Data Analyzing by Attention to Weighted Multicollinearity in Logistic Regression Applicable in Industrial Data

Marzieh Shahmandi*[1], Mohammad Mehdi Gharahbeigi[2], Tooraj Ghaffary[2] and Leila Shahmandi[2]

[1]Department of Mathematics, Isfahan University of Technology, Isfahan, Iran
[2]Young Researchers club, Shiraz Branch, Islamic Azad University, Shiraz, Iran

ABSTRACT

The Middle East́s largest industrial complex produces flat steel sheets with specific properties such as low thickness, high strength and suitable formability in order to reduce the vehicle weight and fuel consumption and prevention of environmental pollution. The aim of this study is to investigate the effect of some important explanatory variables on suitable formability of manufacturing steel sheets according to primary data set. Existence or lack of existence of crack on steel sheet is considered as a binary response variable. It is determined by bending test with the angle of zero degree. Existence of multicollinearity between mentioned explanatory variables has an effect on the probability of crack existence. Because of special condition of the response variable, which is binary, the suitable regression is logistic, and correction techniques based on least squares do not work. Developments in weighted multicollinearity diagnostics are used to assess maximum likelihood logistic regression parameter estimates. Then principal component, a biased estimation method, is used in a way that it has additional scaling parameter which can accommodate a spectrum of explanatory variable standardizations. After that, by this scale parameter α, other biased estimation methods such as partial least squares, ridge and Stein are explained. They can considerably reduce the variance of the parameter estimation.

Keywords: logistic regression; partial least squares; principal component; quasistandardization; ridge; Stein; weighted multicollinearity
2010 Mathematics Subject Classification: 62H35; 62J86; 62J07; 62p30; 97K80; 46N30

* Tel.: 00982602966.
E-mail address: marzieh.shahmandi@gmail.com.

1. INTRODUCTION

Extending previous research on multicollinearity and regression, this study analyzes a primary data set according to special condition of response variable which is binary. Least squares method is not proper Rather, it is proposed that logistic regression is appropriate model for this research. This data set is related to the Middle East́s largest producer of flat steel. Before fitting model, data are checked because of multicollinearity by mentioned indicators in [Marx and Smith (1990a,b)]. Multicollinearity makes model unstable, and the estimated parameters will be inaccurate. Thus the interpretation of the relation between the response and each explanatory variable in terms of odds ratios may be erroneous. It is also proposed some unbiased methods to solve this problem and to estimate the parameters of this model, i.e. principal component (PC), partial least squares (PLS), ridge and Stein.

Principal component analysis (PCA) was explained by [Hotelling (1933)]. [Gower (1966)] evaluated the relation between PCA and some statistical techniques. [Hawkins (1973)] recognized an error in multivariate data by PCA. [Fomby et.al (1978)] used the properties of PCA in least squares constrains. [De Leeuw (1986)] explained nonlinear PCA. [Marx (1992)], introduced a spectral of scale explanatory variables that is defined by scale parameter α and is named quasistandardization. Scaling parameter values between zero and one lead to an interpolation between correlation and covariance matrices. [Marx (1992)] used PC and quasistandardization methods for a mine data set. [Boente et.al (2010)] focused on detecting influential observations in PC method and its structure. [Boik (2013)] applied PC method by paying attention to constraints on correlation matrix.

Ridge regression was explained by [Hoerl and Kennard (1970)], and [Schaefer et.al (1984)] used ridge estimator in logistic regression. [Kibria and Saleh (2012)] and [Roozbeh and Arashi (2013)] applied it in a probit regression model and partially linear model.

Stein estimator was introduced by [Stein (1960)] and [Schaefer (1986)] used it in multiple logistic regression. Also [Marx and Smith (1990b)] used ridge and Stein methods for a data set from lake acidification. [Fisher and Sun (2011)] explained improved Stein-type shrinkage estimators in multicollinearity condition.

[Wold (1984)] introduced PLS. [Escofier and Page's (1988)] evaluated the relation between PLS regression and multiple factor analysis and [Pages and Tenenhaus (2001)] continued it. [Bastien et.al (2005)] applied PLS method for a data set of Bordeaux wines because of multicollinearity. [Bjorkstrom (2010)] used Krylov sequences to compare PC and PLS methods in some aspects. [Fujiwara et.al (2012)] introduced a new methodology to select variables for PLS method based on the nearest correlation spectral clustering. [Zerzucha et.al (2012)] discussed dissimilarity PLS applied to nonlinear modeling.

The study, therefore, uses quasistandardization method and PC logistic regression models after identifying multicollinearity. Then, considering assessment indicators such as deviance and sum of coefficients variance, the best α and the best model are selected. Then by using this α, other methods such as PLS, ridge and Stein are applied to estimate model parameters. Finally, according to this data set, the best method is identified to estimate the model parameters.

This article consists of 3 sections. Section 1 is an introduction and gives a brief overview of logistic regression, introduces weighted multicollinearity diagnostics, and defines quasistandardization of explanatory variables. Section 2 explains logistic regression biased estimation methods such as PC, PLS, ridge and Stein methods. Section 3 compares these methods with a primary data set of the largest industrial company in the Middle East.

1.1 Logistic Regression

There are many fields of study such as medicine and epidemiology, in which it is very important to predict a binary response variable, or equivalently the probability of occurrence of an event (success), in term of the values of a set of explanatory variables related to it.

Let $X_1, X_2, ..., X_p$ be a set of continuous variables observed without error and let us consider n times of observation of such variables that will be resumed in the matrix $\aleph = (x_{ij})_{n\times p}$. Let $Y = (y_1, y_2, ..., y_n)'$ be a random sample of a binary response variable Y associated with the observation in□ $\aleph$, that is, $y_i \in \{0,1\}, i = 1,...,n$. then, the logistic regression model is given by

$$y_i = \pi_i + \varepsilon_i \qquad i = 1,2,...n \tag{1.1}$$

,where π is the expectation of Y given $(X_1 = x_{i1}, X_2 = x_{i2}, ..., X_p = x_{ip})$ that is modelized as

$$\pi_i = P\{Y = 1 | X_1 = x_{i1}, X_2 = x_{i2}, ..., X_p = x_{ip}\} = \frac{e^{\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j}}{1 + e^{\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j}} \tag{1.2}$$

where $\beta_0, \beta_1, ..., \beta_p$ are the parameters of the model and ε_i are zero mean independent errors whose variances are given by $Var[\varepsilon_i] = \pi_i(1 - \pi_i), \; i = 1,2,...n$.

Once the model has been estimated, its goodness of fit must be tested. The most usual method to solve the test

$$\left\{\begin{array}{ll} H_0 : l_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j & i = 1,2,...,n \\ H_1 : l_i \neq \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j & some\ i \end{array}\right\}$$

is based on the Wilks statistic (Deviance) defined as $-2\ln\Lambda$, with Λ that is the usual likelihood-ratio statistic. The deviance is given by

$$G^2(M) = 2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{\hat{\pi}_i}\right) + (1 - y_i)\ln\left(\frac{1 - y_i}{1 - \hat{\pi}_i}\right)\right] n \xrightarrow{H_0} \infty \chi^2_{n-p-1} \tag{1.3}$$

This statistic has approximately a chi-squared distribution.

The diagonal matrix V contains variance of the estimated. □Y values. The matrix $\Phi = X'VX$ is named the information matrix. Denoted $\hat{\Phi} = X'\hat{V}X$ as estimated information matrix, in other words $\hat{\Phi} = \hat{S}'\hat{S}$ that $\hat{S} = V^{\frac{1}{2}}X$. Then we have $\hat{V}ar(\hat{\beta}) = \hat{\Phi}^{-1}$.

1.2 Weighted Multicollinearity Diagnostics for Logistic Regression

The logistic model becomes unstable when strong dependence exists among explanatory variables, so it seems that no variable is important when all others are in the model(multicollinearity). To develop suitable diagnostics for multicollinearity and have a standard of comparison, scaling of the information matrix is preferred. These diagnostics were mentioned in [Marx and Smith (1990a,b)].

Weighted Condition Number

Consider $\lambda_0^*, ..., \lambda_p^*$ as the ordered eigenvalues of $\hat{\Phi}^* = \hat{S}^{*'}\hat{S}^*$,so that

$$\hat{S}_{ij}^* = \frac{\hat{S}_{ij} - \bar{S}_j}{\sqrt{\sum_{i=1}^{n}\left(\hat{S}_{ij} - \bar{S}_j\right)^2}} \tag{1.4}$$

and $\bar{S}_j = \frac{\sum_{i=1}^{n}\hat{S}_{ij}}{n}$, condition numbers are defined as

$$k_j = \left(\frac{\lambda_{\max}^*}{\lambda_j^*}\right)^{\frac{1}{2}}, j = 1,2,...,p$$

Large values of k_j (≥ 30) indicate ill conditioning.

Weighted Variance Proportion

Let m_{ju} as the juth member of eigenvectors matrix of $\hat{\Phi}$. The weighted proportion of variance for the jth estimated coefficient can be expressed as

$$\omega_{uj} = \frac{m_{ju}^2 / \lambda_u^*}{C_{jj}}$$

That $C_{jj} = \sum_{u=0}^{p} \lambda_u^{*-1} m_{ju}^2$. A small eigenvalue (relative to the maximum eigenvalue) responsible for at least two large proportions suggests that weighted multicolinearity is damaging desirable properties of the logistic regression. For example, if ω_{32} and ω_{34} are large (near one), it will be related to multicollinearity where $\hat{\beta}_4$ and $\hat{\beta}_2$ variances will inflated.

1.3 Quasistandardization of Explanatory Variables

[Marx (1992)] introduced a class of PC estimators for generalized linear regression defined by scaling parameter. The additional parameter allows a spectrum of standardized explanatory variables which can result in interpolation between correlation and covariance matrices. Choice of the scaling parameters depends on the researcher's objectives for the model. Consider $\aleph = \left(x_{ij}\right)_{n \times p}$ as a matrix of continuous explanatory variables, then define:

$$x_{\alpha ij} = q_j^{-\alpha} (n-1)^{\frac{-1}{2}} \left(x_{ij} - \bar{x}_j\right)$$

$$q_j^2 = (n-1)^{-1} \sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j\right)^2$$

Denoted $\aleph_\alpha = \left(x_{\alpha ij}\right)_{n \times p}$, $X_\alpha = [1|\aleph_\alpha]$. The parameter α allows a spectrum of scaling. He indicated that in practice it may seem unnatural to use parameter values outside the unit interval.

2. Biased Logistic Regression Estimators

Using Taylor series arguments, it can be shown that the maximum likelihood (ML) parameter estimates are asymptotically unbiased. In making certain adjustments to ML, asymptotically biased parameter estimates can be constructed. PC, PLS, ridge and Stein, asymptotically biased estimators, are presented in this article.

2.1 A Continuum of Principal Component Estimators

Sample principal components (PCS) are orthogonal linear spans with maximum variance of the X_α matrix columns, denoted by $Z_{\alpha j} = X_\alpha m_{\alpha j}$, where $m_{\alpha 1}, m_{\alpha 2}, \ldots, m_{\alpha p}$ are the eigenvectors of the sample information matrix $\hat{\Phi}_\alpha = X_\alpha{}' \hat{V}_\alpha X_\alpha$, which are associated with corresponding eigenvalues $\lambda_{\alpha 1} \geq \lambda_{\alpha 2} \geq \ldots \geq \lambda_{\alpha p}$ of the $\hat{\Phi}_\alpha$.

The logistic regression can be expressed in terms of PCS.

$$L_\alpha = X_\alpha \beta_\alpha = Z_\alpha M'_\alpha \beta_\alpha = Z_\alpha \gamma_\alpha$$

As a result of the invariance property of ML estimates we have:

$$\hat{\beta}_\alpha^{*pc} = M_\alpha \hat{\gamma}_\alpha^{*pc}$$

then, the prediction equation will be $Y=\hat{\prod}_{\alpha}^{*pc}$ where $\hat{\prod}_{\alpha}^{*pc}=\left(\hat{\pi}_{\alpha 1}^{*pc},\hat{\pi}_{\alpha 2}^{*pc},...,\hat{\pi}_{\alpha n}^{*pc}\right)$. This model in terms of a specific subset (s) of principal components is,

$$L_{\alpha(s)}=Z_{\alpha(s)}\gamma_{\alpha(s)}=X_{\alpha}M_{\alpha(s)}\gamma_{\alpha(s)}=X_{\alpha}\beta_{\alpha(s)}$$

Where we have

$$\hat{\beta}_{\alpha(s)}^{*pc}=M_{\alpha(s)}\hat{\gamma}_{\alpha(s)}^{*pc}.$$

For different values of α we can have different PC estimators.

This method is able to handle multicollinearity among the explanatory variables. Also it can make for stronger predictions. On the other hand, there is a difficulty to interpret the coefficients of the new PC components. And This method is sensitive to the scales of explanatory variables, they need to be normalized before computing the PC components.

2.2 Partial Least Squares Logistic Regression Estimator

Partial Least Squares Regression

PLS regression is used to study the relationship between a numerical response variable and a set of k explanatory variables in situations in which multiple regression is unstable or not feasible at all (strong multicollinearity, small number of observation compared to the number of variables, missing data). We can encounter the same kind of problems also in logistic regression and, more generally when using a generalized linear model.

PLS regression defines PLS components given by linear spans of the explanatory variables and uses them as new explanatory variables of regression model.

PC regression and PLS regression differ in the methods used in production of new components. PC regression produces the PC given by the covariance structure between the explanatory variables, while PLS regression produces the PLS components given by covariance structure between the explanatory and response variables.

PLS method is able to model multiple response variables as well as multiple explanatory variables. And it can handle multicollinearity among the explanatory variables. Also it is robust in face of missing data and it can make for stronger predictions. On the other hand, it is difficult to interpret the coefficients of the new PLS components. And because the distributional properties of estimates are not known, the researcher can not assess significance except through bootstrap induction. Also there is no test model statistic [Pirouz (2006)].

PLS Generalized Linear Regression (PLS-GLR)

With this constraint that PLS components $t_h=\sum_{j=1}^{m}w_{hj}^{*}x_j$ are orthogonal, PLS generalized linear regression of Y on $x_1,x_2,...,x_p$ with m components is written as

$$g(\Theta)=\sum_{h=1}^{m}c_h\left(\sum_{j=1}^{m}w_{hj}^{*}x_j\right)$$

Where w_{hj}^{*} are achieved by the covariance structure between Y and x_j. The parameter Θ may be either the mean of a continuous Y, or the probability vector of the values taken by a discrete variable Y. The link function g is chosen by the user according to the probability distribution of Y and the model goodness of fit to the data.

PLS-GLR Algorithm

The algorithm consist of four steps:

1- Computation of the m PLS components $t_h \left(h = 1,2,...,m\right)$.

2- Generalized linear regression of Y on the m retained PLS components.

3- expression of PLS-GLR in terms of the original explanatory variables

4- Bootstrap validation of coefficients in the final model of PLS-GLR

All these steps were expressed in [Bastien et.al (2005)].

2.3 Ridge Logistic Regression Estimator

[Schaefer (1986)] suggested:

$$\hat{\beta}_{Ridge}(k) = \left(X'_\alpha \hat{V}_\alpha X_\alpha + kI\right)^{-1} X'_\alpha \hat{V}_\alpha X_\alpha \hat{\beta}_\alpha \qquad (2.3)$$

The choice of the k is subjective, however [Schaefer (1986)] recommended a harmonic mean method, $k = \dfrac{p+1}{\hat{\beta}'\hat{\beta}}$.

Ridge method has fast and simple computations and interpretation of the coefficients is clear.

2.4 Stein Logistic Regression Estimator

[Schaefer (1986)] suggested an extension of the [Stein (1960)] estimator for logistic regression. Consider shrinking the ML estimate as follows:

$$\hat{\beta}_{Stein} = c\hat{\beta}_{ML} \qquad (2.4)$$

Where $0 \prec c \prec 1$. The purpose of Stein estimation is to shrink both the estimated parameter vector, and the associated standard errors, by a simple scaling technique. c is chosen, which minimizes the $E\left(L^2\right) = \left(c\hat{\beta} - \beta\right)' \left(c\hat{\beta} - \beta\right)$ criterion (with respect to c) it will be:

$$c = \frac{\hat{\beta}'\hat{\beta}}{\hat{\beta}'\hat{\beta} + trace\left(\hat{\Phi}_\alpha^{-1}\right)}$$

Properties of Stein method is similar to the ridge method. And also there is no main disadvantage for both of them.

3. Example

The objective of the study is to predict the suitable formability of steel sheets based on five explanatory variables according to a primary data set of the Middle East´s largest industrial complex.

x_1:yield strength (N/mm^2)

x_2: final tensile strength (N/mm^2)

x_3: silicon (percent)

x_4: aluminum

x_5:nitrogen gas (percent)

Formality is checked by bending test with the angle of zero degree, and if there will no cracks on the steel sheet it will be a success. The steel sheet data set includes 50 observations. At firs, the data set is evaluated about multicollinearity. The result is given in table (1).

Table 1: Variance proportion decomposition

E − value	k_j	intercept	x_1	x_2	x_3	x_4	x_5
5.2711	1.0000	0.00013	0.0006	0.00020	0.0048	0.0055	0.0037
0.3508	3.8765	0.00004	0.0014	0.00003	0.2099	0.2646	0.0272
0.2096	5.0154	0.00000	0.0018	0.00002	0.2902	0.0895	0.3153
0.1449	6.0313	0.00523	0.0078	0.00703	0.0746	0.3774	0.1869
0.0215	15.6736	0.00330	0.5303	0.08123	0.0783	0.1091	0.0792
0.0022	49.0917	0.99130	0.4581	0.91149	0.3422	0.1539	0.3877

The decomposition matrix has a last row of variables some are nearly one and also there is a large condition number (≥ 30) with the smallest eigenvalue of the information matrix. Both of them indicate multicollinearity. In table (2) and (3) the effects of multicollinearity can be seen on ML parameter estimators.

Table 2: Analysis of maximum likelihood estimates

parameter	df	estimate	standard	wald	pr > chisq
			error	chi − square	
intercept	1	23.1101	12.8375	3.2407	0.0718
X_1	1	−0.0256	0.0400	0.4095	0.5222
X_2	1	−0.0688	0.0326	4.4596	0.0347
X_3	1	−0.1397	0.1306	1.1451	0.2846
X_4	1	0.0430	0.0457	0.8836	0.3472
X_5	1	0.1050	0.0618	2.8825	0.0895

Table 3: Cross-table of observed and predicted response

observed	predicted		
response	0	1	sum
0	2	6	8
1	2	40	42
sum	4	46	50

According to p-value, only the explanatory variables x_2, x_5 are statistically significant at the risk level of 0.1. Also we have a high percentage of misclassified responses (16%), that we compute them by assigning to the most probable level of response. Then we compute PC, PLS, ridge, Stein estimators.

For every $\alpha \in [0,1.2]$, PC estimator with one, two,…PCS with maximum variances and also forward stepwise estimator are calculated, for example, in figure (1) and (2), it can be seen. Deviance value and sum of coefficient variances are displayed for $\alpha \in \{0.3,0.9,1,1.2\}$.

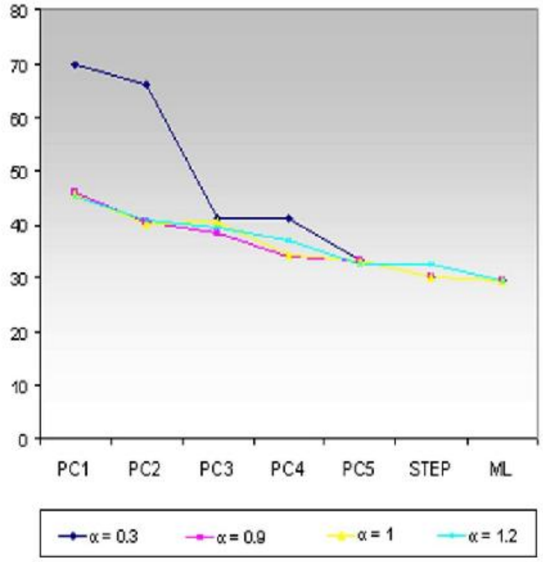

Figure 1: deviance of estimators in terms of α

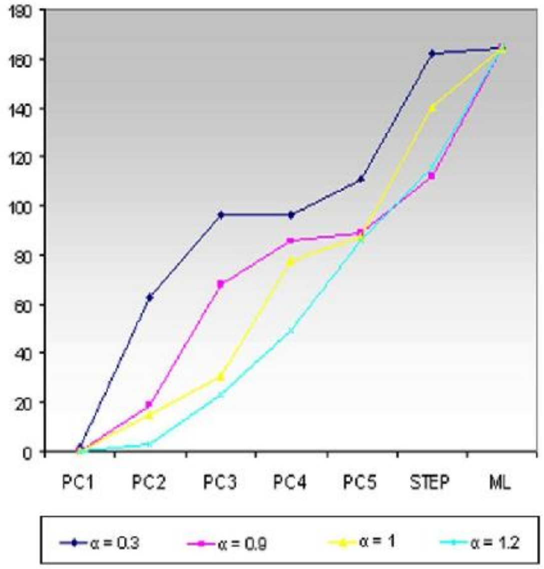

Figure 2: sum of coefficients variance for estimators in terms of α

By increasing every $\alpha \in [0,1.2]$, the reduction in the amount of sum of coefficients variance and deviance is evident. We can show two previous figures in other shapes (3) and (4).

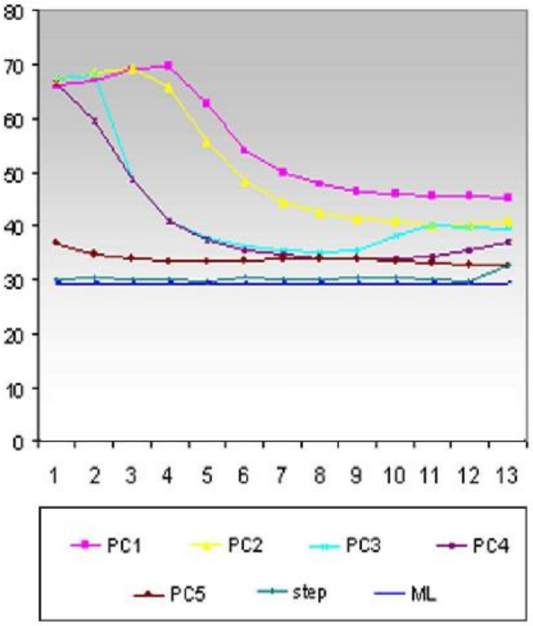

Figure 3: deviance of estimators

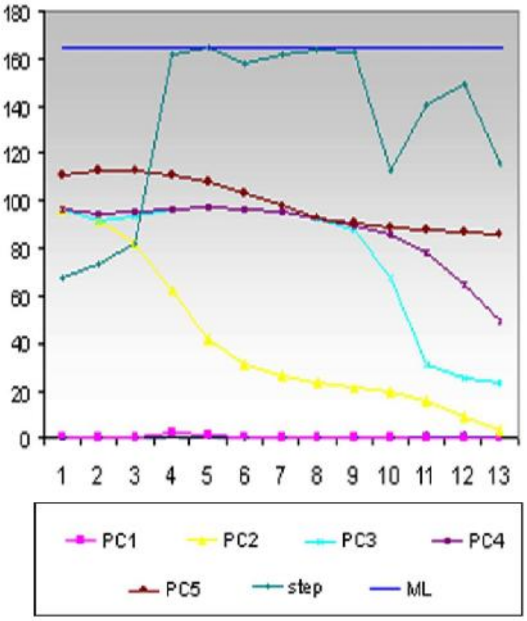

Figure 4: sum of coefficients variance for estimators

It can show PC estimators with 4 and 5 PCS have deviances near to deviance of ML estimators. The sum of coefficients variances of these two estimators is a bit more than other PC estimators, but these variances are less than deviances of ML estimator. Then we select PC estimators with 4 and 5 PCS with $\alpha = 1$ as candidates.

Also [Aucott et.al (1984)] considers that the best estimator for the parameter vector is before a sudden increase of this variance. Then by paying attention to information in table (4) we can select PC estimators with 3 PCS as a another candidate.

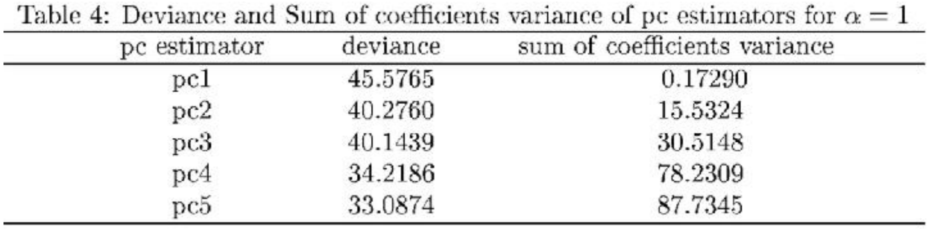

Table 4: Deviance and Sum of coefficients variance of pc estimators for $\alpha = 1$

pc estimator	deviance	sum of coefficients variance
pc1	45.5765	0.17290
pc2	40.2760	15.5324
pc3	40.1439	30.5148
pc4	34.2186	78.2309
pc5	33.0874	87.7345

Furthermore, we obtain PLS, ridge and Stein estimator with $\alpha = 1$.

PLS Logistic Estimator

At first we compute PLS components. We should fit regression of response variable on every explanatory variable. Due to the results of these regressions, in table (5), all explanatory variables are significant at the risk level of 0.1.

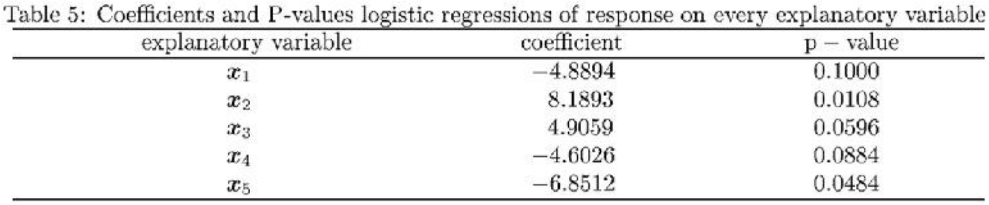

Table 5: Coefficients and P-values logistic regressions of response on every explanatory variable

explanatory variable	coefficient	p − value
x_1	−4.8894	0.1000
x_2	8.1893	0.0108
x_3	4.9059	0.0596
x_4	−4.6026	0.0884
x_5	−6.8512	0.0484

Then we have component t_1:

$$t_1 = \frac{-4.8894x_1 + 8.1893x_2 + 4.9059x_3 - 4.6026x_4 - 6.8512x_5}{\sqrt{\left(4.8894^2 + 8.18932^2 + 4.9059^2 + 4.6026^2 + 6.8512^2\right)}}$$
$$= -0.3613x_1 + 0.6051x_2 + 0.3625x_3 - 0.3401x_4 - 0.5062x_5$$

For computing t_2, we should fit regression of response variable on t_1 and every explanatory variable. The results of these regressions are reported in table (6).

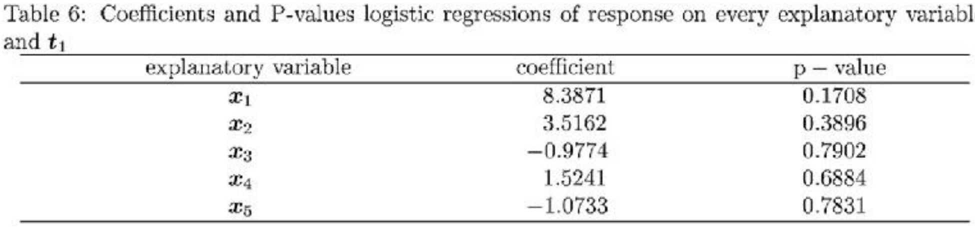

Table 6: Coefficients and P-values logistic regressions of response on every explanatory variable and t_1

explanatory variable	coefficient	p − value
x_1	8.3871	0.1708
x_2	3.5162	0.3896
x_3	−0.9774	0.7902
x_4	1.5241	0.6884
x_5	−1.0733	0.7831

Considering p– values, none of the explanatory variables is significant for t_2 structure. Then the model has one component. After fitting regression of response variable on t_1, we rewrite t_1 based on explanatory variables. The result is reported in table (7).

Table 7: Regression of response variable on $\boldsymbol{t}_1$ component

variable	df	parameter	standard	wald	p − value
		estimation	error	chi − squre	
intercept	1	2.1695	0.5477	15.6928	< 0.0001
$\boldsymbol{t}_1$	1	-8.3026	3.1436	6.9756	0.0083

Moreover we apply mentioned non-parametric validation with *B* = 1000 for coefficients of PLS logistic regression according to the steel sheets data set. It is displayed in figure (5).

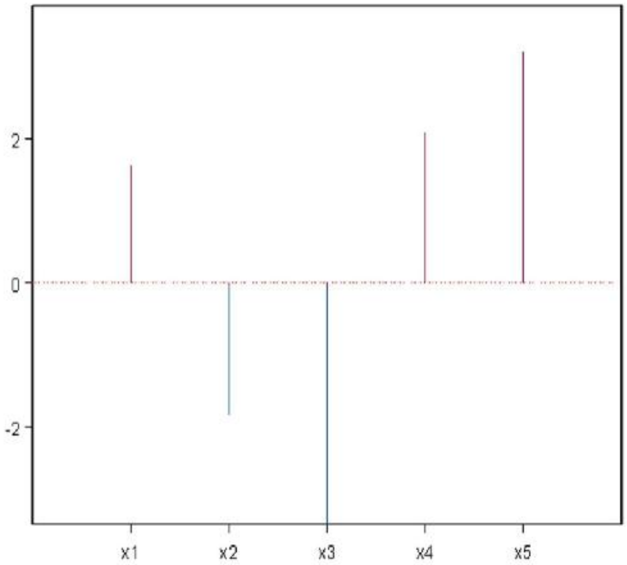

Figure 5: 95% Randomized/balanced bootstrap confidence intervals for parameters

Regarding confidence intervals and having no zero in these intervals, it can show that all explanatory variables are significant. Finally all estimators of parameters become non Quasistandardized.

In tables (8) and (9) the results such as estimated parameters and standard deviation of estimators based on all mentioned methods are reported, also in table (10), we have deviance and sum of coefficients variance for estimators.

Table 8: Estimated parameters by different methods

parameter vector	ML	pc3	pc4	pc5
$\hat{\beta}_0$	23.1101	-4.4603	6.2763	8.7784
$\hat{\beta}_1$	-0.0256	0.0173	0.0263	0.0332
$\hat{\beta}_2$	-0.0688	0.0045	-0.0379	-0.0459
$\hat{\beta}_3$	-0.1397	-0.0277	0.0286	-0.0194
$\hat{\beta}_4$	0.0430	0.0207	0.0336	0.0057
$\hat{\beta}_5$	0.1050	0.0197	0.0119	0.0284

	stepwise pc	ridge	stein	pls logistic
$\hat{\beta}_0$	21.3561	11.3581	15.0609	7.9961
$\hat{\beta}_1$	-0.0331	0.0011	-0.0167	0.0234
$\hat{\beta}_2$	-0.0616	-0.0397	-0.0448	-0.0415
$\hat{\beta}_3$	-0.1061	-0.0619	-0.0911	-0.0989
$\hat{\beta}_4$	0.0714	0.0225	0.0280	0.0318
$\hat{\beta}_5$	0.0822	0.0504	0.0684	0.0572

Table 9: Estimated standard deviation by different methods

standard deviation vector	ML	pc3	pc4	pc5
$SE(\hat{\beta}_0)$	12.8375	5.5233	8.8442	9.3658
$SE(\hat{\beta}_1)$	0.0400	0.0139	0.0151	0.0174
$SE(\hat{\beta}_2)$	0.0326	0.0050	0.0277	0.0294
$SE(\hat{\beta}_3)$	0.1306	0.0825	0.0901	0.1078
$SE(\hat{\beta}_4)$	0.0457	0.0179	0.0197	0.0396
$SE(\hat{\beta}_5)$	0.0618	0 0344	0.0348	0.0403

	stepwise pc	ridge	stein	pls logistic
$SE(\hat{\beta}_0)$	11.8460	7.5051	8.3663	10.6281
$SE(\hat{\beta}_1)$	0.0390	0.0193	0.0261	0.0366
$SE(\hat{\beta}_2)$	0.0307	0.0209	0.0212	0.0277
$SE(\hat{\beta}_3)$	0.0895	0.0811	0.0851	0.1215
$SE(\hat{\beta}_4)$	0.0300	0.0280	0.0298	0.0403
$SE(\hat{\beta}_5)$	0.0486	0.0340	0.0403	0.0532

Table 10: Deviance and sum of coefficients variance by different methods

	ML	pc3	pc4	pc5
deviance	29.2748	40.2739	34.2186	33.0874
sum of coefficients variance	164.8277	30.5148	78.2309	87.7345

	stepwise pc	ridge	stein	pls logistic
deviance	30.0249	31.1422	31.6766	31.6144
sum of coefficients variance	140.3416	56.3354	70.0054	112.9771

4. Conclusion

- In the section (4), table (9) shows standard deviation for β_0 of all estimators still are inflated, especially for forward stepwise and PLS logistic estimators, that are the same as it is for ML estimator.

- Table (8) shows that forward stepwise and Stein have the same sign. Also PLS logistic, PC with 5 PCS and ridge estimators have another same sign.

- Table (10) shows by order, forward stepwise, ridge, PLS logistic, Stein and PC estimators with 5, 4, 3 PCS have the less deviances after ML estimator. Also by order ML, forward stepwise, PLS logistic, PC estimators with 5, 4 PCS, Stein, ridge and PC estimator with 3 PCS have the maximum sum of coefficients variances.

Choice of which method is better depends on the purpose of the model. Good parameter estimates and good prediction are two different aspects of the model. With complex data, we do not expect a single model to be the best for all purposes.

According to this steel sheet data set, PC estimator with 3 PCS, ridge and Stein with regard to deviance, have deviances near to ML estimator›s deviance and their sum of coefficients variances are much less than that of ML estimator.

Multicollinearity may lead to have parameter estimation with sign that has conflict with expected sign in reality. Based on this steel data set, multicolinearity has no effect on the sign of ML estimator. For example, according to expert opinions we expect x_1(yield strength) and x_2(final tensile strength) appear with negative signs in model. Consequently, it seems Stein estimator is the most reliable one among the three mentioned estimators (Stein, ridge and PC estimator with 3 PCS).

All these methods can substantially reduce the variance of the estimated coefficients and prediction variance for future observations outside the mainstream of weighted multicollinearity.

References

Aucott, L. S., Garthwaite, P. H., Currall, J. (1984). Regression methods for high dimensional multicollinear data. Commun. Stat., 29, 1021-1037.

Bastien, P., Vinzi, V. E., Tenenhaus, M. (2005). Pls generalized linear regression. Comput. Stat. Data An., 48, 17-46.

Bjorkstrom, A. (2010). Krylov sequences as a tool for analyzing iterated regression algorithms. Scand. J. Statist., 37, 166-175.

Boente, G., Pires, A. M., Rodrigues, I., M. (2010). Detecting influential observations in principal components and common principal components. Comput. Stat. Data An., 54, 2967-2975.

Boik, R. J. (2013). Model-based principal components of correlation matrices. J. Multivariate. Anal.,116, 310-331.

De Leeuw, J. (1986). Nonlinear principal component analysis. In Comp. stat. 82, (Eds. Caussinus H.,Eltinger P., Tomassone R.), Wien: Physica-Verlag, 77-86.

Escofier, B., Page's, J. (1988). Analysis factorielles multiple. *Paris: Dunod*.

Fisher, T. J., Sun, X. (2011). Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix. Comput. Stat. Data An., 55, 1909-

1918.

Fomby, T. B., Hill, R. C., Johnson, S. R. (1978). An optimal property of principal components in the context of restricted least squares. J. Amer. Statist. Assoc., 73, 181-193.

Fujiwara, K., Sawada, H., Kano, M. (2012). Input variable selection for PLS modeling using nearest correlation spectral clustering. Chemometr. intell. Lab., 118, 109-119.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. Bimetrica, 53, 325-338.

Hawkins, D. M. (1973). The detection of errors in multivariate data using principal components. J. Amer. Statist . Assoc., 69, 340-344.

Hoerl, A. E., Kennard, R. W. (1970). Ridge regression: applications to non orthogonal problems. Technometrics, 12, 69-82.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. J. Educ. Psychol., 24, 417-441, 498-520. [Stein (1960)]

Hwang, J. T., Nettleton, D. (2000). Principal components regression with data-chosen components and related methods. Iowa State University, Ames Iowa 50011-1210.

Kibria, B. M., Saleh, A. K. E. (2012). Improved the estimators of the parameters of a probit regression model: A ridge regression approach. J. Stat. Plan. Infer., 142, 1421-1435.

Marx, B. D. (1992)., A continuum of principal component generalized linear regression. Comput. Stat. Data An., 13, 385-393.

Marx, B. D., Smith, E. P. (1990a). Principal component estimators for generalized linear regression. Biometrica, 77, 23-31.

Marx, B. D., Smith, E. P. (1990b). Weighted multicollinearity in logistic regression: diagnostics and biased estimation techniques with an example from lake acidification. Can. J. Fish. Aquat. Sci., 47, 1128-1135.

Pages, J., Tenenhaus, M. (2001). Multiple factor analysis combined with PLS path modeling, Application to the analysis of relationship between physicochemical variables, sensory profiles and hedonic judgments. Chemometr. Intell. Lab., 58, 261-273.

Pirouz, D. M. (2006). An overview of partial least squares, University of Western Ontario, The Richard Ivey School of Business.

Roozbeh, M., Arashi, M. (2013). Feasible ridge estimator in partially linear models. J. Multivariate. Anal.,116, 35-44.

Schaefer, R. L. (1986). Alternative estimators in logistic regression when the data are collinear. Stat. Comput. Simul., 25, 79-91.

Schaefer, R. L., Roi, L. D., Wolf, R. A. (1984). A ridge logistic estimator. Commun. Stat. Theor. M., 13, 99-113.

Stein, C. M. (1960). Multiple regression. *Contribution to probability and statistics, Stanford university press*.

Wold, H. (1984). PLS regression, Encyclopedia of Statistical Sciences, Academic Press, New York, 6, 581-591.