Hacettepe Journal of
Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 54 (3) (2025), 1187 – 1205
DOI : 10.15672/hujms.1638426

Research Article

The r–d class estimator under exact linear restrictions in generalized linear models: Theory, simulation and application

Atif Abbasi*[1], M. Revan Özkale[2]

[1] *The University of Azad Jammu and Kashmir Muzaffarabad, Department of Statistics, 13100, AJK/Pakistan*

[2] *Çukurova University, Faculty of Science and Letters, Department of Statistics, 01330, Adana/TÜRKYE*

Abstract

In this paper, we propose a restricted $r-d$ class estimator in generalized linear models by combining Liu and Principal component regression estimators, when exact linear restrictions are available as prior information along with the sample data. In addition, Particle Swarm Optimization is introduced and utilized to estimate the biasing parameter d of the newly constructed restricted estimator. In the presence of multicollinearity problem, the new estimator is compared with the current estimators that are maximum likelihood, principal components regression and $r-d$ class estimators, respectively. The performance of the proposed estimators is examined through simulation studies and a numerical example, considering response variables that follow Poisson, Binomial, and Negative binomial distributions. The evaluation is based on the scalar mean square error and the estimated mean square error criteria. The results indicate that the proposed estimator consistently outperforms all competing estimators considered in this study, both in simulation experiments and the numerical example, for suitably chosen values of the biasing parameter d.

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

Keywords. Exact restrictions, logistic regression, mean square error, negative binomial response, Poisson response, particle swarm optimization, principal components regression

1. Introduction

The maximum likelihood (ML) technique is often used to estimate the parameters in generalized linear models (GLMs) and it is a known fact that multicollinearity, which happens when explanatory variables are correlated, is a typical issue in GLMs. Multicollinearity has a substantial impact on ML estimates; for example, it raises its variance and causes some coefficients to become unstable. Various studies have been carried out in the literature to tackle the issue of multicollinearity, some of which make use of prior information in the form of exact or stochastic linear restrictions in addition to the sample data. Some of these studies are listed as [4, 11, 12, 19, 22, 23].

*Corresponding Author.
Email addresses: atifbbs3@gmail.com/atif.abbasi@ajku.edu.pk (A. Abbasi), mrevan@cu.edu.tr (M.R. Özkale)
Received: 12.02.2025; Accepted: 15.05.2025

The principal component regression (PCR) estimator has been combined with other estimators to handle the multicollinearity problem in a few studies. These include Özkale [21] proposed an estimator of the r–d class by combining the Liu and PCR estimators in GLM. Abbasi and Özkale [3] developed an $r-k$ class estimator by combining the ridge and PCR estimators in GLM, respectively. Özkale and Arıcan [20] introduced a first-order $r-d$ class estimator for logistic regression by combining the PCR and Liu logistic estimators. The PCR estimator has also been used in conjunction with other estimators in situations where additional information is available, such as in the form of stochastic linear restrictions; Abbasi and Özkale [1] have proposed a stochastic restricted r–d class estimator by combining the PCR and Liu estimators under stochastic linear restrictions in GLM. Abbasi and Ozkale [2] have developed an iterative stochastic restricted r–k class estimator by combining the ridge and PCR estimators under stochastic linear restrictions in GLMs.

Although several studies in the literature utilize the PCR estimator in combination with other estimators, relatively few incorporate prior information. To the best of our knowledge, no existing work has investigated the integration of the PCR estimator with another estimator under the simultaneous availability of sample information in GLMs and exact linear restrictions. The key contributions of this study are as follows:

- Propose a new estimator within the framework of generalized linear models (GLMs) that integrates the PCR estimator with the Liu estimator, incorporating additional information in the form of exact linear restrictions along with the sample data. This estimator is known as the restricted r–d class estimator.
- Using the particle swarm optimization (PSO) technique to estimate the optimal value of the biasing parameter d.

The PSO algorithm is inspired by social behavior, where a group of individuals (called particles) adjust their movements by moving toward areas that have worked well before [27]. The PSO is better than other heuristic algorithms in many ways, including its straightforward implementation, minimal parameter requirements, short computation times, steady convergence feature, less dependence on initial points than other approaches, and it is resilient. Moreover, the PSO has been effectively used in regression [26] and is useful to prevent ill-conditioning, which results from near linear relationships between explanatory variables. Compared with traditional approaches, this novel approach offers a number of advantages, including being simple to implement, having no function derivatives, require few parameters, quickly compute the results and exhibits stable convergence [29]. Therefore, the main advantages of this study are the introduction of a new iterative restricted r–d class estimator under exact linear restriction by combining PCR and Liu estimators, as well as the implementation of a new optimization technique for estimating the biasing parameter d of the proposed estimator.

The remaining sections of this study are arranged as follows: Some mathematical notations are shown in Section 2. An iterative restricted $r-d$ class estimator is derived in Section 3, along with the first-order approximated (FOA) form. The sampling distribution of the proposed estimator is given in Section 4. The new estimator's mean square error (MSE) is explained in Section 5. The overview of PSO and how it is used to estimate the biasing parameter of the proposed estimator are provided in Section 6. The numerical example is given in Section 7. Simulation studies are presented in Section 8. Finally, the study is concluded in Section 9.

2. The preliminary GLMs

Considering the sample data (y, X), where $y = (y_1, ..., y_n)^\top$ is a $n \times 1$ vector of observations of the response variable Y and $x_i^\top = (x_{i1}, ..., x_{ip}), i = 1, ..., n$ denotes the $i-$th observations of the $n \times p$ matrix of explanatory variable $X = [X_1, ..., X_p]$. Assume the response variable Y has a probability distribution that belongs to an exponential family of the type:

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

where θ_i is a canonical parameter, ϕ is a nuisance or dispersion parameter and $a(.)$, $b(.)$ and $c(.)$ are the known functions corresponding to the type of probability density function under consideration. The mean and variance of Y_i are represented by $E(Y_i) = \mu_i = \frac{\partial}{\partial\theta_i} b(\theta_i)$ and $var(Y_i) = a(\phi)\frac{\partial^2}{\partial\theta_i^2} b(\theta_i)$. Where, μ_i is related to the set of explanatory variables $X_1, ..., X_p$ through the link function $g(\mu_i) = \eta_i$, and $\eta_i = x_i^\top \beta$ is the systematic component or linear predictor and $g(.)$ is a monotonic and differentiable function.

The ML approach is commonly used to estimate parameters in GLMs by maximizing the log-likelihood function:

$$l(y_i, \theta_i, \phi) = \sum_{i=1}^{n}\left[\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right], i = 1, 2, ..., n. \tag{2.1}$$

To estimate the parameters, partial derivatives of the log-likelihood function given in (2.1) are set equal to zero, which are also referred to as the score functions. The score equations in general are solved by the iterative process of Fisher's scoring technique. Then, the ML estimate of β is obtained as

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

where $W = diag(1/w_{ii})$ is an $n \times n$ diagonal matrix having weights $w_{ii} = var(Y_i)[g'(\mu_i)]^2$, $g'(\mu_i) = \frac{\partial\eta_i}{\partial\mu_i}$ denotes the first-order derivative of the link function, $z^{(t)}$ is a working response of order $n \times 1$ consists of the elements $z_i^{(t)} = \sum_{j=1}^{p} x_{ij}\hat{\beta}_j^{(t)} + (y_i - \hat{\mu}_i^{(t)})\frac{\partial\eta_i^{(t)}}{\partial\mu_i^{(t)}}, i = 1, ..., n$, where t denotes the iteration step and $z_i^{(t)}$, $\hat{\mu}_i^{(t)}$ and $\frac{\partial\eta_i^{(t)}}{\partial\mu_i^{(t)}}$ are evaluated at $\hat{\beta}^{(t)}$. The asymptotic form of the ML estimator is obtained as t approaches infinity:

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

Generally, for a large sample, $\hat{\beta}_{ML} \sim N\left(\beta, (X^\top \hat{W}_{ML} X)^{-1}\right)$.

3. Iterative restricted r–d class estimator in GLMs

A singular value decomposition (SVD) approach is applied to obtain a new estimator because the SVD provides a basis that allows reconstructing the model matrix in a low-rank matrix so that it factorizes the matrix into a rotation. With this rotation, the components that are not important for the model are separated and discarded from the model.

The linear predictor $\eta = X\beta$ is expressed as $\eta = XUU^\top\beta = V\alpha$ where $U = [U_1, \cdots, U_p]$ is a $p \times p$ orthogonal matrix and U_j are the eigenvectors while $V = XU$ and $\alpha = U^\top\beta$. Therefore, by virtue of the SVD, the information matrix $X^\top \hat{W}_{ML} X$ has the form $U^\top X^\top \hat{W}_{ML} XU = V^\top \hat{W}_{ML} V = \Lambda = diag(\lambda_j)$ is a $p \times p$ diagonal matrix consisting of the eigenvalues ($\lambda_1 = \lambda_{\max} \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p = \lambda_{\min}$). For the i-th observation of the matrix V the linear predictor η_i is given as $\eta_i = x_i^\top UU^\top\beta = V_i^\top\alpha$ and $V_i^\top = x_i^\top U$ denotes the row vector of the matrix V and is called the principal components (PCs).

The PCs can be split into two parts, one of which consists of maximum information having larger eigenvalues and it retains in the model, while the other part is removed from the model that contains minimum information having small eigenvalues. This means that the V matrix and α vector can be partitioned as $V = \begin{bmatrix} V_r & V_{p-r} \end{bmatrix}$ and $\alpha = \begin{bmatrix} \alpha_r^\top & \alpha_{p-r}^\top \end{bmatrix}$

while the component $V_r = XU_r (r \leqslant p)$ consists of PCs having large eigenvalues and assumed to be stand in the model. Likewise, the matrices U and Λ are also decomposed as $U = \begin{bmatrix} U_r & U_{p-r} \end{bmatrix}$ and $\Lambda = \begin{bmatrix} \Lambda_r & 0 \\ 0 & \Lambda_{p-r} \end{bmatrix}$, where $\Lambda_r = V_r^\top \hat{W}_{ML} V_r = U_r^\top X^\top \hat{W}_{ML} X U_r$ and $\Lambda_{p-r} = V_{p-r}^\top \hat{W}_{ML} V_{p-r} = U_{p-r}^\top X^\top \hat{W}_{ML} X U_{p-r}$. Therefore, we take into account the component consisting of maximum information such as $\eta_{r,i} = V_{ri}^\top \alpha_r$ where $V_{r,i}^\top$ is the row vector of the matrix V_r and $\alpha_r = U_r^\top \beta$.

By using the reduced set of PCs the PCR, $r-d$ and $r-k$ class estimators have been developed by [3, 21, 25] as

$$\hat{\beta}_r^{(t+1)} = U_r (U_r^\top X^\top \hat{W}_{ML}^{(t)} X U_r)^{-1} U_r^\top X^\top \hat{W}_{ML}^{(t)} \hat{z}_r^{(t)}, \tag{3.1}$$

$$\hat{\beta}_{rd}^{(t+1)} = U_r (U_r^\top X^\top \hat{W}_{ML}^{(t)} X U_r + I_r)^{-1} (U_r^\top X^\top \hat{W}_{ML}^{(t)} X U_r z_{rd}^{(t)} + d U_r^\top \hat{\beta}_{ML}), \tag{3.2}$$

$$\hat{\beta}_{rk}^{(t+1)} = U_r \left(U_r^\top X^\top \hat{W}_{ML}^{(t)} X U_r + k I_r \right)^{-1} U_r^\top X^\top \hat{W}_{ML}^{(t)} \hat{z}_{rk}^{(t)} \tag{3.3}$$

where

$$\begin{aligned}
\hat{z}_r^{(t)} &= X U_r U_r^\top \hat{\beta}_r^{(t)} + \hat{D}_r^{-1} (y - \hat{\mu}_r^{(t)}), \\
\hat{z}_{rk}^{(t)} &= X U_r U_r^\top \hat{\beta}_{rk}^{(t)} + \hat{D}_{rk}^{-1} (y - \hat{\mu}_{rk}^{(t)}), \\
\hat{z}_{rd}^{(t)} &= U_r^\top \hat{\beta}_{rd}^{(t)} + (U_r^\top X^\top \hat{W}_{ML} X U_r) U_r^\top X^\top \hat{W}_{ML} \hat{D}_{rd}^{-1} (y - \hat{\mu}_{rd}^{(t)}).
\end{aligned}$$

and $\hat{D}_r$, $\hat{\mu}_r$, $\hat{D}_{rd}$, $\hat{\mu}_{rd}$, $\hat{D}_{rk}$, $\hat{\mu}_{rk}$ are evaluated at the corresponding estimator.

Now in order to develop the restricted $r-d$ class estimator in GLMs, we impose exact linear restrictions on the parameter β such as $h = H\beta$ where h is a known $q \times 1$ vector and H is a known $q \times p$ known matrix with $rank(H) = q$. By imposing a set of q linearly independent restrictions on the parameters, we have $H = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_q \end{bmatrix}$ where $H_i = [H_{i1}, H_{i2}, \cdots, H_{ip}]$. This constraint must be successfully handled for that subspace in the reduced form model if we are working in a parameter subspace that is: $h = H_r \alpha_r$, where $\alpha_r = U_r^\top \beta$ and $H_r = HU_r$ denote a matrix $q \times p$ that has $rank(H_r) = q$. To construct the restricted $r-d$ class estimator by combining the sample and prior information, we consider the following objective function:

$$\begin{aligned}
F(\alpha_r; y, h, d) \; = \; & \sum_{i=1}^{n} \left\{ \frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} - \frac{1}{2} \sum_{i=1}^{p} (\alpha_{r,j} - d\hat{\alpha}_{r,j})^2 \\
& - \frac{1}{2} \sum_{i=1}^{q} \zeta_i (h_i - H_{r,ij} \alpha_{r,j})^2,
\end{aligned} \tag{3.4}$$

where $\zeta_1, ..., \zeta_q$ are the lagrange multipliers, d is a biasing parameter in the interval $(0, 1)$ and $\hat{\alpha}_r = U_r^\top \hat{\beta}_{ML}$ is the ML estimator of the reduced model.

Using Eq. (3.4), we estimate α_r with elements $\alpha_{r,j}$, $j = 1, 2 \cdots, r$, by calculating the derivatives of $F(\alpha_r; y, h, d)$ with respect to $\alpha_{r,j}$. Applying the chain rule, we have

$$\begin{aligned}
\frac{\partial F(\alpha_r; y, h, d)}{\partial \alpha_{r,j}} \; = \; & \sum_{i=1}^{n} \frac{y_i - \mu_i}{w_{ii}} g'(\mu_i) v_{r,ij} - \sum_{j=1}^{p} (\alpha_{r,j} - d\hat{\alpha}_{r,j}) \\
+ \; & \sum_{i=1}^{q} \zeta_i H_{r,ij} (h_i - H_{r,ij} \alpha_{r,j})
\end{aligned} \tag{3.5}$$

where $v_{r,ij}$ is the ij-th element of the V_r matrix. In matrix notation, Eq. (3.5) can be written as

$$S(\alpha_r, d) = [V_r^\top W D^{-1}(y-\mu) - \alpha_r + d\hat{\alpha}_r + H_r^\top \mathcal{L}_r(h - H_r\alpha_r)],$$

where $D = diag(\frac{\partial \eta_i}{\partial \mu_i}) = diag(g'(\mu_i))$ and $L_r = Diag(\zeta_1, ..., \zeta_q)$. Now taking the derivative of Eq. (3.5) with respect to $\alpha_{r,k}$ we get

$$\begin{aligned}\frac{\partial^2 F(\alpha_r; y, h, d)}{\partial \alpha_{r,j}\partial \alpha_{r,k}} &= \sum_{i=1}^{n}(y_i-\mu_i)\frac{\partial}{\partial \alpha_{rk}}\frac{1}{w_{ii}}g'(\mu_i)v_{r,ij} - \sum_{i=1}^{n}\frac{v_{r,ij}v_{r,ik}}{w_{ii}} \\ &\quad -\delta_{jk} - \sum_{i=1}^{q}\zeta_i H_{r,ij}H_{r,ik}\end{aligned}$$

since $\delta_{jk} = 1$ if $j = k$ and 0 otherwise. Minus times the expected value of the second-order derivative gives

$$Q_{jk}(\alpha_{rj}, d) = -E\left[\frac{\partial^2}{\partial \alpha_{r,j}\partial \alpha_{r,k}}F(\alpha_r; y, h, d)\right] = \sum_{i=1}^{n}\frac{v_{r,ij}v_{r,ik}}{w_{ii}} + \delta_{jk} + \sum_{i=1}^{q}\zeta_i H_{r,ij}H_{r,ik}$$

which can be expressed in matrix form as:

$$Q(\alpha_r, d) = (V_r^\top W V_r + I_r) + H_r^\top L_r H_r$$

where I_r is the identity matrix of order r. Following [3, 21, 25], we set the weights in the ML estimates and using the technique of Fisher's scoring algorithm, we get

$$\hat{\alpha}_{R-rd}^{(t+1)} = \hat{\alpha}_{R-rd}^{(t)} + \left\{[Q(\alpha_r, d)]^{-1}\right\}_{W=\hat{W}_{ML}} \{[S(\alpha_r, d)]\}_{\alpha_r=\hat{\alpha}_{R-rd}^{(t)}}.$$

Premultiplying both sides by $[Q(\alpha_r, d)]_{W=\hat{W}_{ML}}$, replacing the values of the Q and S matrices and then simplifying some notations, we get

$$\begin{aligned}\hat{\alpha}_{R-rd}^{(t+1)} &= [(V_r^\top \hat{W}_{ML}V_r + I_r) + H_r^\top L_r H_r]^{-1}[V_r^\top \hat{W}_{ML}V_r\hat{\alpha}_{R-rd}^{(t)} \\ &\quad + V_r^\top \hat{W}_{ML}(D_{R-rd}^{(t)})^{-1}(y - \hat{\mu}_{R-rd}^{(t)}) + d\hat{\alpha}_r + H_r^\top L_r h].\end{aligned}$$

By applying the inverse formula† on $[(V_r^\top \hat{W}_{ML}V_r + I_r) + H_r^\top L_r H_r]^{-1}$ and solving

$$\begin{aligned}&[(V_r^\top \hat{W}_{ML}V_r + I_r) + H_r^\top L_r H_r]^{-1}H_r^\top L_r h = (V_r^\top \hat{W}_{ML}V_r + I_r)^{-1}H_r^\top \\ &\times\{I - [L_r^{-1} + H_r(V_r^\top \hat{W}_{ML}V_r + I_r)^{-1}H_r^\top]^{-1}H_r(V_r^\top \hat{W}_{ML}V_r + I_r)H_r^\top\}L_r^{-1}h \\ = \; &(V_r^\top \hat{W}_{ML}V_r + I_r)^{-1}H_r^\top[L_r^{-1} + H_r(V_r^\top \hat{W}_{ML}V_r + I_r)^{-1}H_r^\top]^{-1}h, \qquad (3.6)\end{aligned}$$

we get

$$\begin{aligned}&\hat{\alpha}_{R-rd}^{(t+1)} = (V_r^\top \hat{W}_{ML}V_r + I_r)^{-1}\left[V_r^\top \hat{W}_{ML}V_r\hat{\alpha}_{R-rd}^{(t)} + V_r^\top \hat{W}_{ML}(D_{R-rd}^{(t)})^{-1}(y - \hat{\mu}_{R-rd}^{(t)}) + d\hat{\alpha}_r\right] \\ &-(V_r^\top \hat{W}_{ML}V_r + I_r)^{-1}H_r^\top[L_r^{-1} + H_r(V_r^\top \hat{W}_{ML}V_r + I_r)^{-1}H_r^\top]^{-1} \\ &\times\{H_r(V_r^\top \hat{W}_{ML}V_r + I_r)^{-1}[V_r^\top \hat{W}_{ML}V_r\hat{\alpha}_{R-rd}^{(t)} + V_r^\top \hat{W}_{ML}(D_{R-rd}^{(t)})^{-1}(y - \hat{\mu}_{R-rd}^{(t)}) + d\hat{\alpha}_r] - h\}.\end{aligned}$$

Transforming back to the original parameters and as $\zeta_1, \ldots, \zeta_q \to \infty$, we obtain an iterative restricted $r-d$ class estimator in GLMs as:

$$\begin{aligned}&\hat{\beta}_{R-rd}^{(t+1)} = U_r(U_r^\top X^\top \hat{W}_{ML}XU_r + I_r)^{-1}(U_r^\top X^\top \hat{W}_{ML}z_{R-rd}^{(t)} + dU_r^\top \hat{\beta}_{ML}) - U_r \\ &\times(U_r^\top X^\top \hat{W}_{ML}XU_r + I_r)^{-1}H_r^\top[H_r(U_r^\top X^\top \hat{W}_{ML}XU_r + I_r)^{-1}H_r^\top]^{-1} \\ &\times[H_r(U_r^\top X^\top \hat{W}_{ML}XU_r + I_r)^{-1}(U_r^\top X^\top \hat{W}_{ML}z_{R-rd}^{(t)} + dU_r^\top \hat{\beta}_{ML}) - h], \qquad (3.7)\end{aligned}$$

where $z_{R-rd}^{(t)} = XU_rU_r^\top \hat{\beta}_{R-rd}^{(t)} + (D_{R-rd}^{(t)})^{-1}(y - \hat{\mu}_{R-rd}^{(t)})$ and $\hat{\mu}_{R-rd}^{(t)}$ and $D_{R-rd}^{(t)}$ are evaluated at $\hat{\beta}_{R-rd}^{(t)}$.

†Following [24], the inverse formula is as: If A_{pxp}, B_{pxn}, C_{nxn} and D_{nxp} then $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$

Since the term $U_r(U_r^\top X^\top \hat{W}_{ML} X U_r + I_r)^{-1}(U_r^\top X^\top \hat{W}_{ML} X U_r z_{R-rd}^{(t)} + dU_r^\top \hat{\beta}_{ML})$ is in the form of the iterative $r-d$ class estimator in GLMs given by Eq. (3.2). Eq. (3.7) is a restricted form estimator of the $r-d$ class estimator [‡]

3.1. The first-order approximated form

From Eq. (3.7), the first-order approximated (FOA) form of the iterative restricted $r-d$ class estimator is given as

$$\begin{aligned}\hat{\beta}_{R-rd}^{(1)} &= U_r(U_r^\top X^\top W^{(0)} X U_r + I_r)^{-1}(U_r^\top X^\top W^{(0)} z^{(0)} + dU_r^\top \hat{\beta}^{(1)}) - U_r \\ &\times (U_r^\top X^\top W^{(0)} X U_r + I_r)^{-1} H_r^\top [H_r(U_r^\top X^\top W^{(0)} X U_r + I_r)^{-1} H_r^\top]^{-1} \\ &\times [H_r(U_r^\top X^\top W^{(0)} X U_r + I_r)^{-1}(U_r^\top X^\top W^{(0)} z^{(0)} + dU_r^\top \hat{\beta}^{(1)}) - h] \qquad (3.8)\end{aligned}$$

where $\hat{\beta}^{(1)} = (X^\top W^{(0)} X)^{-1} X^\top W^{(0)} z^{(0)}$ is the FOA ML estimator, $z^{(0)} = X\beta^{(0)} + (D^{(0)})^{-1}(y - \mu^{(0)})$ is the initial working response and $D^{(0)}$ and $\mu^{(0)}$ are evaluated at the initial value $\beta^{(0)}$.

The $\hat{\beta}_{R-rd}^{(1)}$ is a general estimator that contains different estimators under particular conditions;

- If $d = 0$, $r = p$ and $H = 0$, we get the FOA ML estimator.
- If $d = 0$ and $H = 0$, we get the FOA PCR estimator proposed by [25] as $\hat{\beta}_r^{(1)} = U_r(U_r^\top X^\top W^{(0)} X U_r)^{-1} U_r^\top X^\top W^{(0)} z^{(0)}$.
- If $H = 0$; that is no prior information, we obtain the FOA $r-d$ class estimator given by [21] as $\hat{\beta}_{rd}^{(1)} = U_r(U_r^\top X^\top W^{(0)} X U_r + I_r)^{-1}(U_r^\top X^\top W^{(0)} z^{(0)} + dU_r^\top \hat{\beta}^{(1)})$.

Since the initial working responses for all FOA estimators are the same, we can also write Eq. (3.8) as

$$\begin{aligned}\hat{\beta}_{R-rd}^{(1)} &= \hat{\beta}_{rd}^{(1)} - U_r(U_r^\top X^\top W^{(0)} X U_r + I_r)^{-1} U_r^\top H^\top \\ &\times [HU_r(U_r^\top X^\top W^{(0)} X U_r + I_r)^{-1} U_r^\top H^\top]^{-1}(H\hat{\beta}_{rd}^{(1)} - h), \qquad (3.9)\end{aligned}$$

where $\hat{\beta}_{rd}^{(1)}$ is the FOA $r-d$ class estimator that can also be denoted as

$$\hat{\beta}_{rd}^{(1)} = U_r(U_r^\top X^\top W^{(0)} X U_r + I_r)^{-1}(U_r^\top X^\top W^{(0)} X U_r + dI_r) U_r^\top \hat{\beta}_r^{(1)} \qquad (3.10)$$

with the FOA principal components regression (PCR) estimator obtained from Eq. (3.1).

4. Sampling distribution of the $r-d$ class estimator under exact restriction

Özkale [21] obtained the asymptotic sampling distribution of the $r-d$ class estimator in GLMs. In this section, inspired by the studies of [6, 21, 22] we will investigate the asymptotic properties of the $r-d$ class estimator under the exact restrictions. A Taylor series approximation of $S(\alpha_r, d)$ around the true parameter value α_r, divided by the square root of the number of observations gives

$$n^{-1/2} S(\alpha_r, d) = n^{-1/2} S(\hat{\alpha}_{R-rd}, d) + n^{-1} \frac{\partial}{\partial \alpha_r} S(\alpha_r, d)\Big|_{\alpha_r = \hat{\alpha}_{R-rd}} n^{1/2}(\alpha_r - \hat{\alpha}_{R-rd}) + rem$$

where $\frac{\partial}{\partial \alpha_r} S(\alpha_r, d)$ is the $p \times p$ matrix of derivatives of S evaluated at α_r, rem is a reminder term that tends to zero as n tends to infinity and $\hat{\alpha}_{R-rd}$ is the restricted $r-d$ class estimator

‡We mean that Eq. (3.7) has a form similar to $\hat{\beta}_{R-rd}^{(t+1)} = \hat{\beta}_{rd}^{(t)} - U_r(U_r^\top X^\top \hat{W}_{ML} X U_r + I_r)^{-1} U_r^\top H^\top [HU_r(U_r^\top X^\top \hat{W}_{ML} X U_r + I_r)^{-1} U_r^\top H^\top]^{-1}(HU_r\hat{\beta}_{rd}^{(t)} - h)$, but the computation of $\hat{\beta}_{rd}^{(t)}$ is different from Eq.(3.2) because of the working response.

at convergence. By definition, the first term on the right hand side equals zero; that is, $S(\hat{\alpha}_{R-rd}, d) = 0$.

By the law of large numbers $n^{-1} \frac{\partial}{\partial \alpha_r} S(\alpha_r, d)\Big|_{\alpha_r = \hat{\alpha}_{R-rd}}$ converges to its expectation, which is minus times $Q(\alpha_r, d)|_{\alpha_r = \hat{\alpha}_{R-rd}} = Q(\hat{\alpha}_{R-rd}, d)$.

By the central limit theorem $n^{-1/2} S(\alpha_r, d)$ converges to a $p-$variate normally distributed vector with expectation $[(d-1)\alpha_r + H_r^\top L_r(h - H_r\alpha_r)]$ and variance $(V_r^\top \hat{W}_{ML} V_r + dI_r)(V_r^\top \hat{W}_{ML} V_r)^{-1}(V_r^\top \hat{W}_{ML} V_r + dI_r)$. This implied that $n^{1/2}(\hat{\alpha}_{R-rd} - \alpha_r)$ has asymptotically a normal distribution with expectation $[Q(\hat{\alpha}_{R-rd}, d)]^{-1}[(d-1)\alpha_r + H_r^\top L_r(h - H_r\alpha_r)]$ and variance $[Q(\hat{\alpha}_{R-rd}, d)]^{-1}(V_r^\top \hat{W}_{ML} V_r + dI_r)(V_r^\top \hat{W}_{ML} V_r)^{-1}(V_r^\top \hat{W}_{ML} V_r + dI_r)[Q(\hat{\alpha}_{R-rd}, d)]^{-1}$.

When the transformation is made to the original parameter space, we get

$$\begin{aligned} n^{1/2}(\hat{\beta}_{R-rd} - \beta) \approx\ & AN(U_r [Q(\hat{\alpha}_{R-rd}, d)]^{-1} [(d-1)U_r^\top \beta + U_r^\top H^\top L_r(h - HU_rU_r^\top \beta)] \\ & \times U_r [Q(\hat{\alpha}_{R-rd}, d)]^{-1} (V_r^\top \hat{W}_{ML} V_r + dI_r)(V_r^\top \hat{W}_{ML} V_r)^{-1} \\ & \times (V_r^\top \hat{W}_{ML} V_r + dI_r) [Q(\hat{\alpha}_{R-rd}, d)]^{-1} U_r^\top) \end{aligned}$$

where "AN" refers to asymptotic normality. Furthermore, application of the inverse formula on $[Q(\hat{\alpha}_{R-rd}, d)]^{-1}$ given by Eq. (3.6) and as $\zeta_1, \ldots, \zeta_q \to \infty$, we get

$$[Q(\hat{\alpha}_{R-rd}, d)]^{-1} = M_r(1) = S_r(1)^{-1} - S_r(1)^{-1} U_r^\top H^\top P_r^{-1} H U_r S_r(1)^{-1}$$

where $S_r(1) = (V_r^\top \hat{W}_{ML} V_r + I_r)$, $S_r(d) = (V_r^\top \hat{W}_{ML} V_r + dI_r)$ and $P_r = H_r S_r(1)^{-1} H_r^\top$. By Eq. (3.6), we have

$$[Q(\hat{\alpha}_{R-rd}, d)]^{-1} U_r^\top H^\top = S_r(1)^{-1} U_r^\top H^\top (L_r^{-1} + P_r)^{-1} L_r^{-1}$$

which is $[Q(\hat{\alpha}_{R-rd}, d)]^{-1} U_r^\top H^\top L_r = S_r(1)^{-1} U_r^\top H^\top P_r^{-1}$ as $\zeta_1, \ldots, \zeta_q \to \infty$. At the end of these calculations, we get

$$\begin{aligned} n^{1/2}(\hat{\beta}_{R-rd} - \beta) \approx\ & \mathrm{AN}\Big((d-1)\, U_r M_r(1) U_r^\top \beta + U_r S_r(1)^{-1} U_r^\top H^\top P_r^{-1}(h - HU_rU_r^\top \beta), \\ & U_r M_r(1)(V_r^\top \hat{W}_{ML} V_r + dI_r)(V_r^\top \hat{W}_{ML} V_r)^{-1}(V_r^\top \hat{W}_{ML} V_r + dI_r) M_r(1) U_r^\top\Big) \end{aligned}$$

This result reduces to the sampling distribution of the $r-d$ class estimator given by [21] when there is no exact restriction.

5. The MSE of the FOA restricted $r-d$ class estimator

This section shows the MSE of the FOA restricted $r-d$ class estimator along with the bias and variance of the restricted $r-d$ class estimator. In view of Eq. (3.9), we can write the bias

$$Bias(\hat{\beta}^{(1)}_{R-rd}) = E(\hat{\beta}^{(1)}_{R-rd}) - \beta = [(d-1)U_r M_r^{(0)}(1) U_r^\top - U_{p-r} U_{p-r}^\top]\beta, \quad (5.1)$$

and the variance

$$\begin{aligned} var(\hat{\beta}^{(1)}_{R-rd}) &= [I_p - U_r S_r^{(0)}(1)^{-1} U_r^\top H^\top (P_r^{(0)})^{-1} H] var(\hat{\beta}^{(1)}_{rd}) [I_p - U_r S_r^{(0)}(1)^{-1} U_r^\top H^\top (P_r^{(0)})^{-1} H]^\top \\ &= U_r M_r^{(0)}(1) S_r^{(0)}(d) \Lambda_r^{-1} S_r^{(0)}(d) M_r^{(0)}(1) U_r^\top, \quad (5.2) \end{aligned}$$

where $M_r^{(0)}(1) = S_r^{(0)}(1)^{-1} - S_r^{(0)}(1)^{-1} U_r^\top H^\top (P_r^{(0)})^{-1} H U_r S_r^{(0)}(1)^{-1}$. With the help of [19], which gives the expectation of $\hat{\beta}^{(1)}_{rd}$, we obtain

$$\begin{aligned} E(\hat{\beta}^{(1)}_{R-rd}) = \ & U_r S_r^{(0)}(1)^{-1} S_r^{(0)}(d) U_r^\top \beta - U_r S_r^{(0)}(1)^{-1} U_r^\top H^\top (P_r^{(0)})^{-1} \\ & \times [H U_r S_r^{(0)}(1)^{-1} S_r^{(0)}(d) U_r^\top \beta - H U_r U_r^\top \beta], \end{aligned}$$

where $S_r^{(0)}(1) = (U_r^\top X^\top W^{(0)} X U_r + I_r)$, $S_r^{(0)}(d) = (U_r^\top X^\top W^{(0)} X U_r + dI_r)$ and $P_r^{(0)} = H_r S_r^{(0)}(1)^{-1} H_r^\top$.

Thus, the MSE of the $\hat{\beta}^{(1)}_{R-rd}$ is obtained by writing Eqs. (5.1) and (5.2) in

$$MSE(\hat{\beta}^{(1)}_{R-rd}) = var(\hat{\beta}^{(1)}_{R-rd}) + [Bias(\hat{\beta}^{(1)}_{R-rd})][Bias(\hat{\beta}^{(1)}_{R-rd})]^{\top}.$$

Remark. Since the PCR estimator and its variants are obtained under the condition $U^{\top}_{p-r}\beta = 0$, if this condition is also taken into account, it will be easily seen that the variance and expected value found in Section 5 are the same as those found in the sampling distribution of $\hat{\beta}^{(1)}_{R-rd}$ in Section 4.

6. Estimating the biasing parameter of the restricted $r-d$ class estimator via particle swarm optimization

The PSO was introduced by [13] as an evolutionary computing technique based on swarm intelligence. This algorithm is developed by emulating the social interactions observed in flocks of birds and schools of fish. In PSO, every potential solution is referred to as a particle. A particle in the PSO resembles a fish or bird that glides over the search field. Potential solutions to the problems are given by the positions of the particles. Because each particle has a velocity vector, it may explore the space and look for the optimal position. The best particle in the swarm, known as the global best (Gbest) at each generation, and each particle's best position (Pbest) determine the route each particle will follow. The stochastic structure of the particle is expanded throughout this process, which also quickly converges to the optimum solution. The PSO technique has been used in many studies like;

Sancar and Inan [26] used PSO to estimate the biasing parameter of an estimator that depends on two biasing parameters. Kareem and Algamal [5] used the PSO to estimate the biasing parameter of the generalized ridge estimator, using the MSE as the objective function. Wiktorowicz et al. [30] proposed a method for TakagiSugeno fuzzy systems using Sparse regressions and particle swarm optimization. Uslu et al. [28] acquired an optimal value for the shrinkage parameter in ridge regression by optimizing particle swarm. İnan et al. [9] used the PSO technique to estimate the shrinkage and bias parameters of the Liu-type estimator.

6.1. How PSO works?

The PSO algorithm can be described as follows, based on the approaches described in [29] and [9]. The stages involved in the optimization process can be described as follows:

Step 1: The position vectors of each particle in D-dimensional space are randomly initialized and store in vector X_i that is; $X_i = (x_{i1}, x_{i2}, \cdots, x_{id}, \cdots, x_{iD})$.

Step 2: The velocities are randomly initialized and stored in a velocity vector that is $V_i = (v_{i1}, v_{i2}, \cdots, v_{id}, \cdots, v_{iD})$.

Step 3: Based on the objective function, the personal best Pbest and global best Gbest are calculated using the following equations:

$$P_i = (p_{i1}, p_{i2}, \cdots, p_{id}, \cdots, p_{iD})$$

$$P_g = (p_{g1}, p_{g2}, \cdots, p_{gd}, \cdot p_{gD}).$$

Step 4: The values of the PSO parameters that include the inertia weight w, the cognitive and social coefficients c_1 and c_2 are considered to be $w = 0.90$, $c_1 = c_2 = 2$, the number of particles equal to 30 and the number of iterations is 200. The parameter values

are selected on the basis of the literature [7].
Step 5: The following formulas are used to update the velocity and position, respectively

$$v_{i,t+1}^d = w * v_{i,t}^d + c_1 * rand * (p_{i,t}^d - x_{i,t}^d) + c_2 * rand * (p_{g,t}^d - x_{i,t}^d)$$

$$x_{i,t+1}^d = x_{i,t}^d + v_{i,t+1}^d$$

where *rand* are random numbers in (0, 1) which are uniformly and independently distributed.
Step 6: Until a predefined maxit iteration number *maxit* is achieved, steps 3 to 5 are repeated.

In our study, the convergence of the PSO algorithm is evaluated by running the algorithm for a fixed number of iterations, specifically 200 ($maxit = 200$). The solution obtained in the final iteration is considered the convergent estimate of the optimal value of the biasing parameter d. Stochastic optimization commonly uses this fixed iteration method, which gives the particles enough time to explore and settle in the search space. This strategy ensures consistent termination and is well supported in the foundational literature of the PSO ([7, 9, 26, 27]).

6.2. Implementation of PSO to estimate the biasing parameter of the restricted $r-d$ class estimator

We developed an objective function for our PSO-based algorithm that minimizes multicollinearity, lowers model bias, and improves predictive performance in accordance with the goals of the study. Our PSO-based algorithm's objective function consists of three components:

$$Min\{PMSE + \Phi(d)\},$$

where $PMSE$ is the prediction mean square error (PMSE) defined as:

$$PMSE = \sum_{i=1}^{n} \frac{1}{n}(y_i - \tilde{\mu}_i)^2$$

with y_i is the ith component of y and $\tilde{\mu}_i = g^{-1}(x_i^\top \tilde{\beta})$ (See[14]) and $\Phi(d)$ is the the condition number (CN)

$$\Phi(d) = \begin{cases} 0 & \text{if } CN > 30 \\ CN & \text{otherwise.} \end{cases}$$

The CN of the restricted $r-d$ class estimator is defined as for the particular case when $H = 0$:

$$CN = \sqrt{\frac{\lambda_{\max}(U_r \Lambda_r S_r(d)^{-1} S_r(1) U_r^\top)}{\lambda_{\min}(U_r \Lambda_r S_r(d)^{-1} S_r(1) U_r^\top)}}$$

where $\lambda_{\max}(.)$ and $\lambda_{\min}(.)$ show the maximum and minimum eigenvalues of the matrix $X^\top W^{(0)} X$.

The motivation for applying the PSO technique in this study lies in its straight forward implementation, independence from function derivatives, minimal parameter requirements, as well as fast and stable convergence.

7. Numerical example

This section examines the performance of the restricted $r-d$ class, ML, $r-d$, and PCR estimators using a real-world data set. Myers [17] first reported on the data set, which included 44 observations about mines in the Appalachian region of western Virginia. Marx [15] also used this data set, which includes the number of fractures or injuries as the dependent variable y and four explanatory variables; (x_1) denotes the thickness of the inner burden in feet; (x_2), the lower seam height; (x_3), the percentage of extraction from the lower previously mined seam; and (x_4), the number of years the mine has been in operation. Myers [17] and Marx [15] used a generalized linear regression model to examine how these explanatory factors affected the dependent variable y, assuming that the observations y_i belonged to the Poisson distribution with a logarithmic link function, also Kurtoğlu and Özkale [12] used this data set with a log link function. The regression model is

$$log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i},$$

where μ_i is the approximate total number of fractures or upper seam injuries at the i-th coal mining site. Our goal is to determine whether these factors have an effect on the number of fractures or injuries (y) that occur at the top seams of the mines.

To acquire our results, we employ the MATLAB programming language. Before computing the findings, the intercept term is included in the model once the explanatory variables have been normalized using unit length scaling. The findings are obtained using an iterative re-weighted least square (IRLS) approach, and the convergence criterion is $||\hat{\beta}^{(t+1)} - \hat{\beta}^{(t)}|| \leq 1 \times 10^{-6}$. An ordinary least-squares (OLS) estimator is considered as an initial estimate. For $X^\top \hat{W}_{ML} X$, the eigenvalues are computed as $\lambda_1 = 98.6908, \lambda_2 = 2.2452, \lambda_3 = 1.6254, \lambda_4 = 1.2299, \lambda_5 = 0.9730$. The existence of multicollinearity is indicated using condition number (CN) defined as; $CN_{ML} = \sqrt{\frac{\lambda_{\max}(X^\top \hat{W}_{ML} X)}{\lambda_{\min}(X^\top \hat{W}_{ML} X)}} = 101.4284$. A serious multicollinearity problem in the data is indicated by the CN value, which is significantly greater than 30.

The performance of the estimators in the presence of multicollinearity is evaluated by imposing exact linear constraints on the parameters. Since we do not have prior knowledge of the data, we must establish a constraint on the parameter, taking into account the ML estimator given in Table 1. The restriction is established using the ML estimator in the final iteration as $\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 = 1.8373$, which is approximately true for $\hat{\beta}_{ML}$. This restriction gives; $H = [11111], h = 1.8373$. The percentage of total variation (PTV) approach, $PTV = \frac{\sum_{j=1}^{r} \lambda_j}{\sum_{j=1}^{q} \lambda_j} \times 100$, is used to determine the number of PCs to retain where r is the number of PCs that the model will retain. Although there is no exact threshold value for selecting the PTV, it is randomly selected to be 0.95, resulting in $r = 2$.

To choose the optimal value of the biasing parameter d, we used the PSO technique explained in Section 6, which gives $d = 0.4528$.

Table 1 presents the results of iteratively obtained coefficients of the estimators along with their scalar mean square error (SMSE)[§] values. It is evident that the proposed restricted $r - d$ class estimator has a lower SMSE value than its counterparts, suggesting that it outperforms other estimators when multicollinearity is present. Thereafter, the ML estimator has smaller SMSE value followed by the $r - d$ class estimator. In addition, Figure 1 is provided to evaluate the performance for all d values in the interval $(0, 1)$. Figure 1 illustrates that the proposed restricted $r - d$ class estimator performs best and achieves the least SMSE values when the biasing parameter d has a value between 0 and 0.78 for this data set.

[§]Scalar mse is the trace of the matrix mse where the parameter β is replaced with the ML estimator obtained at the end of the convergence to achieve unbiasedness

Table 1. Iteratively estimated coefficients and the SMSE values of the FOA estimators when $d = 0.4528$ for the mine data

Coefficients	ML	PCR	$r-d$	$Rest.-rd$
β_0	0.5646	2.0298	0.9759	0.9834
β_1	-1.5241	-0.4849	-0.2188	0.1653
β_2	4.6499	-0.0941	-0.0371	0.1816
β_3	-0.3114	-0.7174	-0.3222	0.2850
β_4	-1.5417	-0.5658	-0.2550	0.2010
SMSE	2.9115	25.9473	12.7531	2.0846

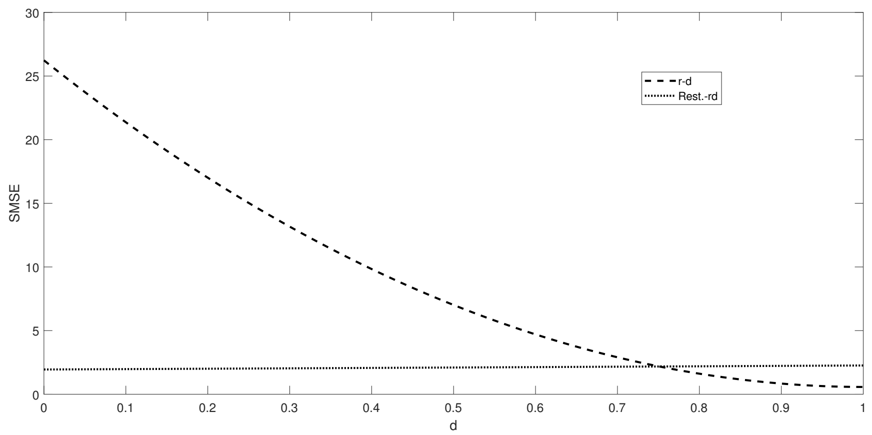

Figure 1. Graph of the SMSE values of the $r-d$ and $Rest.-rd$ class estimators for mine data

This numerical example demonstrates that the optimal performance of the proposed estimator is achieved for appropriately selected values of the biasing parameter d.

8. Simulation Studies

In this section, simulation studies are conducted for the Poisson, Binomial, and Negative binomial response variables to evaluate the performance of the estimators; ML, PCR, $r-d$ class, and restricted $r-d$ class estimators. The performance of these estimators is assessed using the estimated mean squared error (EMSE) criterion, calculated as follows:

$$EMSE(\tilde{\beta}) = \frac{1}{MCN} \sum_{m=1}^{MCN} \left(\tilde{\beta}_{(m)} - \beta\right)^{\top} \left(\tilde{\beta}_{(m)} - \beta\right)$$

where $\tilde{\beta}_{(m)}$ is the estimate of β, and MCN is the number of replications in the Monte Carlo simulation experiment, which is repeated up to 500 times. The subscript m indicates the m-th replication of the simulation experiment. The MATLAB programming language is used to analyze our results. The subsequent steps required to perform the Monte Carlo simulation experiments for Poisson, Binomial, and Negative binomial response variables are outlined in Subsections 8.1, 8.2 and 8.3.

8.1. Experiment 1: Poisson response

This experiment is designed for the Poisson response. The necessary steps are detailed as follows:

1. The number of explanatory variables utilized is $p = 4, 8$ and the sample size is $n =$ 25, 50, 100, 200, 300, and 500.

2. By following [18], the explanatory variables are created as:

$$x_{ij} = (1 - \gamma^2)^{1/2} \upsilon_{ij} + \gamma \upsilon_{i,p+1}, i = 1, \cdots, n, j = 1, \cdots, p,$$

where υ_{ij} are independent standard normal pseudo-random numbers and γ^2 indicates the degree of multicollinearity between any two explanatory variables. Prior to computing the findings, the explanatory variables are standardized using the unit length standardization technique.

3. $\gamma^2 = 0.90$, 0.95, and 0.99 are the values of multicollinearity degree that are taken into consideration.

4. To ensure that $\beta^\top \beta = 1$, the parameter vector β is computed as a normalized eigenvector that corresponds to the largest eigenvalue of the $X^\top X$ matrix (see [10]).

5. By following [16] the exact linear restrictions for $p = 4$ and $p = 8$ are considered as $H = \begin{bmatrix} 1 & 0 & -2 & 1 \end{bmatrix}$, $h = [0]$, $H = \begin{bmatrix} 1 & 0 & -2 & 1 & -3 & 1 & 1 \end{bmatrix}$, $h = [0]$.

6. The PSO approach described in Section 6 is used to find the values of the biasing parameter d.

7. We used the percentage of total variation (PTV) method to calculate the number of PCs.

8. The Poisson distribution yields the response variable $y_i \sim P(\mu_i)$ with the log-link function $\mu_i = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip})$.

9. Initially OLS estimator $\hat{\beta}_{ols} = \beta^{(0)} = \left(X^\top X\right)^{-1} X^\top y$ is taken as an estimate of β.

The findings are given in Tables 2 and 3, and described as follows:

i) The findings indicate that the EMSE values of the restricted $r-d$ class estimator are less than those of its counterparts for every parameter taken into account in the simulation research. Thus, the performance of the restricted $r-d$ class estimator is better compared to its counterparts for all factors included in the simulation analysis.

ii) For $p = 4$, the restricted $r-d$ class estimator has smaller EMSE values than those for $p = 8$.

iii) The $r-d$ class estimator outperforms both the ML and PCR estimators, while the PCR estimator outperforms the ML estimator.

iv) The EMSE values of the restricted $r-d$ class estimator increase with the degree of multicollinearity, with the exception of $p = 4$ and $\gamma^2 = 0.99$. In contrast, when the degree of multicollinearity γ^2 increases, the EMSE values of all other estimators increase.

8.2. Experiment 2: Binomial response

In this section, a simulation experiment is carried out for a binomially distributed response variable to assess the applicability and performance of the estimators. The response variable for the logistic regression model is generated as $y_i \sim Bern(\mu_i)$, where $\mu_i = \frac{\exp(\beta_1 x_{i1} + \cdots + \beta_p x_{ip})}{1+\exp(\beta_1 x_{i1} + \cdots + \beta_p x_{ip})}$, $i = 1,\ldots,n$ and β_j, $j = 1, \ldots, p$ as defined in Step 4. The initial weight matrix for the $i-$th observation is defined as $W^{(0)} = diag(\mu_i^{(0)}(1 - \mu_i^{(0)}))$. The remaining steps follow the same procedure used for the Poisson response case.

The simulation results are presented in Tables 4 and 5 and discussed below.

i) The restricted $r-d$ class estimator consistently outperforms its counterparts in all the parameters considered, as it yields the smallest EMSE values.

ii) When $p = 4$ EMSE values of the proposed restricted $r - d$ class estimator increase as the degree of multicollinearity increases from 0.90 to 0.95, but then decreases as it reaches 0.99. A similar trend is observed for $p = 8$ and sample sizes $n = 25$ and $n = 50$, where EMSE values increase from a multicollinearity level of 0.90 to 0.95, and then decrease as it approaches 0.99. However, for the remaining sample sizes, it is observed that as the degree of multicollinearity increases from 0.90 to 0.99, the EMSE values of the proposed estimator increase consistently.

iii) On the other hand, when $p = 4$, the EMSE values of the other estimators increase as the degree of multicollinearity rises. This pattern also holds for $p = 8$ in the case of the ML and PCR estimators across all sample sizes, while for the $r - d$ class estimator, it holds for all sample sizes except $n = 25$ and $n = 50$.
iv) It is observed that the EMSE values of the estimators are small for $p = 4$ compared to $p = 8$.
v) As the sample size changes, the EMSE values of the estimators are found to follow an increasing or decreasing trend.

8.3. Experiment 3: Negative binomial response

In this experiment response variable follows a Negative binomial distribution and is generated as $y_i \sim NB\left(\mu_i, \mu_i + \alpha\mu_i^2\right)$ where $\mu_i = \exp\left(\beta_1 x_{i1} + \cdots + \beta_p x_{ip}\right)$, $i = 1, \ldots, n$, we choose $\alpha = 1$ by following [8]¶. The weight matrix for the $i-$th observation is computed as $\hat{W} = diag(\mu_i + \alpha\mu_i^2)$. The remaining steps are consistent with those employed for the Poisson and binomial responses.

The corresponding results are presented in Tables 6 and 7 and discussed as follows.

i) The findings reveal that the restricted $r - d$ class estimator demonstrates superior performance, achieving the lowest EMSE values across all scenarios considered in the simulation experiment.
ii) As multicollinearity increases from 0.90 to 0.95, the EMSE values of the restricted $r - d$ class estimator generally decrease, except for the case when $n = 25$ and $p = 4$. At a higher multicollinearity level of 0.99, the EMSE values display a mixed behavior, with both increases and decreases observed. For $p = 8$, however, the EMSE values consistently decrease as multicollinearity rises from 0.90 to 0.99, except when $n = 300$ and $n = 500$.
iii) For the PCR and $r - d$ class estimators, it is observed that their EMSE values generally decrease with increasing multicollinearity when $p = 4$, except in the case of $n = 25$; a similar pattern is noted for $p = 8$. In contrast, the EMSE values of the ML estimator consistently increase as the degree of multicollinearity increases for both $p = 4$ and $p = 8$.
iv) With an increase in the number of variables, the EMSE values of all estimators generally increase, except for the proposed restricted $r - d$ class estimator, which shows a varying trend decreasing or increasing.
v) The sample size does not exhibit a consistent pattern, as the EMSE values fluctuate, showing either an increasing or decreasing trend.

¶Alternatively, we can also choose $\alpha = 2$ however, the results for $\alpha = 2$ follow the same pattern as those for $\alpha = 1$

Table 2. The EMSE values of the estimators when response is Poisson and $p = 4$

n	ρ^2	ML	PCR	$r-d$	$Rest.-rd$
25	0.90	16.5459	16.5245	5.9896	**3.8747**
	0.95	31.6022	28.6427	10.4592	**6.7716**
	0.99	150.4122	31.9371	13.4802	**2.5389**
50	0.90	19.3317	19.3053	7.7935	**5.0963**
	0.95	38.3345	38.3340	14.4728	**9.5887**
	0.99	178.6639	55.8333	21.5054	**0.2163**
100	0.90	17.7548	17.7544	6.5862	**4.6129**
	0.95	34.4845	34.4841	12.4235	**7.9209**
	0.99	159.7253	51.5610	17.8851	**0.2628**
200	0.90	18.7464	18.7460	7.1991	**4.9793**
	0.95	34.3699	34.3609	11.4953	**7.5452**
	0.99	179.8283	50.7470	17.3068	**0.2137**
300	0.90	17.1506	17.1503	6.8819	**4.5848**
	0.95	32.6015	32.6010	11.9064	**7.6334**
	0.99	168.2958	50.0797	16.1720	**0.2158**
500	0.90	17.2189	17.2180	6.6679	**4.5034**
	0.95	34.9920	34.7900	12.4191	**8.1064**
	0.99	165.8687	47.1490	16.0628	**0.2212**

Table 3. The EMSE values of the estimators when response is Poisson and $p = 8$

n	ρ^2	ML	PCR	$r-d$	$Rest.-rd$
25	0.90	68.9429	30.2201	11.5976	**9.8546**
	0.95	138.6119	36.3581	13.1340	**10.6035**
	0.99	677.4506	25.7664	2.1115	**1.0228**
50	0.90	51.5821	36.9222	14.0309	**12.0081**
	0.95	94.0140	57.4648	23.1516	**19.2515**
	0.99	490.4639	109.8739	24.5933	**8.8887**
100	0.90	33.2936	33.2869	12.6595	**11.4244**
	0.95	64.4132	49.3777	18.3520	**15.6584**
	0.99	315.1606	110.0131	37.9837	**27.6827**
200	0.90	31.2520	31.2510	12.2906	**10.3662**
	0.95	59.6384	53.2279	19.7053	**16.4349**
	0.99	295.5698	124.6166	41.8809	**27.7601**
300	0.90	40.1266	40.1266	15.2064	**13.2830**
	0.95	83.1414	68.4822	24.2125	**20.1188**
	0.99	381.1883	127.1909	41.9745	**26.0196**
500	0.90	36.6968	36.3280	14.5676	**12.6982**
	0.95	71.0364	67.6588	24.8313	**21.0111**
	0.99	326.7051	132.0536	46.2041	**32.1742**

Table 4. The EMSE values of the estimators when response is Binomial and $p = 4$

n	ρ^2	ML	PCR	$r-d$	$Rest.-rd$
25	0.90	188.4481	188.5039	60.3603	**39.4919**
	0.95	369.3163	369.9402	118.1156	**75.9407**
	0.99	1819.6149	216.3051	105.2365	**1.3695**
50	0.90	118.8548	118.8549	44.5335	**30.1796**
	0.95	228.9720	228.9721	84.2046	**56.2669**
	0.99	1109.5102	208.7408	84.8605	**0.7557**
100	0.90	85.6377	85.6379	29.9775	**18.8872**
	0.95	166.2209	166.2211	56.9087	**36.2202**
	0.99	810.0375	188.5690	62.7268	**0.7147**
200	0.90	79.5157	79.5158	26.8746	**17.4624**
	0.95	152.2272	152.2273	49.9723	**32.1026**
	0.99	737.0525	190.1574	59.6078	**0.6996**
300	0.90	71.8987	71.8988	26.1721	**16.8905**
	0.95	141.4898	141.4899	50.6768	**32.0691**
	0.99	698.1147	200.8021	68.4204	**0.6629**
500	0.90	73.3194	73.3195	29.1282	**21.2918**
	0.95	141.6855	141.6856	54.9378	**39.8005**
	0.99	684.5109	192.0477	72.9669	**0.7393**

Table 5. The EMSE values of the estimators when response is Binomial and $p = 8$

n	ρ^2	ML	PCR	$r-d$	$Rest.-rd$
25	0.90	3832.5605	22.3145	54.9806	**5.4799**
	0.95	8053.4831	45.5218	127.6206	**6.0366**
	0.99	38566.1217	224.7185	84.7523	**1.6138**
50	0.90	380.0216	241.0176	92.6872	**79.5904**
	0.95	757.1590	477.4244	180.9736	**155.5399**
	0.99	4191.5981	512.8195	124.0465	**63.1969**
100	0.90	178.3174	178.3175	62.9701	**55.9302**
	0.95	341.5112	258.3903	92.5900	**79.5015**
	0.99	1644.7933	450.1622	170.4996	**115.8137**
200	0.90	140.9834	140.9835	49.3190	**41.9409**
	0.95	273.3667	218.7013	75.1796	**61.6703**
	0.99	1325.5761	447.2790	150.7823	**104.8773**
300	0.90	189.7297	189.7298	64.9465	**56.9265**
	0.95	369.0343	295.8827	100.2390	**84.8610**
	0.99	1803.6438	599.0036	201.7927	**130.6163**
500	0.90	157.1340	157.1340	56.0308	**47.4147**
	0.95	303.7187	303.7188	106.0373	**89.6855**
	0.99	1486.7207	541.0760	183.8575	**127.9551**

Table 6. The EMSE values of the estimators when response is Negative binomial, $p = 4$ and $\alpha = 1$

n	ρ^2	ML	PCR	$r-d$	$Rest.-rd$
25	0.90	69.4873	43.1531	8.6476	**7.6212**
	0.95	162.9744	52.6549	16.7881	**15.9914**
	0.99	708.0939	34.1665	15.0108	**5.4217**
50	0.90	83.0232	41.1175	11.3281	**2.6799**
	0.95	158.6553	14.4143	4.2874	**0.5912**
	0.99	647.6859	10.7762	2.7652	**0.3999**
100	0.90	81.0309	35.7251	11.9641	**2.4192**
	0.95	155.6804	6.5941	2.7355	**0.5863**
	0.99	722.3650	6.2276	2.7586	**0.5860**
200	0.90	99.7542	32.5093	12.3103	**1.6208**
	0.95	189.4241	6.5331	3.2064	**0.5966**
	0.99	900.5217	6.4502	3.1421	**0.6015**
300	0.90	88.6472	37.3344	14.4888	**3.9345**
	0.95	170.2598	6.2827	3.5228	**0.6259**
	0.99	822.4616	6.1366	3.4863	**0.6330**
500	0.90	92.3056	41.2066	16.4297	**5.4871**
	0.95	173.3877	5.9433	4.0116	**0.6602**
	0.99	816.7623	5.8526	3.9852	**0.6652**

Table 7. The EMSE values of the estimators when response is Negative binomial, $p = 8$ and $\alpha = 1$

n	ρ^2	ML	PCR	$r-d$	$Rest.-rd$
25	0.90	572.8496	59.4823	24.0669	**10.7441**
	0.95	1169.2343	50.7189	26.7411	**7.5421**
	0.99	1719.6149	116.3051	55.2365	**3.3695**
50	0.90	221.0540	82.8581	16.3831	**10.9310**
	0.95	411.6614	19.7837	4.7562	**1.8126**
	0.99	2060.8819	9.7820	3.1759	**0.7557**
100	0.90	149.7130	85.3499	21.3825	**21.1825**
	0.95	279.1197	21.6654	5.6251	**1.0958**
	0.99	1363.4064	8.6090	2.0891	**0.4463**
200	0.90	148.7087	112.2574	36.5111	**35.8360**
	0.95	280.2627	19.0209	6.9569	**0.8321**
	0.99	1348.8521	7.1567	2.9124	**0.5359**
300	0.90	193.7352	92.1232	29.0334	**21.6329**
	0.95	373.2256	5.9993	3.3852	**0.5618**
	0.99	1897.2516	5.8532	3.2694	**0.5638**
500	0.90	189.1891	96.1145	36.2378	**25.5231**
	0.95	361.7255	5.3087	3.0706	**0.6607**
	0.99	1750.7123	5.2572	3.0301	**0.6667**

9. Conclusion

This study develops a novel estimator that is subject to exact linear restrictions and is called the restricted $r - d$ class estimator. Furthermore, to estimate the biasing parameter d, a PSO algorithm is used. Through real-life data with Poisson response and simulation studies from Poisson, binomial, and negative binomial distributions, the novel estimator is compared to other estimators that currently exist that are ML, PCR and $r-d$ class estimators, respectively. The performance evaluation criteria are SMSE and EMSE, respectively, for the numerical example and the simulation study.

By comparing the proposed estimator with others considered in this study, it shows that in the numerical illustration, the proposed restricted $r - d$ class estimator acquires smaller SMSE values and outperforms its counterparts, particularly when the value of d lies in the range from 0 to 0.78. When simulation experiments are taken into account to compare the proposed estimator with other estimators, it is seen that the proposed restricted $r - d$ class estimator has the best performance compared to its counterparts, as it obtains smaller EMSE values in all aspects considered in the simulation experiments.

The findings obtained in this study highlight that imposing exact restriction on the parameters, i.e. incorporating prior knowledge along with the sample data, significantly improves the performance of the estimators over those available in the literature to combat the multicollinearity problem, thereby improving the model performance.

Acknowledgements

The authors thank the editor and referees for their valuable comments and suggestions which have significantly improved the quality of this article.

Author contributions. Both authors contributed equally to the preparation of this manuscript.

Conflict of interest statement. The authors declare that they have no conflicts of interest or competing financial interests.

Funding. The authors declare that no funds or grants were received during the preparation of this manuscript.

Data availability. The data used in the numerical example can be found from [17].

References

[1] A. Abbasi and M.R. Özkale, 2024. *Iterative stochastic restricted $r-d$ class estimator in generalized linear models: application to binomial, Poisson and negative binomial distributions*, Hacet. J. Math. Stat. **53**(2), 1419-1437, 2024.

[2] A. Abbasi and M.R. Özkale, 2025. *Iterative Stochastic Restricted r-k Class Estimator in Generalized Linear Models: Application on Logistic Regression*, Iran. J. Sci. **49**, 357367, 2025.

[3] A. Abbasi and M.R. Özkale, 2021. *The r-k class estimator in generalized linear models applicable with simulation and empirical study using a Poisson and Gamma responses*, Hacet. J. Math. Stat. **50**(2), 594-611, 2021.

[4] A. Abbasi and M.R. Özkale, 2023. *Restricted Liu estimator under stochastic linear restrictions in generalized linear models: theory and applications*, Commun. Stat. Simul. Comput. **54**(5), 4001422.

[5] Q. Abdul Kareem and Algamal, Z. 2020. *Generalized ridge estimator shrinkage estimation based on particle swarm optimization algorithm*, Iraqi Journal of Statistical Sciences, **17**(2), 26-35.

[6] A.J. Dobson and A. G. Barnett, An introduction to generalized linear models. Chapman and Hall/CRC, 2018.

[7] R. C. Eberhart and Y.Shi, Particle swarm optimization: developments, applications and resources.In Proceedings of the 2001 congress on evolutionary computation (IEEE Cat. No. 01TH8546) (Vol. 1, pp. 81-86). IEEE, 2021

[8] J. Huang and H. Yang, A two-parameter estimator in the negative binomial regression model. J. Stat. Comput. Simul, 84(1), 124-134, 2014.

[9] D. Inan, E. Egrioglu, B. Sarica, S. Eichberg, O.E., Askin and M. Tez, *Particle swarm optimization based Liu-type estimator.* Commun. Stat. Theory Methods., **46**(22), 11358-11369, 2017.

[10] B.M.G. Kibria, *Performance of some new ridge regression estimators.* Commun. Stat. Simul. Comput 32(2), 419-435, 2003.

[11] F. Kurtoğlu and M. R. Özkale, *Restricted ridge estimator in generalized linear models: Monte Carlo simulation studies on Poisson and binomial distributed responses*, Commun. Stat. Simul. Comput. **48**(4), 1-28, 2019.

[12] 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.

[13] J. Kennedy and R. Eberhart, *Particle swarm optimization. In Proceedings of ICNN'95-international conference on neural networks.* ISPRS int. j. geo-inf., 1942-1948, 1995.

[14] Y. Lamari, B. Freskura, A.Abdessamad, S. Eichberg and S. de Bonviller, *Predicting spatial crime occurrences through an efficient ensemble-learning model*, ISPRS int. j. geo-inf, **9**(11), 645, 2020.

[15] B.D. Marx, *A continuum of principal component generalized linear regressions.*Comput. Stat. Data Anal. **13**(4), 385-393, 1992.

[16] K. Mansson and B.MG. Kibria, *Estimating the unrestricted and restricted Liu estimators for the Poisson regression model: Method and application*, Comput. Econ., **58**, 311-326, 2021.

[17] R.H. Myers, *Classical and modern regression with applications*, Belmont, CA: Duxbury press, 1990.

[18] G.C. McDonald, D.I. Galarneau, *A monte carlo evaluation of some ridge-type estimators*, J. Am. Stat. Assoc. **70**(350), 407-416, 1975.

[19] H. Nyquist, *Restricted Estimation of Generalized Linear Models*, J. R. Stat. Soc.Ser.C. **40**(1), 133-141, 1991.

[20] M.R. Özkale and E. Arıcan, *First-order* $r-d$ *class estimator in binary logistic regression model*, Stat Probabil. Lett. 106, 19-29, 2016.

[21] 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.

[22] M.R. Özkale and H. Nyquist, *The stochastic restricted ridge estimator in generalized linear models*, Stat. Pap. **62**(3), 1421-1460, 2021.

[23] M.R. Özkale and A. Abbasi, *Iterative restricted OK estimator in generalized linear models and the selection of tuning parameters via MSE and genetic algorithm*, Stat. Pap. 1979-2040, 2022.

[24] Rao and Toutenburg, *Linear Models: Least Squares and Alternatives* Springer, New York, 1995.

[25] 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.

[26] N. Sancar and D. Inan, *A new alternative estimation method for Liu-type logistic estimator via particle swarm optimization: an application to data of collapse of Turkish commercial banks during the Asian financial crisis*, Appl. Stat. **48**, 2499-2514, 2021.

[27] M. Settles, *An introduction to particle swarm optimization.* Department of Computer Science, University of Idaho, **2** 12, 2005.
[28] V.R. Uslu, E. Egrioglu, and E. Bas, *Finding optimal value for the shrinkage parameter in ridge regression via particle swarm optimization*, Am. J. Intell. Syst., **4**, 142-147, 2014.
[29] D. Wang, D. Tan and L. Liu, *Particle swarm optimization algorithm: an overview.*, Soft Comput. **22**(2), 387-408, 2018.
[30] K. Wiktorowicz, T. Krzeszowski and K. Przednowek, *Sparse regressions and particle swarm optimization in training high-order TakagiSugeno fuzzy systems*, Neural Comput. Appl., **33**(7), 2705-2717, 2021.