experimental set up the response variable follows a Poisson distribution with the log link function. The procedure for achieving the Monte Carlo simulation results are discussed under the following scenarios:

- (1) To perceive the impact of the sample sizes on the performance of the estimators, we selected $n = 200, 400, 600$ and 800 , whereas the number of explanatory variables are to be selected as $p = 8, 10$ and 12 .
- (2) The explanatory variables are produced by using the equation given by [15] as $x_{ij} = (1 - \gamma^2)^{1/2} z_{ij} + \gamma z_{i,(p+1)}$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$, where z_{ij} are independent standard normal pseudo-random numbers and γ is specified by which the correlation among any of two explanatory variables is assumed to be γ^2 . Then explanatory variables are standardized by using unit length standardization scheme for the purpose that $X^T X$ becomes a matrix of correlation.
- (3) To examine the influence of the correlation in the simulation study, we considered the values $\gamma^2 = 0.90, 0.95$ and 0.99 .
- (4) For every set of explanatory variables, parameter vector β is selected as an eigenvector corresponding to the substantially large eigenvalue of $X^T X$ matrix.
- (5) By using the pseudo-random numbers the response variable is generated from a Poisson distribution that is $y_i \sim P(\mu_i)$ having the log-link function, which might be expressed as $\mu_i = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})$, where $i = 1, 2, \dots, n$ and β_j , $j = 1, \dots, p$.
- (6) The initial values are set to be $\mu_i^{(0)} = \exp(X \hat{\beta}_{ols})$, where $\hat{\beta}_{ols}$ is the OLS estimator. And the initial working response for the i th observation is considered as $z_i^{(0)} = X \hat{\beta}_{ols} + (W^{(0)})^{-1} (y_i - \mu_i^{(0)})$ for the Poisson distribution, whereas $W^{(0)}$ is a weight matrix and calculated as diagonal elements of $(\mu_i^{(0)})$.
- (7) The ML estimator is acquired by using an iterative algorithm along with the weight matrix \hat{W} , which remains to be fixed at final iteration during the process of finding the iterative estimators. The value 1×10^{-6} is to be used as a convergence criterion for achieving the ML estimator and the iteration is carried on until the desired convergence criterion is met, that is $\|\hat{\beta}^{(t+1)} - \hat{\beta}^{(t)}\| \leq 1 \times 10^{-6}$.
- (8) For selecting the number of PCs we used the method of PTV. An arbitrarily selected cut-off point value is chosen as 0.90 .
- (9) The value of k for the first order approximated ridge estimator is to be found by using $k = \frac{r}{\hat{\beta}_{ols}^T M_r M_r^T \hat{\beta}_{ols}}$.
- (10) In the Monte Carlo simulation study the estimated MSE (EMSE) characteristics of the estimators are found for every choice of n, p and γ^2 and a program to evaluate the EMSE characteristics is made in MATLAB programming language which is replicated up to 1000 times. The formula for calculating the EMSE is given below:

$$EMSE(\tilde{\beta}) = \frac{1}{1000} \sum_{i=1}^{1000} (\tilde{\beta}_{(r)} - \beta)^T (\tilde{\beta}_{(r)} - \beta),$$

where the subscript r denotes the r th replication and $\tilde{\beta}_{(r)}$ is the estimate of β in the aforementioned replication.

The simulated results of the EMSE are shown in Tables 2, 3 and 4. Tables 2-4 reveal the performance of ML, ridge, PCR and $r - k$ class estimators in terms of their EMSE values. Reduction rate (RR) is also computed with respect to the $r - k$ class estimator which is computed as

$$RR = \frac{EMSE(\tilde{\beta}) - EMSE(\hat{\beta}_{rk})}{EMSE(\tilde{\beta})},$$

where $\tilde{\beta}$ is any estimator to be compared with the $r - k$ class estimator. RR is used to quantify the strength of the $r - k$ class estimator over the competitors.

Table 2. RR, EMSE values of the ML, ridge, PCR and $r - k$ class estimators when $p = 8$ and Poisson response

n	γ^2	ML	Ridge	PCR	$r - k$
200	0.90	37.525551600	6.634948370	14.816844830	3.620857272
	RR	0.903509552	0.454274989	0.755625620	—
	0.95	74.898961730	9.025821874	1.134143020	1.126853379
	RR	0.984955020	0.875152269	0.006427444	—
	0.99	357.633127400	12.956143160	1.125907322	1.122861521
	RR	0.996860298	0.913333659	0.002705197	—
400	0.90	36.757518980	6.475868688	14.669363070	3.652012230
	RR	0.900645845	0.436058326	0.751044936	—
	0.95	75.525567000	9.015851375	1.143600032	1.133592781
	RR	0.984990595	0.874266696	0.008750657	—
	0.99	358.991520900	13.734417470	1.124515272	1.120107231
	RR	0.99687985	0.918445232	0.003919948	—
600	0.90	37.379177570	6.320171343	16.199015530	3.360099991
	RR	0.910107707	0.468353023	0.792573815	—
	0.95	74.889655430	8.413768773	1.142149944	1.140927777
	RR	0.984765215	0.864397536	0.001070059	—
	0.99	373.446567200	13.571149240	1.132983179	1.128861009
	RR	0.996977182	0.916819056	0.003638333	—
800	0.90	36.430436340	6.212917176	15.0342819700	3.378795924
	RR	0.907253487	0.456165948	0.7752605720	—
	0.95	69.644517380	7.633279444	1.137273251	1.124261111
	RR	0.983857148	0.852715845	0.011441525	—
	0.99	344.454171700	12.227834970	1.120663201	1.119735478
	RR	0.996749247	0.908427331	0.000827834	—

Table 3. RR, EMSE values of the ML, ridge, PCR and $r - k$ class estimators when $p = 10$ and Poisson response

n	γ^2	ML	Ridge	PCR	$r - k$
200	0.90	46.450003420	8.207749747	18.066915660	4.316719839
	RR	0.907067395	0.474067805	0.761070460	—
	0.95	88.607949890	10.529122080	1.113766959	1.107497599
	RR	0.987501149	0.894815770	0.005628969	—
	0.99	428.426512200	16.199528900	1.104810291	1.102194135
	RR	0.997427344	0.931961346	0.002367968	—
400	0.90	49.314254300	9.062197277	15.740118510	4.047476215
	RR	0.917924822	0.553367016	0.742856052	—
	0.95	94.107980950	11.422366900	1.110830353	1.104596150
	RR	0.988262460	0.903295336	0.005612202	—
	0.99	459.319592000	17.9122536100	1.099700932	1.097529787
	RR	0.997610531	0.938727431	0.001974305	—
600	0.90	51.333409800	8.852738585	18.092092880	4.196684320
	RR	0.918246531	0.525945076	0.768037653	—
	0.95	99.445072440	11.649305330	1.106013423	1.100245662
	RR	0.988936147	0.905552681	0.0052149110	—
	0.99	480.176492300	17.885007240	1.100365920	1.097872610
	RR	0.997713606	0.938614919	0.002265891	—
800	0.90	47.878570820	8.2520536760	21.898757710	4.482281784
	RR	0.906382298	0.456828329	0.795317988	—
	0.95	94.975872200	11.138258900	1.108479909	1.104095402
	RR	0.988374991	0.900873609	0.003955424	—
	0.99	468.697664700	17.957672400	1.098806610	1.096708442
	RR	0.997660094	0.938928141	0.001909497	—

Tables 2-4, demonstrated that the ML estimator $\hat{\beta}_{ML}$ have largest EMSE values for all the sample sizes, as well as for all levels of the multicollinearity and number of explanatory variables when compared to the other biased estimators. It is to be noted that, as the levels of multicollinearity increased the EMSE values of the ML estimator also increased. This

Table 4. RR, EMSE values of the ML, ridge, PCR and $r - k$ class estimators when $p = 12$ and Poisson response

	γ^2	ML	Ridge	PCR	$r - k$
200	0.90	61.515596600	11.760815610	20.006249900	5.743763706
	RR	0.906629147	0.511618590	0.712901532	—
	0.95	118.990357100	15.455487480	1.094932081	1.09167392
	RR	0.990825526	0.929366581	0.002975674	—
	0.99	597.959631000	25.211374200	1.088032985	1.086625068
	RR	0.998182779	0.956899411	0.001294002	—
400	0.90	58.598507160	11.383937040	19.626232310	5.331225031
	RR	0.909021146	0.531688816	0.728362278	—
	0.95	112.529613400	14.439576700	1.090225373	1.086905414
	RR	0.990341161	0.924727335	0.003045205	—
	0.99	551.727960700	23.056078060	1.082692108	1.081275818
	RR	0.998040201	0.953102353	0.001308119	—
600	0.90	56.206296330	11.507211350	26.659439460	6.593502267
	RR	0.882691038	0.427011283	0.752676635	—
	0.95	112.478571500	15.6504103500	1.097124998	1.093592122
	RR	0.990277329	0.930123741	0.003220122	—
	0.99	537.647396700	24.231441960	1.076318516	1.075055926
	RR	0.998000444	0.955633844	0.001173064	—
800	0.90	57.145665640	10.889680990	27.232295920	6.190831589
	RR	0.891665772	0.431495597	0.772665823	—
	0.95	111.711436400	13.8972465400	1.096952756	1.092815095
	RR	0.990217518	0.921364632	0.003771959	—
	0.99	554.175703600	23.293587300	1.077799657	1.076469580
	RR	0.998057530	0.953786870	0.001234067	—

dramatic change in the EMSE values of ML estimator can also be observed as the number of explanatory variables increased. This shows that ML estimator is positively related with levels of multicollinearity and the number of explanatory variables. However, the role of the sample size is not considered to be a significant factor in affecting the performance of ML estimator. Although, a small amount of difference in the EMSE values of the ML estimator is observed by a change of sample size. This difference is to be seen either in increasing or decreasing form. Likewise, RR of the ML estimator is observed to be almost 90% at lowest level of multicollinearity that is 0.90, and it increased as the levels of multicollinearity, the number of explanatory variables and sample size are increased. This increment almost reached a value of 100% RR as the level of multicollinearity moved from 0.90 to 0.99. Therefore, RR of ML estimator is positively related with the levels of multicollinearity, number of explanatory and the sample size. Moreover, the RR of ML estimator is seen to be highest as compared to ridge and PCR estimators in all aspects.

Similarly, the ridge estimator is also affected by changing the levels of multicollinearity and the number of explanatory variables. It is to be perceived that as the levels of multicollinearity increased, the EMSE values of the ridge estimator increased as well. This increment can also be noticed as the number of explanatory variables increased. In so far as a ridge estimator is to be concerned it reacted almost very similar to the ML estimator against multicollinearity, regardless the number of explanatory variables and the sample size as well. That is the EMSE values of $\hat{\beta}(k)$ are seen to be increased as the number of explanatory variables and the levels of multicollinearity increased. Thus, EMSE of $\hat{\beta}(k)$ is observed to be positively related with the levels of multicollinearity and the number of explanatory variables. Whereas, the effect of sample size is to be monitored either in the increasing or decreasing form as in the case of ML estimator. As long as, the RR of ridge estimator is to be concerned, it is a bit low at lowest level of multicollinearity and then increased gradually as the number of explanatory variables and levels of multicollinearity are increased. It is to be noticed that the RR is approximately 95%, in the presence of high multicollinearity and large number of explanatory variables. Nevertheless, the sample

size is not considered to be as an important factor in affecting the RR of ridge estimator. Although, a minor change is to be observed in RR either in increasing or decreasing form by change of a sample size.

For the PCR estimator, it is conceived that at lowest level of multicollinearity the EMSE values of the PCR estimator are high. Furthermore, as the number of explanatory variables and sample size increased the EMSE of the PCR estimator increased as well at the aforementioned level of multicollinearity. So, the number of explanatory variables and sample size are positively related to the EMSE of PCR estimator at that point of collinearity. As, the levels of multicollinearity and the number of explanatory variables increased the EMSE values of the PCR estimator decreased quickly. Therefore, an inverse relationship is to be investigated between the EMSE values of PCR estimator with high levels of multicollinearity and large number of explanatory variables. Similarly, the RR for PCR estimator is seemed to be high at lowest level of multicollinearity and it increased as the sample size and the explanatory variables are increased at this level, and showed almost 77% level of RR. But, as the levels of multicollinearity and number of explanatory variables increased, the RR of PCR estimator became very low, which means that it did not differ very much from $r - k$ class estimator in the presence of high multicollinearity and large number of explanatory variables. The role of sample size is again not considered to be significant in affecting the RR, although a small change, either in increasing or decreasing form is to be observed in RR.

Here, we are discussing the behavior of our new proposed $r - k$ class estimator under different scenarios. From the tabular results, it is explored that at the lowest level of multicollinearity the values of EMSE of $\hat{\beta}_{rk}$ are high and seemed to be gradually increased as the number of explanatory variables increased. Thus, $\hat{\beta}_{rk}$ is positively related to the number of explanatory variables at that particular point of multicollinearity. Nevertheless, as the levels of multicollinearity and number of explanatory variables increased, the EMSE values of $\hat{\beta}_{rk}$ decreased rapidly which displays that the high levels of multicollinearity and large number of explanatory variables are inversely related to the EMSE of $\hat{\beta}_{rk}$. Beside this, no serious effect of the sample size is realized on the behavior of $\hat{\beta}_{rk}$.

Now, we make comparisons of the newly proposed $r - k$ class estimator against its counterparts, which are $\hat{\beta}_{ML}$, $\hat{\beta}(k)$ and $\hat{\beta}_r$ respectively in terms of EMSE criterion. It is apparent that, the $r - k$ class estimator performed efficiently as compared to all of its competitors regardless of the level of multicollinearity, the number of explanatory variables and the sample size. The proposed estimator $\hat{\beta}_{rk}$ contains smallest EMSE values as compared to its counterparts. Hence, the newly proposed estimator outclasses all of its contestants.

6.2. Simulation study with Gamma response

In this experiment, response variable follows a gamma distribution with a reciprocal link function. While, the rest of the procedure is illustrated underneath:

- (1) To monitor the impact regarding number of observations on the estimators, we take the same sample sizes as in the case of Poisson response.
- (2) The explanatory variables are produced by using the equation given by [15] as $x_{ij} = (1 - \gamma^2)^{1/2} z_{ij} + \gamma z_{i,(p+1)}$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$, where z_{ij} are independent standard normal pseudo-random numbers.
- (3) The same set of correlation values are to be considered as in the Poisson response to observe the effects of multicollinearity.
- (4) For every set of explanatory variables, parameter vector β is selected as an eigenvector corresponding to the substantially large eigenvalue of $X^T X$ matrix.

- (5) By applying the pseudo-random number, the response variable of the gamma model is generated as $\text{Gamma}(v = \mu^2/\text{var}, u = \text{var}/\mu)$, where $\mu = \exp(X\beta)$, and var denotes μ^2 .
- (6) The initial fitted values are set to be as: $\mu_i^{(0)} = \frac{(y+\bar{y})}{2}$ to compute IRLS, and then computed the weight matrix $W^{(0)}$ that is $\hat{W}^{(0)} = \text{diag}(\mu_i^{(0)})^2$. In addition to this, the initial working response z for the i th observation is considered as:

$$z_i^{(0)} = \frac{1}{\mu_i^{(0)}} - \left(W^{(0)}\right)^{-1} \left(y_i - \mu_i^{(0)}\right).$$

The rest of steps are followed similarly as in the case of Poisson response considered in Section 6.1 except that cut-off point value, which is arbitrarily chosen as 0.80 instead of 0.90.

The simulated results for the Gamma distribution are reported in Tables 5-7. These results showed that the ML estimator has larger EMSE values as compared to other biased estimators for all the aspects that are considered. The values of RR for ML estimator are also higher as compared to ridge and PCR estimators, which means that it differs substantially from proposed $r - k$ class estimator. The ridge estimator also contained large EMSE values as compared to PCR and $r - k$ class estimators. The RR values of the ridge estimator are also higher than the PCR estimator. Furthermore, it is evident that the proposed $r - k$ class estimator consisted of smallest EMSE values as compared to its counterparts for all the different scenarios, which are considered in the simulation study. Although, the RR values of PCR estimator are seemed to be small, which means that the EMSE values of PCR estimator are not much deviated from the EMSE values of $r - k$ class estimator. Nevertheless, the newly developed $r - k$ class estimator performed efficiently as compared to its counterparts according to the EMSE criterion at all levels of multicollinearity, number of explanatory variables as well as for each sample size.

Table 5. RR, EMSE values of the ML, ridge, PCR and $r - k$ class estimators when $p = 8$ and Gamma response

n	γ^2	ML	Ridge	PCR	$r - k$
200	0.90	6.470187	5.26132	1.694170	1.692702
	RR	0.738384	0.678277	0.000867	—
	0.95	7.037896	5.259769	2.509046	2.505413
	RR	0.644011	0.523665	0.001448	—
	0.99	8.937555	4.320989	1.856880	1.854712
	RR	0.792481	0.570767	0.00116	—
400	0.90	4.840330	4.533189	3.360391	3.357198
	RR	0.306411	0.259418	0.000950	—
	0.95	4.816353	4.414173	3.690117	3.684973
	RR	0.234904	0.165195	0.001394	—
	0.99	10.250065	6.677664	2.659756	2.655296
	RR	0.740948	0.602362	0.001677	—
600	0.90	4.557798	4.390382	2.563674	2.563325
	RR	0.437596	0.416150	0.000136	—
	0.95	5.223324	4.923769	4.841758	4.837445
	RR	0.073876	0.017532	0.000891	—
	0.99	8.053276	7.299779	6.891482	6.879726
	RR	0.145723	0.057543	0.001706	—
800	0.90	4.119247	4.029864	3.217761	3.217077
	RR	0.219013	0.201691	0.000212	—
	0.95	5.379205	5.146413	3.689900	3.688674
	RR	0.314272	0.283253	0.000332	—
	0.99	11.442052	9.059892	3.385485	3.383178
	RR	0.704321	0.626576	0.000681	—

Table 6. RR, EMSE values of the ML, ridge, PCR and $r - k$ class estimators when $p = 10$ and Gamma response

n	γ^2	ML	Ridge	PCR	$r - k$
200	0.90	6.725315	4.879247	2.012484	2.011344
	RR	0.700929	0.587776	0.000566	—
	0.95	10.996022	6.556518	1.919557	1.916631
	RR	0.825698	0.707675	0.001524	—
	0.99	18.696256	6.604226	1.355227	1.355139
	RR	0.927518	0.794807	0.000065	—
400	0.90	4.660879	4.350649	2.347800	2.347003
	RR	0.496446	0.460540	0.000340	—
	0.95	5.345370	4.782703	3.523911	3.520888
	RR	0.341320	0.263829	0.000858	—
	0.99	8.580232	5.563526	3.445426	3.439806
	RR	0.599101	0.381722	0.001631	—
600	0.90	4.650520	4.458327	2.596699	2.596115
	RR	0.441758	0.417693	0.000225	—
	0.95	5.316252	4.933863	3.463860	3.461015
	RR	0.348975	0.298518	0.000821	—
	0.99	10.645504	7.741610	2.947297	2.943354
	RR	0.723512	0.619801	0.001338	—
800	0.90	4.655214	4.508783	3.186016	3.185615
	RR	0.315689	0.293465	0.000126	—
	0.95	5.402928	5.113568	3.857560	3.855624
	RR	0.286383	0.246001	0.000502	—
	0.99	8.918563	7.141924	2.508333	2.507115
	RR	0.718888	0.648958	0.000486	—

Table 7. RR, EMSE values of the ML, ridge, PCR and $r - k$ class estimators when $p = 12$ and Gamma response

n	γ^2	ML	Ridge	PCR	$r - k$
200	0.90	6.844589	4.885149	2.333045	2.329927
	RR	0.659596	0.523059	0.001336	—
	0.95	11.680592	6.570061	1.269806	1.268581
	RR	0.891394	0.806915	0.000965	—
	0.99	29.221371	7.414583	1.335053	1.334850
	RR	0.954319	0.819970	0.000153	—
400	0.90	5.943753	5.274643	2.075884	2.075177
	RR	0.650864	0.606575	0.000341	—
	0.95	5.772376	4.909312	3.577512	3.572440
	RR	0.381115	0.272314	0.001418	—
	0.99	8.982236	5.243724	2.418714	2.415251
	RR	0.731108	0.539402	0.001432	—
600	0.90	5.147963	4.841649	2.218392	2.217934
	RR	0.569163	0.541905	0.000207	—
	0.95	5.543631	5.061856	2.627509	2.626005
	RR	0.526302	0.481217	0.000572	—
	0.99	9.198872	6.611784	2.207942	2.206136
	RR	0.760173	0.666333	0.000818	—
800	0.90	4.689527	4.502824	2.599765	2.599365
	RR	0.445708	0.422726	0.000154	—
	0.95	4.892099	4.624351	3.759408	3.758130
	RR	0.231796	0.187317	0.000340	—
	0.99	9.470270	7.376626	2.329244	2.327379
	RR	0.754244	0.684493	0.000801	—

7. Conclusion

In the presence of multicollinearity it is obvious that the performance of the ML estimator is poor in GLMs. Like its coefficients become unstable, along with an overestimated covariance matrix, particularly, when the explanatory variables are linearly related to each

other. In order to alleviate the alarming effects of multicollinearity in GLMs, a new estimator is presented in this study by combining the PCR and ridge estimators, which is named as the $r - k$ class estimator in GLMs. The performance of the newly developed estimator is evaluated via simulation studies having response variables from Gamma and Poisson distributions, along with a numerical example having a Poisson response. The performance evaluation criterion was scalar MSE. From the results of numerical example, it is observed that the performance of the proposed estimator and ridge estimator is affected by the choice of the k value. However, it is to be noticed that, there have been always an existing value of k , at which the $r - k$ class estimator outperformed all of its counterparts in terms of scalar MSE criterion. While the simulation results indicated that, the proposed $r - k$ class estimator is superior to its counterparts with respect to all the aspects considered in the simulation studies. Consequently, it is observed that the levels of multicollinearity and the number of explanatory variables are the most important factors, which affect the performance of estimators. Whereas, the number of observations do not seemed to be an important factor in determining the performance of estimators, when compared to the levels of multicollinearity and number of the explanatory variables in the study. Therefore, it is recommended that in the presence of multicollinearity, researchers might use the biased estimators instead of ML estimator. Further, it is emphasized that PCR and the proposed $r - k$ class estimators can serve as better alternatives to ML estimator in the presence of high multicollinearity.

References

- [1] A. Agresti, *An Introduction to Categorical Data Analysis*, John Wiley and Sons, New Jersey, 2007.
- [2] F.S.M. Batah, M.R. Özkale and S.D. Gore, *Combining unbiased ridge and principal component regression estimators*, *Comm. Statist. Theory Methods* **38** (13), 2201-2209, 2009.
- [3] M.R. Baye and D.F. Parker, *Combining ridge and principal component regression: A money demand illustration*, *Comm. Statist. Theory Methods* **13** (2), 197-205, 1984.
- [4] P.J. Green, *Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives*, *J. Roy. Statist. Soc. Ser. B* **46** (2), 149-170, 1984.
- [5] A.E. Hoerl and R.W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, *Tecnometrics* **12** (1), 55-67, 1970.
- [6] A.E. Hoerl, R.W. Kennard and K.F. Baldwin, *Ridge regression: Some simulations*, *Comm. Statist. Theory Methods* **4** (2), 105-123, 1975.
- [7] D. Inan, *Combining the Liu-type estimator and the principal component regression estimator*, *Statist. Papers* **56** (1), 147-156, 2015.
- [8] F. Kurtoğlu and M.R. Özkale, *Liu estimation in generalized linear models: Application on gamma distributed response variable*, *Statist. Papers* **57** (4), 911-928, 2016.
- [9] F. Kurtoğlu and M.R. Özkale, *Restricted ridge estimator in generalized linear models: Monte Carlo simulation studies on Poisson and binomial distributed responses*, *Comm. Statist. Simulation Comput.* **48** (4), 1-28, 2017.
- [10] F. Kurtoğlu and M.R. Özkale, *Restricted Liu estimator in generalized linear models: Monte Carlo simulation studies on gamma and Poisson distributed responses*, *Hacet. J. Math. Stat.* **48** (4), 1191-1218, 2019.
- [11] M.J. Mackinnon and M.L. Puterman, *Collinearity in generalized linear models*, *Comm. Statist. Theory Methods* **18** (9), 3463-3472, 1989.
- [12] K. Månsson, B.M.G. Kibria and G. Shukur, *On Liu estimators for the logit regression model*, *Econ. Model.* **29** (4), 1483-1488, 2012.

- [13] K. Månsson and G. Shukur, *A Poisson ridge regression estimator*, Econ. Model. **28** (4), 1475-1481, 2011.
- [14] B.D. Marx and E.P. Smith, *Trust principal component estimation for generalized linear regression*, Biometrika **77** (1), 23-31, 2019.
- [15] G.C. McDonald and D.I. Galarneau, *A Monte Carlo evaluation of some ridge-type estimators*, J. Amer. Statist. Assoc. **70** (350), 407-416, 1975.
- [16] J.A. Nelder and R.W.M. Wedderburn, *Generalized linear models*, J. Roy. Statist. Soc. Ser. A **135** (3), 370-384, 1972.
- [17] H. Nyquist, *Restricted estimation of generalized linear models*, J. R. Stat. Soc. Ser. C. Appl. Stat. **40** (1), 133-141, 1991.
- [18] M.R. Özkale, *The r - d class estimator in generalized linear models: Applications on gamma, Poisson and binomial distributed responses*, J. Stat. Comput. Simul. **89** (4), 615-640, 2019.
- [19] M.R. Özkale, *The red indicator and corrected VIFs in generalized linear models*, Comm. Statist. Simulation Comput., Published Online, 2020.
- [20] M.R. Özkale and E. Arican, *A new biased estimator in logistic regression model*, Statistics **50** (2), 233-253, 2016.
- [21] M.R. Özkale and H. Nyquist, *The stochastic restricted ridge estimator in generalized linear models*, Statist. Papers, Published Online, 2019, DOI:<https://doi.org/10.1007/s00362-019-01142-7>.
- [22] M. Qasim, B.M.G. Kibria, K. Månsson and P. Sjölander, *A new Poisson Liu Regression Estimator: Method and application* J. Appl. Stat., Published Online, 2019, DOI:10.1080/02664763.2019.1707485
- [23] B. Segerstedt, *On ordinary ridge regression in generalized linear models*, Comm. Statist. Theory Methods **21** (8), 2227-2246, 1992.
- [24] E.P. Smith and B.D. Marx, *Ill-conditioned information matrices, generalized linear models and estimation of the effects of acid rain*, Environmetrics **1** (1), 57-71, 1990.
- [25] S. Türkan and G. Özel, *A new modified Jackknifed estimator for the Poisson regression model*, J. Appl. Stat. **43** (10), 1892-1905, 2016.
- [26] N. Varathan and P. Wijekoon, *Optimal stochastic restricted logistic estimator*, Statist. Papers doi: 10.1007/s00362-019-01121-y, 2019.
- [27] L.A. Weissfeld and S.M. Sereika, *A multicollinearity diagnostic for generalized linear models*, Comm. Statist. Theory Methods **20** (4), 1183-1198, 1991.
- [28] J. Wu and Y. Asar, *On almost unbiased ridge logistic estimator for the logistic regression model*, Hacet. J. Math. Stat. **45** (3), 989-998, 2016.
- [29] J. Wu and H. Yang, *On the principal component Liu-type estimator in linear regression*, Comm. Statist. Simulation Comput. **44** (8), 2061-2072, 2015.