Journal of Statistical and Econometric Methods, vol.2, no.4, 2013, 161-171
ISSN: 2241-0384 (print), 2241-0376 (online)
Scienpress Ltd, 2013

A New Logistic Ridge Regression Estimator Using Exponentiated Response Function

U.P. Ogoke[1], E.C. Nduka[2] and M.E. Nja[3]

Abstract

The logistic ridge regression estimator was designed to address the problem of variance inflation created by the existence of collinearity among the explanatory variables in logistic regression models. To reduce the bias introduced by the logistic ridge estimator, and at the same time achieve variance reduction, a modified generalized logistic ridge regression estimator is proposed in this paper. By exponentiating the response function, the weight matrix is enhanced, thereby reducing the variance. The modified estimator is jackknifed to achieve bias reduction.

Mathematics Subject Classification: 62P
Keywords: Jackknife estimator, variance inflation, bias, weight matrix, biasing constant, modified logistic

1 Introduction

Regression methods have become an integral component of data analysis in describing the relationship between a response variable and one or more

[1] Department of Mathematics and Statistics, University of Port Harcourt, Rivers State. Nigeria.
[2] Department of Mathematics and Statistics, University of Port Harcourt, Rivers State. Nigeria.
[3] Department of Mathematics, Federal University, Lafia, Nigeria.

Article Info: *Received* : October 7, 2013. *Revised* : November 10, 2013.
Published online : December 1, 2013.

explanatory variables. When the response variable is binary or dichotomous taking in two possible values, the Logistic regression model becomes the standard method of analysis. However, when there is exact collinearity among the regressors, the IWLS update fails because in that case, the information matrix assumes singularity.When there exist collinearity(ies) in the explanatory variables, the Ridge i.e. in this case, Logistic Ridge Regression is applied since collinearity inflates the variance, bias and indeterminable parameter estimates. Now, when the components of the explanatory variables are both categorical and continuous, it necessitates a generalized logistic ridge regression to handle. If an information matrix $I_m=X^{'}WX$ is ill-conditioned, it means that there is the existence of collinearity among the explanatory variables. Where there is exact collinearity, the information matrix assumes singularity and the iterative weighted least squares method collapses. The ill-conditioning of a nonsingular information matrix can be detected using its condition number? (Lesaffre and Marx, 1993). Let $\lambda_1, \lambda_2, \ldots, \lambda_t$ be the eigenvalues of the information matrix I_m in descending order. The condition number of I_m is given as: $k=(\lambda_{max}/\lambda_{min})^{1/2}$. When there is no collinearity at all, the condition number is equal to one. As collinearity increases, the condition number increases. When there is exact collinearity, the information matrix assumes singularity. Lesaffre and Marx (1993) showed within the logistic regression framework that exact collinearity among the regressors and non-existence of the maximum likelihood estimators are the only causes of singularity for the information matrix. An ill-conditioning situation results in inflated variances of the estimated regression parameters.

Note: let $\pi(x) = E(Y|X = x)$ represents the conditional mean of Y given x when the logistic distribution is used, i.e.

$$\pi(x) = \frac{e^{\beta_o+\beta_{1x}}}{1+\ e^{\beta_o+\beta_{1x}}} \tag{1a}$$

A simple logit transformation, defined in terms of π(x) is

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \cdots, -\infty < x < \infty \tag{1b}$$

The logit g(x) is linear in its parameter and may be continuous, hence $-\infty < x < \infty$.

$$g(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p = X^1\beta \tag{1c}$$

is MultipleLinear Regression Model (MLRM) .

The logistic regression model is a special case of the Generalized Linear model in the exponential family,

$$F(y;\theta,\varphi) = exp\left[\frac{\{x\theta + b(\theta)\}}{\alpha(\varphi)} + c(x,\varphi)\right] \tag{1d}$$

for some specific functions *a(.), b(.) and c(..)* , where $X = (x_0, x_1, \ldots, x_k)$ be an *n×(k+1)* design matrix and *y* is an *n×1* response vector. A Generalized Linear model is one in which each component of the response variable Y has a distribution in the exponential family, taking the form

$$F(y;\theta,\varphi) = exp\left[\frac{\{x\theta + b(\theta)\}}{\alpha(\varphi)} + c(x,\varphi)\right]$$

for some specific functions *a(.), b(.) and c(..)* .

It is also stated as

$$z_i = \sum_{j=1}^{p} x_{ij}\beta_j + e_i h'(\mu_i), \quad i = 1, 2, \dots, n \qquad \text{(1e)}$$

where z_i = adjusted dependent variate

X_{ij} = *(i,j)*thelement of the design matrix.

β_j = jthparameter effects.

$h'(\mu_i)$ = link function, e_i = residual error

μ is the response function. When the link function $h(\mu)$ is defined as the logit, the model (1d) is called the logistic regression model.

The problem of collinearity in a given set of data is responsible for the proposed modified Logistic Ridge regression estimator where the response function is exponentiated.A proposed jackknife estimator provides reduction in bias normally associated with Ridge regression.

2 Methodology

2.1 The Iterative Weighted Least Squares Method

Maximum likelihood methods are often used in the solution of Generalized Linear models for which the Logistic regression is a special case. Principal among them is the Iterative Weighted Least Squares (IWLS) method.

The IWLS update is stated as

$$\hat{\beta} = (X'WX)^{-1}X'WZ \qquad \text{(2.1)}$$

where

$$\{X'WX\}_{rs} = -E\left(\frac{\partial^2 \iota}{\partial\beta_r\partial\beta_s}\right)$$

$$l(\beta, y) = \sum\sum y_i x_{ij}\beta_j - \sum m_i \log\left(1 + exp\sum x_{ij}\beta_j\right) \qquad \text{(2.2)}$$

$$Z = \eta + (y-\mu)\frac{d\eta}{d\mu} \qquad \text{(2.3)}$$

is the adjusted dependent variate for η, the systematic component of the model. However, when there is exact collinearity among the regressors, the IWLS update fails because in that case, the information matrix assumes singularity. Where collinearity is not exact, the variances of parameter estimates are inflated in accordance with the degree of collinearity. To overcome this problem, the logistic ridge regression estimator was first suggested (Lesaffre and Marx, 1993) by Schaefer, Roi and Wolfe (1984). Earlier, Hoerl and Kennard (1970) had proposed the ridge regression estimator in the context of the General linear model.

The Logistic Ridge regression estimator is a modification of the IWLS estimator where small amount of weights are added to the diagonal elements of the information matrix in order to prevent the singularity of the matrix. By so doing, the IWLS algorithm can be executed even for cases of exact collinearity.

2.2 The Logistic Ridge Regression Estimator

The Logistic Ridge regression estimator is given as

$$\hat{\beta} = (X'WX + CI)^{-1}X'WZ \tag{2.4}$$

Where C is an identity of biasing constants. In the case of Ordinary Logistic Ridge regression estimator, the elements of C are all equal and chosen by successive trials. By some canonical transformation, the Ordinary Logistic Ridge regression estimator is stated as

$$\hat{\beta}_{ORE}^{(k)} = V\hat{\alpha}_{ORE}^{(k)} = V(I - CF_c^{-1})\hat{\alpha}^{(k)}$$

Where $\hat{\alpha}^{(k)} = V'\hat{\beta}^{(k)} = \hat{\alpha}$ at the kth-iteration. V = a p×p matrix whose columns are eigenvectors of X'X

$$F_c = diag(\lambda_1 + c, \lambda_2 + c, \dots, \lambda_k + c)$$

$c_1 = c_2 = \dots = c_1 = c, c \geq 0, c$ is a biasing constant.The Generalized Logistic Ridge estimator is obtained by generalizing the biasing constant c, which is stated in canonical form as

$$\hat{\beta}_{GRE}^{(k)} = V\hat{\alpha}_{GRE}^{(k)} = V(I - CF^{-1})\hat{\alpha}^{(k)}$$

where $\hat{\alpha}^{(k)} = V'\hat{\beta}^{(k)} = \hat{\alpha}$ at the kth – iteration as earlier defined.

$$F_c = diag(\lambda_1 + c, \lambda_2 + c, \dots, \lambda_k + c_k), c_i \geq 0, c_i \quad (i = 1, 2, \dots, k)$$

is an element of $C = diag(c_1 c_2, \dots, c_k)$ which is also obtained by successive guesses. λ_i is the ith eigenvalue of $(X'WX + CI)$. In General Linear models c_i has been estimated by several authors, including Khalaf and Shukur (2005).

Quenonille (1956) introduced a Jackknife technique which was later used by Singh et al (1986) to propose an almost unbiased ridge estimator as a method for reducing the bias created by the Ridge method. Starting with the multiple linear regression model.

$$Y = X\beta + e$$

where Y is an $(n\times 1)$ vector of observed responses. X is an $(n\times p)$ design matrix, β is a $(p\times 1)$ vector of regression coefficients and e is an $(n\times 1)$ vector of residual errors, such that $E(e) = 0$ and $E(ee') = \sigma^2 I$. Singh et al (1986) obtained its canonical transformation as

$$Y = Z\alpha + e \tag{2.5}$$

where
$Z = XV$ and

$$\alpha = V'\beta \tag{2.6}$$

$Z'Z = J = diag(\lambda_1\lambda_2, \dots, \lambda_k), \lambda_i$ is the *ith* eigenvalue of $X'X$.
By this transformation, they obtained the Generalized Estimator $\hat{\alpha}_{GRE}$ as

$$\hat{\alpha}_{GRE} = (I - CF^{-1})\hat{\alpha} \tag{2.7}$$

where

$$C = diag(c_1 c_2, \dots, c_k), c_i > 0$$

$F_c = J + C = diag(\lambda_1 + c_1, \lambda_2 + c_2, \dots, \lambda_k + c_k)$ and $\hat{\alpha}$ is the OLS estimator of α .

$$\hat{\beta}_{GRE} = V\hat{\alpha}_{GRE} = V(I - CF^{-1})\hat{\alpha} \tag{2.8}$$

Following the lines of Hinkley (1977), Singh et al (1986) derived the Jackknifed form of (2.5), i.e. of $\hat{\alpha}_{GRE}$ as

$$\hat{\alpha}_{JRE} = [I - (CF^{-1})^2]\hat{\alpha} \tag{2.9}$$

so that

$$\hat{\beta}_{JRE} = V\hat{\alpha}_{JRE} = V[I - (CF^{-1})^2]\hat{\alpha} \tag{2.10}$$

Nja (2013) extended Jackknife estimation to the logistic Ridge Regression model by redefining models (2.4) and (2.5) respectively as

$$Z = X\beta + eh'(\mu) \tag{2.11}$$

and

$$Z = S\alpha + eh'(\mu) \tag{2.12}$$

where $S = XV$ and V is the matrix whose columns are the eigenvectors of the information matrix as $(X'WX + CI)$ and $h(\mu)$ is as earlier defined.

Batah (2011) showed in the context of the General Linear Model that the bias of the Modified Jackknife Ridge estimator is smaller than the bias of the Generalized Ridge estimator. The theorem and proof by Khurana, M. Chaubey, Y.P; and Chandra, S. (2012) shown will help to buttress this fact.

2.3 The Proposed Estimator

The proposed Modified Logistic Ridge regression estimator is stated as

$$\hat{\beta} = (X'W_{\sqrt{1+\gamma}}X + CI)^{-1}X'W_{\sqrt{1+\gamma}}Z_{\sqrt{1+\gamma}} \tag{2.13}$$

where $W_{\sqrt{1+\gamma}} = diag\left[m_i\mu_i^{\sqrt{1+\gamma}}(1 - \mu_i^{\sqrt{1+\gamma}}\right]$

$m_i = ith$ group sub total

$\mu_i = ith$ group response function

For $0 \le \gamma \le 1$

i = number of groups

$Z_{\sqrt{1+\gamma}}$ is a column vector with elements

$$Z_i = \eta + (\frac{y_i}{m_i} - \mu_i^{\sqrt{1+\gamma}})\frac{1}{\mu_i^{\sqrt{1+\gamma}(1-\mu_i^{\sqrt{1+\gamma}})}} \tag{2.14}$$

where

$$\eta_i = x_{ij}\hat{\beta}_0 + x_{ij}\hat{\beta}_1 + \cdots + x_{ij}\hat{\beta}_{p-1} \tag{2.15}$$

p = number of parameters to be estimated.

yi number of favourable outcomes.

Using equations (2.11) and (2.12) and borrowing from the Jackknife method adopted by Singh et al (1986), we propose the following Jackknife Logistic Ridge regression estimator.

$$\hat{\alpha}_{JMLR} = [I - (CF^{-1})^2]\hat{\alpha} \tag{2.16}$$

where $F = diag(\lambda_i + c_i), \lambda_\iota$ is the ith eigenvalue of the matrix $(X'W_{\sqrt{1+\gamma}}X + CI)$ and c_i is the *i*th element of the matrix C of the generalized biasing constants and α is as previously defined.

3 Numerical Illustration

The table below is that of 4 subpopulations characterized by sex, Diastolic Blood Pressure (DBP) and Body Mass Index (BMI) showing the number of people whose survival time from the date of medication is greater than or equal to 10 years and the number of people whose survival time is less than 10 years. The probability of a person surviving 10 or more extra years is modelled using the modified Logistic Ridge regression proposed in this work.

Table 1: Survival Data

Sex x_1	Diastolic B.P x_2	BMI x_3	No.Surviving ≥10 yrs Y_i	No.surviving <10 yrs	Total
M	< 90	20.1	9	6	15
M	≥90	25.3	6	10	16
F	< 90	18.3	7	6	13
F	≥ 90	37	8	10	18
			30	32	62

3.1 Results

The results of parameter estimates and their variances are presented for the IWLS, the Logistic Ridge estimator and the modified Logistic Ridge estimator. Also presented are results for the variances and bias of estimates for the Logistic Ridge, the modified logistic ridge parameters and those of their corresponding Jackknife parameter estimates. Variances and bias of their parameter estimates and those of their corresponding Jackknife models are also presented.

Table 2: β parameter estimates

Estimator	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
IWLS	-1.700	0.179	1.124	0.040
Logistic. Ridge	-0.3551	-0.1642	-1.0824	0.0372
Jackknife logistic Ridge	-0.3551	-0.1642	-1.0824	0.0372
Modified logistic ridge	-0.6846	-0.1299	-0.9207	-0.0316
Jackknife Modified Logistic Ridge	-0.6846	-0.1299	-0.9207	-0.0316

Table 3: α parameter estimates

Estimator	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
Logistic Ridge	0.9086	0.5440	-0.3919	-0.0042
Jackknife logistic Ridge	0.9317	0.5484	- 0.3951	-0.0042
Modified logistic ridge	1.0647	0.2303	-0.2920	-0.0170
Jackknife Modified Logistic Ridge	1.0917	0.2321	-0.2943	-0.0170

Table 4: Variances of β Parameter Estimates

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
IWLS	7.0969	0.4502	1.1236	0.0059
Logistic Ridge	0.6637	0.1045	0.0698	0.0000
Modified logistic ridge	0.6638	0.1045	0.0698	0.0000

Table 5: Variances of α parameter estimates

	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
Logistic Ridge	0.6304	0.1028	0.0687	0.0000
Jackknife logistic Ridge	0.6629	0.1045	0.0698	0.0000
Modified logistic Ridge	0.6305	0.1028	0.0687	0.0000
Jackknife Modified Logistic Ridge	0.0030	0.1045	0.0698	0.0000

Table 6: Bias α estimates

	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
Logistic Ridge	-0.0237	-0.0045	0.0032	0.0000
Jackknife logistic Ridge	-0.0006	-0.0000	0.0000	0.0000
Modified logistic Ridge	-0.0278	-0.0019	0.0024	0.0000
Jackknife Modified Logistic Ridge	-0.0007	-0.0000	0.0000	0.0000

3.2 Discussion

From the results, it can be seen that the Modified Generalized Logistic Ridge estimator is superior to the Generalized Logistic Ridge estimator in terms of bias. Both estimators especially have approximately the same variances of parameter estimates in both β and α estimates (see Tables 4, and 5). By introducing the Jackknifed estimator to the Generalized Logistic Ridge and the Modified Generalized Logistic Ridge estimators, the bias of the Jackknife estimators drop respectively (Table 6). The purpose of modifying the response function is to enhance the weight matrix so that by jackknifing the modified estimator bias reduction can be achieved.

Theorem 3.1
Let *K* be a *(p×p)* diagonal matrix with non-negative entries, then the difference of total squared biases of the modified Jackknife Ridge (MJR) and Generalized Ridge estimators (GRE) of β as given by

$$D_2 = \sum\left\{\left|Bias(\hat{\beta}_{MJR})\right|_i^2 - \left|Bias(\hat{\beta}_{GRE})\right|_i^2\right\} \qquad (3.1a)$$

is positive.
Proof: using the expression for MJR given as

$$\hat{\alpha}_{MJR} = [I - (CF^{-1})^2][I - CF^{-1}]\hat{\alpha} = [I - (F^{-1}\phi C)] \qquad (3.1b)$$

where
$\phi = (I + CF^{-1} - C^*F^{-1})$ and $C^* = CF^{-1}C$.
we have

$$Bias(\hat{\alpha}_{GRE}) = -(CF^{-1})|\alpha| \qquad (3.1c)$$

Also using the expression for GRE of α given as

$$\widehat{\alpha}_{GRE} = [I - CF^{-1}]\hat{\alpha} \tag{3.1d}$$

we have

$$Bias(\hat{\alpha}_{GRE}) = -(CF^{-1})|\alpha|.$$

Comparing (3.1b) and (3.1c) component – wise, we have

$$\left|Bias(\hat{\alpha}_{MJR})\right|_i - |Bias(\hat{\alpha}_{GRE})|_i = \frac{c_i\phi_i}{\lambda_i + c_i}|\hat{\alpha}_i| - \frac{c_i}{\lambda_i + c_i}|\hat{\alpha}_i|$$

$$= \frac{c_i\left[1 + \frac{c_i}{\lambda_i + c_i} - \frac{c_i^2}{(\lambda_i + c_i)^2}\right]}{\lambda_i + c_i}|\hat{\alpha}_i| - \frac{c_i}{\lambda_i + c_i}|\hat{\alpha}_i| = \frac{\lambda_i c_i^2}{(\lambda_i + c_i)^3}|\hat{\alpha}_i|$$

which is a positive quantity. This proves the result.
Both theorem and proof are extended to the Logistic Ridge estimator by redefining F as $F = diag(\lambda_i + c_i)$ where λ_i is the *ith* eigenvalue of the matrix

$$(X'WX + CI)$$

and $\hat{\alpha}$ is defined as $\hat{\alpha} = V'\hat{\beta}$ with

$$\hat{\beta} = (X'WX + CI)^{-1}X'WZ.$$

The bias of an estimator in which the weight function has been modified differs only slightly from that of a Logistic Ridge estimator in favour of the modified estimator. This is also true when the modification is on the response function. To achieve a significant reduction in bias for any type of Logistic Ridge estimator, the Jackknife procedure should be applied as demonstrated in this paper. The Jackknife procedure is applied to the Generalized Logistic Ridge and the modified response function Logistic Ridge estimators. In both cases, it is observed that there is an enormous reduction in bias in favour of the Jackknife estimators. We used the above illustrative example to demonstrate this.

4 Conclusion

The jackknife modified Logistic Ridge estimator is superior to both the Logistic Ridge and the modified Logistic Ridge estimators in terms of variance and bias reduction.

References

[1] A.E. Hoerl and R.W. Kennard, Ridge regression: Biased estimation for non-orthogonal problems *Communications in Statistics-Theory and Methods*, **4**, (1970), 105-123.

[2] B. Singh, Y.P Chaubey and T.D Divivedi, An almost unbiased ridge estimator, *Sankhya* B, **48**, (1986), 342-360.

[3] E. Lesaffre, B.D. Max, Collinearity in Generalized Linear Regression, *Communications in Statistics-Theory and Methods*, **22**(7), (1993), 1933-1952.

[4] F.S.M. Batah, A new estimator by Generalized Modified Jackknife Regression Estimator, *Journal of Basarah Researches* (Sciences), **37**(A), (2011), 138-149.

[5] G. Khalaf and G. Shukur, Choosing ridge parameters for regression problems, *Communications in Statistics-Theory and Methods*, **34**, (2005), 1177-1182.

[6] Hinkley, Jackknifing in Unbalanced Situations, *Technometrics*, **19**(3), (1977), 285-292.

[7] M.E. Nja, Jackknifing the Logistic Ridge regression estimator, (2013), (to appear).

[8] M.H. Quenonille, Notes on bias in estimation, *Biometrika*, **43**, (1956), 353-360.

[9] M. Khurana, Y.P Chaubey and S Chandra, Jackknifing the Ridge Regression Estimator, A revisit: *Technical report*, **12**(1), (2012), 1-19.

[10] R.L Schaefer, L.D Roi and R.A Wolfe, A ridge Logistic estimator. *Communications in Statistics-Theory and Methods*, **13**, (1984), 99-113.