Journal of Statistical and Econometric Methods, vol. 2, no.3, 2013, 7-15
ISSN: 2051-5057 (print version), 2051-5065(online)
Scienpress Ltd, 2013

A New Estimation Procedure for Generalized Linear Regression Designs with Near Dependencies

Mbe Egom Nja[1]

Abstract

As a further improvement on Ridge regression estimation in Generalized Linear Models where near dependencies exists among explanatory variables, a new estimation procedure is here proposed. The new procedure perturbs the weighted matrix directly to enlarge the eigenvalues of the information matrix, thereby yielding smaller variances of parameter estimates. The method combines the idea of Iterative Weighted Least Squares and the Ridge Regression methods. The new method proves to be superior to the existing Ridge methodby further reducing the variances of parameter estimates and the residual variance.

Keywords: Inexact collinearity, information matrix, residual variance, singular value decomposition, condition number, condition index.

1 Introduction

The Ridge regression estimation method is the traditional and well acclaimed estimation procedure in both General and Generalized Linear Models where collinearity or near dependence (near collinearity) is an issue. This is because collinearity among the explanatory variables inflates the variances of parameter estimates and the well known estimation methods which include the Newton-Raphson, Fisher's scoring and the Iterative Weighted Least Squares do not have the capacity to deal with it. To solve the problem of inflated variances, the Ridge estimator was introduced by Hoerl and Kennard [3]. The proposed method further reduces variances of parameter estimates thus making it superior to the Ridge regression technique. For the Ridge regression model, the variance-covariance matrix is obtained as $V(\hat{\beta}) = \sigma^2(X'WX + KI)^{-1}$ and for the proposed method it is $V(\hat{\beta}) = \sigma^2[X'(W + KI)X]^{-1}$. The Ridge estimation method for obtaining the parameter estimates $\hat{\beta}$ is given as:

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

[1] Department of Mathematics, Federal University Lafia, Nigeria, Tel: +2347036507635

Article Info: *Received :* April 8, 2013. *Revised :* May 2, 2013.
Published online : September 9, 2013

where$\sigma^2 = \frac{SS_R}{n-k}$ for sample size n and k number of samples. $SS_R$is the regression sum of squares. This derives from the Iterative Weighted Least Squares update given as:

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

In both General Linear and Generalized Linear Models, collinearity exists among two or more independent variables when such variables are highly correlated (Mason, [5]). The effect of this is to produce regression estimates with inflated variances. Collinearity is said to exist among columns of $X = (x_1, x_2, \dots, x_p)$ if for a suitably predetermined $e_n > 0$ there exist constants$c_1, c_2, \dots, c_p$, not all zero, such that $c_1x_1 + c_2x_2 + \cdots + c_px_p = S$ with $||S|| < e_n||c||$ (Gunst [2]).

If the goal is simply to predict Y from a set of X variables, then collinearity is not a problem. The predictions will still be accurate, and the overall R^2(or adjusted R^2) quantifies how well the model predicts the Y values (Motulskey [8]). If the goal is to understand how the various X variables impact on Y, then collinearity is a big problem. One problem is that the individual P values can be misleading (a P value can be high, even though the variable is important). The second problem is that the confidence intervals on the regression coefficients will be very wide. This renders both the regression coefficients and their standard errors unstable.

Collinearity can be exact (perfect) or inexact (near dependency). When there is an exact (perfect) collinearity in the nxp data matrix:

$$X = (x_1, x_2, \dots, x_p)$$

we find a set of values
$C = (c_1, c_2, \dots, c_p)$, not all zero such that the linear combination:

$$c_1x_1 + c_2x_2 + \cdots + c_px_p = 0$$

otherwise, we have various degrees of collinearity in descending order.

When there is no exact collinearity but some near dependencies in the design matrix, one can find one or more non-zero vector v such that:

$$Xv = a$$

with $a \neq 0$ but small (close to 0). That is a near dependency exists if the length of vector a, $||a||$ is small.

$||a||$is the positive square root of the smallest eigenvalue of $X'X$. Near dependencies exist when the condition indices of X and $X'X$ are high.

Collinearity and its various degrees can be detected using variance inflation factor, dependency of explanatory variables, condition number and by performing the singular value decomposition on the information matrix.

The condition number is the square root of the highest condition index. The condition index η_k is defined as $\eta_k = \lambda_{max}/\lambda_k, k = 1, \dots, p$ where λ_k is one of the singular values of X. The condition number is defined as $k = (\lambda_{max}/\lambda_{min})^{1/2}$. When there is no collinearity at all, the eigenvalues, condition index and condition number willall be equal to zero. As collinearity increases, the condition number increases. The condition index can be used to signal the existence of “near” dependencies (near collinearity) in the data matrix X (Belsley et al [1]). Near dependence occurs when the correlation matrix C is near-singular (Sundberg [9]).

2 Singular Value Decomposition

To assess the extent to which “near dependencies degrade the estimated variance of each regression coefficient, Belsley et al [1] decomposed a coefficient variance into a sum of terms each of which is associated with a singular value. In Ordinary Least Square estimation, the model Variance-Covariance matrix of the estimator $\hat{\beta}$ is:

$$Var_M\left(\widehat{\beta}\right) = \sigma^2(X'X)^{-1}$$

The Singular value decomposition of X is U_1DU_2'

Thus $Var_M\left(\widehat{\beta}\right) = \sigma^2\left[\left(U_1DU_2'\right)'(U_1DU_2')\right]^{-1}$

$$= \sigma^2U_2D^{-2}U_2' \tag{1}$$

where $D = diag(\lambda_1, \lambda_2, \dots, \lambda_p)$ is a $(p \times p)$ non-negative diagonal matrix of singular values of X. U_1 is $(n \times p)$ and column orthogonal. $U_2\ is\ (p \times p)$ and both row and column orthogonal.

The kth diagonal element of $Var_M(\hat{\beta})$ is the estimated variance for the kth coefficient $\widehat{\beta}_k$.

The diagonal elements of $U_2D^{-2}U_2'$ are:

$$\sum_{j=1}^{p}\frac{u_{2kj}^2}{\lambda_j^2}\quad,\qquad j = 1,2,\dots.p$$

$$U_2 = (u_{2kj})_{(p\times p)}\quad,\quad Var_M\left(\widehat{\beta_k}\right) = \sigma^2\sum_{j=1}^{p}\frac{u_{2kj}^2}{\lambda_j^2}$$

where λ_j is jth singular value.

It is obvious that a small λ_j leads to a large component of $Var_M(\hat{\beta}_k)$. If both u_{2kj} and λ_j are small, $Var_M(\hat{\beta}_k)$ may not be affected.

In Generalized Linear Models, the interest is in the collinear relations among the columns in the matrix $\bar{X} = U_1DU_2'$

where $D = diag\left(\lambda_1, \lambda_2, \dots, \lambda_p\right)$ a diagonal matrix of singular matrix values of $W^{1/2}X$, U_1, $U_2$are as previously defined.

$$K(\bar{X}) = {}^{\lambda_{max}}/_{\lambda_{min}}$$

Where λ_{max} and λ_{min} are maximum and minimum singular values of $\overline{X}$. Based on the extrema of the ratio of quadratic forms Liao and Valliant [4] bounded the condition number in the range

$$\frac{w_{min}^{1/2}}{w_{max}^{1/2}} K(X) \leq K(\bar{X}) \leq \frac{W_{max}^{1/2}}{W_{min}^{1/2}}$$

where w_{max} = max weight , w_{min} = min weight

3 Estimation Methods for Generalized Linear Models in the Presence of Collinearity

3.1 The Ridge Procedure

This procedure is derived from the Iterative Weighted Least Squares which is a maximum likelihood method McCullagh and Nelder [7]. It is given as:

$$\beta = (X'WX + KI)^{-1}X'WZ \tag{2}$$

Where K is a biasing constant, I is an identity matrix, W is a diagonal matrix of weights with diagonal elements:

$$w_i = m_i\mu_i(1 - \mu_i) \tag{3}$$

Where$\mu_i = \frac{\exp(\sum X_{ij}\beta_j)}{1+\exp(\sum X_{ij}\beta_j)}$ is the response probability for the logistic regression model (a special case of the Generalized Linear Model).
Z is the working vector with components:

$$z_i = \eta_i + \frac{y_i - m_i\mu_i}{m_i}\frac{d\eta_i}{d\mu_i}$$

Where$\eta = \beta_o + \sum X_{ij}\beta_j$ is the linear predictor for the binary regression model. The differential operator $\frac{d\eta}{d\mu}$ is the derivative of the link function $h(\mu)$.

3.2 The Proposed Estimation Method

This method borrows from the Iterative Weighted Least Squares and the Ridge regression methods and is given as follows:

$$\beta = \left[X'(W + KI)X\right]^{-1}X'(W)Z \tag{4}$$

Where KI is the Tikhonov matrix for the biasing constant K. Starting with an initial guess values for K, we continue to iterate until the deviance is sufficiently small. Usually, the initial value of K is a very small number. The new method is an improvement of the

Ridge regression method in terms of variance reduction and within the framework of near dependency.

4 Comparison of the Ordinary Ridge Estimator and the Proposed Estimator

The Generalized Linear Model can be written as:

$$Z = X\beta + eh'(\mu) \qquad (5)$$

where$\widehat{\beta}$ =$(X'WX)^{-1}X'WZ$ (6)

is an Iterative Weighted Least Squares estimator. In canonical form (5) can be written as:

$$Z = C\alpha + eh'(\mu) \qquad (7)$$

where$C = XT,T$ is a $(p \times p)$ orthogonal matrix consisting of eigenvectors of $(X'WX + KI)$

J=diag $(\lambda_1,\ \lambda_2, \dots, \lambda_p)$

withλ_i as the ith eigenvalue of $\left(X'W + kI\right)X$

$\hat{\alpha}_{IWLS} = T'\beta$

C' C=T' X' XT=A=diag$(\lambda_1, \lambda_2, \dots, \lambda_p)$

$\hat{\beta}_{IWLS} = T'\alpha_{IWLS}$

$\hat{\alpha}_{ORR} = (I - kA^{-1})\hat{\alpha}_{IWLS}$, $\quad \hat{\beta} = T'\hat{\alpha}_{ORR}$

Where $k = k_1,\ k_2, \dots\ k_p$ is a fixed biasing constant.

To demonstrate that the proposed procedure (4) is superior to the existing Ridge method (2) by variance reduction, the following theorem is developed and proven

THEOREM: Let K be a $(p \times p)$ symmetric positive definite matrix. Then the proposed logistic regression (PLR) estimator has smaller variance than the Ordinary Logistic Ridge estimator.

Proof:

Let $V(\hat{\alpha}_{ORR})$ be variance of the Ordinary Ridge estimator and $V(\hat{\alpha}_{PLR})$ be the variance of the proposed Logistic estimator. It is enough to show that

$V(\hat{\alpha}_{ORR}) - V(\hat{\alpha}_{PLR}) > 0$.

$$V(\hat{\alpha}_{ORR}) - V(\hat{\alpha}_{PLR}) = \sigma^2 W_{ORR} A^{-1} W'_{ORR} - \sigma^2 W_{PLR} A^{-1} W'_{PLR}$$

$$= \sigma^2 \left[(I - kA^{-1}k_{ORR})A^{-1}k_{OR}(I - kA^{-1}k_{ORR})' \right] - \sigma^2 \left[(I - kA^{-1}k_{PLR})A_{PLR}{}^{-1}(I - kA^{-1}k_{PLR})' \right]$$

$$= \sigma^2 M$$

where:

$$M = \left[(I - kA^{-1}k_{ORR})A^{-1}k_{OR}(I - kA^{-1}k_{ORR})' \right] - \sigma^2 \left[(I - kA^{-1}k_{PLR})A_{PLR}{}^{-1}(I - kA^{-1}k_{PLR})' \right]$$

$$(I - rA^{-1}k_{(ORR)})^2 = diag\left[\frac{\lambda^2_{1(ORR)}}{(\lambda_{1(ORR)} + k)^2}, \frac{\lambda^2_{2(ORR)}}{(\lambda_{2(ORR)} + k)^2}, \dots, \frac{\lambda^2_{p(ORR)}}{(\lambda_{p(ORR)} + k)^2} \right]$$

$$(A^{-1}k_{(ORR)}) = diag\left[\frac{1}{(\lambda_{1(ORR)} + k)}, \frac{1}{(\lambda_{2(ORR)} + k)}, \dots, \frac{1}{(\lambda_{p(ORR)} + k)} \right]$$

$$(I - kA^{-1}k_{(ORR)})^2 A^{-1}k_{(ORR)} = \left[\frac{\lambda^2_{1(ORR)}}{(\lambda_{1(ORR)} + k)^3}, \frac{\lambda^2_{2(ORR)}}{(\lambda_{2(ORR)} + k)^3}, \dots, \frac{\lambda^2_{p(ORR)}}{(\lambda_{p(ORR)} + k)^3} \right]$$

and

$$(I - kA^{-1}k_{(PLR)})^2 A^{-1}k_{(PLR)} = \left[\frac{\lambda^2_{1(PLR)}}{(\lambda_{1(PLR)} + k)^3}, \frac{\lambda^2_{2(PLR)}}{(\lambda_{2(PLR)} + k)^3}, \dots, \frac{\lambda^2_{p(PLR)}}{(\lambda_{P(PLR)} + k)^3} \right]$$

$$\therefore \quad M = \left[\frac{\lambda^2_{1(ORR)}}{(\lambda_{1(ORR)} + k)^3} - \frac{\lambda^2_{1(PLR)}}{(\lambda_{1(PLR)} + k)^3}, \dots, \frac{\lambda^2_{p(ORR)}}{(\lambda_{p(ORR)} + k)^3} - \frac{\lambda^2_{p(PLR)}}{(\lambda_{p(PLR)} + k)^3} \right]$$

$\lambda_{iORR)}$ is the ith eigenvalue of $(X'WX + KI)$

and $\lambda_{i(PLR)}$ is the ith eigenvalue of $X'(W + KI)X$

It is enough to show that:

$\lambda_{i(PLR)} > \lambda_{i(OLR)}$

The eigenvalues of a (2×2) weighted matrix $X'WX$ are given as

$$\lambda = \frac{1}{2}[[(w_1x_{11}^2 + w_2x_{21}^2) + (w_1x_{12}^2 + w_2x_{22}^2)^2] \pm \{[(w_1x_{11}^2 + w_2x_{21}^2) + (w_1x_{12}^2 + w_2x_{22}^2)^2$$
$$- 4[(w_1x_{11}^2 + w_2x_{21}^2)(w_1x_{12}^2 + w_2x_{22}^2) - (w_1x_{11}x_{12} + w_2x_{21}x_{22})(w_1x_{12}x_{11} + w_2x_{22}x_{21})]]\}^{1/2}]$$

These eigenvalues can be increased by increasing the diagonal elements of $X'WX$, *i.e* by increasing $(w_1x_{11}^2 + w_2x_{21}^2)$ *and* $(w_1x_{12}^2 + w_2x_{22}^2)$

Since $x_{11}, x_{12}, x_{21}, x_{22} \geq 0$, increasing $(w_1x_{11}^2 + w_2x_{21}^2)$ *and* $(w_1x_{12}^2 + w_2x_{22}^2)$ implies increasing w.

This can be generalized to any $(p \times p)$ weighted matrix. But $w^{\sqrt{1+\delta}} > w \quad for\ 0 \leq \delta \leq 1$.

Thus $\lambda_{i\sqrt{1+\delta}} > \lambda_i$. *i.e* $\lambda_{i(PLR)} > \lambda_{i(OLR)}$.

Hence $\dfrac{\lambda^2_{i(PLR)}}{(\lambda_{i(PLR)} + k)^3} < \dfrac{\lambda^2_{i(ORR)}}{(\lambda_{i(ORR)} + k)^3}$

$\therefore$ M is positive definite.

and $V(\hat{\alpha}_{ORR}) - V(\hat{\alpha}_{PLR}) > 0$

5 Illustrative Examples

The explanatory variables in each of the six examples were illustrated by Mbachu, Nduka&Nja [6]. The variables are sex (X_1), race (X_2) and percentage fat intake (X_3). In each case, the response variable is the number of obesed persons. The condition indicies of the design matrices in all the six examples are high, indicating the existence of near dependencies.

The design matrices for the six examples are given as follows:

$$X_1 = [1, 0, 0, 40; 1, 0, 1, 30; 1, 1, 0, 20; 1, 1, 1, 30]$$
$$X_2 = [1, 0, 0, 30; 1, 0, 1, 18; 1, 1, 0, 12; 1, 1, 1, 18]$$
$$X_3 = [1, 0, 1, 50; 1, 0, 1, 45; 1, 1, 0, 30; 1, 1, 1, 40]$$
$$X_4 = [1, 0, 0, 22; 1, 0, 1, 30; 1, 1, 0, 13; 1, 1, 1, 26]$$
$$X_5 = [1, 0, 0, 26; 1, 0, 1, 21; 1, 1, 0, 15; 1, 1, 1, 21]$$
$$X_6 = [1, 0, 0, 34; 1, 0, 1, 23; 1, 1, 0, 17; 1, 1, 1, 23]$$

Using the MATLAB Software the following parameter estimates are obtained for both the existing Ridge estimation method and for the proposed method:

Table 1: parameter estimates for Ridge and proposed methods

		β_0	β_1(sex)	β_2(race)	β_3(fat)	σ^2_{res}
Example 1	Ridge	5.1581	-1.6894	0.4441	-0.1508	0.000521
	Proposed	5.5351	-1.8160	0.4203	-0.1609	0.00019
Example 2	Ridge	4.1875	-1.6990	-0.0726	-0.1685	0.0164
	Proposed	4.3966	-1.7873	-0.1158	-0.1758	0.0002296
Example 3	Ridge	8.1530	2.4085	0.9260	-0.1793	0.0015
	Proposed	10.0544	-2.9562	0.9863	-0.2198	0.00012
Example 4	Ridge	8.7163	-2.9417	4.9335	-0.4244	0.00724
	Proposed	13.1068	-4.3502	7.0757	-0.6351	0.0002
Example 5	Ridge	6.2588	-1.7207	0.6342	-0.2736	0.00597
	Proposed	6.6087	-1.7931	0.5519	-0.2877	0.000229
Example 6	Ridge	3.8662	-1.6991	-0.4758	-0.1003	0.0565
	Proposed	4.2528	-1.7974	-0.4871	-0.1150	0.0243

Table 2: Variances of parameter estimates for Ridge and proposed methods

		$Var(\beta_0)$	$Var(\beta_1)$	$Var(\beta_2)$	$Var(\beta_3)$
Example 1	Ridge	3.46×10^{-3}	5.6×10^{-4}	2.9×10^{-4}	2.7×10^{-6}
	Proposed	1.4×10^{-3}	2.15×10^{-4}	1.05×10^{-4}	1.06×10^{-6}
Example 2	Ridge	7.39×10^{-2}	1.76×10^{-2}	1.02×10^{-2}	1.06×10^{-4}
	Proposed	1.09×10^{-3}	2.57×10^{-4}	1.45×10^{-4}	1.6×10^{-6}
Example 3	Ridge	2.66×10^{-2}	2.8×10^{-3}	9×10^{-4}	1×10^{-5}
	Proposed	2.38×10^{-3}	2.57×10^{-4}	5.35×10^{-5}	1.02×10^{-6}
Example 4	Ridge	1.9×10^{-1}	2.3×10^{-2}	5.29×10^{-2}	4.42×10^{-4}
	Proposed	7.46×10^{-3}	9.4×10^{-4}	2.16×10^{-3}	1.9×10^{-5}
Example 5	Ridge	2.3×10^{-1}	2.7×10^{-2}	1.4×10^{-2}	3.9×10^{-4}
	Proposed	2.3×10^{-3}	2.5×10^{-4}	1.3×10^{-4}	4.2×10^{-6}
Example 6	Ridge	3.7×10^{-1}	6.04×10^{-2}	3.45×10^{-2}	4.06×10^{-4}
	Proposed	1.72×10^{-1}	2.7×10^{-3}	1.5×10^{-2}	1.8×10^{-4}

6 Discussion

A new estimation update for Generalized Linear Models where near dependencies is an issues has been developed in this study. The design is a midwife between the Iterative weighted Least Squares and the Ridge regression methods.

Parameter estimates which depict the effect of the factors under investigation on the response are obtained in table 1. Also shown in table 1 are the variances of the effects. The variances show clear superiority of the new method over the existing Ridge method in terms of variance reduction. There is gross reduction in the variances of the parameter estimates for all the six illustrative examples. For instance, the variances for example 1 are 5.2*10^-4 for the Ridge method and 1.9*10^-4 for the proposed method and for example 2 the variances are 1.64*10^-2 for the Ridge method and 2.29*10^-4 for the proposed method.

A theorem to demonstrate that the proposed procedure is superior to the existing Ridge method in terms of variance reduction is developed and proven.

7 Conclusion

When there is near dependency among explanatory variables in Generalized Linear Models, this proposed method of estimation should be adopted. This is because of the drastic reduction in the variances of parameter estimates as shown by the illustrative examples and the developed theorem. Additionally the new method is superior to the existing Ridge method in term of residual variance.

References

[1] Belsley, D.A; Edwin, K; Welsch, R.E. (19980): Regression Diagnostics Identifying influential data and sources of collinearity. Wiley, New York.

[2] Gunst, R.F. (1984): Toward a Balanced Assessment of Collinearity Diagnostics. American Statistician. **38**, 79-82.

[3] Hoerl, A.E., Kennard, R.W. (1970): Ridge regression biased estimation for non-orthogonal problems.

[4] Liao, D., Valliant, R. (2011): Collinearity Diagnostics for Complex Survey Data. A Ph.D thesis from the University of Maryland, College Park.

[5] Mason, G. (1987): Coping with collinearity. The Canadian Journal of Program Evaluation.

[6] Mbachu, H.I., Nduka, E.C., Nja, M.E. (2012): Residual Analysis in the presence of Collinearity in Generalized Linear Models. An unpublished Ph.D thesis from the University of Port Harcourt Nigeria.

[7] McCullagh, P., Nelder, J.A. (1992): Generalized Lineaar Models. Chapman and Hall. Madras.

[8] Motulsky, H. (2002): Multicollinearity in multiple Regressions. Graphpad software (hmotulsky@graphpad.com

[9] Sundberg, R. (2002): Collinearity. Encyclopedia of Environmetrics, John Wiley & Sons Ltd Chichester, **1**, 365-366