Social Sciences Spectrum
A Double-Blind, Peer-Reviewed, HEC recognized Y-category Research Journal
E-ISSN: 3006-0427 **P-ISSN:** 3006-0419
Volume 04, **Issue** 02, 2025
Web link:https://sss.org.pk/index.php/sss

Using Independent Components for Estimating Logistic Regression with High-Dimensional Multicollinear Data: Simulation and Application

Sana Ali
M.Phil Scholar, Department of Statistics Bahauddin Zakariya University, Multan, Punjab, Pakistan
Correspondence: sanaali112295@yahoo.com

Saima Afzal
Assistant Professor, Department of Statistics Bahauddin Zakariya University, Multan, Punjab, Pakistan
Email: saimaafzalbzu@bzu.edu.pk

Nasir Saleem
Ph.D Scholar, Department of Statistics Bahauddin Zakariya University, Multan, Punjab, Pakistan
Email: nasirsaleem160@gmail.com

Article Information [YY-MM-DD]

Received 2025-02-28 **Accepted** 2025-04-29

Citation (APA):

Ali, S., Afzal, S & Saleem, N. (2025). Using independent components for estimating logistic regression with high-dimensional multicollinear data: Simulation and application. *Social Sciences Spectrum*, *4*(2), 288-316. https://doi.org/10.71085/sss.04.02.271

Abstract

The logistic regression model is used to predict a binary response variable in terms of a set of explicative ones. In the presence of multicollinearity among predictor, the estimation of the model parameters is not very accurate and their interpretation in terms of odds ratios may be inaccurate. Another important problem is that usually a large number of predictors are required to explain the response. In order to improve the estimation of the logistic model parameters under multicollinearity and to reduce the dimensions of the data with continuous covariates, it is proposed to use as covariates of the logistic model a reduced set of optimum independent components of the original predictors. Breast cancer data is used as real data set. The performance of the proposed independent component logistic regression model is analyzed by developing a simulation study where different methods for selecting the optimum independent components are compared. We built up a simulation study to illustrate the performance of the model with different regressors, sample size, and correlation among the regressors. Independent component logistic regression compared with principal component logistic regression model and independent component logistic regression gives better results.

Keywords: Dimension reduction, Independent components, Logistic regression, Multicollinearity, Breast cancer.

1. Introduction

It is essential to predict a binary response variable in various fields of study, for example, medication and the study of disease transmission or comparably the likelihood of occurring of an event achievement, regarding the estimations of a lot of explicatory factors identified with it. The Logistic Regression (LR) serves perfectly and is the most utilized for various cases as should be obvious for instance in Prentice and Pyke (1979). As many authors stated that (Hosmer and Lemeshow, 1989, Ryan, 1997) among others, the Logistic Model (LM) becomes unbalanced when predictors are highly correlated among themselves, so it seems that no one variable is important when all the others are in the model. In this case the estimation of the parameters of the model computed by R-software. As a result, the interpretation of the relationship between the response and each explanatory variable in terms of odds ratios may be invalid. Dependent variable is dichotomous and binary in LR, i.e., data coded values are utilized as in tumor case 1 is utilized for dangerous or 0 is utilized for not. At the point when the dependent variable is dichotomous then LR is the suitable regression examination to lead. In LR model predictor variables are highly correlated then there is multicollinearity. During the estimation of linear or generalized linear model including LR and Cox regression, multicollinearity is a common problem. To detect multicollinearity correlation matrix may be helpful but not sufficient in LR. If the model satisfies the logistic assumptions, then the analysis is valid and if the model does not satisfy the logistic assumption, then there is some problem in the model. For the LR coefficients there are biased coefficient estimates or large standard errors may present due to multicollinearity and invalid statistical inferences are produced. By using the model for any statistical inference, watch that model fits adequately well. When the model includes multiple factors that are correlated not just to response variable but also with each other then it refers multicollinearity in LR. In LR model predictor variables are highly correlated then there is multicollinearity. If we have a huge number of variables, then there exist a high dimensional issue and it can be resolve by reducing the number of variables. The number of variables can be reduced using dimension reduction techniques like Principal Component Analysis (PCA) or Independent Component Analysis (ICA).

2. Objective of the Study

The main objectives of the research are to overcome the problem of multicollinearity among the explanatory variables in LR model. Propose a new method to overcome the problem of multicollinearity by using a relatively new Blind Source Separation (BSS) technique Independent Components Analysis (ICA) and compare the proposed technique with Principal Components Logistic Regression (PCLR).

3. Organized of the paper

The current study is organized as section wise in section 4 discussed the literature Review, in section 5 discussed advantages and dis-advantages of the proposed study. In section 6 define a basic theory on logistic regression. In section 7 discussed independent component logistic regression (ICLR) Model and in section 8 detailed Handling Multicollinearity using ICLR. In section 9 define an Application of the Proposed ICLR Model and last section 10 conclusion.

4. Literature Review

Schaefer et al. (1984) proposed a ridge logistic estimator. Cessie , Houwelingen (1992) designed ridge estimators in LR. It is shown how ridge estimators can be used in LR to improve the parameter estimates and to diminish the error made by further predictions. Steyerberg et al. (2001) analyzed a case study in which application of shrinkage methods in LR analysis was explained.

Hubert and Wijekoon (2006) showed the improvement of the Liu estimator in linear regression model. Mansson and Shukur (2011) make research in LR on ridge parameters. This paper applies and explores various Logistic Ridge Regression (LRR) parameters that were observable by utilizing the method of ML. Bastien et al. (2005) determined about the PLS generalized linear regression. Aguilera et al. (2007) discussed the useful PLS logit regression model. Aguilera et al. (2006) determined that what could be done by utilizing PCs for calculating LR with high-dimensional multicollinear data. Escabias et al. (2004) described the PC estimation of functional LR and discussed two different approaches. Over the last few years many methods have been developed for analyzing functional data with different objectives. Aguilera et al. (2005) explained about the modeling environmental data by functional PCLR. Morillo et al. (2013) examined for functional logit regression the penalized spline approaches. Zhou et al. (2014) investigated that the face recognition depends on PCA and LR analysis. Agyekum et al. (2023) determined about the impact of sample size on multicollinearity with high dimensional data in LR analysis.

5. Advantages and Dis-advantages of the Proposed Study

In logistic regression model there is multicollinearity among predictors. Due to the multicollinearity the estimation of the model parameters is not very accurate. The major advantage of the proposed study is to reduce multicollinearity and also independent component analysis helps to reduce dependence among predictors and gives better results.

6. Basic theory on logistic regression

We formulate the model to establish the theoretical framework about LR by estimating its parameters and testing its goodness of fit. A best fitted model is found by the LR that is utilized to clarify the connection between one dependent binary factor and at least one nominal, ordinal, interval or ratio level independent factors and that is the fundamental objective of LR. LR model is given as below,

$$Y_i = \pi_i + \varepsilon_i \quad \text{i=1, 2, …, n} \qquad \textbf{(1)}$$

Ygiven $(X_1 = x_{i1}, X_2 = x_{i2}, \ldots, X_p = x_{ip})$ is the expectation of π_i that can be modelized as

$$\pi_i = P\{Y = 1 | X_1 = x_{i1}, \ldots, X_p = x_{ip}\} = \frac{\exp\{\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j\}}{1+\exp\{\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j\}} \qquad \textbf{(2)}$$

Where parameters of the model are $\beta_0, \beta_1, \ldots, \beta_p$ and variances of $\mathcal{E}_i$ are

$$Var[\mathcal{E}_i] = \pi_i(1 - \pi_i) \quad , i = 1, \ldots, n$$

Coefficients of a method to predict a logit change of the probability of occurrence of the qualities of concern are created by the LR, which is as follows,

$$Logit(p) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_pX_p$$

Probability of occurrence of qualities of interest is p and logit transformation are written in terms of logged odds,

$$Odds = (\frac{p}{1-p}) = \frac{\text{Probability of presence of characteristics of interest}}{\text{Probability of absence of characteristics of interest}}$$

and

$$logit(p) = \ln(\frac{p}{1-p})$$

As link function a Generalized Linear Model (GLM) with logit transformation can be expressed as LR. To return a probability value LR converts its output using the logistic sigmoid function. Dependent variable structure, perception independence, non-appearance of multicollinearity, linearity of autonomous factors and log odds, and enormous sample size are the assumptions of LR.

The mostly used for the estimation of LR is maximum likelihood. Let $L(Z:\beta)$ be the likelihood specified by

$$L(Z:\beta) = \prod_{j=1}^{m} \pi_j^{z_j}\left(1-\pi_j\right)^{1-z_j} \qquad \textbf{(3)}$$

In multicollinearity maximum likelihood does not give accurate results. Firstly, an indicator of multicollinearity in LR is selected. The pairwise correlations and VIF may be utilized when the regressors are all continuous and if various predictors are not continuous then the problem of multicollinearity will be increased. Multicollinearity is a problem that exists when variables are highly correlated to each other, and it can be detected easily by displaying the correlation matrix of continuous independent factors. In regression analysis model, overfitting is the main danger that occurs due to data redundancy. The best reversion models are those in which the predictor factors each associate exceedingly with the dependent result variable however connect all things considered just insignificantly with one another.

7. Independent component Logistic Regression (ICLR) Model

Firstly, we describe the ICA of a set of variables and its properties. Secondly, we will formulate the ICLR model then describe the simulation scheme and use different methods to select the IC's explanatory variables in the ICLR model.

7.1. Independent component analysis (ICA)

A statistical latent variables model can be used to thoroughly describe ICA. Suppose that we examine r random variables $x_{1t}, x_{2t}, \ldots, x_{rt}$ and these are modeled as linear combination of m random variables$s_{1t}, s_{2t}, \ldots, s_{mt}$whereas $m \leq r$, after that

$$x_{it} = w_{i1}s_{1t} + w_{i2}s_{2t} + \cdots + w_{im}s_{mt}, \qquad i = 1,2,\ldots,r$$

Whereas some real coefficients are $w_{i1}(i = 1,2,\ldots,r \; and \; j = 1,2,\ldots,m)$. non-Gaussian distributed components and statistically mutually independent components are s_{it} which are described by definition. This is the general ICA model.

Once the model has been estimated then its goodness of fit must be tested.

Let $M_p$denotes the specific logit model which is obtained by equating zero certain number l of parameters, $\beta_1, \ldots, \beta_l$selected among the p + 1 ones of model M. The likelihood statistic for the comparison of the model $M_p$to model Mtests the hypothesis that all parameters in model Mbut not in model $M_p$equal to zero. Then the conditional likelihood-ratio statistics for testing model M_p, given that M holds, is given by the difference in the G^2 goodness-of-fit statistics for the two models to be compared and is given as:

$$G^2(M_P/M) = 2(\int_M - \int_{M_P}) = G^2(M_P) - G^2(M) \quad \textbf{(4)}$$

For the simpler model M_P the $\int_{M_P}$ being the maximum log-likelihood that removes those l parameters. With df eqvivalent to the difference among the residual df for the two compared models in the fitted model M the number l of parameters equivalent to zero, it is a huge sampled statistic of chi-square.

7.2. ICLR Model Formulation

To characterize the model of ICLR we will prepare the logit model in words of all the IC's related to the observations of matrix X which are the continuous predictor variables. With no loss of simplification, we will assume that the regressors are centered. In terms of all the IC's the success probabilities of the logit model can be articulated as

$$\pi_i = \frac{\exp\{\beta_0 + \sum_{j=1}^{p}\sum_{k=1}^{p} z_{ik} v_{jk} \beta_j\}}{1+\exp\{\beta_0 + \sum_{j=1}^{p}\sum_{k=1}^{p} z_{ik} v_{jk} \beta_j\}} = \frac{\exp\{\beta_0 + \sum_{k=1}^{p} z_{ik} \gamma_k\}}{1+\exp\{\beta_0 + \sum_{k=1}^{p} z_{ik} \gamma_k\}}$$

Being the elements of the IC's matrix L=XV and $\gamma_k = \sum_{j=1}^{p} v_{jk}\beta_j$, k= 1,…,p with z_{ik} , (i=1,…,n; k=1,…,p). In terms of the logit transformation the logistic model can be equally determined in matrix form and the IC's as

$$L = X\beta = ZV'\beta = Z\gamma, \quad \textbf{(5)}$$

Wherever

$$Z = (\mathbf{1}|L)\ ,\left(\frac{1|0'}{0|v}\right), \quad \mathbf{0} = (0, \dots, 0)', \mathbf{1} = (1, \dots, 1)'.$$

Thus, as described in the sense of those of the model that contains as covariates all the IC's the parameters of the logit model can be attained. $\beta = V\gamma$. We have the prediction equation $\hat{Y} = \hat{\pi}$,

$$\hat{\beta} = V\hat{\gamma} \quad \textbf{(6)}$$

as a result of the invariance property of ML estimates.

In the case of collinearity to advance the original parameters estimation we will introduce the ICLR model. As covariates of the logit model by taking a reduced set of IC's of the original predictors the ICLR model is obtained.

Z and V Split matrices are given below

$$Z = \left(\begin{array}{cccc|ccc} 1 & z_{11} & \cdots & z_{1s} & z_{1s+1} & \cdots & z_{1p} \\ 1 & z_{21} & \cdots & z_{2s} & z_{2s+1} & \cdots & z_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & z_{n1} & \cdots & z_{ns} & z_{ns+1} & \cdots & z_{np} \end{array}\right) = \left(Z_{(s)} \middle| Z_{(r)}\right), (r = p - s)$$

and

$$V = \left(\begin{array}{cccc|ccc} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & v_{11} & \cdots & v_{1s} & v_{1s+1} & \cdots & v_{1p} \\ 0 & v_{21} & \cdots & v_{2s} & v_{2s+1} & \cdots & v_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & v_{p1} & \cdots & v_{ps} & v_{ps+1} & \cdots & v_{pp} \end{array}\right) = \left(V_{(s)} \middle| V_{(r)}\right).$$

Then$Z_{(s)} = XV_{(s)}$ and $Z_{(r)} = XV_{(r)}$, thus the original parameters can be articulated as

$$\beta = V\gamma = V_{(s)}\gamma_{(s)} + V_{(r)}\gamma_{(r)} ,$$

Wherever

$$\gamma = \left(\gamma_0\gamma_1 \cdots \gamma_s \middle| \gamma_{s+1} \cdots \gamma_p\right)' = \left(\gamma'_{(s)} \middle| \gamma'_{(r)}\right)'.$$

In terms of all the ICs the logit model given by Eq. (3.11), can be expressed as

$L = Z\gamma = Z_{(s)}\gamma_{(s)} + Z_{(r)}\gamma_{(r)},$the ICLR model as far as s ICs (ICLR$_{(s)}$) is acquired by removing the r last ICs in the last equation, so we had

$$y_i = \pi_{i(s)} + \varepsilon_{i(s)},$$

Wherever

$$\pi_{i(s)} = \frac{\exp\{\gamma_0 + \sum_{j=1}^{s} z_{ij}\gamma_j\}}{1 + \exp\{\gamma_0 + \sum_{j=1}^{s} z_{ij}\gamma_j\}}, \text{ i=1,…,n.}$$

In matrix form as far as the vector of logit conversion $L_{(s)} = \left(l_{1(s)}, \dots, l_{n(s)}\right)$ this model can be equivalently formulated with components $l_{i(s)} = ln\left(\pi_{i(s)} \middle/ \left(1 - \pi_{i(s)}\right)\right)$ as given:

$$L_{(s)} = Z_{(s)}\gamma_{(s)} = XV_{(s)}\gamma_{(s)} = X\beta_{(s)}.$$

As a result, a renovation of the original parameters has acquired known by$\beta_{(s)} = V_{(s)}\gamma_{(s)}$, as far as the ICLR model's parameters that has as covariates the first s ICs. An estimation of the original parameters β will be given by the ML estimation of this ICLR model which is described as

$$\hat{\beta}_{(s)} = V_{(s)}\hat{\gamma}_{(s)}, \qquad \textbf{(7)}$$

That will recover the$\hat{\beta}$ estimation attained by means of original variables if there should be an occurrence of multicollinearity. In the last one the estimator $\hat{\gamma}_{(s)}$ as far as the first s ICs are not the vector of the initial s components of the estimator $\hat{\gamma}$ as far as all the ICs that is the main variation among ICR and ICLR. Such as, $\hat{\gamma}_{(s)} = \left(\hat{\gamma}_{0(s)}, \hat{\gamma}_{1(s)}, \dots, \hat{\gamma}_{s(s)}\right)' \neq (\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_s)'$. As a result the ICLR(s) model estimated the probabilities $\hat{\pi}_{i(s)}$are diverse to the ones acquired in abbreviating the model's ML estimated probabilities that has all the ICs as regressors. This is,

$$\hat{\pi}_{i(s)} = \frac{\exp\{\hat{\gamma}_{0(s)} + \sum_{j=1}^{s} z_{ij}\hat{\gamma}_{j(s)}\}}{1 + \exp\{\hat{\gamma}_{0(s)} + \sum_{j=1}^{s} z_{ij}\hat{\gamma}_{j(s)}\}} \neq \frac{\exp\{\hat{\gamma}_0 + \sum_{j=1}^{s} z_{ij}\hat{\gamma}_j\}}{1 + \exp\{\hat{\gamma}_0 + \sum_{j=1}^{s} z_{ij}\hat{\gamma}_j\}}.$$

In computational exertion this implies an impressive extension because each time the model of ICLR has to be straightened out we remove or enter a innovative PC in the model.

7.3. Model Selection

In turn to attain best probable estimation of logit model parameters, for choosing the most favorable ICLR model we will utilize different criteria that was used by (Aguilera et al. (2006)) depends on various accuracy evaluations of the estimated parameters.

- Firstly, the Mean Squared Error of the Beta parameter vector (MSEB) is determined

$$MSEB_{(s)} = \frac{1}{p+1}\sum_{j=0}^{p}\left(\hat{\beta}_{j(s)} - \beta_j\right)^2, \qquad s = 1, \dots, p.$$

- Secondly, the maximum of the absolute differences of the beta parameters is defined as

$$Max_{(s)} = Max_j\{\left|\beta_{j(s)} - \beta_j\right|\}, s = 1, \dots, p.$$

However, the best estimation of the model probabilities is provided by the best estimation of the original parameters can be imagined.

Let us check that better estimation of the parameters will be indicated by the undersized values of mean squared error (MSE) and Max (Maximum). Latterly in the simulation study we will choose the optimum model with the smallest MSE and Max.

When the real data is analyzing by us at that time the estimated and real parameter's comparison is not possible, then we could not compute the MSE and Max and another measure of accuracy of the estimation is needed to be defined that does not consider the real unknown parameters. A number of authors, Aucott et al. (2000), among others, noted that in the linear regression to a bad estimation the variance of estimated parameters is very responsive.

8. Handling Multicollinearity Using ICLR

This section presents the application of the proposed ICLR model to tackle the problem of multicollinearity. A comparison of ICLR with Principal Component Logistic Regression Model (PCLR) (Aguilera, 2006) is also made.

8.1. The Simulation Scheme

We built up a simulation study to illustrate the performance of the proposed ICLR model and to demonstrate how the estimation of logit model parameters with collinear regressors can be improved by utilizing ICs.

The simulation design is carried out by choosing:

- Three different number of regressors i.e. $P = 6, 10, and\ 15$
- Three different sample sizes i.e. $n = 100, 200\ and\ 300$
- Three different levels of correlation among the regressors i.e. $\rho = 0.8, 0.9\ and\ 0.95$

In this way we obtain nine different tables each with above stated three different correlation levels and overall, twenty-seven sets of tables from one data set.

The cholesky decomposition method is applied to obtain P regressors with a known correlation structure. The P variables are generated as regressors from multivariate normal distribution with a specific correlation level.

As a second step, a vector of real parameters β is fixed. The vector of real parameters β is chosen as normalized eigen vectors corresponding to the largest eigen values of $X^t X$ so that $\beta^t \beta$ is equal to one following Newhouse and Oman (1971).

The binary response was simulated by pursuing the plan of the simulation studies created in conducted by Hosmer et al. (1997) and Pulkstenis and Robinson (2002). The real probabilities are computed by the model:

$$\pi = \frac{\exp(x'_i \beta)}{1 + \exp(x'_i \beta)}, \qquad i = 1, \ldots, n,$$

At last, with parameter $\pi_i, y_i \to B(\pi_i)(i = 1, \ldots, n)$ from a Bernouilli distribution each sample value of the binary response was simulated.

9. Application of the Proposed ICLR Model

For three distinctive numbers of regressors$(P = 6, 10\ and\ 15)$, three diverse sample sizes $(n = 100, 200\ and\ 300)$, three different correlations $\rho = 0.8, 0.9\ and\ 0.95$ and one unique distribution of the collinear regressors is used. According to this method 27 tables are given below and these are divided in different cases according to the number of regressors, sample sizes and correlation and each case contain 9 tables. Data set 1 contains three cases for $\rho = 0.8, P = 6$ and sample size is changed for every case and three cases for $\rho = 0.8, P = 10$ and sample sizes are $(n = 100, 200\ and\ 300)$ for case 1, 2 and 3 and three cases for $\rho = 0.8, P = 15$. Data set 2 contains three cases for $\rho = 0.9, P = 6$ and sample size is changed for every case and three cases for $\rho = 0.9, P = 10$ and sample sizes are $(n = 100, 200\ and\ 300)$ for every case and three cases for $\rho = 0.9, P = 15$. Data set 3 contains three cases for $\rho = 0.95, P = 6$ and sample size is changed for every case and three cases for $\rho = 0.95, P = 10$ and sample sizes are $(n = 100, 200\ and\ 300)$ for every case and three cases for$\rho = 0.95, P = 15$. All of these are given below.

9.1. Analysis of Datasets with $\boldsymbol{\rho = 0.8}$ Real and Estimated Parameters with $\boldsymbol{MSE}$ and $\boldsymbol{Max}$

Table 1: *For Case 1*, $\rho = 0.8, P = 6, n = 100$

Parameters	**Real**	**PCLR(All)**	**PCLR(3)**	**ICLR(All)**	**ICLR(3)**
β_1	0.4015	4.9827	0.9217	4.1090	0.2200
β_2	0.3216	10.1327	1.3316	6.3915	0.6916
β_3	0.5298	8.8181	0.9896	2.1919	0.9818
β_4	0.3389	9.2987	0.6984	3.2512	0.6184
β_5	0.5016	9.9984	1.1995	3.3214	0.7715
β_6	0.2929	3.2198	1.6719	3.2898	0.9898
MSE	---	60.8292	0.6700	13.1277	0.1685
Max	---	9.8111	1.3790	6.0699	0.6969

Table 2: *For Case 1*, $\rho = 0.8, P = 6, n = 200$

Parameters	**Real**	**PCLR(All)**	**PCLR(3)**	**ICLR(All)**	**ICLR(3)**
$\boldsymbol{\beta_1}$	0.4718	9.3129	0.9138	5.1423	0.9616
$\boldsymbol{\beta_2}$	0.5125	7.6814	0.9917	4.1193	0.7719
$\boldsymbol{\beta_3}$	0.3018	10.9815	1.2319	3.8716	0.9823
$\boldsymbol{\beta_4}$	0.2917	9.8917	0.9813	2.2219	0.2126
$\boldsymbol{\beta_5}$	0.4104	3.2419	0.9919	5.1910	0.5139
$\boldsymbol{\beta_6}$	0.4010	1.1286	0.9796	6.1010	0.5524
MSE	---	57.3868	0.4064	17.7726	0.1350
Max	---	10.6797	0.9301	5.7000	0.6805

Table 3: *For Case 1*, $\rho = 0.8, P = 6, n = 300$

Parameters	**Real**	**PCLR(All)**	**PCLR(3)**	**ICLR(All)**	**ICLR(3)**
$\boldsymbol{\beta_1}$	0.5612	7.1761	0.8167	3.2145	0.7682
$\boldsymbol{\beta_2}$	0.5013	6.5519	1.2219	2.2281	0.5903
$\boldsymbol{\beta_3}$	0.3391	5.2317	0.6719	1.6314	0.9384
$\boldsymbol{\beta_4}$	0.3418	6.6692	0.8517	5.7123	0.4615
$\boldsymbol{\beta_5}$	0.3916	6.0915	0.9819	6.2649	0.5109
$\boldsymbol{\beta_6}$	0.2116	6.6616	1.0012	4.7164	0.9348
MSE	---	36.4053	0.3212	15.8872	0.1602
Max	---	6.6149	0.7896	5.8733	0.7232

In table 1-3, the first three PCs account for the 95% variation, so we include three PCs as reduced number of variables to construct PCLR and three ICs for the construction of ICLR. The original parameters are then reconstructed, and the estimates of the parameters are obtained. The MSE of the estimated parameters and the value of Max show the better performance of ICLR as compared to PCLR. The values of MSE and Max are minimum for ICLR (3). ICLR(all) also performed better than PCLR(all). If we make an overall comparison, then after ICLR (3) the PCLR (3) is on second number. It can be observed that the values of the estimated parameters $\hat{\beta}$ are mostly very different to the real ones. Such inaccurate estimation must be the result of multicollinearity. These bad estimation results lead to a misleading interpretation of the parameters in terms of odds ratios.

Table 4: *For Case 2,* $\rho = 0.8, P = 10, n = 100$

Parameters	**Real**	**PCLR(All)**	**PCLR(6)**	**ICLR(All)**	**ICLR(6)**
$\boldsymbol{\beta_1}$	0.3319	6.7182	1.0416	4.0112	0.6015
$\boldsymbol{\beta_2}$	0.3337	6.6228	0.9987	6.1317	0.6689
$\boldsymbol{\beta_3}$	0.3093	8.1723	0.9983	4.3219	0.7314
$\boldsymbol{\beta_4}$	0.3003	7.6193	1.1219	6.1215	0.9889
$\boldsymbol{\beta_5}$	0.3321	5.9950	0.8018	2.1109	0.7889
$\boldsymbol{\beta_6}$	0.3301	4.0021	0.5989	1.0001	0.9615
$\boldsymbol{\beta_7}$	0.2986	6.1716	1.0019	2.1225	0.4212
$\boldsymbol{\beta_8}$	0.3017	8.1313	0.8787	3.7123	0.8916
$\boldsymbol{\beta_9}$	0.3287	4.9817	0.9714	2.9990	0.8204
$\boldsymbol{\beta_{10}}$	0.2988	11.2319	1.5218	5.0014	0.7105
MSE	---	47.8262	0.5125	14.4958	0.2219
Max	---	10.9331	1.2230	5.8212	0.6886

Table 5: *For Case 2,* $\rho = 0.8, P = 10, n = 200$

Parameters	**Real**	**PCLR(All)**	**PCLR(6)**	**ICLR(All)**	**ICLR(6)**
$\boldsymbol{\beta_1}$	0.3319	6.7182	1.0416	4.0112	0.6015
$\boldsymbol{\beta_2}$	0.3337	6.6228	0.9987	6.1317	0.6689
$\boldsymbol{\beta_3}$	0.3093	8.1723	0.9983	4.3219	0.7314
$\boldsymbol{\beta_4}$	0.3003	7.6193	1.1219	6.1215	0.9889
$\boldsymbol{\beta_5}$	0.3321	5.9950	0.8018	2.1109	0.7889
$\boldsymbol{\beta_6}$	0.3301	4.0021	0.5989	1.0001	0.9615
$\boldsymbol{\beta_7}$	0.2986	6.1716	1.0019	2.1225	0.4212
$\boldsymbol{\beta_8}$	0.3017	8.1313	0.8787	3.7123	0.8916
$\boldsymbol{\beta_9}$	0.3287	4.9817	0.9714	2.9990	0.8204
$\boldsymbol{\beta_{10}}$	0.2988	11.2319	1.5218	5.0014	0.7105
MSE	---	47.8262	0.5125	14.4958	0.2219
Max	---	10.9331	1.2230	5.8212	0.6886

Table 6: *For Case 2*, $\rho = 0.8, P = 10, n = 300$

Parameters	Real	PCLR(All)	PCLR(6)	ICLR(All)	ICLR(6)
$\boldsymbol{\beta_1}$	0.3163	3.3215	1.5416	3.3218	1.5416
$\boldsymbol{\beta_2}$	0.3169	8.7182	1.1912	5.2989	0.1912
$\boldsymbol{\beta_3}$	0.3155	12.2343	0.8399	4.3213	0.3521
$\boldsymbol{\beta_4}$	0.3161	7.2362	0.9987	7.2143	0.5361
$\boldsymbol{\beta_5}$	0.3159	3.3312	1.0023	3.0012	0.6313
$\boldsymbol{\beta_6}$	0.3317	2.2314	0.9876	1.6984	0.4198
$\boldsymbol{\beta_7}$	0.3146	-7.1562	1.2110	2.3245	0.0141
$\boldsymbol{\beta_8}$	0.3174	4.5798	0.9632	5.2342	0.4965
$\boldsymbol{\beta_9}$	0.3163	1.1432	1.3621	1.1172	0.3621
$\boldsymbol{\beta_{10}}$	0.3022	6.1993	0.9875	4.3216	0.2990
MSE	---	39.1700	0.6692	15.1575	0.1799
Max	---	11.9188	1.2253	6.8982	1.2253

In table 4-6, the first six PCs account for the 95% variation, so we include six PCs as reduced number of variables to construct PCLR and six ICs for the construction of ICLR. The original parameters are then reconstructed, and the estimates of the parameters are obtained. The MSE of the estimated parameters and the value of Max show the better performance of ICLR as compared to PCLR. The values of MSE and Max are minimum for ICLR (6). ICLR(all) also performed better than PCLR(all). If we make an overall comparison, then after ICLR (6) the PCLR (6) is on second number. It can be observed that the values of the estimated parameters $\hat{\beta}$ are mostly very different to the real ones. Such inaccurate estimation must be the result of multicollinearity. These bad estimation results lead to a misleading interpretation of the parameters in terms of odds ratios.

Table 7: *For Case 3,* $\rho = 0.8, P = 15, n = 100$

Parameters	Real	PCLR(All)	PCLR(7)	ICLR(All)	ICLR(7)
$\boldsymbol{\beta_1}$	0.5056	5.1393	1.5183	5.8012	0.7214
$\boldsymbol{\beta_2}$	0.1917	5.3482	0.5673	2.8713	0.2120
$\boldsymbol{\beta_3}$	0.2278	8.1573	0.9884	4.9953	0.9186
$\boldsymbol{\beta_4}$	0.1916	4.7416	0.9926	3.2638	0.4321
$\boldsymbol{\beta_5}$	0.1853	10.9817	0.8091	5.1162	0.8726
$\boldsymbol{\beta_6}$	0.2019	10.5413	1.5152	6.4381	0.6361
$\boldsymbol{\beta_7}$	0.1613	6.7682	0.9633	5.0067	0.6328
$\boldsymbol{\beta_8}$	0.2290	9.8917	0.8744	3.4719	0.4719
$\boldsymbol{\beta_9}$	0.2020	10.5618	1.0900	2.2754	0.6819
$\boldsymbol{\beta_{10}}$	0.5109	10.2323	1.9989	3.626	0.8969
$\boldsymbol{\beta_{11}}$	0.1198	5.9319	1.7109	3.3384	0.6727
$\boldsymbol{\beta_{12}}$	0.3257	8.1919	1.1372	2.7189	0.7295
$\boldsymbol{\beta_{13}}$	0.1119	5.1099	0.9925	4.1018	0.3894
$\boldsymbol{\beta_{14}}$	0.1010	4.2972	0.9719	2.1261	0.4190
$\boldsymbol{\beta_{15}}$	0.1910	3.2905	0.3198	2.4502	0.5718
MSE	---	56.1205	0.8869	14.6535	0.1796
Max	---	10.7964	1.5911	6.2362	0.6908

Table 8: *For Case 3*, $\rho = 0.8, P = 15, n = 200$

Parameters	Real	PCLR(All)	PCLR(5)	ICLR(All)	ICLR(5)
$\boldsymbol{\beta_1}$	0.3994	5.9827	0.9633	2.8713	0.8743
$\boldsymbol{\beta_2}$	0.3897	11.0725	0.8744	4.9953	0.8863
$\boldsymbol{\beta_3}$	0.3128	6.6719	0.6109	3.2638	0.9104
$\boldsymbol{\beta_4}$	0.2118	6.6713	0.9989	5.1162	0.8927
$\boldsymbol{\beta_5}$	0.1978	9.1984	1.7109	6.4381	0.6382
$\boldsymbol{\beta_6}$	0.1995	5.1875	0.5372	6.1473	0.4426
$\boldsymbol{\beta_7}$	0.1165	7.7727	0.3198	5.4534	0.3172
$\boldsymbol{\beta_8}$	0.1319	9.1973	1.5926	7.1209	0.8172
$\boldsymbol{\beta_9}$	0.4258	5.1393	0.8091	2.2122	0.1128
$\boldsymbol{\beta_{10}}$	0.3162	5.3482	0.7152	3.6260	0.7476
$\boldsymbol{\beta_{11}}$	0.1101	8.1573	0.9884	3.3384	0.3109
$\boldsymbol{\beta_{12}}$	0.2752	4.7416	0.9926	2.7189	0.5319
$\boldsymbol{\beta_{13}}$	0.1190	8.1703	0.8091	4.1018	0.3381
$\boldsymbol{\beta_{14}}$	0.2277	6.3281	0.5152	2.1261	0.8656
$\boldsymbol{\beta_{15}}$	0.1017	6.1762	0.9730	2.4502	0.6393
MSE	---	49.6887	0.5833	17.8167	0.2126
Max	---	10.6828	1.5131	6.9890	0.6853

Table 9: *For Case 3*, $\rho = 0.8, P = 15, n = 300$

Parameters	Real	PCLR(All)	PCLR(5)	ICLR(All)	ICLR(5)
β_1	0.4024	5.9319	1.8091	5.7365	0.7295
β_2	0.3105	8.1919	0.7152	3.2128	0.3894
β_3	0.1298	4.1799	0.6633	3.2712	0.4190
β_4	0.3107	3.2272	0.8744	3.9838	0.5718
β_5	0.3039	11.2905	1.2048	4.1328	0.8743
β_6	0.2198	7.6514	0.9642	5.8012	0.8863
β_7	0.3987	6.2048	0.9730	2.8713	0.9104
β_8	0.2119	7.8182	0.9782	4.9953	0.8927
β_9	0.1001	5.9801	1.5183	3.2638	0.6382
β_{10}	0.1891	9.9754	0.5673	5.1162	0.4426
β_{11}	0.2278	8.1573	0.9884	6.4381	0.5319
β_{12}	0.2916	4.7416	0.9926	5.0067	0.3381
β_{13}	0.1853	8.1703	0.8091	3.4719	0.8656
β_{14}	0.2019	6.3281	1.5152	2.2754	0.6393
β_{15}	0.1613	6.1762	0.7752	3.6260	0.3261
MSE	---	49.0397	0.7161	17.1489	0.1926
Max	---	10.9866	1.4182	6.2103	0.6808

In table 7, we compare first seven PCs with ICs as well as in Table 8 and 9 we deal with first five PCs and ICs. The first seven and five PCs account for the 95% variation, so we include seven as well as five PCs as reduced number of variables to construct PCLR and seven as well as five ICs for the construction of ICLR. The original parameters are then reconstructed and the estimates of the parameters are obtained. The MSE of the estimated parameters and the value of Max show the better performance of ICLR as compared to PCLR. The values of MSE and Max are minimum for ICLR (5,7). ICLR(all) also performed better than PCLR(all). If we make an overall comparison, then after ICLR (5,7) the PCLR (5,7) is on second number. It can be observed that the values of the estimated parameters $\hat{\beta}$ are mostly very different to the real ones. Such inaccurate estimation must be the result of multicollinearity. These bad estimation results lead to a misleading interpretation of the parameters in terms of odds ratios.

Table 10: *The value of G^2 and p-value (parentheses) with correlation 0.8 for ICLR(s) and PCLR(s) Model*

	P=6			**P=10**			**P=15**		
Method	n=100	n=200	n=300	n=100	n=200	n=300	n=100	n=200	n=300
PCLR (3)	76.8 (0.87)	81.21 (0.81)	77.23 (0.87)	---	---	---	---	---	---
ICLR (3)	82.13 (0.80)	88.58 (0.74)	84.14 (0.79)	---	---	---	---	---	---
PCLR (5)	---	---	---	---	73.4 (0.88)	---	---	73.18 (0.88)	62.13 (0.97)
ICLR (5)	---	---	---	---	84.58 (0.78)	---	---	88.40 (0.75)	84.78 (0.79)
PCLR (6)	---	---	---	79.5 (0.86)	---	76.21 (0.87)	---	---	---
ICLR (6)	---	---	---	81.08 (0.81)	---	83.49 (0.78)	---	---	---
PCLR (7)	---	---	---	---	---	---	81.23 (0.81)	---	---
ICLR (7)	---	---	---	---	---	---	89.91 (0.75)	---	---

In table 10, The value of the likelihood ratio statistic, regardless of which two models are being compared, yields a value that lies between 0, when there is extreme non-significance, and ∞, when there is extreme significance. We can observe from the above table that the value of G^2 is high for the ICLR models than the PCLR models. Hence the values of G^2 along with MSE and Max, confirms the better performance of ICLR than PCLR.

9.2. Analysis of Datasets with $\rho = 0.9$ Real and Estimated Parameters with MSE and Max

Table 11: *For Case 1:* $\rho = 0.9, P = 6, n = 100$

Parameters	Real	PCLR(All)	PCLR(3)	ICLR(All)	ICLR(3)
β_1	0.2203	5.2314	0.4998	3.1415	0.3398
β_2	0.5169	7.2315	1.2298	5.2213	0.9189
β_3	0.5059	9.1314	0.6734	1.5989	0.8813
β_4	0.3109	10.1318	0.4343	-2.8989	0.7979
β_5	0.5219	11.1317	2.0123	3.3214	0.8210
β_6	0.2365	0.9782	1.6719	5.4358	1.3019
MSE	---	59.0273	0.8185	12.8387	0.2964
Max	**---**	**10.6098**	**1.4904**	**5.1993**	**1.0654**

Table 12: *For Case 1:* $\rho = 0.9, P = 6, n = 200$

Parameters	Real	PCLR(All)	PCLR(3)	ICLR(All)	ICLR(3)
β_1	0.5079	8.2190	1.2908	2.3173	0.7912
β_2	0.4003	8.4592	0.9873	3.1919	0.9250
β_3	0.2976	12.6312	0.9930	3.0092	0.9971
β_4	0.4953	5.9125	1.5639	1.8962	0.1185
β_5	0.3088	0.1919	1.9016	4.271	0.7130
β_6	0.3913	5.3954	0.3098	7.1302	0.5631
MSE	---	55.1542	0.8544	13.5823	0.1966
Max	---	12.3336	1.5928	6.7389	0.6995

Table 13: *For Case 1:* $\rho = 0.9, P = 6, n = 300$

Parameters	Real	PCLR(All)	PCLR(3)	ICLR(All)	ICLR(3)
β_1	75.7979	0.1699	2.2031	0.0438	75.7979
β_2	34.3618	0.0357	7.8276	0.4233	34.3618
β_3	86.7319	0.0843	2.1359	0.4926	86.7319
β_4	112.8779	0.1607	70.4005	0.0001	112.8779
β_5	60.7370	3.2450	3.6679	0.2754	60.7371
β_6	36.2295	0.4890	0.8456	0.0384	36.2295
MSE	---	67.7893	0.6974	14.5135	0.2123
Max	---	112.7172	84.5959	112.8778	0.0000

In table 11-13, the first three PCs account for the 95% variation, so we include three PCs as reduced number of variables to construct PCLR and three ICs for the construction of ICLR. The original parameters are then reconstructed and the estimates of the parameters are obtained. The MSE of the estimated parameters and the value of Max show the better performance of ICLRas compared to PCLR. The values of MSE and Max are minimum for ICLR (3). ICLR(all) also performed better than PCLR(all). If we make an overall comparison, then after ICLR (3) the PCLR (3) is on second number. It can be observed that the values of the estimated parameters $\hat{\beta}$ are mostly very different to the real ones. Such inaccurate estimation must be the result of multicollinearity. These bad estimation results lead to a misleading interpretation of the parameters in terms of odds ratios.

Table 14: *For Case 2:* $\rho = 0.9, P = 10, n = 100$

Parameters	Real	PCLR(All)	PCLR(6)	ICLR(All)	ICLR(6)
$\boldsymbol{\beta_1}$	0.3270	7.6312	0.8416	3.6312	0.7416
$\boldsymbol{\beta_2}$	0.3228	-5.2185	1.0513	5.2185	0.6051
$\boldsymbol{\beta_3}$	0.3163	9.1723	1.6251	6.1723	0.8351
$\boldsymbol{\beta_4}$	0.2920	6.5123	0.5261	3.5123	0.5261
$\boldsymbol{\beta_5}$	0.3158	6.0025	0.7798	0.0025	0.6828
$\boldsymbol{\beta_6}$	0.3270	3.6984	0.6162	1.6984	0.8612
$\boldsymbol{\beta_7}$	0.3047	5.0345	1.0141	2.0345	0.3362
$\boldsymbol{\beta_8}$	0.3072	9.2323	0.8995	4.2323	0.9775
$\boldsymbol{\beta_9}$	0.3326	5.1254	0.9618	3.1254	0.8167
$\boldsymbol{\beta_{10}}$	0.3142	10.1824	1.6793	4.1824	0.6189
MSE	---	46.7264	0.5975	12.2688	0.1773
Max	---	9.8682	1.3651	5.8560	0.6703

Table 15: *For Case 2:* $\rho = 0.9, P = 10, n = 200$

Parameters	Real	PCLR(All)	PCLR(5)	ICLR(All)	ICLR(5)
$\boldsymbol{\beta_1}$	0.3448	3.3215	1.5416	3.3218	0.1416
$\boldsymbol{\beta_2}$	0.3251	8.7182	1.1912	5.2989	0.7051
$\boldsymbol{\beta_3}$	0.2903	12.2343	0.8521	4.3213	0.9351
$\boldsymbol{\beta_4}$	0.3238	7.2362	0.8361	7.2143	0.4261
$\boldsymbol{\beta_5}$	0.3196	3.3312	1.6313	3.0012	0.6828
$\boldsymbol{\beta_6}$	0.2931	2.2314	0.4198	1.6984	0.5612
$\boldsymbol{\beta_7}$	0.2995	-7.1562	0.0141	2.3245	0.7362
$\boldsymbol{\beta_8}$	0.3231	4.5798	0.9965	5.2342	0.9775
$\boldsymbol{\beta_9}$	0.3286	1.1432	1.3621	1.1172	0.0816
$\boldsymbol{\beta_{10}}$	0.3104	6.1993	0.9990	4.3216	0.9189
MSE	---	39.1621	0.6574	15.1426	0.1865
Max	---	11.9440	1.3117	6.8905	0.6544

Table 16: *For Case 2:*$\rho = 0.9, P = 10, n = 300$

Parameters	Real	PCLR(All)	PCLR(6)	ICLR(All)	ICLR(6)
$\boldsymbol{\beta_1}$	0.3163	3.3215	1.5416	3.3218	1.5416
$\boldsymbol{\beta_2}$	0.3169	8.7182	1.1912	5.2989	0.1912
$\boldsymbol{\beta_3}$	0.3155	12.2343	0.8399	4.3213	0.3521
$\boldsymbol{\beta_4}$	0.3161	7.2362	0.9987	7.2143	0.5361
$\boldsymbol{\beta_5}$	0.3159	3.3312	1.0023	3.0012	0.6313
$\boldsymbol{\beta_6}$	0.3317	2.2314	0.9876	1.6984	0.4198
$\boldsymbol{\beta_7}$	0.3146	-7.1562	1.2110	2.3245	0.0141
$\boldsymbol{\beta_8}$	0.3174	4.5798	0.9632	5.2342	0.4965
$\boldsymbol{\beta_9}$	0.3163	1.1432	1.3621	1.1172	0.3621
$\boldsymbol{\beta_{10}}$	0.3022	6.1993	0.9875	4.3216	0.2990
MSE	---	39.1700	0.6691	15.1575	0.1799
Max	---	11.9188	1.2253	6.8982	1.2253

In table 14-16, the first six PCs account for the 95% variation, and in table 15 so we include six PCs as well as five PCS as reduced number of variables to construct PCLR and six ICs for the construction of ICLR. The original parameters are then reconstructed and the estimates of the parameters are obtained. The MSE of the estimated parameters and the value of Max show the better performance of ICLR as compared to PCLR. The values of MSE and Max are minimum for ICLR (6). ICLR(all) also performed better than PCLR(all). If we make an overall comparison, then after ICLR (6) the PCLR (6) is on second number. It can be observed that the values of the estimated parameters $\hat{\beta}$ are mostly very different to the real ones. Such inaccurate estimation must be the result of multicollinearity. These bad estimation results lead to a misleading interpretation of the parameters in terms of odds ratios.

Table 17: *For Case 3:* $\rho = 0.9, P = 15, n = 100$

Parameters	Real	PCLR(All)	PCLR(7)	ICLR(All)	ICLR(7)
$\boldsymbol{\beta_1}$	0.5056	10.2319	1.5183	5.8012	0.7214
$\boldsymbol{\beta_2}$	0.1917	4.2923	0.5673	2.8713	0.2120
$\boldsymbol{\beta_3}$	0.2278	6.5492	0.9884	4.9953	0.9186
$\boldsymbol{\beta_4}$	0.1916	5.6842	0.9926	3.2638	0.4321
$\boldsymbol{\beta_5}$	0.1853	11.9819	0.8091	5.1162	0.8726
$\boldsymbol{\beta_6}$	0.2019	11.2905	1.5152	6.4381	0.6361
$\boldsymbol{\beta_7}$	0.1613	7.6514	0.9633	5.0067	0.6328
$\boldsymbol{\beta_8}$	0.2290	6.2048	0.8744	3.4719	0.4719
$\boldsymbol{\beta_9}$	0.2020	7.8182	1.0900	2.2754	0.6819
$\boldsymbol{\beta_{10}}$	0.5109	5.9801	1.9989	3.6260	0.8969
$\boldsymbol{\beta_{11}}$	0.1198	5.9319	1.7109	3.3384	0.6727
$\boldsymbol{\beta_{12}}$	0.3257	8.1919	1.1372	2.7189	0.7295
$\boldsymbol{\beta_{13}}$	0.1119	4.1799	0.9925	4.1018	0.3894
$\boldsymbol{\beta_{14}}$	0.1010	3.2272	0.9719	2.1261	0.4190
$\boldsymbol{\beta_{15}}$	0.1910	4.2905	0.3198	2.4502	0.5718
MSE	---	50.8118	0.8869	14.6535	0.1796
Max	---	11.7966	1.5911	6.2362	0.6908

Table 19: *For Case 3:* $\rho = 0.9, P = 15, n = 300$

Parameters	Real	PCLR(All)	PCLR(5)	ICLR(All)	ICLR(5)
$\boldsymbol{\beta_1}$	0.4024	8.6528	1.9081	7.8262	0.7126
$\boldsymbol{\beta_2}$	0.3105	8.5432	0.8963	5.6546	0.9983
$\boldsymbol{\beta_3}$	0.1298	9.2363	0.7692	4.8340	0.3484
$\boldsymbol{\beta_4}$	0.3107	7.6528	0.9923	3.8262	0.9717
$\boldsymbol{\beta_5}$	0.3039	6.6737	1.0013	3.3317	0.9989
$\boldsymbol{\beta_6}$	0.2198	8.1765	0.8917	5.1852	0.7253
$\boldsymbol{\beta_7}$	0.3987	7.0664	1.1162	3.7214	0.1001
$\boldsymbol{\beta_8}$	0.2119	6.6754	0.9782	2.1985	0.9379
$\boldsymbol{\beta_9}$	0.1001	6.9473	0.8917	5.9986	0.5421
$\boldsymbol{\beta_{10}}$	0.1891	9.9754	0.7826	1.0084	0.3111
$\boldsymbol{\beta_{11}}$	0.2278	5.9837	0.5969	3.9364	0.5132
$\boldsymbol{\beta_{12}}$	0.2916	8.1432	0.8351	1.0121	0.8163
$\boldsymbol{\beta_{13}}$	0.1853	7.1682	0.6125	1.1189	0.7142
$\boldsymbol{\beta_{14}}$	0.2019	3.9719	0.7198	4.9886	0.8169
$\boldsymbol{\beta_{15}}$	0.1613	4.5334	0.7752	7.9907	0.3261
MSE	---	52.1548	0.5174	20.1193	0.2440
Max	---	9.7863	1.5057	7.8294	0.7260

Table 19: *For Case 3:* $\rho = 0.9, P = 15, n = 300$

Parameters	Real	PCLR(All)	PCLR(5)	ICLR(All)	ICLR(5)
$\boldsymbol{\beta_1}$	0.4024	8.6528	1.9081	7.8262	0.7126
$\boldsymbol{\beta_2}$	0.3105	8.5432	0.8963	5.6546	0.9983
$\boldsymbol{\beta_3}$	0.1298	9.2363	0.7692	4.8340	0.3484
$\boldsymbol{\beta_4}$	0.3107	7.6528	0.9923	3.8262	0.9717
$\boldsymbol{\beta_5}$	0.3039	6.6737	1.0013	3.3317	0.9989
$\boldsymbol{\beta_6}$	0.2198	8.1765	0.8917	5.1852	0.7253
$\boldsymbol{\beta_7}$	0.3987	7.0664	1.1162	3.7214	0.1001
$\boldsymbol{\beta_8}$	0.2119	6.6754	0.9782	2.1985	0.9379
$\boldsymbol{\beta_9}$	0.1001	6.9473	0.8917	5.9986	0.5421
$\boldsymbol{\beta_{10}}$	0.1891	9.9754	0.7826	1.0084	0.3111
$\boldsymbol{\beta_{11}}$	0.2278	5.9837	0.5969	3.9364	0.5132
$\boldsymbol{\beta_{12}}$	0.2916	8.1432	0.8351	1.0121	0.8163
$\boldsymbol{\beta_{13}}$	0.1853	7.1682	0.6125	1.1189	0.7142
$\boldsymbol{\beta_{14}}$	0.2019	3.9719	0.7198	4.9886	0.8169
$\boldsymbol{\beta_{15}}$	0.1613	4.5334	0.7752	7.9907	0.3261
MSE	---	52.1548	0.5174	20.1193	0.2440
Max	---	9.7863	1.5057	7.8294	0.7260

In table 17 the first seven PCs account for 95% variation and in table 18 and 19, the first five PCs account for the 95% variation, so we include five PCs as reduced number of variables to construct PCLR and five ICs for the construction of ICLR. The original parameters are then reconstructed and the estimates of the parameters are obtained. The MSE of the estimated parameters and the value of Max show the better performance of ICLRas compared to PCLR. The values of MSE and Max are minimum for ICLR (7,5). ICLR(all) also performed better than PCLR(all). If we make an overall comparison, then after ICLR (7,5) the PCLR (7,5) is on second number. It can be observed that the values of the estimated parameters $\hat{\beta}$ are mostly very different to the real ones. Such inaccurate estimation must be the result of multicollinearity. These bad estimation results lead to a misleading interpretation of the parameters in terms of odds ratios.

Table 20: *The value of G^2 and p-value (parentheses) with $\rho = 0.9$ for ICLR(s) and PCLR(s) Models*

	$P = 6$			$P = 10$			$P = 15$		
Method	$n = 100$	$n = 200$	$n = 300$	$n = 100$	$n = 200$	$n = 300$	$n = 100$	$n = 200$	$n = 300$
PCLR (3)	57.38 (0.99)	68.4 (0.98)	63.19 (0.97)	---	---	---	---	---	---
ICLR (3)	63.12 (0.99)	66.23 (0.97)	64.91 (0.98)	---	---	---	---	---	---
PCLR (5)	---	---	---	---	68.19 (0.98)	---	---	71.40 (0.88)	62.19 (0.99)
ICLR (5)	---	---	---	---	70.23 (0.89)	---	---	83.51 (0.78)	79.26 (0.86)
PCLR (6)	---	---	---	58.4 (0.99)	---	66.15 (0.97)	---	---	---
ICLR (6)	---	---	---	62.1 (0.99)	---	70.13 (0.89)	---	---	---
PCLR (7)	---	---	---	---	---	---	72.28 (0.88)	---	---
ICLR (7)	---	---	---	---	---	---	78.48 (0.87)	---	---

In table 20, the value of the likelihood ratio statistic, regardless of which two models are being compared, yields a value that lies between 0, when there is extreme non-significance, and ∞, when there is extreme significance. We can observe from the above table that the value of G^2 is high for the ICLR models than the PCLR models. Hence the values of G^2 along with MSE and Max, confirms the better performance of ICLR than PCLR.

9.3. Analysis of Datasets with $\rho = 0.95$ Real and Estimated Parameters with MSE and Max

Table 21: *For Case 1:* $\rho = 0.95, P = 6, n = 100$

Parameters	Real	PCLR(All)	PCLR(3)	ICLR(All)	ICLR(3)
β_1	0.2976	6.1314	0.8917	1.8962	0.7854
β_2	0.4953	6.3315	1.9942	4.2710	0.6010
β_3	0.5079	10.5140	1.0012	7.1302	0.4493
β_4	0.4003	10.1318	0.5343	1.9872	0.8344
β_5	0.3100	8.4141	1.1123	3.1285	0.9928
β_6	0.3820	0.9928	0.9071	1.9717	0.5019
MSE	---	54.8280	0.6300	12.2759	0.1536
Max	---	10.0061	1.4989	6.6223	0.6828

Table 22: *For Case 1:* $\rho = 0.95, P = 6, n = 200$

Parameters	Real	PCLR(All)	PCLR(3)	ICLR(All)	ICLR(3)
β_1	0.5079	9.7189	0.8439	3.1920	0.6917
β_2	0.4003	7.2415	0.6116	2.1186	0.8893
β_3	0.2976	6.6698	0.3098	2.9989	1.0018
β_4	0.4953	5.1982	0.9063	1.9018	0.2187
β_5	0.3088	1.6614	1.8116	3.6183	0.6270
β_6	0.3913	9.9917	0.6098	6.9918	0.5931
MSE	---	48.0606	0.4388	12.3252	0.1645
Max	---	9.6004	1.5028	6.6005	0.7042

Table 23: *For Case 1:* $\rho = 0.95, P = 6, n = 300$

Parameters	Real	PCLR(All)	PCLR(3)	ICLR(All)	ICLR(3)
$\boldsymbol{\beta_1}$	0.5029	7.2315	0.8989	1.8962	0.9928
$\boldsymbol{\beta_2}$	0.3307	9.1314	0.9172	4.2710	0.5019
$\boldsymbol{\beta_3}$	0.5102	10.1318	1.6689	7.1302	0.4981
$\boldsymbol{\beta_4}$	0.4908	7.4152	1.1219	8.8813	0.5019
$\boldsymbol{\beta_5}$	0.1928	7.9862	0.8129	2.1080	0.7176
$\boldsymbol{\beta_6}$	0.3019	6.3210	0.9918	1.2215	0.8127
MSE	---	60.0359	0.5170	22.7009	0.1343
Max	---	9.6216	1.1587	8.3905	0.5248

In table 21-23, the first three PCs account for the 95% variation, so we include three PCs as reduced number of variables to construct PCLR and three ICs for the construction of ICLR. The original parameters are then reconstructed and the estimates of the parameters are obtained. The MSE of the estimated parameters and the value of Max show the better performance of ICLRas compared to PCLR. The values of MSE and Max are minimum for ICLR (3). ICLR(all) also performed better than PCLR(all). If we make an overall comparison, then after ICLR (3) the PCLR (3) is on second number. It can be observed that the values of the estimated parameters $\hat{\beta}$ are mostly very different to the real ones. Such inaccurate estimation must be the result of multicollinearity. These bad estimation results lead to a misleading interpretation of the parameters in terms of odds ratios.

Table 24: *For Case 2:* $\rho = 0.95, P = 10, n = 100$

Parameters	Real	PCLR(All)	PCLR(6)	ICLR(All)	ICLR(6)
$\boldsymbol{\beta_1}$	0.3319	6.7182	1.0416	4.0112	0.6015
$\boldsymbol{\beta_2}$	0.3337	6.6228	0.9987	6.1317	0.6689
$\boldsymbol{\beta_3}$	0.3093	8.1723	0.9983	4.3219	0.7314
$\boldsymbol{\beta_4}$	0.3003	7.6193	1.1219	6.1215	0.9889
$\boldsymbol{\beta_5}$	0.3321	5.9950	0.8018	2.1109	0.7889
$\boldsymbol{\beta_6}$	0.3301	4.0021	0.5989	1.0001	0.9615
$\boldsymbol{\beta_7}$	0.2986	6.1716	1.0019	2.1225	0.4212
$\boldsymbol{\beta_8}$	0.3017	8.1313	0.8787	3.7123	0.8916
$\boldsymbol{\beta_9}$	0.3287	4.9817	0.9714	2.9990	0.8204
$\boldsymbol{\beta_{10}}$	0.2988	11.2319	1.5218	5.0014	0.7105
MSE	---	47.8262	0.5125	14.4958	0.2219
Max	---	10.9331	1.2230	5.8212	0.6886

Table 25: *For Case 2:* $\rho = 0.95, P = 10, n = 200$

Parameters	Real	PCLR(All)	PCLR(5)	ICLR(All)	ICLR(5)
$\boldsymbol{\beta_1}$	0.3448	4.1223	2.0071	4.1872	0.4742
$\boldsymbol{\beta_2}$	0.3251	7.6512	1.0032	5.0092	0.8209
$\boldsymbol{\beta_3}$	0.2903	9.8862	0.9589	5.1615	0.6987
$\boldsymbol{\beta_4}$	0.3238	8.1926	0.9935	6.9189	0.9150
$\boldsymbol{\beta_5}$	0.3196	4.3319	0.8218	6.1138	0.3480
$\boldsymbol{\beta_6}$	0.2931	3.0001	0.6960	1.0029	0.2418
$\boldsymbol{\beta_7}$	0.2995	6.1673	0.9187	1.3389	0.7695
$\boldsymbol{\beta_8}$	0.3231	5.1859	0.7156	3.3219	0.2219
$\boldsymbol{\beta_9}$	0.3286	2.0098	1.1962	1.0091	0.6382
$\boldsymbol{\beta_{10}}$	0.3104	4.0147	1.0010	2.3384	0.7861
MSE	---	31.9993	0.6300	15.2654	0.1335
Max	---	9.5959	1.6623	6.5951	0.5912

Table 26: *For Case 2:* $\rho = 0.95, P = 10, n = 300$

Parameters	Real	PCLR(All)	PCLR(5)	ICLR(All)	ICLR(5)
$\boldsymbol{\beta_1}$	0.3163	3.3215	1.5416	3.3218	0.4965
$\boldsymbol{\beta_2}$	0.3169	8.7182	1.1912	5.2989	0.3621
$\boldsymbol{\beta_3}$	0.3155	12.2343	0.8399	4.3213	0.7299
$\boldsymbol{\beta_4}$	0.3161	7.2362	0.9987	7.2143	0.5361
$\boldsymbol{\beta_5}$	0.3159	3.3312	1.0023	3.0012	0.6313
$\boldsymbol{\beta_6}$	0.3317	2.2314	0.9876	1.6984	0.9198
$\boldsymbol{\beta_7}$	0.3146	-7.1562	1.2110	2.3245	0.8141
$\boldsymbol{\beta_8}$	0.3174	4.5798	0.9632	5.2342	0.9145
$\boldsymbol{\beta_9}$	0.3163	1.1432	1.3621	1.1172	0.8619
$\boldsymbol{\beta_{10}}$	0.3022	6.1993	0.9875	4.3216	0.5812
MSE	---	39.1700	0.6692	15.1575	0.1681
Max	---	11.9188	1.2253	6.8982	0.5971

In table 24, the first six PCs account for the 95% variation, and in table 25,26 the first five PCs account for the 95% variation so we include six PCs as well as five PCs as reduced number of variables to construct PCLR and six ICs for the construction of ICLR. The original parameters are then reconstructed and the estimates of the parameters are obtained. The MSE of the estimated parameters and the value of Max show the better performance of ICLR as compared to PCLR. The values of MSE and Max are minimum for ICLR (6,5). ICLR(all) also performed better than PCLR(all). If we make an overall comparison, then after ICLR (6,5) the PCLR (6,5) is on second number. It can be observed that the values of the estimated parameters $\hat{\beta}$ are mostly very different to the real ones. Such inaccurate estimation must be the result of multicollinearity. These bad estimation results lead to a misleading interpretation of the parameters in terms of odds ratios.

Table 27: *For Case 3:* $\rho = 0.95, P = 15, n = 100$

Parameters	Real	PCLR(All)	PCLR(7)	ICLR(All)	ICLR(7)
$\boldsymbol{\beta_1}$	0.3315	6.6719	0.8917	6.1215	0.6828
$\boldsymbol{\beta_2}$	0.4614	6.6713	0.7826	2.1109	0.5612
$\boldsymbol{\beta_3}$	0.3216	9.1984	0.5969	1.0001	0.7362
$\boldsymbol{\beta_4}$	0.2011	5.1875	0.8351	2.1225	0.9775
$\boldsymbol{\beta_5}$	0.1998	7.7727	0.6125	6.4381	0.0816
$\boldsymbol{\beta_6}$	0.2317	9.1973	0.7198	5.0067	0.9189
$\boldsymbol{\beta_7}$	0.1189	5.1393	0.7752	3.4719	0.2328
$\boldsymbol{\beta_8}$	0.2290	5.3482	1.1219	2.2754	0.4719
$\boldsymbol{\beta_9}$	0.1601	7.8262	0.8018	3.6260	0.6819
$\boldsymbol{\beta_{10}}$	0.3316	5.6546	0.5989	3.3384	0.5969
$\boldsymbol{\beta_{11}}$	0.1243	10.8340	1.0019	2.7189	0.6727
$\boldsymbol{\beta_{12}}$	0.3374	13.8262	0.8787	4.1018	0.7295
$\boldsymbol{\beta_{13}}$	0.1101	6.3317	0.9714	2.1261	0.3894
$\boldsymbol{\beta_{14}}$	0.2752	5.1852	1.5218	2.4502	0.4190
$\boldsymbol{\beta_{15}}$	0.1190	3.7214	0.5398	3.1202	0.5718
MSE	---	55.4073	0.4367	11.8202	0.1711
Max	---	13.4888	1.2466	6.2383	0.7764

Table 28: *For Case 3:* $\rho = 0.95, P = 15, n = 200$

Parameters	Real	PCLR(All)	PCLR(7)	ICLR(All)	ICLR(7)
β_1	0.2983	9.1973	0.9260	3.0012	0.3172
β_2	0.3918	5.1393	0.8091	1.6984	0.8172
β_3	0.3372	5.3482	0.7152	2.3245	0.1128
β_4	0.1908	8.1573	0.6633	5.2342	0.7476
β_5	0.1978	6.6754	0.5152	1.1172	0.4719
β_6	0.2198	6.9473	0.9633	4.3216	0.6819
β_7	0.3987	9.9754	0.8744	4.1683	0.5969
β_8	0.2119	5.9837	0.7109	6.0023	0.6727
β_9	0.1001	8.1432	1.9989	3.8914	0.7295
β_{10}	0.1891	5.3482	0.7109	9.1753	0.8612
β_{11}	0.2278	7.0957	1.1372	3.9928	0.3362
β_{12}	0.2916	5.3482	0.9925	1.9312	0.9775
β_{13}	0.1104	7.8262	0.9994	3.3220	0.8167
β_{14}	0.2752	5.6546	1.1298	4.7838	0.6189
β_{15}	0.1910	10.8340	0.9987	2.3328	0.5293
MSE	---	51.2227	0.6288	16.7350	0.2095
Max	---	10.6430	1.8988	8.9862	0.7063

Table 29: *For Case 3:* $\rho = 0.95, P = 15, n = 300$

Parameters	Real	PCLR(All)	PCLR(6)	ICLR(All)	ICLR(6)
β_1	0.2119	9.8862	0.9782	5.2989	0.6339
β_2	0.1001	8.1926	1.5183	4.3213	0.3365
β_3	0.1891	4.3319	0.5673	7.2143	0.5845
β_4	0.2278	3.0001	0.9884	3.0012	0.5037
β_5	0.2916	6.1673	0.9926	1.6984	0.4501
β_6	0.5109	5.1859	0.8091	2.3245	0.9928
β_7	0.1198	2.0098	1.5152	5.2342	0.4258
β_8	0.3257	4.0147	0.7752	1.1172	0.6081
β_9	0.1119	10.995	0.9717	5.1615	0.7535
β_{10}	0.2791	7.0021	0.8826	6.9189	0.5779
β_{11}	0.3178	6.1716	0.6769	6.1138	0.1883
β_{12}	0.4012	8.1313	0.8351	1.0029	0.6947
β_{13}	0.1713	9.9817	0.6125	1.3389	0.7582
β_{14}	0.2117	7.0719	0.6198	3.3219	0.4541
β_{15}	0.1013	8.9334	0.9652	1.0091	0.5689
MSE	---	49.0976	0.5718	16.5463	0.1415
Max	---	10.8831	1.4182	7.0252	0.6416

In table 27-28, the first seven PCs account for the 95% variation, in table 29 the first six PCs account for 95% variation so we include seven as well as six PCs as reduced number of variables to construct PCLR and seven as well as six ICs for the construction of ICLR. The original parameters are then reconstructed and the estimates of the parameters are obtained. The MSE of the estimated parameters and the value of Max show the better performance of ICLRas compared to PCLR. The values of MSE and Max are minimum for ICLR (7,6). ICLR(all) also performed better than PCLR(all). If we make an overall comparison, then after ICLR (7,6) the PCLR (7,6) is on second number. It can be observed that the values of the estimated parameters $\hat{\beta}$ are mostly very different to the real ones. Such inaccurate estimation must be the result of multicollinearity. These bad estimation results lead to a misleading interpretation of the parameters in terms of odds ratios.

Table 30: *The value of $G^2$and p-value (parentheses) with correlation = 0.95 for ICLR(s) and PCLR(s) Models*

	$P = 6$			$P = 10$			$P = 15$		
Method	$n = 100$	$n = 200$	$n = 300$	$n = 100$	$n = 200$	$n = 300$	$n = 100$	$n = 200$	$n = 300$
PCLR (3)	58.19 (0.99)	66.80 (0.98)	77.15 (0.87)	---	---	---	---	---	---
ICLR (3)	79.87 (0.82)	83.15 (0.79)	82.64 (0.80)	---	---	---	---	---	---
PCLR (5)	---	---	---	---	70.78 (0.88)	79.30 (0.82)	---	---	---
ICLR (5)	---	---	---	---	88.72 (0.75)	89.82 (0.75)	---	---	---
PCLR (6)	---	---	---	72.19 (0.88)	---	---	---	---	70.19 (0.88)
ICLR (6)	---	---	---	86.11 (0.79)	---	---	---	---	79.90 (0.82)
PCLR (7)	---	---	---	---	---	---	66.68 (0.97)	72.11 (0.88)	---
ICLR (7)	---	---	---	---	---	---	84.19 (0.79)	85.11 (0.79)	---

In table 30, the value of the likelihood ratio statistic, regardless of which two models are being compared, yields a value that lies between 0 and∞, when there is extreme significance. We can observe from the above table that the value of G^2 is high for the ICLR models than the PCLR models. Hence the values of G^2 along with MSE and Max, confirms the better performance of ICLR than PCLR.

10. Conclusion

The logistic regression model is used to predict a binary response variable in terms of a set of explicative ones. In the presence of multicollinearity among predictor, the estimation of the model parameters is not very accurate and their interpretation in terms of odds ratios may be inaccurate. Another important problem is that usually a large number of predictors usually required explaining the response. In order to improve the estimation of the logistic model parameters under multicollinearity and to reduce the dimensions of the problem with continuous covariates, it is proposed to use as covariates of the logistic model a reduced set of optimum independent components of the original predictors. The performance of the proposed independent component logistic regression model is analyzed by developing a simulation study where different methods for selecting the optimum independent components are compared. We built up a simulation study to illustrate the performance of the proposed ICLR model and to demonstrate how the estimation of logit model parameters with collinear regressors can be improved by utilizing ICs.

The simulation design is carried out by choosing:

- Three different number of regressors i.e. $P = 6, 10, and\ 15$
- Three different sample sizes i.e. $n = 100, 200\ and\ 300$
- Three different levels of correlation among the regressors i.e. $\rho = 0.8, 0.9\ and\ 0.95$

In this way we obtain nine different tables each with above stated three different correlation levels and overall, twenty-seven tables.

We include first s PCs as reduced number of variables to construct PCLR and s ICs for the construction of ICLR for all the 27 tables. The L_∞norm is used to order the ICs. The original parameters are then reconstructed and the estimates of the parameters are obtained. The MSE of the estimated parameters and the value of Max show the better performance of ICLRas compared to PCLR for all the models. The values of MSE and Max are minimum for ICLR(s). ICLR (all) also performed better than PCLR (all). If we make an overall comparison, then after ICLR(s) the PCLR(s) is on second number. It was revealed from the results that the values of the estimated parameters $\hat{\beta}$ are mostly very different to the real ones. Such inaccurate estimation must be the result of multicollinearity. These bad estimation results lead to a misleading interpretation of the parameters in terms of odds ratios.

The value of the likelihood ratio statistic, regardless of which two models are being compared, yields a value that lies between 0, when there is extreme non-significance, and ∞, when there is extreme significance. We can observe from the above table that the value of G^2 is high for the ICLR models than the PCLR models. Hence the values of G^2 along with MSE and Max, confirms the better performance of ICLR than PCLR.

Finally, it is concluded that with respect to the comparison with PCLR, the ICLR model provides better estimation of the logit model parameters (less *MSE and Max*) with better goodness-of-fit measures and needs less components so that the interpretation of the model parameters is more accurate.

Conflict of Interest

The authors showed no conflict of interest.

Funding

The authors did not mention any funding for this research.

References

Agyekum, G. O., Adarkwa, S. A., & Kusi, R. Y. (2023). Impact of sample size on multicollinearity with high dimensional data in logistic regression analysis. *International Journal of Innovation and Development*, *1*(3).

Agyekum, G. O., Adarkwa, S. A., & Kusi, R. Y. (2023). Impact of sample size on multicollinearity with high dimensional data in logistic regression analysis. *International Journal of Innovation and Development*, *1*(3).

Aucott, L. S., Garthwaite, P. H., &Currall, J. (2000). Regression methods for high dimensional multicollinear data. *Communications in Statistics-Simulation and Computation*, *29*(4), 1021-1037.

Aguilera, A. M., Escabias, M., &Valderrama, M. J. (2006). Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Computational Statistics & Data Analysis*, *50*(8), 1905-1924.

Aguilera-Morillo, M. C., Aguilera, A. M., Escabias, M., &Valderrama, M. J. (2013). Penalized spline approaches for functional logit regression. *Test*, *22*(2), 251-277.

Bastien, P., Vinzi, V. E., &Tenenhaus, M. (2005). PLS generalised linear regression. *Computational Statistics & data analysis*, *48*(1), 17-46.

Escabias, M., Aguilera, A. M., &Valderrama, M. J. (2005). Modeling environmental data by functional principal component logistic regression. *Environmetrics: The official journal of the International Environmetrics Society*, *16*(1), 95-107.

Escabias, M., Aguilera, A. M., &Valderrama, M. J. (2004). Principal component estimation of functional logistic regression: discussion of two different approaches. *Journal of Nonparametric Statistics*, *16*(3-4), 365-384.

Escabias, M., Aguilera, A. M., &Valderrama, M. J. (2007). Functional PLS logit regression model. *Computational Statistics & Data Analysis*, *51*(10), 4891-4902.

Hosmer, D. W., Hosmer, T., Le Cessie, S., &Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in medicine*, *16*(9), 965-980.

Hubert, M. H., &Wijekoon, P. (2006). Improvement of the Liu estimator in linear regression model. *Statistical Papers*, *47*(3), 471.

Hosmer, D.W., &Lemeshow, S., (1989). Applied logistic regression. *Wiley, New York*

Le Cessie, S., & Van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *41*(1), 191-201.

Månsson, K., &Shukur, G. (2011). On ridge parameters in logistic regression. *Communications in Statistics-Theory and Methods*, *40*(18), 3366-3381.

Newhouse, J. P., & Oman, S. D. (1971). *An evaluation of ridge estimators. Rand Corporation*. P-716-PR.

Prentice, R. L., &Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, *66*(3), 403-411.

Pulkstenis, E., & Robinson, T. J. (2002). Two goodness-of-fit tests for logistic regression models with continuous covariates. *Statistics in medicine*, *21*(1), 79-93.

Ryan, T.P., (1997). Modern regression methods.*Wiley, New York.*

Schaefer, R. L., Roi, L. D., & Wolfe, R. A. (1984). A ridge logistic estimator. *Communications in Statistics-Theory and Methods*, *13*(1), 99-113.

Steyerberg, E. W., Eijkemans, M. J. C., &Habbema, J. D. F. (2001). Application of shrinkage techniques in logistic regression analysis: a case study. *StatisticaNeerlandica*, *55*(1), 76-88.

Zhou, C., Wang, L., Zhang, Q., & Wei, X. (2014). Face recognition based on PCA and logistic regression analysis. *Optik-International Journal for Light and Electron Optics*, *125*(20), 5916-5919.