

APPLICATION OF RIDGE REGRESSION TO MULTICOLLINEAR DATA

G. R. Pasha¹ and Muhammad Akbar Ali Shah²

¹Department of Statistics, Bahauddin Zakariya University, Multan.

²Department of Statistics, Islamia University, Bahawalpur.

Abstract: The main thrust of this paper is to investigate the ridge regression problem in multicollinear data. The properties of ridge estimator are discussed. Variance inflation factors, eigen values and standardization problem are studied through an empirical comparison between OLS and ridge regression method by regressing number of persons employed on five variables. Methods to choose biasing parameter K are also presented.

Keywords: Biasing parameter, eigen values, inflation factors, multicollinearity, Ridge regression, standardization.

INTRODUCTION

In linear estimation, one postulates a model of the form

$$Y = X\beta + \varepsilon$$

The $n \times p$ matrix X contains the values of p predictor variable at each of n data points. Y is the vector of the observed values, β is the $p \times 1$ vector of the population values and ε is an $n \times 1$ vector of experimental errors having the properties $E(\varepsilon) = 0$ and $E(\varepsilon'\varepsilon) = \sigma^2/n$. For convenience, we assume that the X variables are scaled so that $(X'X)$ has the form of a correlation matrix. The conventional estimator for β is the least squares estimator, $\hat{\beta}$, where $\hat{\beta}$ is chosen to minimize the sum of squares of residuals $\Phi(\hat{\beta})$.

$$\Phi(\hat{\beta}) = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

The two key properties of $\hat{\beta}$ are that it is unbiased, that is, $E(\hat{\beta}) = \beta$ and it has minimum variance among all linear unbiased estimators. The variance matrix is

$$V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

In the development of ridge regression, Hoerl and Kennard [1976] focus attention on the eigen values of $(X'X)$. A serious non-orthogonal or "ill-conditioned" problem is characterized by the fact that the smallest eigen value, $\lambda_{\min}$ is very much smaller than unity. Hoerl and Kennard [1976] have summarized dramatic inadequacy of least squares for nonorthogonal problems by noting that the expected squared length of the coefficient vector is

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 \text{Tr}(X'X)^{-1} = \beta'\beta + \sigma^2 / \lambda_{\min}$$

Thus $\hat{\beta}$, the least squares coefficient vector, is much too long, on the average, for ill conditioned data, since $\lambda_{\min} < 1$. The least squares solution yields coefficients whose absolute values are too large and whose signs may actually reverse with negligible changes in the data [Buonaccorsi 1996].

RIDGE SOLUTION

The ridge estimator is obtained by solving $(X'X + KI)\hat{\beta}^* = g$ to give

$$\beta^{\wedge*} = (X'X + KI)^{-1} g$$

where $g = X'Y$ and K is ridge parameter holds $K \geq 0$. In general, there is an "optimum" value of K for any problem. But it is desirable to examine the ridge solution for a range of admissible values of K . Hoerl gave the name ridge regression to his procedure because of similarity of its mathematics to methods he used earlier, i.e., "ridge analysis", for graphically depicting the characteristics of second order response surface equations in many predictor variables [Cheng and Schneeweiss 1996, Cook 1999].

Key properties applicable to ridge regression are:

- (i) $\beta^{\wedge*}$ minimizes the sum of squared residuals on the sphere centered at the origin whose radius is the length of $\beta^{\wedge*}$. That is, for a given sum of squared residuals, it is the coefficient vector with minimum length.
- (ii) The sum of squared residuals is an increasing function of K .
- (iii) $\beta^{\wedge*} \beta^{\wedge*} < \beta' \beta$, and $\beta^{\wedge*} \beta^{\wedge*} \rightarrow$ as $K \rightarrow \infty$.
- (iv) The ratio of the largest characteristic root of the design matrix $(X'X + KI)$ to the smallest root, called the condition number of the data, is $(\lambda_1 + K)/(\lambda_K + K)$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K$ are the ordered roots of $X'X$ and is a decreasing function of K . The condition number is often taken as a measure of ill-conditioning and thus related to the stability of the coefficient estimates.
- (v) The ridge estimator

$$\beta^{\wedge*} = [I_k + K(X'X)^{-1}]^{-1} \beta^{\wedge} = W \beta^{\wedge}$$
 is a linear transformation of the least square estimator.
- (vi) The mean square error of $\beta^{\wedge*}$ is

$$\begin{aligned} E[(\beta^{\wedge*} - \beta)(\beta^{\wedge*} - \beta)'] &= E[(\beta^{\wedge*} - \beta)' W' W (\beta^{\wedge} - \beta)] + (W\beta - \beta)' (W\beta - \beta) \\ &= \sigma^2 \sum \lambda_i (\lambda_i + K) + K^2 \beta' (X'X + KI)^{-2} \beta \end{aligned}$$
- (vii) There always exist a $K > 0$, such that $\beta^{\wedge*}$ has a smaller MSE than $\beta^{\wedge}$.

MULTICOLLINEARITY

Multicollinearity can cause serious problem in estimation and prediction, increasing the variance of least squares estimator of the regression coefficients and tending to produce least squares estimates that are too large in absolute value [Wethrill 1986]. If the two explanatory variables are involved, there is no guarantee that any of the pair wise correlation coefficients will be large. Variance inflation factor (VIF) is also used to detect multicollinearity. Marquardt and Snee [1970] suggest that VIF greater than 10 indicates multicollinearity. In the class of biased estimators, the most popular is ridge regression. Ridge regression overcomes problem of multicollinearity by adding a small quantity to the diagonal of X' (which is in correlation form) i.e. $\beta^{\wedge*} = (X'X + KI)^{-1} X'Y$ where $X'X$ and $X'Y$ are in correlation form.

In the presence of multicollinearity the ridge estimator is much more stable (i.e. has smaller variance) than the OLS. But the cost of this procedure is that it introduces bias; so the elements of D are chosen to

keep this bias small enough so that the overall mean squared error is still reduced.

DATA STANDARDIZATION IN RIDGE REGRESSION

One source of controversy in ridge regression is the availability of re-centering and rescaling the original data to make $X'X$ a correlation matrix. Standardization is unnecessary for most theoretical results and not advisable for these cases when the investigator is committed to the centering, scaling and MSE computation in original units. Therefore it may be confusing when Hoerl and Kennard [1970a] and Marquardt and Snee [1970] recommended standardization without ruling out these cases. Often there is nothing different in a model where (i) a temperature is measured in °C or °F. (ii) a money variable is measure in U.S. dollars, Pakistani Rs., or a linear combination of prices of various countries reflected by, say, the Special Drawing Rights; and (iii) the base year of an index number deflator variable is 1961 or 1976 etc. For these “essentially similar” models the investigator wants the regression coefficient to be “essentially linear”, i.e. “equivalent”. The advantage of standardization is that it makes the coefficient β_i comparable with each other.

EIGEN VALUES

Let $Y = AX$, where $A = a_{ij}$, ($ij = 1, 2, \dots, n$), be a linear transformation over F . In general, the transformation carries a vector $X = x_1, x_2, \dots, x_n$ is a vector $Y = [y_1, \dots, y_n]'$ whose only connection with X is through the transformation. We shall investigate here the possibility of certain vectors X being carried by the transformation into λX , where λ is either a scalar or F or of some field of which F is subfield. Any vector, which by the transformation is carried out into λX , that is, any vector X for

$$AX = \lambda X$$

is called invariant vector under the transformation. The character equation we have

$$X - AX = (I - A)X = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{12} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ \lambda - a_{n1} & -a_{n2} & \dots & -a_{nn} \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{vmatrix} = 0$$

The system of equation has non-trivial solution if and only if $(\lambda I - A) = 0$. The expansion of this determinant yields a polynomial, $\Phi(\lambda)$ of degree in λ which is known as the characteristic polynomial of the transformation or of matrix A . The equation $\Phi(\lambda) = 0$ is called the characteristic roots of A . Those characteristic roots are called eigen values and characteristic vectors are called eigen vectors.

VARIANCE INFLATION FACTOR (VIF)

The VIF is related to the results of auxiliary regression and is measured by the diagonal elements of $(X'X)^{-1}$. The j th diagonal elements of $(X'X)^{-1}$ can be represented as $(1-R_{j2})^{-2}$, where R_{j2} is the coefficient of determination from the auxiliary regression of the j th independent variable on the other $(K-1)$ regressors. When $R_{j2}' Y = (X'X + KI)^{-1} X' Y \rightarrow 1$, the j th diagonal element of $(X'X)^{-1}$, which when multiplied by σ^2 gives the sampling variance of the j th element of β , gets very large and hence the name VIF. Despite the intuitive nature of this measure, however, it has a serious flaw that prevents it from becoming a useful indicator of ill conditioning. If there is a single, near-linear dependency, the large diagonal elements of $(X'X)^{-1}$ indicate which variable are involved. However, if two or more linear dependencies are present, it becomes difficult to sort them out by inspecting $(X'X)^{-1}$.

SELECTION OF VARIABLES IN RIDGE REGRESSION BY RIDGE TRACE

Variable selection procedures often do not perform well when the predictor variables are highly correlated [Velilla 1998, Hsich 1997]. Marquardt and Snee [1970] point out that when the data is highly multicollinear, the maximum variance inflation completely destabilizes all the criteria obtained from the least squares estimates. Hoerl and Kennard [1970b] suggest that the ridge trace can be used as a guide for variable selection. They propose the following procedure for eliminating predictor variables from the full model.

- (i) Eliminate predictor variables that are stable but have small predicting power; that is those with small standardized regression coefficient.
- (ii) Eliminate predictor variables with unstable coefficients that do not hold their predicting power because the coefficients tend to zero as K increases.
- (iii) Eliminate one or more of the remaining predictor variables that have small coefficients. The subset of the remaining predictor variable is used in the final model.

RIDGE ESTIMATORS

The purpose of ridge trace is to give the analyst a compact picture of the effect of the non orthogonality of $X'X$ on the estimation of β . Hoerl *et al.* [1975] recommended $K_{HKB} = P \sigma^2 / \beta' \beta$ as general rule where the parameters are estimated from the full equation least squares fit. Their studies suggest that the resulting ridge estimator yields coefficient estimates with smaller mean squared error than that obtained from least squares. In a latter paper Hoerl and Kennard [1975] suggest an iterative procedure where $K_{HKB} = P \sigma^2 / \beta_i^{\wedge'} \beta_i^{\wedge}$ where $\beta_i^{\wedge} = \beta_R \sim (K_i)$. Farebrother [1975] suggested $K = \sigma^2 / \beta^{\wedge'} \beta^{\wedge}$, which for the Gonman-Toman data, yields $K = 0.003$ with this formula. It is of interest to note that

in the case $X'X = I$, the choice of K which will minimize $E[L_j^2(K)]$ is $K = t \sigma^2 / \beta' \beta$. Marquardt and Snee [1970] suggested using the value of K for which the maximum variance inflation factor is "between one and ten and close to one". Mallows [1973] extended the concept of C_p -plots to C_K -plots, which may be used to determine K . Specifically, he suggested plotting C_K , versus V_K where

$$C_K = (RSS_K / \sigma^2) - n + 2 + 2 \operatorname{tr}(XL)$$

$$V_K = 1 + \operatorname{tr}(X'XLL')$$

and $L = (X'X + KI)^{-1}X'$.

Here RSS_K is the residual sum of squares as a function of K . The suggestion is to choose K to minimize C_K . Newhouse and Oman [1971] conducted a simulation study of ridge estimators. Their study was restricted to the case of two predictors for two different values of X , the correlation between two predictors and a number of schemes for choosing K . Their conclusions indicated the least for the case $P = 2$ that ridge estimator may be worse than OLS and in general, fail to establish any superiority. They further suggested that there is nothing to suggest that results in the higher dimensions ($P > 2$) would be substantially different. The basic idea is to choose K so that

$$\tilde{\beta}_R' \tilde{\beta}_R = \beta' \beta - \sigma^2 \sum 1/\lambda_i$$

If the RHS is negative, they suggest two modifications. Although neither method was better than least squares in all cases, they concluded, based on optimal rule for choosing K , that there is a sufficient potential improvement to warrant further investigations of ridge estimators. They also considered $p = 2$ and found that their results were comparable to those of Newhouse and Oman [1971]. They suggest that there may be some advantage to ridge estimators in higher dimensions which is not available for $p = 2$. This is consistent with the results of Stein [1960]. They also report that the values of K chosen by their optimal rules were originally less than those obtained from the ridge trace. Lawless and Wang [1976] proposed first orthogonalizing the $X'X$ matrix by finding a matrix P such that $P'X'XP = \Lambda$, where Λ is the diagonal matrix whose elements are the eigenvalues λ_i of $X'X$ and setting

$$K_{LW} = P\sigma^2 / \sum \lambda_i g_i^2$$

where g_i is the transformed coefficients, $g = P' \beta$. Lawless and Wang [1976] found their estimator did well. λ_i are the eigen values.

COMPARISON BETWEEN THE OLS AND RIDGE REGRESSION

We try to justify that ridge regression is better than OLS method. The following regression model is fitted to the data in which number of persons employed are regressed on 5 predictor variables

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + e_i$$

Y = No. of persons employed (million), X_1 = Land cultivated (million hectares), X_2 = Inflation rate %, X_3 = No. of establishment, X_4 = Population (million), X_5 = Literacy rate (%).

	Y	X1	X2	X3	X4	X5
1974-75	20.30	19.55	0.2671	3286	68.924	22.2
	20.08	19.82	0.1166	3248	71.033	22.5
	21.89	19.76	0.1178	3373	73.205	22.8
	22.73	20.10	0.0779	3676	75.444	23.2
	23.62	19.98	0.0663	3715	77.516	23.4
	24.15	20.23	0.1072	3750	80.130	23.7
	24.70	20.30	0.1237	3815	82.580	24.0
	25.27	20.42	0.1000	3882	84.254	26.2
	25.85	20.31	0.0448	3931	87.758	26.5
	26.40	20.33	0.0836	4047	90.480	26.9
	26.96	20.61	0.0746	4423	93.286	27.2
	27.93	20.67	0.0483	4349	96.180	27.5
	28.70	20.92	0.0387	4544	99.162	27.9
	28.99	20.66	0.3884	4573	102.230	28.0
	29.99	20.73	0.3087	4595	105.409	28.1
	30.82	20.73	0.3854	4543	108.678	28.3
	31.78	20.77	0.3886	4589	111.938	28.6
	31.78	20.96	0.2910	4656	111.938	34.9
	31.94	21.06	0.4112	4849	113.610	36.0
	32.45	21.40	0.2129	4809	116.470	37.2
	33.29	21.51	0.6121	4852	119.390	38.4
	33.60	21.55	0.4291	4998	122.361	39.6
	34.42	21.68	0.1231	5072	125.387	40.9
	36.84	21.98	0.5120	4992	128.421	42.2
	37.73	21.96	0.4001	4924	131.510	43.6
	38.59	21.93	0.4014	4992	134.511	45.0
	40.40	21.99	0.4423	5081	137.512	47.1
2000-01	41.20	21.99	0.4328	5128	140.473	52.0

Source: Economic Survey of Pakistan 2000-01.

EMPIRICAL RESULTS

Uncorrected sum of squares and cross product matrix

	X ₁	X ₂	X ₃	X ₄	X ₅	Y
X ₁	1.62x10 ⁴					
X ₂	8.00x10 ¹	8.72x10 ¹				
X ₃	2.16x10 ⁶	1.59x10 ⁴	4.72x10 ⁸			
X ₄	3.98x10 ⁴	3.67x10 ¹	7.51x10 ⁶	1.24x10 ⁵		
X ₅	1.98x10 ⁴	3.06x10 ¹	2.54x10 ⁶	5.76x10 ⁴	1.45x10 ⁴	
Y	1.25x10 ⁴	8.65x10 ¹	2.21x10 ⁶	5.21x10 ⁴	2.21x10 ⁴	1.51x10 ⁴

Correct sum of squares and cross products matrix

	X ₁	X ₂	X ₃	X ₄	X ₅	Y
X ₁	1.34x10 ¹					
X ₂	1.00x10 ⁰	3.44x10 ¹				
X ₃	8.54x10 ³	7.45x10 ²	6.34x10 ⁶			
X ₄	3.56x10 ²	3.78x10 ¹	2.87x10 ⁵	9.23x10 ³		
X ₅	6.41x10 ¹	5.81x10 ⁰	3.98x10 ⁴	1.72x10 ³	2.91x10 ²	
Y	1.20x10 ⁴	8.60x10 ¹	2.62x10 ⁶	5.24x10 ⁴	2.87x10 ⁴	1.58x10 ⁴

$$X'Y = \begin{vmatrix} 0.949 \\ 0.565 \\ 0.924 \\ 0.996 \\ 0.973 \end{vmatrix}$$

Eigen values = [4.1352 0.6939 0.1223 0.0327 0.0143]

$$(X'X)^{-1} = \begin{vmatrix} 15.9031 & 1.8922 & 1.8866 & -6.2106 & -11.9521 \\ & 2.2547 & 0.9698 & -5.8593 & -1.8765 \\ & & 8.0819 & -6.6858 & -3.2455 \\ & & & 40.4973 & -24.4707 \\ & & & & 38.5121 \end{vmatrix}$$

Coefficient of determination (R-square) = 0.995, adjusted $R^2 = 0.993$. SE of estimate = 0.395. Estimate of $\sigma^2 = 0.1549$. Now computing OLS estimates.

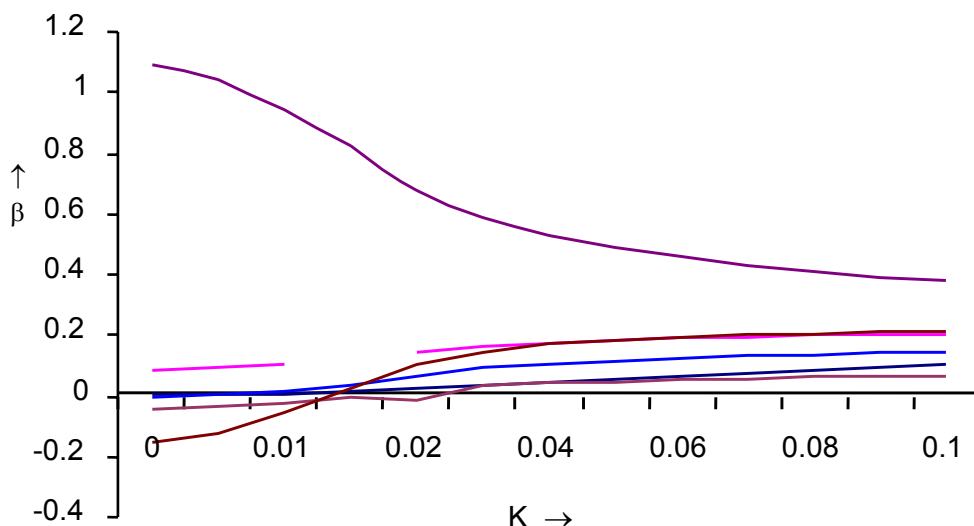
$$\beta^{\wedge} = \begin{vmatrix} 0.0839 \\ -0.0489 \\ -0.0109 \\ 1.1380 \\ -0.1855 \end{vmatrix}$$

As most of the variation is found in first three eigen values and maximum VIF is 40.4973, which is greater multicollinearity. We also see that R^2_{xixj} , R^2 are very high, and least square estimates are unstable. The predictor variables are correlated so we can apply ridge regression techniques to get stable set of coefficients.

i) Ridge Trace i.e. $(X'X)^{-1}, \beta^{\wedge*} = (X'X + KI)^{-1} X'Y$

- | | | | |
|-----------------|----------------|-----------------|-----------------|
| (i) K = 0.001 | (ii) K = 0.002 | (iii) K = 0.005 | (iv) K = 0.01 |
| (v) K = 0.02 | (vi) K = 0.03 | (vii) K = 0.04 | (viii) K = 0.05 |
| (ix) K = 0.06 | (x) K = 0.07 | (xi) K = 0.08 | (xii) K = 0.09 |
| (xiii) K = 0.10 | | | |

Plotting the values of β on graph and finding the value of K i.e. $0.05 < K < 0.06$ i.e. $K = 0.055$



$$\beta^{\wedge*} = \begin{vmatrix} 0.1870 \\ 0.0500 \\ 0.1200 \\ 0.4700 \\ 0.1880 \end{vmatrix}$$

MSE($\beta^{\wedge}$) = 6.2532		K= 0.055	MSE($\beta^{\wedge*}$) = 2.5006
ii)	K_{HKB}	= 0.578	
iii)	K_F	= 0.1165	
iv)	K_{LW}	= 0.4231	
v)	K_{PZ}	= 0.0765	
vi)	K_{PZ}	by VIF	

$0.06 \leq K \leq 0.50$. We can take any value in this range. The results are summarized in the following Table.

Table: Comparison of coefficient and MSE's for different Ridge Regression method and OLS method

Coefficient	K_{tr}	K_{HKB}	K_F	K_{LW}	K_{PZ}	OLS
β_1^*	0.1870	0.205	0.209	0.211	0.200	0.839
β_2^*	0.0501	0.901	0.069	0.088	0.061	-0.049
β_3^*	0.1203	0.181	0.154	0.182	0.139	-0.139
β_4^*	0.4702	0.239	0.363	0.257	0.467	1.138
β_5^*	0.1882	0.208	0.215	0.216	0.206	-0.186
MSE	2.5006	1.280	1.715	1.366	1.991	6.253

DISCUSSION AND CONCLUSIONS

The number of persons employed in Pakistan is regressed on 5 predictor variables. All the symptoms of applying ridge regression are found in this problem. So we adopted six procedures for the choosing biasing ridge parameter. We note that all methods are better than OLS as it is clear from the table. The six procedures are discussed as follows:

First method is Ridge trace in which we start from $K=0$ and then after taking three values 0.001, 0.002, 0.005 for K , we give the equal space of 0.01. We plot the regression coefficient against K . The system has been stabilized at $0.05 < K < 0.06$ i.e. $K=0.055$ is the ridge parameter. But a very frequent criticism on ridge regression is that it is subjective approach. This technique is also very useful for selection of variables. We observe following points from the graph:

- The variable X_1 and X_2 has not so much variation after introducing bias in the system and they are stable for very high value of K . So they have no predictive power and we exclude them from the model.
- The variable X_3 has smaller value for $K=0$ and when K increases and we can not exclude it from the model.
- The variable X_4, X_5 are the coefficients which are varying very rapidly. The variable X_5 has the most smaller value but with the increasing K it move upward from zero and obtain its predictive power. So we include this in our model. The variable X_4 loose its prediction power at the initial stage but then it go toward zero. So no complete statement can be made about the variable X_4 . We are including this variable because it has still highest predictive power. Our selected model by Ridge trace is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + e_i$$

The second method for the selection of $K=0.116$ is given by Farebrother [1975] but Hoerl [1962], Kennard and Baldwin [1985] showed that their estimator is better than Farebrother [1975] and made simulation in the

favor of their estimator and prove that it has smaller MSE in more than 50% problems. We observe that his $K_{HKB} = 0.58$ gives the smaller MSE, but Wichn and Wahba [1988] showed that K_{HKB} is doing it poorly. Lawless and Wang [1976] modify their estimator and they make use of eigen values also and found that their estimator is doing well. This gives $K_{LW} = 0.42$. The technique developed in this communication seems to be very reasonable because of having smaller bias and smaller MSE. This method gives value of K_{PZ} very close to ridge trace. So we can say that either the K_{HKB} or K_{PZ} is the best one. Table shows that Ridge regression by any method is best one than OLS method.

References

- Buonaccorsi, J.P. (1996) "A modified estimating equation approach to correcting for measurement error in regression", *Biometrika*, 83, 433-440
- Cheng, C.-L. and Schneeweiss, H. (1996) "The polynomial regression with errors in variables", Discussion Paper 42, Sonderforschungsbereich 386, University of Munich, Munich.
- Cook, R.D. (1999) "Applied Regression Including Computing and Graphics", Wiley, New York.
- Farebrother, (1975) "Principal component estimators and minimum mean squares criteria in regression analysis", *Review of Econometrics and statistics*, 54, 332-336.
- Hoerl, A.E. and Kennard, R.W. (1975) "Ridge regression; some simulations", *Commun. Statist. Theor. Meth.*, 5(4), 307-323.
- Hoerl, A.E. (1962) "Application of Ridge Analysis to regression problems", *Chemical Engineering Progress*. 58, 54-59.
- Hoerl, A.E. and Kennard, R.W. (1970a) "Ridge regression; biased estimation for non orthogonal problems", *Technometrics* 12, 55-67.
- Hoerl, A.E. and Kennard, R.W. (1970b) "Ridge regression: an application to non orthogonal problems", *Technometrics*, 12, 69-82.
- Hoerl, A.E. and Kennard, R.W. (1976) "Ridge regression: iterative estimation of the biasing parameter", *Commun. Statist. Theor. Meth.* 5(1), 77-88.
- Hoerl, A.E., Kennard, R.W. and Hoerl, R.W. (1975) "Practical use of Ridge regression; a challenge met", *The Journal of the Royal Statistical Society C*, 34(2), 114-120.
- Hsich, F. (1997) "Estimation in homoscedastic linear regression models with censored data: an empirical process approach", *Ann. Statist.*, 25, 2665-2681.
- Kennard, R.W. and Baldwin (1985) "An examination of criticisms of ridge regression estimation methodology", Technical Report A, Department of Statistics, University of Melbourne, Australia.

- Lawless and Wang, P. (1976) "A simulation study of Ridge and other Regression Estimators", *Communications in Statistics, Part A-Theory and Method* 5, 307-323.
- Mallows, H. (1973) "Modern Factor Analysis", University of Chicago Press
- Marquardt, D.W. and Snee, R.D. (1970) "Generalized in inverses, Ridge regression, Biased linear estimation", *Tecchnometrics*, 12, 591-612.
- Marquardt, D.W. and Snee, R.D. (1970) "Ridge regression in practice". *The American Stat.*, 29, 3-19.
- Newhouse, W. and Oman (1971) "A comparison of ridge estimators", *Technometrics*, 20, 301-311.
- Stein, J. (1960) "A critical view of ridge regression", *The Statistician*, 22, 181-187.
- Velilla, S. (1998) "Accessing the number of linear components in a general regression problem", *J. Amer. Stat. Assoc.*, 93, 1088-1098.
- Wethrill (1986) "Evaluation of ordinary ridge regression", *Bulletin of Mathemetical Statistics*, 18, 1-35.
- Wichern, D.W. and Churchill, G.A. (1978) "A comparison of ridge estimators", *Techonometrics* 20, 301-311.
- Wichn, D. and Wahba (1988) "Applied Regression Analysis: A Research Tool", Pacific Gross, California, Wordsworth and Brooks.