INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals

A Two-Parameter Ridge Estimator for Handling Extreme Multicollinearity Problems in Logistic Regression

David I. Oladapo[a*], Biodun T. Efuwape[b] and Timothy O. Olatayo[b]

[a]*Department of Mathematics, Adeleke University, Ede, Osun state, Nigeria.*

[b]*Department of Statistics, Olabisi Onabanjo University, Ago-Iwoye, Ogun state, Nigeria*

ARTICLE INFO

Article history:
Received 17 August 2024
Received in revised form 15 October 2024
Accepted 29 October 2024

Keywords:
Multicollinearity, Logistic Regression, Robust Estimator, Mean Square Error, Biasing Parameter

MSC 2020 Subject classification:
62P05

ABSTRACT

This paper introduces a robust two-parameter ridge estimator that is customized for logistic regression models, which tend to be sensitive to extreme multicollinearity problems. Inflated standard errors and unreliability in the results stem from the problem of multicollinearity characterized by high correlations among predictor variables in logistic regression model. Traditional approaches like the Maximum Likelihood Estimator (MLE) and one-parameter ridge-type estimators often perform poorly under these settings, thus calling for the development of more robust approaches. This new proposal is called New Biased Two Parameter (NBTP), which extends the ridge regression framework by introducing additional biasing parameters customized for an extreme multicollinearity problem. The paper combines a theoretical analysis with extensive Monte Carlo simulations and real application to Pena data. It demonstrates that the new estimator, New Biased Two Parameter (NBTP), provides much more stable and accurate parameter estimates than previous methods. The results underscore the importance of using robust estimation methods within logistic regression, especially when multicollinearity may be widespread in fields such as medical research, finance, and the social sciences.

1. Introduction

Logistic regression is a statistical methodology that is applied broadly in modeling binary outcomes using experimental data from different fields, such as medicine, finance, social sciences, and others. Logistic regression is one of the most popular methodologies due to its capability of being very simple for interpretation and flexible enough during the analysis of the relationship existing between a binary dependent variable and several predictor variables. Multicollinearity is, however, always a critical problem in logistic regression if the predictor variables are highly correlated—that is, it has a high correlation coefficient. The occurrence of multicollinearity will seriously affect the reliability and accuracy of logistic regression estimates. It will cause a problem of inflated standard errors in logistic regression, thus complicating the task of determining how much each of the predictors contributes individually. Logistic models then are: The presence of multicollinearity in the data structure offers instability and unreliability in the estimates derived from the traditional estimation methods, such as the Maximum Likelihood Estimator (MLE). This is problematic because it may make its inferences misleading: among others, it can lead to the mistaken conclusion that important variables are insignificant, as well as confuse its inferences with false signs for the coefficients (Menard, 2002).

The MLE also assumes that predictors are not highly correlated with each other; however, if the assumption is violated, even a very small change in the data may have an important effect on the estimates (Wooldridge, 2010). Ridge regression was proposed as a method of biased estimation that adjusted the MLE by including a penalty term in the regression coefficients in order to reduce the effects of multicollinearity (Hoerl and Kennard, 1970). Ridge regression is one way of overcoming the variance inflation problem by shrinking the coefficients' size to have better estimate stability, at the cost of some bias (Tibshirani, 1996). Although ridge regression is quite efficiently better than traditional estimators, the standard approach, where there is only a single biasing parameter, may be too weak in extreme scenarios of multicollinearity where it obtains an unreasonably high predictor correlation (Obenchain 1977; Marquardt and Snee,

* Corresponding author. Tel.: +2348067096726
E-mail address: Ifeoluwa.oladapo@gmail.com (David I. Oladapo)
https://doi.org/10.62054/ijdm/0104.15
© 2024 Department of Mathematics, Modibbo Adama University.

1975). Due to the inadequacy of ridge regression, particularly under extremely high multicollinearity conditions, this paper introduces a new two-parameter ridge-type estimator named New Biased Two Parameter (NBTP), where the second parameter is strategically adjusted to account for the magnitude of multicollinearity existing in the dataset, hence offering a more profound approach to balance between bias and variance (Muniz and Kibria, 2009).

2. Methodology

2.1 Development of the New Biased Two Parameter (NBTP) Estimator

The New Biased Two Parameter (NBTP) estimator is developed as a special extension that can increase the performance of the logistic regression models in case of extreme multicollinearity. Multicollinearity itself is a definition in terms of logistic regression, which is dealing with binary response variables and the usual way of estimating the coefficients is through the Maximum Likelihood Estimator (MLE). However, in cases where the predictor variables are highly correlated, the MLE may give rise to unreliable estimates because of the inflated variances of the coefficient estimates. There are also some other references related to the derivation of the new estimator, such as Menard (2002) and Dormann *et al*. Ridge regression, first proposed by Hoerl and Kennard (1970), is one approach to solving multicollinearity. It does this by introducing a parameter which causes the coefficients to be biased toward zero and at the same above the ridge regression framework. This further allows flexibility in adjustment for the severity of multicollinearity to offer a robust solution where time penalizes their size. Since it reduces the level of variance of the estimates by accepting some bias, it is not sufficient for very highly conditioned matrices (X'WX) in case of extreme multicollinearity (Marquardt and Snee, 1975; Obenchain, 1977).

In this regard, the New Biased Two Parameter (NBTP) estimator remedies such constraints by introducing a second biasing parameter over and standard ridge regression with one biasing parameter is insufficiently used. Alkhamisi *et al*. [2006]. The reason the New Biased Two Parameter (NBTP) estimator attains a better bias-variance tradeoff over highly collinear data is that its two biasing parameters are adapted to every specific level of multicollinearity, ensuring more optimal balance, suggests Muniz and Kibria (2009).
A new two parameter ridge type estimator is proposed in this study following the works of Dawoud *et al* (2023) in linear regression. The estimator is expressed as thus:

$$\hat{\beta}_{NBTP} = (M + k(1+d)I)^{-1}(M - kI)\hat{\beta}_{ML} \tag{1}$$

2.2 Properties of $\hat{\boldsymbol{\beta}}_{NBTP}$

In this section, the property of the newly proposed estimator is discussed. For the newly developed estimator will be compared with some existing estimators, namely: Maximum Likelihood Estimator (MLE), Logistic Ridge Estimator (LRE), Logistic Liu Estimator (LLE), Logistic Kibra-Lukman Estimator (LKLE), and Logistics Modified Ridge Type Estimator (LMRTE).

Equations (2) to (7) are the listed estimators used in this work for theoretical comparison as thus;

$$\hat{\beta}_{MLE} = (M)^{-1} X'Wz \tag{2}$$

$$\hat{\beta}_{LRE} = (M + kI)^{-1} X'Wz \tag{3}$$

$$\hat{\beta}_{LLE} = (M + I)^{-1}(M + d)\hat{\beta}_{MLE} \tag{4}$$

$$\hat{\beta}_{LKLE} = (M + kI)^{-1}(M - kI)\hat{\beta}_{MLE} \tag{5}$$

$$\hat{\beta}_{LMRTE} = (M + k(1+d)I)^{-1} M\hat{\beta}_{MLE} \tag{6}$$

$$M = X'\hat{W}X \tag{7}$$

Recall that the logistic regression model is expressed by Auguilera *et al* in matrix form in terms of the logit transformation as $L = X\beta = XTT'\beta = Z\alpha$ where T= $[t_1,\dots,t_p]$ shows an orthogonal matrix with $Z'WZ = T'X'WXT = \Lambda$ *and* $\Lambda = diag(\lambda_1,\dots,\lambda_p)$, the λ_i are the ordered Eigen values of $X'WX$.

© 2024 Department of Mathematics, Modibbo Adama University.

The mean square error matrix (MSEM) of an estimator $\hat{\beta}$ is defined as

$$\text{MSEM}\left(\hat{\beta}\right)=Cov\left(\hat{\beta}\right)+bias\left(\hat{\beta}\right)\left(bias\left(\hat{\beta}\right)\right)' \tag{8}$$

where $Cov\left(\hat{\beta}\right)$ is the dispersion matrix and $bias\left(\hat{\beta}\right)=E\left(\hat{\beta}\right)-\beta$

The bias and dispersion matrix of $\hat{\alpha}_{prop3}$ can be computed as follows:

$$Bias(\hat{\alpha}_{NBTP}) = [(\Lambda + k(1+d)I)^{-1}(\Lambda - kI) - I]\alpha \tag{9}$$

$$Cov\left(\hat{\alpha}_{prop3}\right)=\left(\Lambda+k\left(1+d\right)I\right)^{-1}\left(\Lambda-kI\right)\Lambda^{-1}\left(\Lambda+k\left(1+d\right)I\right)^{-1}\left(\Lambda-kI\right) \tag{10}$$

The MMSE and MSE in terms of eigenvalues are defined respectively as

$$MMSE\left(\hat{\alpha}_{prop3}\right)=Cov\left(\hat{\alpha}_{prop3}\right)+Bias\left(\hat{\alpha}_{prop3}\right)Bias\left(\hat{\alpha}_{prop3}\right)' \tag{11}$$

$$=\begin{matrix}\left(\Lambda+k\left(1+d\right)I\right)^{-1}\left(\Lambda-kI\right)\Lambda^{-1}\left(\Lambda+k\left(1+d\right)I\right)^{-1}\left(\Lambda-kI\right)+\\ \left[\left(\Lambda+k\left(1+d\right)I\right)^{-1}\left(\Lambda-kI\right)-I\right]\alpha\alpha\left[\left(\Lambda+k\left(1+d\right)I\right)^{-1}\left(\Lambda-kI\right)-I\right]\end{matrix} \tag{12}$$

$$MSE\left(\hat{\alpha}_{prop3}\right)=tr\left(MMSE\left(\hat{\alpha}_{prop3}\right)\right)$$

$$MSE\left(\hat{\alpha}_{prop3}\right)=\sum_{i=1}^{p}\left[\frac{\left(\lambda_i-k\right)^2}{\lambda_i\left(\lambda_i+k\left(1+d\right)\right)^2}\right]+\sum_{i=1}^{p}\left[\frac{\left(\left(2+d\right)k\right)^2}{\left(\lambda_i+k\left(1+d\right)\right)^2}\right]\alpha_i^2 \tag{13}$$

2.2.1 Comparisons of $\hat{\alpha}_{prop3}$ with Existing Estimators using the MSEM Criterion

2.2.1.1 Comparison of $\hat{\alpha}_{prop3}$ and $\hat{\alpha}_{MLE}$

$$\hat{\alpha}_{MLE}=\left(M_1\right)^{-1}X'Wz \text{ with } MSEM\left(\hat{\alpha}_{MLE}\right)=\Lambda^{-1} \tag{14}$$

Theorem 1: $\hat{\alpha}_{prop3}$ is better than $\hat{\alpha}_{MLE}$ if

$$\alpha'\left(\left(\Lambda+k\left(1+d\right)I\right)^{-1}\left(\Lambda-kI\right)-I\right)'\left[\left(\Lambda^{-1}-\Lambda^{-1}\left(\Lambda+k\left(1+d\right)I\right)^{-2}\left(\Lambda-kI\right)^2\right)\right]^{-1}$$
$$\left(\left(\Lambda+k\left(1+d\right)I\right)^{-1}\left(\Lambda-kI\right)-I\right)\alpha<1$$

Proof

The difference of the dispersion is

$$Cov\left(\hat{\alpha}_{MLE}\right)-Cov\left(\hat{\alpha}_{prop3}\right)=\left(\Lambda^{-1}-\Lambda^{-1}\left(\Lambda+k\left(1+d\right)I\right)^{-2}\left(\Lambda-kI\right)^2\right)$$

$$=diag\left[\frac{1}{\lambda_i}-\frac{\left(\lambda_i-k\right)^2}{\lambda_i\left(\lambda_i+k\left(1+d\right)\right)^2}\right]_{i=1}^{p}$$

It is observed that $\left(\Lambda^{-1}-\Lambda^{-1}\left(\Lambda+k\left(1+d\right)I\right)^{-2}\left(\Lambda-kI\right)^2\right)$ is positive definite since for 0 < d < 1 and k >

© 2024 Department of Mathematics, Modibbo Adama University.

0, $\left(\lambda_i + k\left(1+d\right)\right) - \lambda_i\left(\lambda_i - k\right) > 0$. Hence, by lemma 2 the proof is completed.

2.2.1.2 Comparison of $\hat{\alpha}_{prop3}$ and $\hat{\alpha}_{LRE}$

$$\hat{\alpha}_{LRE} = \left(M_1 + kI\right)^{-1} M_1 \hat{\alpha}_{MLE} \tag{15}$$

And the MSE is given as:

$$MSE\left(\hat{\alpha}_{LRE}\right) = \sum_{j=1}^{p} \frac{\lambda_j}{\left(\lambda_j + k\right)^2} + k^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{\left(\lambda_j + k\right)^2}$$

Theorem 2: $\hat{\alpha}_{prop3}$ is better than $\hat{\alpha}_{LRE}$ if

$$\alpha'\left(\left(\Lambda + k\left(1+d\right)I\right)^{-1}\left(\Lambda - kI\right) - I\right)'\left[R_1 + \left(\left(\Lambda + k\right)^{-1} - I\right)\alpha\alpha'\left(\left(\Lambda + k\right)^{-1} - I\right)\right]^{-1}$$

$$\left(\left(\Lambda + k\left(1+d\right)I\right)^{-1}\left(\Lambda - kI\right) - I\right)\alpha < 1$$

$$R_1 = \left(\Lambda\left(\Lambda + k\right)^{-2} - \Lambda^{-1}\left(\Lambda + k\left(1+d\right)I\right)^{-2}\left(\Lambda - kI\right)^2\right)$$

Proof

The difference of the dispersion is

$$Cov\left(\hat{\alpha}_{LRE}\right) - Cov\left(\hat{\alpha}_{prop3}\right) = \left(\Lambda\left(\Lambda + k\right)^{-2} - \Lambda^{-1}\left(\Lambda + k\left(1+d\right)I\right)^{-2}\left(\Lambda - kI\right)^2\right)$$

$$= diag\left[\frac{\lambda_i}{\left(\lambda_i + k\right)^2} - \frac{\left(\lambda_i - k\right)^2}{\lambda_i\left(\lambda_i + k\left(1+d\right)\right)^2}\right]$$

It is observed that $\left(\Lambda\left(\Lambda + k\right)^{-2} - \Lambda^{-1}\left(\Lambda + k\left(1+d\right)I\right)^{-2}\left(\Lambda - kI\right)^2\right)$ will be positive definite if and only if $\lambda\left(\lambda_i + k\left(1+d\right)\right) - \left(\lambda_i - k\right)\left(\lambda_i + k\right) > 0$. For $0 < d < 1$ and $k > 0$, by lemma 3 the proof is completed.

2.2.1.3 Comparison of $\hat{\alpha}_{prop3}$ and $\hat{\alpha}_{LLE}$

$$\hat{\alpha}_{LLE} = \left(M_1 + I\right)^{-1}\left(X'Wz + d\hat{\alpha}_{MLE}\right) \tag{16}$$

The Mean Squared Error (MSE) of LLE is defined as:

$$MSE\left(\hat{\beta}_{LLE}\right) = \sum_{i=1}^{p} \frac{\left(\lambda_i + d\right)^2}{\lambda_i\left(\lambda_i + 1\right)^2} + \left(1 - d\right)^2 \sum_{i=1}^{p} \frac{\alpha_i^2}{\left(\lambda_i + 1\right)^2}$$

Theorem 3: $\hat{\alpha}_{prop3}$ is better than $\hat{\alpha}_{LLE}$ if

$$\alpha'\left(\left(\Lambda + k\left(1+d\right)I\right)\left(\Lambda - kI\right) - I\right)'\left[R_2 + \left(\left(\Lambda + 1\right)^{-1}\left(\Lambda + d\right) - I\right)\alpha\alpha'\left(\left(\Lambda + 1\right)^{-1}\left(\Lambda + d\right) - I\right)\right]^{-1}\left(\left(\Lambda + k\left(1+d\right)I\right)^{-1}\left(\Lambda - kI\right) - I\right)\alpha < 1$$

$$R_2 = \left(\Lambda^{-1}\left(\Lambda + d\right)^2\left(\Lambda + 1\right)^{-2} - \Lambda^{-1}\left(\Lambda + k\left(1+d\right)I\right)^{-2}\left(\Lambda - kI\right)^2\right)$$

Proof

The difference of the dispersion is

© 2024 Department of Mathematics, Modibbo Adama University.

$$Cov\left(\hat{\alpha}_{LLE}\right)-Cov\left(\hat{\alpha}_{prop3}\right)=\left(\Lambda^{-1}\left(\Lambda+d\right)^{2}\left(\Lambda+1\right)^{-2}-\Lambda^{-1}\left(\Lambda+k\left(1+d\right)I\right)^{-2}\left(\Lambda-kI\right)^{2}\right)$$

$$=diag\left[\frac{\left(\lambda_i+d\right)^2}{\lambda_i\left(\lambda_i+1\right)^2}-\frac{\left(\lambda_i-k\right)^2}{\lambda_i\left(\lambda_i+k\left(1+d\right)\right)^2}\right]$$

$\Lambda^{-1}\left(\Lambda+d\right)^{2}\left(\Lambda+1\right)^{-2}-\Lambda^{-1}\left(\Lambda+k\left(1+d\right)I\right)^{-2}\left(\Lambda-kI\right)^{2}$ will be positive definite if and only if $\left(\lambda_i+k\left(1+d\right)\right)\left(\lambda_i+d\right)-\left(\lambda_i-k\right)\left(\lambda_i+1\right)>0$ for k > 0 and 0 < d < 1. Hence, by lemma 3 the proof is completed.

2.2.1.4 Comparison of $\hat{\alpha}_{prop3}$ and $\hat{\alpha}_{LKLE}$

$$\hat{\alpha}_{LKLE}=\left(M_1+kI\right)^{-1}\left(M_1-kI\right)\alpha_{MLE} \tag{17}$$

The Mean Squared Error (MSE) of LKLE is defined as:

$$MSE\left(\hat{\beta}_{LKLE}\right)=\sum_{i=1}^{p}\frac{\left(\lambda_i-k\right)^2}{\lambda_i\left(\lambda_i+k\right)^2}+4k^2\sum_{i=1}^{p}\frac{\alpha_i^2}{\left(\lambda_i+k\right)^2}$$

Theorem 4: $\hat{\boldsymbol{\alpha}}_{NBTP}$ is better than $\hat{\boldsymbol{\alpha}}_{LKLE}$ if

$$\alpha'\left(\left(\Lambda+k\left(1+d\right)I\right)^{-1}\left(\Lambda-kI\right)-I\right)'\left[R_3+\left(\left(\Lambda+k\right)^{-1}\left(\Lambda-k\right)-I\right)\alpha\alpha'\left(\left(\Lambda+k\right)^{-1}\left(\Lambda-k\right)-I\right)\right]^{-1}\left(\left(\Lambda+k\left(1+d\right)I\right)^{-1}\left(\Lambda-kI\right)-I\right)\alpha<1$$

$$R_3=\left(\Lambda^{-1}\left(\Lambda-k\right)^{2}\left(\Lambda+k\right)^{-2}-\Lambda^{-1}\left(\Lambda+k\left(1+d\right)I\right)^{-2}\left(\Lambda-kI\right)^{2}\right)$$

Proof

The difference of the dispersion is

$$Cov\left(\hat{\alpha}_{LKLE}\right)-Cov\left(\hat{\alpha}_{prop3}\right)=\left(\Lambda^{-1}\left(\Lambda-k\right)^{2}\left(\Lambda+k\right)^{-2}-\Lambda\left(\Lambda+k\left(1+d\right)I\right)^{-2}\left(\Lambda-kI\right)^{2}\right)$$

$$=diag\left[\frac{\left(\lambda_i-k\right)^2}{\lambda_i\left(\lambda_i+k\right)^2}-\frac{\left(\lambda_i-k\right)^2}{\lambda_i\left(\lambda_i+k\left(1+d\right)\right)^2}\right]$$

$\Lambda^{-1}\left(\Lambda-k\right)^{2}\left(\Lambda+k\right)^{-2}-\Lambda^{-1}\left(\Lambda+k\left(1+d\right)I\right)^{-2}\left(\Lambda-kI\right)^{2}$ will be positive definite if and only if $\left(\lambda_i+k\left(1+d\right)\right)\left(\lambda_i-d\right)-\left(\lambda_i-k\right)\left(\lambda_i+k\right)>0$ for k > 0 and 0 < d < 1. Hence, by lemma 3 the proof is completed.

2.2.1.5 Comparison of $\hat{\alpha}_{prop3}$ and $\hat{\alpha}_{LMRTE}$

$$\hat{\alpha}_{LMRTE}=\left(M_1+k\left(1+d\right)I\right)^{-1}M_1\hat{\alpha}_{MLE} \tag{18}$$

And the MSE is given as:

$$MSE\left(\hat{\alpha}_{LMRTE}\right)=\sum_{i=1}^{p}\frac{\lambda_i+\left(k\left(1+d\right)\right)^2\alpha_i^2}{\left(\lambda_i+k\left(1+d\right)\right)^2}$$

© 2024 Department of Mathematics, Modibbo Adama University.

Theorem 5 $\hat{\alpha}_{prop3}$ is better than $\hat{\alpha}_{LMRTE}$ if

$$\alpha'\left(\left(\Lambda+k(1+d)I\right)^{-1}(\Lambda-kI)-I\right)'\left[R_4+\left(\left(\Lambda+k(1+d)\right)^{-1}-I\right)\alpha\alpha'\left(\left(\Lambda+k(1+d)\right)^{-1}-I\right)\right]^{-1}\left(\left(\Lambda+k(1+d)I\right)^{-1}(\Lambda-kI)-I\right)\alpha<1$$

$$R_4=\left(\left(\Lambda+\left(k(1+d)\right)\right)^2\Lambda-\Lambda^{-1}\left(\Lambda+k(1+d)I\right)^{-2}\left(\Lambda-kI\right)^2\right)$$

Proof

The difference of the dispersion is

$$Cov\left(\hat{\alpha}_{LMRTE}\right)-Cov\left(\hat{\alpha}_{prop3}\right)=\left(\Lambda\left(\Lambda+k(1+d)\right)^{-2}-\Lambda^{-1}\left(\Lambda+k(1+d)I\right)^{-2}\left(\Lambda-kI\right)^2\right)$$

$$=diag\left[\frac{\lambda_i}{\left(\lambda_i+k(1+d)\right)^2}-\frac{(\lambda_i-k)^2}{\lambda_i\left(\lambda_i+k(1+d)\right)^2}\right]$$

It is observed that $\left(\Lambda\left(\Lambda+k(1+d)\right)^{-2}-\Lambda^{-1}\left(\Lambda+k(1+d)I\right)^{-2}\left(\Lambda-kI\right)^2\right)$ will be positive definite if and only if $\lambda_i-(\lambda_i-d)>0$. For 0 < d < 1 and k > 0, by lemma 3 the proof is completed.

2.3 Selection of biasing parameters k and d for $\hat{\alpha}_{NBTP}$

The selection of the biasing parameters k and d is crucial to the performance of the New Biased Two Parameter (NBTP) estimator. The parameter (k) is chosen to minimize the Mean Square Error (MSE) of the estimates, balancing the trade-off between bias and variance. The parameter (d) is introduced to provide additional flexibility in the penalty, allowing the estimator to adapt to varying levels of multicollinearity.

$$\text{MSE}\left(\alpha(k,d)\right)=E\left[\left(\hat{\alpha}(k,d)-\alpha\right)'\left(\alpha(k,d)-\alpha\right)\right] \tag{19}$$

$$\text{g(k,d)}=MSE\left(\hat{\alpha}(k,d)\right)=tr\left[MSEM\left(\hat{\alpha}(k,d)\right)\right]$$

$$MSE\left(\hat{\alpha}_{prop3}\right)=\sum_{i=1}^{p}\left[\frac{(\lambda_i-k)^2}{\lambda_i\left(\lambda_i+k(1+d)\right)^2}\right]+\sum_{i=1}^{p}\left[\frac{\left((2+d)k\right)^2}{\left(\lambda_i+k(1+d)\right)^2}\right]\alpha_i^2 \tag{20}$$

Considering d to be fixed, an optimal value of k is the value that minimizes $MSE\left(\hat{\alpha}_{prop2}\right)$.

Then, by differentiating g(k, d) w.r.t. k and equating to 0, we have

$$k=\frac{\lambda_i}{\lambda_i(d+2)\alpha_i^2+1} \tag{21}$$

However, k depends on the unknown α_i. For practical purposes, it will be replaced by its unbiased estimator $\hat{\alpha}_i$. Hence, this will be obtained

$$\hat{k}=\frac{\lambda_i}{\lambda_i(d+2)\hat{\alpha}_i^2+1} \tag{22}$$

Equation (3.48) returns the biasing parameter for the KL estimator when d =0, which is defined as follows:

$$\hat{k}=\frac{\lambda_i}{1+2\lambda_i\hat{\alpha}_i^2} \tag{23}$$

© 2024 Department of Mathematics, Modibbo Adama University.

$$\hat{k}_{MAX} = Maximum\left(\frac{\lambda_i}{\lambda_i\left(d+2\right)\hat{\alpha}_i^2+1}\right) \tag{24}$$

$$\hat{k}_{MIN} = Minimum\left(\frac{\lambda_i}{\lambda_i\left(d+2\right)\hat{\alpha}_i^2+1}\right) \tag{25}$$

$$\hat{k}_{MED} = Median\left(\frac{\lambda_i}{\lambda_i\left(d+2\right)\hat{\alpha}_i^2+1}\right) \tag{26}$$

$$\hat{k}_{AM} = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{\lambda_i}{\lambda_i\left(d+2\right)\hat{\alpha}_i^2+1}\right) \tag{27}$$

$$\hat{k}_{HM} = p\sum_{i=1}^{p}\left(\frac{\lambda_i}{\lambda_i\left(d+2\right)\hat{\alpha}_i^2+1}\right) \tag{28}$$

$$\hat{k}_{MR} = \frac{1}{2}\left(k_{MAX} + k_{MIN}\right) \tag{29}$$

2.4. Simulation of Study

2.4.1 Design of Simulation Experiments

A full Monte Carlo simulation study was conducted to assess the behavior of the estimator. Several simulation designs were considered assuming logistic regression models. Three levels of multicollinearity specified by the correlation coefficient p among the predictor variables: 0.8, 0.9, 0.95, and 0.99 have been considered, thus comparing moderate and severe multicollinearity. The simulation was also done for varying numbers of predictor variables, 10, 11, and 12 and the size of the sample varies from 150 up to 1000. These are of such settings that they represent a wide spectrum of practical situations for both small data with many predictors and larger data with fewer predictors. Simulation runs were carried out for each combination of (p), (p), and (n). In each run, the performance of the New Biased Two Parameter (NBTP) estimator compared to that of the MLE and other ridge-type estimators, such as the Logistic Ridge Estimator (LRE), Logistic Liu Estimator (LLE), Logistic-Kibra-Lukman (K-L), and Logistic Modified Ridge type was evaluated.

2.4.2 Performance Metrics

The primary metric for evaluating the performance of the estimators was the Mean Square Error (MSE) of the estimated coefficients. The MSE was calculated as the average of the squared differences between the true and estimated coefficients across all simulation runs. Additional metrics included bias, variance, and the condition number of the predictor matrix \ X'X \.

3.1. Results and Discussion

The Logistic estimator was proposed alongside with different four (4) biasing parameter (k) and compared with other four existing estimators namely the LLE, LRE, LKLE and LMRTE with the MLE itself inclusive. The findings from the simulation are provided and discussed below.

© 2024 Department of Mathematics, Modibbo Adama University.

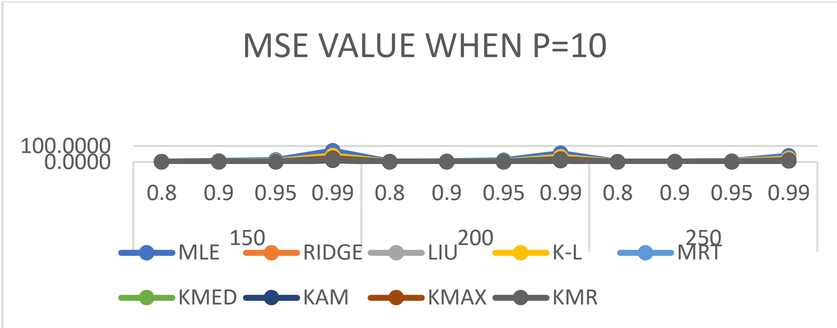

Figure 1: MSE values of estimators at sample sizes 150, 200 and 250

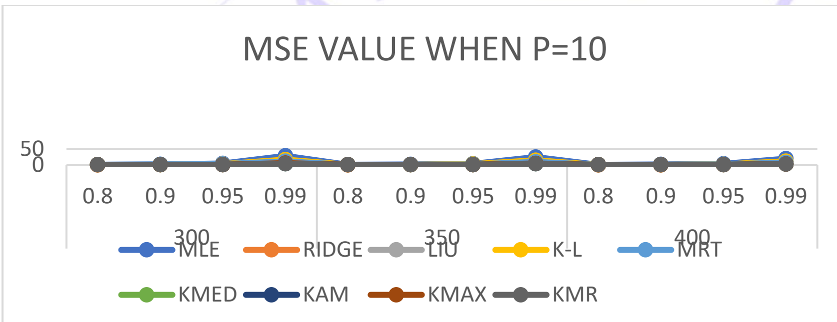

Figure 2: MSE values of estimators at sample sizes 300, 350 and 400

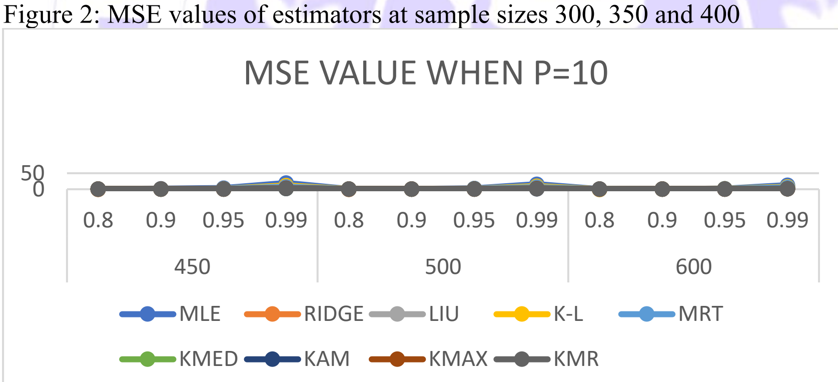

Figure 3: MSE values of estimators at sample sizes 450, 500 and 600

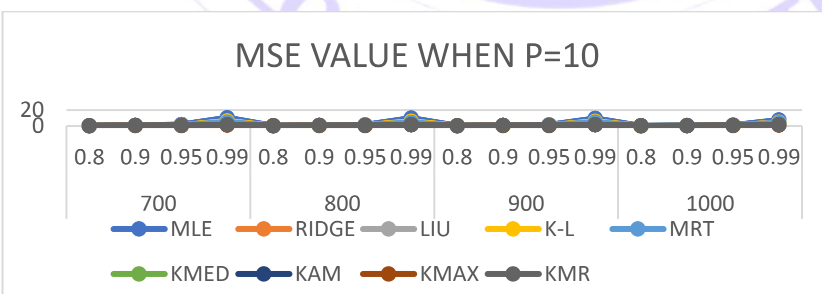

Figure 4: MSE values of estimators at sample sizes 700, 800, 900 and 1000

The simulation results demonstrate that the New Biased Two Parameter (NBTP) estimator consistently achieves

© 2024 Department of Mathematics, Modibbo Adama University.

the smallest MSE across all levels of multicollinearity and sample sizes. Particularly at the highest level of multicollinearity p = 0.99, NBTP significantly outperforms the MLE and other ridge-type estimators, with a substantial reduction in MSE.

For example, at (p = 0.99) with (p = 12) and (n = 150), the MSE for New Biased Two Parameter (NBTP) was observed to be more than 30% lower than that of the MLE, indicating a substantial improvement in estimation accuracy. The bias introduced by was minimal, particularly when compared to the reduction in variance. This resulted in a lower overall MSE, confirming the theoretical advantages of New Biased Two Parameter (NBTP) in handling severe multicollinearity. The results also showed that the performance of New Biased Two Parameter (NBTP) improves with increasing sample size, further stabilizing the estimates.

4.2 Real-life dataset Application of New Biased Two Parameter (NBTP)

The New Biased Two Parameter (NBTP) estimator was applied to the pena dataset, and its performance was compared with that of the MLE and other ridge-type estimators under this study. The focus was on evaluating the accuracy and stability of the estimated coefficients, as well as the overall fit of the logistic regression model.

Table 1: Analysis result using New Biased Two Parameter (NBTP) estimator on real life data

	MLE	**RIDGE**	**LIU**	**K-L**	**MRT**	**NBTPMED**	**NBTPAM**	**NBTPMAX**	**NBTPMR**
β_1	1.0834	0.0040	0.8473	0.8445	0.7895	0.5461	0.5801	0.4629	0.6164
β_2	-0.0543	-0.0276	-0.0509	-0.0508	-0.0417	-0.0396	-0.0406	-0.0373	-0.0416
β_3	0.0529	0.0155	0.0546	0.0547	0.0252	0.0309	0.0321	0.0281	0.0334
β_4	-0.4255	-0.0063	-0.3581	-0.3576	-0.2713	-0.2109	-0.2239	-0.1795	-0.2378
SMSE	0.3005	0.2196	0.2010	0.2002	0.2168	0.1111	3.9927	0.0067	0.0003

4.3 Discussion

The application of New Biased Two Parameter (NBTP) to the real-life dataset yielded several notable improvements over traditional methods. The MSE of the NBTP estimates was consistently lower than that of the MLE and other ridge-type estimators, confirming the robustness of New Biased Two Parameter (NBTP) in handling severe multicollinearity. In particular, the standard errors of the New Biased Two Parameter (NBTP) estimates were significantly smaller, indicating greater stability and reliability. This reduction in standard errors allowed for more accurate interpretation of the predictors' effects on the health outcome. For instance, variables that were deemed insignificant by the MLE due to high standard errors were found to be significant when estimated using New Biased Two Parameter (NBTP), providing more meaningful insights into the relationships between predictors and the outcome.

Moreover, the New Biased Two Parameter (NBTP) estimator produced more reasonable coefficient estimates, with signs and magnitudes that aligned better with prior expectations and the underlying theory. This suggests that New Biased Two Parameter (NBTP) not only improves the accuracy of the estimates but also enhances their interpretability, making it a valuable tool for researchers dealing with complex, multicollinear data.

5. Conclusion

This paper introduces and evaluates a robust two-parameter ridge-type estimator, New Biased Two Parameter (NBTP), designed to handle severe multicollinearity in logistic regression models. Through both theoretical analysis and extensive simulation studies, New Biased Two Parameter (NBTP) has been shown to outperform traditional estimators, particularly in extreme multicollinear scenarios. The real-life application further demonstrated its practical utility, offering more stable and accurate estimates that are crucial for reliable decision-making in fields like medicine, finance, and the social sciences.

The findings of this study highlight the importance of developing specialized estimation techniques to address the challenges posed by multicollinearity in logistic regression. NBTP provides a significant advancement in this area,

© 2024 Department of Mathematics, Modibbo Adama University.

offering a flexible and effective solution that can be tailored to different levels of multicollinearity. Future research could explore the extension of New Biased Two Parameter (NBTP) to other types of regression models, as well as its application in high-dimensional data settings.

References

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J. R., Gruber, B., Lafourcade, B., Leitao, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27-46.

Gujarati, D. N. (2004). Basic econometrics (4th ed.). McGraw-Hill.

Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.

Marquardt, D. W., and Snee, R. D. (1975). Ridge regression in practice. The American Statistician, 29(1), 3-20.

Menard, S. (2002). *Applied logistic regression analysis (2nd ed.)*. Sage.

Muniz, G., and Kibria, B. M. G. (2009). On some ridge regression estimators: An empirical comparisons. *Communications in Statistics-Simulation and Computation*, 38(3), 621-630.

Obenchain, R. L. (1977). Classical F-tests and confidence regions for ridge regression. *Technometrics*, 19(4), 429-439.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data (2nd ed.)*. MIT Press.

© 2024 Department of Mathematics, Modibbo Adama University.