International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 5 Issue 8, August - 2021, Pages: 43-51

Selecting Optimal Rank in Reduced Rank Regression by AIC

Saad Abed Madhi[1] and Samira Faisal Abushilah[2]

[1]Department of Medical Instrumentation Technologies Engineering, Hilla University College, Babylon, Iraq.
[2]Department of Mathematics, Faculty of Education for Girls, University of Kufa, Iraq.

***Abstract**: This paper presents a method for determining the actual rank of the coefficients matrix in the reduced rank multivariate regression model. The method is constructed using the singular value decomposition and the Akaike's Information Criterion (AIC). Some illustrative examples are given to verify this method.*

Keywords: Reduced Rank; AIC; Singular Value Decomposition; More-Pernorse inverse; Least-squares.

1. Introduction

Consider the multivariate regression model

$$y_i^T = c_i^T + x_i^T M + e_i^T, i = 1, 2, \dots, n, \qquad (1.1)$$

where $x_i(r \times 1)$ is a vector of dependent variables, $y_i(t \times 1)$ is the vector of dependent variables, $M(r \times 1)$ is a matrix whose columns are the individual unknown regression coefficients for each dependent variable on the best of the independent variables, $e_i(t \times 1)$ is a vector of unknown response specific constants and $e_i(t \times 1)$ is a vector of stochastic errors assumed to be independently distributed with zero mean and unit variance. T indicates transpose of a matrix or a vector. In matrix form, Equation (1.1) can be written as

$$Y = C + XM + E, \ \ where\ C = Ie_i^T. \qquad (1.2)$$

Suppose the rank of M is s, where $s \leq t$. The full rank regression coefficient matrix occurs when $s = t$. The reduced rank regression coefficient matrix occurs (due to some linear restrictions on the regression coefficients) when $s < t$. Such models have been studied by several authors. Izenman (1975) considered the problem of estimating the regression coefficient matrix having (known reduced) rank [6]. Alvarez et al. (2016) present a procedure for coefficient estimation in multivariate model for reduced rank in the presence of multicollinearity [2]. Davies et al. (1982) gave a method for estimating the coefficient matrix, which is justified by a least-square analysis employing singular value decomposition and the Eckart-Young theorem [3]. Izenman (1975) showed that canonical variable and principal component are special cases of a reduced rank regression model [6].

One problem that arises in the estimation of the reduced rank regression coefficient matrix is the choice of s, the assumed maximum rank of M. Madhi (1981) employed cross validation criterion to determine the rank in reduced rank regression [8].

In this paper, we propose a method for determining the actual rank of the coefficient matrix in the reduced rank regression model (RRRM) employing the method of estimation proposed by Davies et al. (1982) [3] and the Akaike's information criterion (AIC) [1]. Some illustrative examples are also given.

2. Preliminaries

2.1. Singular Value Decomposition of a Matrix (SVD) (Mandel 1982 [9])

If A is a $m \times n$ matrix, of rank k, it can be expressed as

$$A = UDW^T, \qquad (2.1)$$

where

U is an $m \times n$ orthogonal matrix,

W is an $n \times n$ orthogonal matrix,

D is a $m \times n$ diagonal matrix with non-negative elements, $D = diag(\sigma_1, \dots, \sigma_k, 0, \dots, 0)$, $\sigma_1 \geq \cdots \geq \sigma_k > 0$.

If Q and V consist of the first k columns of U and W respectively, and Σ is a $k \times k$ diagonal matrix with positive diagonal elements, $\sigma_1 \geq \cdots \geq \sigma_k > 0$, then

$$A = Q\Sigma V^T, \qquad (2.2)$$

where $Q^TQ = V^TV = I_k$ and $\Sigma = diag(\sigma_1, \sigma_2, \dots, \sigma_k)$ with $\sigma_1 \geq \sigma_2, \dots \geq \sigma_k > 0$ is called basic diagonal (Green, 1976) [5].

The basic diagonal part of the decomposition is always unique regardless of whether A is of full rank, square, or rectangular. Each one of Equations (2.1) and (2.2) is equivalent to

$$A = \sigma_1 q_1 v_1^T + \sigma_2 q_2 v_2^T + \ldots + \sigma_k q_k v_k^T. \qquad (2.3)$$

That is, the sum of k matrices of rank 1. The column vectors $\{q_i\}_{i=1}^k$ of Q are orthonormal (orthogonal and each of length 1) and each has m components. The row vectors $\{v_i\}_{i=1}^k$of V^T are orthonormal and each has n components.

The numbers $\sigma_1, \sigma_2, \ldots$ are the singular values of A. The vectors $[q_1, q_2, \ldots]^T$ and $[v_1, v_2, \ldots]^T$ are respectively the left and right singular vectors. When A is square and symmetric the singular decomposition reduces to known spectral decomposition, where the left and right singular vectors are identical and reduce to eigenvectors.

The orthogonal sets $\{q_1, q_2, \ldots, q_k\}$ and $\{v_1, v_2, \ldots, v_k\}$ can be completed to sets $\{q_1, q_2, \ldots, q_m\}$ and $\{v_1, v_2, \ldots, v_m\}$. A complete decomposition of A is then (if $m \leq n$ without loss of generality)

$$\sum_{i=1}^m \sigma_i q_i v_i^T, \text{ with } \sigma_{k+1} = \cdots = \sigma_m = 0. \qquad (2.4)$$

The singular decomposition (2.3) is of course equal in numerical value to the complete singular decomposition (2.1). The SVD is closely related to the eigenvalue decomposition, since

$$AA^T = Q\Sigma^2 Q^T, \qquad (2.5)$$

where $\Sigma^2 = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$ and $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ are the non-zero eigenvalues of the $m \times n$ matrix AA^T and the columns of Q are the corresponding eigenvectors of AA^T. Furthermore,

$$AA^T = V\Sigma^2 V^T, \qquad (2.6)$$

where $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ are also the non-zero eigenvalues of the $n \times n$ matrix $AA^T$and the columns of V are the corresponding eigenvectors. Hence, the singular values of A are the square roots of the common positive eigenvalues of the $m \times m$ matrix AA^T and the $n \times n$ matrix AA^T.

Rank Determination

If A is a $m \times n$ matrix ($m \geq n$), of rank $k < n$, then the singular values of A satisfy

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > \sigma_{k+1} = \cdots = \sigma_n = 0.$$

Conversely, if $\sigma_k \neq 0$ and $\sigma_{k+1} = \cdots = \sigma_k = 0$ then A is of rank k. Thus the singular value decomposition can be used to determine the rank of matrix.

2.2 Generalized Inverse in Linear Statistical Model

Generalized inverse can be very useful in the regression model, where inverses arise naturally. For example, computer solutions for models involving inverses can be simplified if generalized inverses are used in place of ordinary inverses. There are many types of generalized inverses, each obeying a particular set of properties (Green, 2014) [5]. Also, different ways have been developed to define these inverses. The basic types of generalized inverse are discussed here.

G-Inverse

Let A be an $m \times n$ matrix of any rank, a generalised inverse (or a G-Inverse) of A is an $n \times m$ denoted byA^-, such that

$$AA^-A = A.$$

A generalized inverse always exists, but it is not necessarily unique. One way of illustrating the existence of A^- and its non-uniqueness is by using SVD. For $m \times n$ matrix, write

$$A = U\Sigma V^T,$$

then the general G-Inverse (Good, 1969) [4] is

$$A^- = \mathrm{V}\Sigma^- U^T = \sigma_1^- v_1 u_1^T + \cdots + \sigma_p^- v_p u_p^T,$$

where $p = \min(m, n)$, $\sigma_1, \sigma_2, \ldots, \sigma_p$ are all the singular values of A (non-negative square roots of all the eigenvalues of $A^T A$ if $m \leq n$, or $A^T A$ if $m \geq n$), and σ^- means σ^{-1}if $\sigma \neq 0$ and is otherwise arbitrary. Rao (1965, p.25) [10] shows that A^- exists such that rank $A^- = \min(m, n)$ irrespective of the value of rank A. This is obvious in terms of the SVD by taking all $\sigma_i^- \neq 0$.

Moore-Penrose Generalized Inverse

Moore and Penrose (Good, 1969) [4] defined a particular generalized inverse often called (the pseudo inverse) as a matrix $A^+(n, m)$ to distinguish it from a general g-inverse A^-, satisfying the properties:

- $AA^+A = A$
- $A^+AA^+ = A^+$
- $(AA^+)^T = AA^+$
- $(A^+A)^T = A^+A$

Such an inverse always exists and is unique. For an arbitrary $m \times n$ matrix A, of rank k, write the SVD as

$$A = U\Sigma V^T,$$

then the Moore-Penrose inverse is

$$A^+ = U\Sigma^{-1}U^T = \sigma_1^{-1}v_1u_1^T + \cdots + \sigma_p^{-1}v_pu_p^T. \qquad (2.7)$$

It is clear that A^+ can be uniquely defined as the the g-inverse of minimum rank, since the rank of A^+ is simply the number of σ_j's that do not vanish.

2.3 Eckart-Young Theorem

Given $Y(m, n)$, of rank $s = \min(m, n)$, the matrix $H^+(m, n)$ of rank at most $(r < \min(m, n))$ that best approximate Y, i.e. H^+ satisfies

$$\min_{H\ rank\ H \leq r} ||Y - H||^2,$$

is given by the partial sum of the first r terms of the SVD of Y. That is, if

$$Y = \sum_{i=1}^{s=\min\{m,n\}} \sigma_i u_i v_i^T,$$

then

$$H^* = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

2.4 Akaike's Information Criterion (AIC)

Akaike (1973) defined an information criterion (AIC) [1]. This is very general criterion based on information theoretic concepts and can be used in a wide range of contexts. Snipes and Taylor (2014) [11] utilized AIC to compare different models. Karagrigorious and others (2011) investigated the asymptotic properties of AIC with applications [7]. It is based in part on the likelihood function. When a model involving k independently adjusted parameters is fitted to data, the AIC is defined by

$$AIC = (-2)\log(\mathrm{L}(\hat{\theta}/\mathrm{data})) + 2k,$$

$\hat{\theta}$ is the estimated parameters, and k is the number of parameters in model. $L(\hat{\theta}, data)$, the likelihood at its maximum point of the model estimated, log, denotes the natural logarithm. When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in overfitting. The AIC resolves this problem by interdicting the penalty term for the number of

parameters in the model. If we plot AIC(k) against k we can seek a minimum value of k at which this minimum value is attained. This minimum value of AIC is called MAIC (minimum AIC).

3. Estimation of the Coefficient Matrix Rank in RRRM

In this section, we discuss the method of estimation of the coefficient matrix which has been developed by Davies and Tso (1982). A solution employing matrix singular value decomposition was proposed and justified by the Eckart-Young theorem. This solution has the feature of generality as we can use it even when the regression coefficients are under determined. Generalized inverses have been used to achieve this generality.

Consider the model (1.1). Suppose that there are $s < \min(r, t)$ linear combinations of the independent variables $n_1 = x^T\alpha_1, \dots, n_s = x^T\alpha_s$ (i.e. $n^T = (n_1, \dots, n_s)$) normalized such that ${\alpha_i}^T\alpha_i = J_{ij}$, such that all the variation in y is due to only linear combinations of x plus stochastic error. This reduce the set of independent variables $x_1, x_2, \dots, x_r$ to a new set of independent variables $n_1, \dots, n_s$. Consequently, the rank of M will be less than or equal to s. Again consider the model (1.2). We now center the data by subtracting the column means from each variables of X and Y such that

$$I^TX = O^T, I^TY = 0^T.$$

Hence, the model (1.2) can be written as

$$Y = XM + E,$$

where $X(n \times r)$ and $Y(n \times t)$ are matrices whose n rows contain, respectively, independent and dependent data, and whose columns each sum to zero, E is the matrix of stochastic errors which are assumed to be uncorrelated row-wise. We have lost, c parameters but there is a corresponding loss in the data since quantities

$$y_{ij} - \bar{y}_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, t$$

represents only $(n - 1) \times t$ separate pieces of information to the fact that their sum to zero, whereas

$$y_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, t$$

represent $n \times t$ separate pieces of information. Effectively, the lost pieces of information have been used to enable the proper adjustments to be made to the model so that the C term can be removed. This transformation of the origin data to corrected data is consistent with least-square estimation of the vector of response constants c. The problem now is to estimate the unknown matrix of regression coefficients $M(r \times t)$ subject to the rank constraint

$$rank(M) \leq s \leq \min(r, t).$$

The first step is to determine the unconstrained least-squares estimate of $\widehat{M}$ by minimizing

$$||Y - XM||^2 \qquad (3.1)$$

The unique least-squares solution can be written in generalized matrix form as

$$\widehat{M} = X^+Y,$$

when X is of full column rank, this is equivalent to the ordinary least-squares estimator,

$$\widehat{M} = (X^TX)^{-1}X^TY$$

Hence, the corresponding unconstrained fitted of Y are obtained by

$$\hat{Y} = X\widehat{M}$$

The next step is to consider the estimation of the matrix M when it is constrained to have rank at most s. First of all, let us decompose (3.1) as

$$||Y - XM||^2 = ||Y - \hat{Y}||^2 + ||\hat{Y} - XM||^2. \qquad (3.2)$$

The second term only, in the above decomposition, varies as M varies. We may choose M to satisfy $XM = (\hat{Y})_s$, where $\hat{Y}$, is the partial sum to the s terms of the SVD of $\hat{Y}$. By the Eckart-Young theorem M must minimise the second term in (3.2) and therefore must minimise the least-squares criterion (3.1).

Finding $\hat{Y}$, perform the singular value decomposition of $\bar{Y}$ as follows:

$$\hat{Y} = Q\Sigma V^T,$$

where $Q = (q_1, \dots, q_t)$; $\Sigma = diag(\sigma_1, \dots, \sigma_t)$; $V^T = (v_1, \dots, v_t)^T$.

Now, let M_s^* denote the optimal reduced rank regression coefficient matrix of rank s; then M_s^* can be computed by one of the following two methods:

Method 1

Evaluate the reduced-rank fitted values $Y_s^* = (\hat{Y}_s)$ by taking the partial sum to the s terms of the SVD of $\hat{Y}$, and then premultiplying it by X^+, the generalized inverse of X, we get

$$M_s^* = X^+ Y^* = X^T(\hat{Y})_s.$$

Method 2

We construct the $t \times s$ matrix V_s by taking the first s columns of the matrix V, i.e.

$$V_s = (v_1, \dots, v_s)$$

then evaluate M_s^* by

$$M_s^* = \hat{M} V_s V_s^T.$$

The corresponding reduced-rank fitted values of Y will be given by

$$Y_s^* = \hat{Y}_s = XM_s^*.$$

Confirming that both method 1 and method 2 are equivalent numerically. The residual sum of squares resulting from a rank s fit is then

$$||Y - \hat{Y}||^2 + \sigma_{s+1}^2 + \cdots + \sigma_t^2$$

i.e. then residual resulting from an unconstrained fit plus the contribution from the least significant singular values of $\hat{Y}$.

4. Determination of the Actual Rank of the Coefficient Matrix in the RRRM

Akaike's Information Criterion (AIC) may be extended to the case of order determination for the multivariate linear model. Here, we try to find a formula, using AIC, that can help us to determine the actual rank of the coefficient matrix in the reduced-rank regression model.

Let us consider the model:

$$Y = XM + E \text{ with } I^T X = 0^T \text{ and } I^T Y = 0^T$$

then the likelihood function can be given by

$$L(M) = (2\pi)^{-nt/2} exp\left\{\frac{-1}{2} tr(Y - XM)(Y - XM)^T\right\},$$

and the log-likelihood function is

$$\log L = I(M) = \frac{-nt}{2} log 2\pi - \frac{1}{2} tr\{(Y - XM)(Y - XM)^T\}, \qquad (4.1)$$

when M_s^* is the rank s maximum likelihood estimator of M, (4.1) becomes

$$I(M_s^*) = \frac{-nt}{2} log2\pi - \frac{1}{2} tr(Y - YM_s^*)(Y - YM_s^*)^T, \qquad (4.2)$$

letting $Y_s^* = XM_s^*$, (4.2) can be written as

$$\begin{aligned} I(M_s^*) &= \frac{-nt}{2} log2\pi - \frac{1}{2} tr\{(Y - Y_s^*)(Y - Y_s^*)^T\} \\ &= \frac{-nt}{2} log2\pi - \frac{1}{2}\{trYY^T - trY_s^* Y_s^{*T}\} \\ &= \frac{-nt}{2} log2\pi - \frac{1}{2} trYY^T + \frac{1}{2} trY_s^* Y_s^{*T} \\ &= \frac{-nt}{2} log2\pi - \frac{1}{2} trYY^T + \frac{1}{2}\sum_{i=1}^{s} \lambda_i, \end{aligned}$$

where $\{\lambda_i | \, i = 1, 2, \dots, t\}$are the eigenvalues of $\hat{Y}\hat{Y}^T$. The number of independently adjusted parameters within the model can be obtained as follows:

The matrix $M(r \times t)$ has s linearly independent columns. The remaining columns are linear combinations of these columns. Thus the total number of independently adjustable parameters is

$$rs + (t - s)s = s(r + t - s).$$

Therefore, the AIC for our model can be given by

$$AIC(s) = \frac{-2}{M_s^*} + 2s(r + t - s)$$

$$= -2\left\{\frac{-nt}{2} log2\pi - \frac{1}{2} trYY^T + \frac{1}{2}\sum_{i=1}^{s} \lambda_i\right\} + 2s(r + t - s),$$

we drop the constant and take the criterion to be

$$AIC(s) = -\sum_{i=1}^{s} \lambda_i + 2s(r + t - s), \qquad (4.3)$$

where s is the rank of the coefficient matrix.

5. Applications

In this section, we present two examples illustrating the application of the formula (4.3). A program was written to compute the singular values of the fitted values of Y, needed in the computation of the reduced rank regression coefficients matrix, and AIC values. In addition, it computes M_s^* and the residual sum of squares resulting from a rank s fit is also computed.

1. Example 1

Let

$$X = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \\ 4 & 5 & 3 & 2 & 1 \\ 2 & 3 & 1 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 \\ 4 & 5 & 6 & 3 & 1 \\ 2 & 4 & 5 & 4 & 4 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 6 & 9 & 12 \\ 4 & 3 & 2 & 1 \\ -1 & 3 & 7 & 11 \\ 8 & 6 & 4 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & -2 & 3 & -4 \\ 1 & -2 & 3 & -4 \\ 1 & -2 & 3 & -4 \\ 1 & -2 & 3 & -4 \\ 1 & -2 & 3 & -4 \\ 1 & -2 & 3 & -4 \\ 1 & -2 & 3 & -4 \end{bmatrix}$$

The second row in M is equal to three times the first row. Row four is equal to the difference of rows two and three, and row five is twice the third row. The only linearly independent rows are the first and third, and these from a basis for the five vectors. Thus the rank of M is 2.

The singular values of $\hat{Y}$ are:

$$[46.14312 \quad 24.54751 \quad 2.39542 \quad 1.45948]$$

and coefficient matrices M_s^* are

$$M_1^* = \begin{bmatrix} 2.18694 & 2.57638 & 2.94694 & 3.48654 \\ 6.44121 & 7.58823 & 8.67965 & 10.26894 \\ 1.64044 & 1.93256 & 2.21052 & 2.61528 \\ 4.83434 & 5.69522 & 6.51436 & 7.70718 \\ 3.48694 & 4.10788 & 4.69872 & 5.55907 \end{bmatrix}, \quad M_2^* = \begin{bmatrix} -1.25697 & 1.01572 & 3.76999 & 6.10432 \\ 4.03679 & 6.49863 & 9.25428 & 12.09658 \\ 3.54042 & 2.79357 & 1.75645 & 1.17106 \\ -0.93804 & 3.07937 & 7.89389 & 12.09487 \\ 6.71487 & 5.57067 & 3.92728 & 3.10546 \end{bmatrix}$$

$$M_3^* = \begin{bmatrix} -0.86736 & 0.22906 & 4.32260 & 5.97416 \\ 3.63478 & 7.31032 & 8.68408 & 12.23089 \\ 3.93039 & 2.00620 & 2.30955 & 1.04077 \\ -1.44702 & 4.10706 & 7.17197 & 12.26492 \\ 7.11446 & 4.76386 & 4.49404 & 2.97196 \end{bmatrix}, \quad M_4^* = \begin{bmatrix} -0.28343 & -0.25499 & 3.40292 & 6.74292 \\ 3.61149 & 7.32962 & 8.72077 & 12.20023 \\ 3.94275 & 1.99595 & 2.29008 & 1.05705 \\ -1.15959 & 3.86878 & 6.71926 & 12.64334 \\ 7.43854 & 4.49521 & 3.98363 & 3.39862 \end{bmatrix}$$

The values of s and its AIC values

s	AIC values
1	−2113.18719
2	−2703.7674*
3	−2701.50542
4	−2699.6355

Minimum AIC value (MAIC) is attained with $s = 2$. We conclude that the rank of M is 2.

2. Example 2

Let

$$X = \begin{bmatrix} 1 & 2 & -2 & 3 & 1 \\ -3 & 1.2 & 4 & 3 & 2 \\ -1 & -4 & 5 & 1.4 & 2 \\ 4 & 5 & 1.7 & 1.3 & -3 \\ 1 & 5 & -5 & -1.2 & 2 \\ 5 & 8 & 2 & 3 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 2 & 4 \\ 3 & 5 & 3 \\ 2 & 1 & 0 \\ 5 & 6 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & -1.2 & 4 \\ 1 & -1.2 & 4 \\ 1 & -1.2 & 4 \\ 1 & -1.2 & 4 \\ 1 & -1.2 & 4 \\ 1 & -1.2 & 4 \\ 1 & -1.2 & 4 \end{bmatrix}$$

The largest minor that can be formed from M is of order 3. There is at least one minor which is not equal to zero. Thus, the rank of M is 3.

The singular values of $\hat{Y}$ are:

$$[95.00338 \quad 16.02472 \quad 8.21635]$$

and coefficient matrices M_s^* are

$$M_1^* = \begin{bmatrix} 0.91999 & 1.20066 & 0.63667 \\ 2.52632 & 3.29705 & 1.74832 \\ 3.87028 & 5.05102 & 2.67840 \\ 1.12046 & 1.46229 & 0.77540 \\ 4.56855 & 5.96231 & 3.16163 \end{bmatrix}, \quad M_2^* = \begin{bmatrix} 1.07091 & 2.04556 & -1.17474 \\ 2.34097 & 2.25934 & 3.97311 \\ 3.85051 & 4.94035 & 2.91567 \\ 1.20311 & 1.92499 & -0.21661 \\ 4.54725 & 5.84305 & 3.41731 \end{bmatrix},$$

$$M_3^* = \begin{bmatrix} 0.09619 & 2.67886 & -0.96956 \\ 2.41063 & 2.21408 & 3.95780 \\ 3.31948 & 5.28537 & 3.03236 \\ 2.22623 & 1.26025 & -0.44143 \\ 4.93014 & 5.59428 & 3.33317 \end{bmatrix}$$

The values of s and its AIC values

s	AIC values
1	−9011.64137
2	−9258.43314
3	−9319.9416*

Here, MAIC occurs with $s = 3$. Hence, the rank of M is 3.

6. Conclusion

The numerical examples show that AIC is effective in determining the actual rank in the reduced rank regression model.

References

[1] Akaike, H. (1973), 'Information theory as an extension of the maximum likelihood theory', in *'Second International Symposium on Information Theory'*, pp. 267-281.

[2] Alvarez, W. and Griffin, V.J. (2016), 'Estimation procedure for reduced rank regression', *Statistics, Optimization & Information Computing* 4(2), 107-117.

[3] Davies, P. and Tso, M. K.-S. (1982), 'Procedures for reduced-rank regression', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 31(3), 244-255.

[4] Good, I. J. (1969), 'Some applications of the singular decomposition of a matrix', Technometrics 11(4), 823-831.

[5] Green, P. E. (2014), *Mathematical Tools for Applied Multivariate Analysis*, Academic Press.

[6] Izenman, A. J. (1975), 'Reduced-rank regression for the multivariate linear model', *Journal of multivariate analysis* 5(2), 248-264.

[7] Karagrigorious, A., Matheou, K. and Vonta, I. (2011), 'On Asymptotic properties of AIC Variants with Applications'. *American open Journal of Statistics*, 105-109.

[8] Madhi, S. A. (1981), *Reduced-rank Regression Model Estimation and Rank Determination*, University of Manchester.

[9] Mandel, J. (1982), 'Use of the singular value decomposition in regression analysis', *The American Statistician* 36(1), 15-24.

[10] Rao, C.R. (1973), *Linear Statistical Inference and Its Applications*, Wiley New York.

[11] Snipes, M. and Taylor, D.C. (2014), Model selection and Akaike information criterion.