



# Estimation Procedure for Reduced Rank Regression, PLSSVD

Willin Álvarez \*, Victor Griffin

*Departamento de Matemáticas, Facultad de Ciencia y Tecnología. Universidad de Carabobo, Venezuela.*

(Received: 23 July 2015; Accepted: 30 January 2016)

**Abstract** This paper presents a procedure for coefficient estimation in a multivariate regression model of reduced rank in the presence of multicollinearity. The procedure permits the prediction of the dependent variables taking advantage of both Partial Least Squares (PLS) and Singular Value Decomposition (SVD) methods, which is denoted by PLSSVD. Global variability indices and prediction error sums are used to compare the results obtained by classical regression with reduced rank (OLSSVD) and the PLSSVD procedure when applied to examples with different grades of multicollinearity (severe, moderate and low). In addition, simulations to compare the methods were performed with different sample sizes under four scenarios. The new PLSSVD method is shown to be more effective when the multicollinearity is severe and especially for small sample sizes.

**Keywords** Reduced Rank, Multivariate Regression, Partial Least Squares, Singular Value Decomposition, PLSSVD, Multicollinearity

**AMS 2010 subject classifications** 62H25,65F22

**DOI:** 10.19139/soic.v4i2.146

## 1. Introduction

A new procedure for regression coefficient estimation under reduced rank conditions is developed using partial least squares (PLS) and compared with classical methods which use ordinary least squares (OLS). Hoskuldsson [11] asks the question, “What situations are typical of those where PLS methods can be expected to be good for modelling purposes?” and gives the reply “They are the ones where there are many variables but not necessarily many samples or observations. This is a common situation in many laboratories. Typically it may take a lot of time to get a new sample, but each sample may give a large amount of information (variables).” He also replies to another important question, “Why can one expect PLS regression methods to perform better than multiple linear regression, ridge regression and other well known regression techniques?”. The answer involves the increased stability of predictors derived from PLS methods. It turns out that the essential criteria for the predictability of models is to keep the number of variables as low as possible. In PLS, components are selected that give ‘maximal’ reduction in the covariance  $X^T Y$  of the data. In that sense PLS will give the minimum number of necessary variables.

Xu et al [30] also note that principal component regression (PCR) is similar to PLS in constructing new variables (components): each new component is a linear combination of  $x_1, x_2, \dots, x_m$ , where  $x_i$  is the  $i$ th column of data matrix  $X$ . Both of them shrink the estimated coefficient vector by leaving some of the components out of the model. The major difference between PCR and PLS is that PCR’s components are determined only by  $X$ , whereas

---

\*Correspondence to: Willin Álvarez (Email: walvarez@uc.edu.ve/willin.el.aprendiz@gmail.com. Departamento de Matemáticas, Universidad de Carabobo, Venezuela.

PLS's components are determined for both  $X$  and  $Y$ . Thus PLS takes into account all of the available data. In this paper these observations have been taken into account in that we have restricted the discussion to OLS and PLS regression methods, singular value decomposition, use only small real life samples and include simulations involving small sample sizes.

A summary of recent procedures for coefficient estimation under reduced rank regression (RRR) was given in the article of Heinen and Rengifo [9]. We present for our purposes in this introduction only that of Davies and Tso[5], a pioneer work which relates reduced rank regression with redundancy analysis and an important modification of their work by Ter Braak and Looman [21] which involves the biplots theory of Gabriel [7]. It provides the classical regression procedure with reduced rank referred to here as OLSSVD as it combines OLS with Singular Value Decomposition (SVD) methods. In section 2, a new procedure PLSSVD is presented based on PLS regression and SVD. Section 3 uses global variability indices and prediction error sums to compare the OLSSVD and PLSSVD procedures when applied to examples taken from the literature having different grades of multicollinearity: severe, moderate and low. Also a brief simulation study uses the mean square error to compare the two procedures using simulated data. The conclusion of the study, described in the final section, is that the new PLSSVD procedure is more stable than OLSSVD and is most effective when multicollinearity is severe.

### 1.1. The Basic Model

Multivariate regression with reduced Rank (RRR), a special case of the classic multivariate regression model derived from the works of Anderson [4], Izenman [12] and Davies and Tso[5], is based on the model:

$$Y = X\Theta + E \quad (1)$$

in which  $Y$  is a  $(n \times q)$  matrix,  $X$  is a  $(n \times p)$  matrix,  $\Theta$  is a  $(p \times q)$  matrix of parameters or regression coefficients and  $E$  is the  $(n \times q)$  matrix of errors or random perturbations.

### 1.2. Reduced Rank Regression Procedure of Davies and Tso (1982)

Davies and Tso (1982) [5] base their procedure on the reduced rank regression model  $Y = X\Theta + E$  assuming  $r = \text{rank}(\Theta)$ , with  $r \leq p$ . The essential steps of their procedure are:

1. Form data matrices  $X$  ( $n \times p$ ) and  $Y$  ( $n \times q$ ) containing values for regressor and response variable on  $n$  samples. Scale each response variable and subtract sample means from each variable.
2. Determine the unconstrained least-squares estimate  $\hat{\Theta}$  by  $\hat{\Theta} = X^+Y$  where  $X^+$  is the Moore-Penrose generalized inverse of  $X$  (see, for example, Graybill, 1969 [8], Theorem 7.4.1 or Davies and Tso, 1982, p.245 [5]). When  $X$  has full column rank this is equivalent to use the standard formula:  $\hat{\Theta}_{OLS} = (X'X)^{-1}X'Y$
3. Form the corresponding matrix of fitted values  $\hat{Y} = X\hat{\Theta}$
4. Perform Singular Value Decomposition of  $\hat{Y}$ , that is,  $\hat{Y} = \sum_{i=1}^{\tau} \lambda_i u_i v_i' = UDV'$  (where  $U$  is a matrix of order  $n \times q$  whose columns are the eigenvectors associated with the nonzero eigenvalues of the matrix  $\hat{Y}\hat{Y}'$ ;  $D$  is a diagonal matrix of order  $q$  which contains the square roots of the nonzero eigenvalues of  $\hat{Y}\hat{Y}'$  or  $\hat{Y}'\hat{Y}$  and  $V$  is an orthogonal matrix of order  $q$  whose columns are the eigenvectors of  $\hat{Y}'\hat{Y}$ . The matrices  $U$  and  $V$  satisfy the properties  $U'U = I$  and  $V'V = I$ , and  $\tau$  is rank of  $(\hat{Y})$ ). Evaluate the matrix  $P_s = \sum_{i=1}^s \lambda_i v_i v_i'$  ( $P_s$  is projection matrix de rank  $s$ , with  $s < \tau$ )
5. Evaluate the reduced-rank fitted values  $Y_s^* = \hat{Y}_s = X\hat{\Theta}P_s$
6. Compute the reduced rank matrix of coefficients given by  $\Theta^* = \hat{\Theta}P_s$

This procedure by Davies and Tso has been the object of much subsequent research. For example, Anderson [3] compared the asymptotic distribution of the coefficient matrix with complete rank with that of reduced rank. The estimation model with reduced rank was shown to be superior in general. The same author in 2002 studied the influence of the rank in both models finding that the model with reduced rank is most efficient when the rank is small. In the next section a different approach based on Ter Braak and Looman (1994) [21] which uses a biplot of reduced rank regression is presented. The need to factor the regression coefficient matrix using the techniques of Gabriel (1971) [7] obliged them to develop reduced rank regression using OLS and SVD.

### 1.3. The Reduced Rank Regression Procedure of Ter Braak and Looman (OLSSVD)

Just as Davies and Tso did in 1982[5], Ter Braak and Looman in 1994 [21] based their reduced rank regression model on the classical linear regression model  $Y = X\Theta + E$ , under the assumption that the matrix  $\Theta$  of order  $(p \times q)$  and rank  $r$  does not have complete column rank, that is,  $r < q$ . It is thus possible to factor:

$$\Theta = AB' \quad (2)$$

where  $A$  and  $B$  are matrices of order  $(p \times r)$  and  $(q \times r)$  respectively. Inserting (2) in (1) we have:

$$Y = XAB' + E \quad (3)$$

This model can be written as in factorial analysis (see Mardia et.al, 1979 [15]) as:

$$Y = GB' + E \quad (4)$$

where  $G = XA$  is a matrix of order  $(n \times r)$  that contains the scores of the factors and  $B$  the matrix that contains the factorial weights (Ter Braak and Looman, 1994 [21]).

As in Davies and Tso [5], we have assumed errors in the matrix  $E$  to be uncorrelated for different samples (i.e. across rows), but correlation between the errors in responses measured on the same sample may need to be considered. If the errors are distributed normally with unknown covariance matrix, it has been shown that canonical correlation analysis is applicable (Tso, 1981 [23]). Given an independent estimate covariance matrix  $\Sigma$ , then maximum-likelihood estimation of  $\Theta$  requires only a slight modification of the given least-squares procedure.

Let  $\Sigma$  have eigenanalysis  $\Sigma = VD^2V' = \Gamma\Gamma'$ , say, where  $\Gamma = VD^{-1}$  and  $D$  is understood to have positive elements along its main diagonal. Ter Braak and Looman (1994)[21] take this into consideration and replace the criteria for least squares by:

$$\text{tr}(Y - X\Theta)\Sigma^{-1}(Y - X\Theta)' = \text{tr}(Y - X\Theta)\Gamma\Gamma'(Y - X\Theta)' = \|(Y - X\Theta)\Gamma\|^2. \quad (5)$$

The consequence of this observation is the following. If in (5),  $\Gamma$  is a conveniently chosen symmetric positive definite matrix, the principal component analysis (PCA) and canonical correlation analysis (CCA) can be derived from RRR as shown by Izenman [12]. The estimation in regression with reduced rank takes into account the observation when (3) is introduced in (5) to obtain the orthogonal sum:

$$\|(Y - XAB')\Gamma\|^2 = \|(Y - \hat{Y})\Gamma\|^2 + \|(\hat{Y} - X\Theta)\Gamma\|^2 \quad (6)$$

where  $\hat{Y}$  is the matrix of values adjusted by separate multiple regressions (equivalent to an adjustment by multivariate regression for each of the  $p$  independent variables) and  $\Gamma$  a symmetric positive definite matrix selected conveniently according to the context. Considering (4), we obtain:

$$\|(Y - XAB')\Gamma\|^2 = \|(Y - \hat{Y})\Gamma\|^2 + \|(\hat{Y} - GB')\Gamma\|^2. \quad (7)$$

According to the decomposition (6), in order to do the estimation in reduced rank regression it is necessary to find the minimum of  $\|(Y - XAB')\Gamma\|^2$ . The solution by least squares is thus obtained by minimizing the right hand term of (7) using singular value decomposition (SVD), (see Eckart and Young [6]). In other words:

$$\hat{Y}\Gamma = UDV' \quad (8)$$

where  $U$  and  $V$  are orthogonal matrices of order  $(n \times t)$  and  $(q \times t)$  respectively, where  $t = \min(n, q)$ , that contain the singular values of  $\hat{Y}\Gamma(\hat{Y}\Gamma)'$  and  $(\hat{Y}\Gamma)'(\hat{Y}\Gamma)$  respectively, and  $D$  is a diagonal matrix of order  $q$  formed by the eigenvalues of these matrices in non-increasing order ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_t \geq 0$ ). Remember also that  $\Gamma$  is a positive definite matrix.

The minimum of (7), when the rank of  $Y$  is  $r < q$ , is the sum of the eigenvalues obtained by singular value decomposition:  $\lambda_{r+1}^2 + \dots + \lambda_t^2$  using

$$\hat{G} = U_r \text{ and } \hat{B}_{SVD} = \Gamma^{-1}[VD]_r \quad (9)$$

where the matrices  $U_r$  and  $[VD]_r$  contain only the first  $r$  columns of the matrices  $U$  and  $VD$  of (8), respectively. By (4),  $G = XA$  and by (9),  $\hat{G} = U_r$ , so that it is possible to estimate  $A$  by ordinary least squares using the regression  $\hat{G} = XA$  which gives:

$$\hat{A}_{OLS} = (X'X)^{-1}X'\hat{G}. \quad (10)$$

The matrix of regression coefficients in the model (3) is thus:

$$\hat{\Theta}_{OLS-SVD} = \hat{A}_{OLS}\hat{B}'_{SVD}. \quad (11)$$

Observe that the matrix of regression coefficients is thus obtained by a procedure that combines ordinary least squares estimation and singular value decomposition. The steps of this OLSSVD procedure for the RRR model  $Y = X\Theta + E$  can be summarized as follows, Ter Braak and Looman (1994) [21]:

1. Centralize or standardize the matrices  $X$  ( $n \times p$ ) and  $Y$  ( $n \times q$ ).
2. Express the multivariate regression model of (1) as:  $Y = XAB' + E$ , where  $\Theta = AB'$ .
3. Denote  $G = XA$  and express  $Y = GB' + E$ .
4. Minimize  $\|(\hat{Y} - GB')\Gamma\|$  by singular value decomposition  $\hat{Y} = UDV'$ .
5. Assign  $\hat{G} = U_r$  and  $\hat{B}_{SVD} = \Gamma^{-1}[VD]_r$ .
6. As  $G = XA$  and  $\hat{G} = U_r$  estimate  $A$  by ordinary least squares by  $\hat{A}_{OLS}^{-1} = (X'X)^{-1}X'\hat{G}$
7.  $\hat{\Theta}_{OLS-SVD} = \hat{A}_{OLS}\hat{B}'_{SVD}$ .
8. Perform the estimations using  $\hat{Y}_{OLS-SVD} = X\hat{\Theta}_{OLS-SVD}$ .

This algorithm is very important since by factorizing the matrix of regression coefficients with reduced rank as  $\hat{\Theta}_{OLS-SVD} = \hat{A}_{OLS}\hat{B}'_{SVD}$  it is possible to interpret graphically by means of a biplot following the ideas of Ter Braak and Looman [21] (see for example Alvarez and Cardenas [1]). Ter Braak and Looman also note that in this algorithm, if  $\Gamma = I$ , the solutions can be obtained by a principal components analysis of the values  $\hat{Y}$ . Reduced-rank regression inherits from multiple regression the property that it is invariant for linear rescaling of the regressors, but from principal components analysis we know that it is not invariant for rescaling of the response variables (Jolliffe [14]). There are three methods to overcome this latter problem:

1. Standardize each response variable to zero mean and unit variance and use  $\Gamma = I$ .
2. Weight responses variables with the inverse of the error covariance matrix  $\hat{\Sigma}_p$ , estimated by multivariate regression, *i.e.*, use  $\hat{\Sigma}_p^{-1/2}$ .
3. Weight each response variable with the inverse of its error variance  $\hat{\Sigma}_p(\hat{\sigma}_k^2, k = 1, 2, \dots, q)$ , estimated by multiple regression, *i.e.*, use  $\Gamma = \text{diag}(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_q^2)$ .

Method (1) leads to redundancy analysis (Van den Wollenberg, 1977 [25]), whereas method (2) leads to canonical correlation analysis as was first noted by Velu et. al. (1986)[27]. Method (3) was proposed by Ter Braak (1990)[22] as an intermediate between redundancy analysis and canonical correlation.

## 2. Estimation in Multivariate Reduced Rank Regression using Partial Least Squares and Singular Value Decomposition

It was shown above that the coefficient matrix in reduced rank regression is given by  $\hat{\Theta}_{OLS-SVD} = \hat{A}_{OLS}\hat{B}'_{SVD}$ .  $\hat{A}_{OLS}$  and  $\hat{B}_{SVD}$  can be expanded using (9) and (10) to obtain

$$\hat{\Theta}_{OLS-SVD} = (X'X)^{-1}XU_rD_rV_r'\Gamma^{-1}. \quad (12)$$

When multicollinearity exists, if it is perfect then it is mathematically impossible to estimate  $\hat{\Theta}$  since  $(X'X)^{-1}$  does not exist. If the multicollinearity is not perfect but strong, the estimations obtained will have high standard errors which result in values far from reality. Remember that, in the presence of multicollinearity, the variance, the covariance and the absolute values of the regression coefficients are very large.

## 2.1. PLS Regression

Partial Least Squares (PLS) methods were introduced by Wold H. [28] and PLS regression by Wold S., Martens H. and Wold H. (1983) [29]. Since then many algorithms and applications have been proposed, chiefly in the fields of chemometrics, medicine, biology, petroleum, agriculture and continuous industrial processes. PLS regression (PLSR) is a technique that generalizes and combines the characteristics of principal component analysis and multivariate regression. It is particularly useful when it is necessary to predict a set of dependent variables in terms of a set of independent variables.

As noted previously, the classical multivariate regression procedures have serious stability problems in the presence of severe or strong multicollinearity. PLS regression is a good alternative in such circumstances since it uses basic interactive concepts to adjust bilinear models in several blocks of variables giving a solution to multivariate statistics in the presence of many collinear variables in the data and a modest number of observations. The stability problems in identification and convergence as described by Martens et al in [16] can be thus avoided.

## 2.2. The PLS Regression Model

Many estimation algorithms have been developed for this type of regression as noted for example in Wold et al [29]. One of the most simple algorithms is that of Hoskuldsson [11] described in high level as follows in terms of the weights of the independent variables  $w_h$ , the weights of the dependent variables  $c_h$ , the score vectors of the independent variables  $t_h$ , the score vectors of the dependent variables  $u_h$  and also the weight vectors  $p_h$  and  $c_h$ . These vectors resulting from the algorithm give the PLS estimation. We base the calculations on the modified PLSR algorithm described in Tenenhaus (1998) [20] and emphasize that our objective is centered on calculating the matrix of PLS regression coefficients. Before showing the modified algorithm, we show the NIPALS (Nonlinear iterative Partial Least Squares) algorithm which is the basis for PLSR. It begins by standardizing the sets  $X$  and  $Y$ .

1. Stage 1:  $X_0 = X ; Y_0 = Y$ .
2. Stage 2: for  $h = 1, 2, \dots, a$  :
3. Stage 2.1:  $u_h =$  the first column of  $Y_{h-1}$
4. Stage 2.2: Repeat until convergence of  $w_h$ :
5. Stage 2.2.1:  $w_h = X'_{h-1}u_h / (w'_h u_h)$ .
6. Stage 2.2.2: Scale  $w_h$  to be of length one.
7. Stage 2.2.3:  $t_h = X_{h-1}w_h / (w'_h w_h)$ .
8. Stage 2.2.4:  $c_h = Y'_{h-1}t_h / (t'_h t_h)$ .
9. Stage 2.2.5:  $u_h = Y_{h-1}c_h / (c'_h c_h)$ .
10. Stage 2.3:  $p_h = X'_{h-1}t_h / (t'_h t_h)$ .
11. Stage 2.4:  $X_h = X_{h-1} - t_h p'_h$
12. Stage 2.5:  $Y_h = Y_{h-1} - t_h c'_h$

In Tenenhaus (1998)[20], it is shown that the weight vectors,  $w_h$  and  $c_h$ , and the latent variables  $t_h$  and  $u_h$  are obtained as solutions of the following equations associated with the highest common eigenvalue  $\lambda_h$

$$\begin{aligned} \left(\frac{1}{n-1}X'_{h-1}Y_{h-1}\right)\left(\frac{1}{n-1}Y'_{h-1}X_{h-1}\right)w_h &= \lambda_h w_h \\ \left(\frac{1}{n-1}X_{h-1}X'_{h-1}\right)\left(\frac{1}{n-1}Y_{h-1}Y'_{h-1}\right)t_h &= \lambda_h t_h \\ \left(\frac{1}{n-1}Y'_{h-1}X_{h-1}\right)\left(\frac{1}{n-1}X'_{h-1}Y_{h-1}\right)c_h &= \lambda_h c_h \\ \left(\frac{1}{n-1}Y_{h-1}Y'_{h-1}\right)\left(\frac{1}{n-1}X_{h-1}X'_{h-1}\right)u_h &= \lambda_h u_h \end{aligned}$$

The PLS regression principle can be summarized as obtaining the components  $t_h = X_{h-1}w_h / (w'_h w_h)$  and  $u_h = Y_{h-1}c_h / (c'_h c_h)$  of the  $h$ -th step beginning with the first stage. One of the interesting properties of the previous

algorithm is that the components  $t_h$  are orthogonal which implies that the matrices  $X_h$  and  $Y_h$  are the respective residuals of the regressions of  $X$  and  $Y$  on the components  $t_1, t_2, \dots, t_h$ . The following modifications give origin to the PLS regression algorithm:

$$\begin{aligned}w_h &= X'_{h-1} Y c_h / \|X'_{h-1} Y c_h\| \\c_h &= Y' t_h / t'_h t_h \\p_h &= X' t_h / t'_h t_h\end{aligned}$$

The calculation of  $Y_h$  residues is unnecessary for the determination of PLS components. Rather, these residues are used for choosing the number of components. Let us therefore look at the modified version of the PLS regression algorithm leading to the same vector  $w_h, t_h, c_h$  and  $p_h$  as the classical algorithm. The components  $u_h = Y_{h-1} c_h / c'_h c_h$  are replaced by the components  $\tilde{u}_h = Y c_h / c'_h c_h$ . The modified PLS algorithms is as follows:

1. Stage 1:  $X_0 = X$  ;  $Y_0 = Y$ .
2. Stage 2: for  $h = 1, 2, \dots, a$  :
3. Stage 2.1:  $\tilde{u}_h$  = the first column of  $Y$
4. Stage 2.2: Repeat until convergence of  $w_h$ :
5. Stage 2.2.1:  $w_h = X'_{h-1} \tilde{u}_h / (\tilde{u}'_h \tilde{u}_h)$ .
6. Stage 2.2.2: Scale  $w_h$  to be of length one.
7. Stage 2.2.3:  $t_h = X_{h-1} w_h / (w'_h w_h)$ .
8. Stage 2.2.4:  $c_h = Y' t_h / (t'_h t_h)$ .
9. Stage 2.2.5:  $\tilde{u}_h = Y c_h / (c'_h c_h)$ .
10. Stage 2.3:  $p_h = X' t_h / (t'_h t_h)$ .
11. Stage 2.4:  $X_h = X_{h-1} - t_h p'_h$ .
12. Stage 2.5:  $Y_h = Y_{h-1} - t_h c'_h$ .

The latent variables  $t_h$  and  $u_h$  are obtained as solutions of the following equations associated with the greatest common eigenvalue  $\lambda_h$

$$\begin{aligned}\left(\frac{1}{n-1} X'_{h-1} Y\right) \left(\frac{1}{n-1} Y' X_{h-1}\right) w_h &= \lambda_h w_h \\ \left(\frac{1}{n-1} X_{h-1} X'_{h-1}\right) \left(\frac{1}{n-1} Y Y'\right) t_h &= \lambda_h t_h \\ \left(\frac{1}{n-1} Y' X_{h-1}\right) \left(\frac{1}{n-1} X'_{h-1} Y\right) c_h &= \lambda_h c_h \\ \left(\frac{1}{n-1} Y Y'\right) \left(\frac{1}{n-1} X_{h-1} X'_{h-1}\right) \tilde{u}_h &= \lambda_h \tilde{u}_h.\end{aligned}$$

The solution of each stage permits the construction of the following matrices

$$\begin{aligned}T_h &= [t_1, \dots, t_h] \\ W_h &= [w_1, \dots, w_h] \\ P_h &= [p_1, \dots, p_h] \\ \tilde{W}_h &= [\tilde{w}_1, \dots, \tilde{w}_h]\end{aligned}$$

To obtain the regression coefficient matrix and the regression equation we use the following mathematical expression relating  $T_h$  and  $\tilde{W}_h$ :

$$\begin{aligned}T_h &= X \tilde{W}_h \\ \tilde{W}_h &= W_h (P'_h W_h)^{-1}.\end{aligned}$$

To express the regression of  $Y$  on the PLS components  $t_1, \dots, t_H$  on variables  $X$

$$Y = t_1 c'_1 + t_2 c'_2 + \dots + t_H c'_H + Y_H \quad (13)$$

we can define  $C_H$  as

$$C_H = [c_1, c_2, \dots, c_H]$$

and the regression equation becomes

$$Y = T_H C_H' + Y_H$$

which in turn can be expressed as

$$Y = T_H C_H' + Y_H = X \tilde{W}_H C_H' + Y_H. \quad (14)$$

Finally, we can write the equation for the regression coefficient matrix  $\Theta_{PLS}$  as

$$\Theta_{PLS} = \tilde{W}_H C_H'. \quad (15)$$

The matrix of regression coefficients of (15) is determined from the components  $t_h$  which are mutually orthogonal thus assuring that the multicollinearity does not effect this matrix of parameters contrasting with the estimation of regression coefficients using ordinary least squares. The idea to take advantage of this fact will now be presented.

### 2.3. Reduced Rank Regression Procedure PLSSVD

As a contribution to reduced rank regression in the presence of multicollinearity the estimation procedure denoted by PLSSVD is now presented. It is an integration of the PLS method and the procedure of Ter Braak and Looman (1994) [21]. The integrated procedure assures that the estimation of the regression coefficients are affected as little as possible by problems of multicollinearity.

The procedure consists of the following steps:

1. Begin with the multivariate regression model (with the variables of the two sets  $X$  and  $Y$  centered or standardized):

$$Y = X\Theta + E.$$

2. Factor the coefficient matrix as  $\Theta = AB'$  so that the multivariate reduced rank regression model becomes

$$Y = XAB' + E.$$

As in the equation (3), it is then possible to express the model as  $Y = GB' + E$ , where  $G = XA$ .

3. Minimize  $\|(Y - XAB')\Gamma\|^2 = \|(Y - \hat{Y})\Gamma\|^2 + \|(\hat{Y} - GB')\Gamma\|^2$  using as an initial estimation for  $Y$  that obtained by SVD as was done in the equation (8) obtaining

$$\hat{Y}\Gamma = UDV'$$

where  $U$  and  $V$  are orthogonal matrices of order  $(n \times t)$  and  $(q \times t)$  with  $t = \min(n, q)$ , that contain the eigenvectors of  $(\hat{Y}\Gamma)(\hat{Y}\Gamma)'$  and  $(\hat{Y}\Gamma)'(\hat{Y}\Gamma)$  respectively;  $D$  is a diagonal matrix of order  $q$  containing the eigenvalues of these matrices in non-increasing order ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_t \geq 0$ ); and  $\Gamma$  is a positive definite matrix.

4. Since  $\hat{G} = U = XA$ , it is possible to make a comparison with the regression model with  $U$  representing the matrix of dependent variables,  $X$  the matrix of independent variables and  $A$  the matrix of regression coefficients. Suppose that the matrix of independent variables presents high multicollinearity so that it is not convenient to do the estimation using the OLS model and that the application of the PLS estimation for  $A$  is proposed. According to the described PLS method, the estimation given in the equation (15) is

$$\hat{A}_{PLS} = \tilde{W}C'.$$

5. The matrix of regression coefficients is obtained by combining the PLS estimation of  $A$  with the estimation of  $B$  given by SVD so that the reduced rank regression coefficient matrix given by PLSSVD is

$$\hat{\Theta}_{PLS-SVD} = \hat{A}_{PLS}\hat{B}'_{SVD}.$$

6. Finally, the estimations of  $Y$  are given by

$$\hat{Y}_{PLS-SVD} = X\hat{\Theta}_{PLS-SVD} = X\hat{A}_{PLS}\hat{B}'_{SVD}.$$

Note that the calculation of  $(X'X)^{-1}$  is not present in the PLSSVD procedure so that multicollinearity does not affect the estimation of the coefficient matrix.

### 3. Comparison of the Procedures

In this section, comparisons are made between the reduced rank regression estimation procedure of Ter Braak and Looman [21], called hereafter OLSSVD, and the estimation procedure proposed in this article called PLSSVD. The comparison is done using data already studied by other authors. Following the classification by condition index given in Montgomery and Peck [17], three scenarios are presented according to the grade of multicollinearity found in the data: severe multicollinearity, moderate to strong multicollinearity and light multicollinearity (see Table 1). All the analysis is performed using standardized data and weight matrices  $\Gamma = I$  and  $\Gamma = R_{yy}^{1/2}$ .

Table 1. Condition Indices for the three data sets

| Data Set                   | Condition Index    | Grade of Multicollinearity |
|----------------------------|--------------------|----------------------------|
| Hervé Abdi (2003)[10]      | $5.2298 * 10^{16}$ | Severe                     |
| Alvin Rencher (2002)[19]   | 224.1492           | Moderate to Strong         |
| Valencia et. al.(2004)[24] | 59.5997            | Light                      |

The estimations are evaluated using indices which measure the global variability of the residual matrix as explained by Peña (2002,p.77) [18]. The first index is the trace of the covariance matrix of the residual matrix ( $T$ ); however, realizing that this measure does not take into account the dependence structure among the variables, a second index called generalized variance, (GV), defined as the determinant of the covariance matrix of the residuals is also used. A third index, called generalized standard deviation (GSD), is obtained by taking the square root of the generalized variance. A fourth and last index called the effective variance (EV) is defined as the generalized variance elevated to the  $1/q$ . The mathematical expressions of each of these index measures are found in Table 2.

Table 2. Global Variability Measures for the covariance matrix of the residuals

| Measure of global variability       | INDEX                  |
|-------------------------------------|------------------------|
| Traza ( $T$ )                       | $T = tr(S)$            |
| Generalized Variance (GV)           | $GV = det(S)$          |
| Generalized Standard Deviation(GSD) | $GSD = [det(S)]^{1/2}$ |
| Effective Variance (EV)             | $EV = [det(S)]^{1/q}$  |

$S = cov(\hat{E}); \hat{E}$ =matrix of residuals generated by the estimation procedure

#### 3.1. Data with Severe Multicollinearity

The first set of variables, taken from Hervé Abdi (2003, p.3 [10]), consists of measurements of five types of wine considering price ( $x_1$ ), quantity of sugar ( $x_2$ ), alcohol ( $x_3$ ) and acidity ( $x_4$ ) as independent variables and classification by experts indicating the pleasure received in drinking them and categorizing them in 3 groups ( $y_1$ :Pleasure,  $y_2$ : Meat and  $y_3$ : Dessert) according to their use in parties or to accompany meat or desserts. Table 1 shows that this data presents severe multicollinearity.

**3.2. Data with Moderate to Strong Multicollinearity**

The second set of data is given by 19 observations of an experiment involving a chemical reaction (Box and Youle in Rencher Alvin, 2002 p. 340 [19]). As shown in Table 1, this data presents moderate to strong multicollinearity. The variables are identified as follows: Percentage of the initial material preserved or unchanged ( $x_1$ ), Percentage converted to the desired product ( $x_2$ ), Percentage of flawed product ( $x_3$ ), Temperature ( $y_1$ ), Concentration ( $y_2$ ), Time ( $y_3$ ).

**3.3. Data with light Multicollinearity**

The third and last set of data (Table 1) taken from Valencia et. al. (2004, p.73 [24]), consists of three dependent variables ( $y_1, y_2, y_3$ ), three independent variables ( $x_1, x_2, x_3$ ) and seven observations. The condition index of this data, (see Table 1), shows that the independent variables do not present serious problems of multicollinearity so that the data set is said to have light multicollinearity.

**3.4. Comparison of the PLSSVD and OLSSVD procedures using the weight matrices  $\Gamma = I$  and  $\Gamma = R_{yy}^{-1/2}$**

Table 3 shows the correlations of the independent variables of the three sets of data. In the data with severe multicollinearity, the variable price ( $x_1$ ) has a strong negative correlation with variables alcohol and acidity ( $x_3$  and  $x_4$ ) and a positive correlation of 0.32 with the quantity of sugar ( $x_2$ ). The quantity of sugar has no relation with the other variables. Regarding the data with moderate multicollinearity,  $x_1$  and  $x_3$  have the highest correlation among the independent variables followed by a moderate correlation between  $x_1$  and  $x_2$  and lastly  $x_2$  and  $x_3$  present a very low correlation. The data with light multicollinearity presents independent variables with high correlations.

Table 3. Correlation of the Independent variables of the three data sets

| Severe M. |       |       |       | Moderate to Stronge M. |       |       | Light M. |       |       |
|-----------|-------|-------|-------|------------------------|-------|-------|----------|-------|-------|
| $x_1$     | $x_2$ | $x_3$ | $x_4$ | $x_1$                  | $x_2$ | $x_3$ | $x_1$    | $x_2$ | $x_3$ |
| 1         | 0.32  | -0.9  | -0.95 | 1                      | -0.61 | -0.88 | 1        | 0.86  | -0.95 |
| 0.32      | 1     | 0     | 0     | -0.61                  | 1     | 0.18  | 0.86     | 1     | -0.84 |
| -0.9      | 0     | 1     | 0.95  | -0.88                  | 0.18  | 1     | -0.95    | -0.84 | 1     |
| -0.95     | 0     | 0.95  | 1     |                        |       |       |          |       |       |

M.= Multicollinearity

Table 4 gives the global variability measures already described for the matrix of residuals obtained by each of the procedures and for each of the two weight matrices. These measures show that both procedures give exactly the same results for light and moderate to strong multicollinearity whereas for severe multicollinearity the PLSSVD procedure provides residuals with much less global variability than the OLSSVD procedure for both weight matrices. For both procedures the variability indices are greater when  $\Gamma = R_{yy}^{-1/2}$  than when  $\Gamma = I$ .

Table 4. Global Variability Indices of the residuals for the three data sets

| Method | Severe M. |       |      |      | Moderate to Strong M. |       |      |      | Light M. |      |      |      | $\Gamma$      |
|--------|-----------|-------|------|------|-----------------------|-------|------|------|----------|------|------|------|---------------|
|        | $T$       | $GV$  | $ED$ | $EV$ | $T$                   | $GV$  | $ED$ | $EV$ | $T$      | $GV$ | $ED$ | $EV$ |               |
| PLSSVD | 0.31      | 0     | 0    | 0    | 39.1                  | 46.5  | 6.8  | 3.6  | 5.1      | 0.06 | 0.24 | 0.39 | $I$           |
| OLSSVD | 42.82     | 0.63  | 0.79 | 0.86 | 39.1                  | 46.5  | 6.8  | 3.6  | 5.1      | 0.06 | 0.24 | 0.39 | $I$           |
| PLSSVD | 5.32      | 0.603 | 0.78 | 0.84 | 39.7                  | 78.20 | 8.84 | 4.28 | 17       | 12.9 | 3.59 | 2.34 | $R_{yy}^{-1}$ |
| OLSSVD | 67.02     | 8.46  | 2.91 | 2.04 | 39.7                  | 78.20 | 8.84 | 4.28 | 17       | 12.9 | 3.59 | 2.34 | $R_{yy}^{-1}$ |

M.= Multicollinearity

### 3.5. Comparison of the predictive power of the two procedures for the given data

Another important aspect to consider is the predictive power of the two procedures. In order to compare the predictive power, the sum of squares of the prediction errors (PRESS) was calculated for each dependent variable and as the variables were standardized, the sum of the Press (Total Press) was also calculated to provide a single measure for the global predictive power of each procedure for each data set following the ideas of Valencia et.al. (2004). These indices were calculated for each of the three situations or data sets. The PLSSVD method has higher predictive power than the OLSSVD method when the data presents severe multicollinearity as evidenced by the fact that for  $\Gamma = I$ , Total PRESS PLSSVD is 82.59 against Total PRESS OLSSVD equal to 253.18. For the other two grades of multicollinearity the OLSSVD procedure performs better. Similar results are seen for  $\Gamma = R_{yy}^{-1/2}$ .

### 3.6. Comparison of the predictive power of the two procedures for simulated data

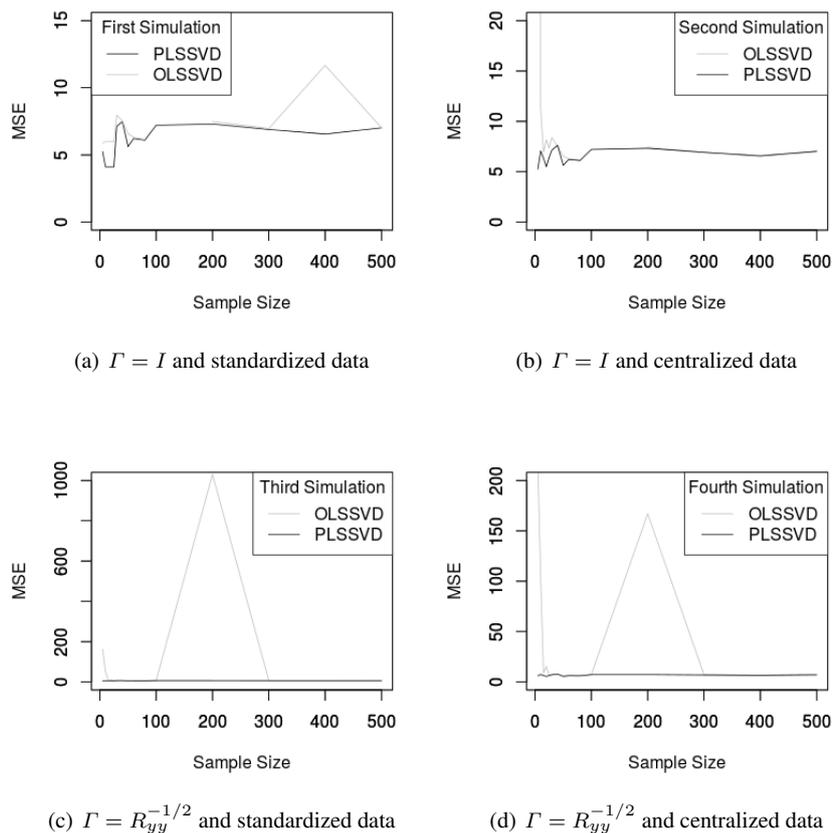


Figure 1. Simulations of severe collinearity

The mean square error (MSE) was used to compare the OLSSVD and PLSSVD procedures for simulated data with severe multicollinearity. Multiple samples of different sizes, from 5 to 500, were compared in four different scenarios. The first using  $\Gamma = I$  and standardized data (Figure 1(a)); the second using  $\Gamma = I$  and centralized data (Figure 1(b)); the third using  $\Gamma = R_{yy}^{-1/2}$  and standardized data (Figure 1(c)) and the fourth using  $\Gamma = R_{yy}^{-1/2}$  and centralized data (Figure 1(d)). The results for each of the simulations are shown in the following figures. In the first simulation (Figure 1(a)), the PLSSVD procedure is seen to have lower MSE than the OLSSVD procedure. Furthermore, the OLSSVD procedure in some cases is not able to make estimates due to the multicollinearity and

although OLSSVD improves somewhat for large samples, it is not stable. In the second scenario, (Figure 1(b)), the MSE for PLSSVD is again observed to be lower than those of OLSSVD for sample sizes less than 100 whereas the behavior of the two procedures is similar for larger sample sizes. For the third and fourth scenarios (Figures 1(c) and 1(d)), similar behavior including the instability of OLSSVD is again observed.

#### 4. Conclusions

The comparisons of the PLSSVD and OLSSVD procedures show that in the presence of severe multicollinearity of the independent variables, the PLSSVD estimation procedure results in lower residual values than the OLSSVD procedure however when the multicollinearity of the independent variables decreases, the opposite is true. For both  $\Gamma = I$  and  $\Gamma = R_{yy}^{-1/2}$  and for both the real data and simulated data, the PLSSVD has better predictive power than OLSSVD when the data presents severe multicollinearity. Even though in some cases the OLSSVD procedure presents good estimations, it is not consistently stable in the presence of severe multicollinearity.

#### REFERENCES

1. Álvarez W. and Cárdenas O., 2010. Un enfoque basado en la descomposición en valores singulares generalizada para el biplot de coeficientes de regresión. Ingenier Industrial. Actualidad y Nuevas Tendencias. A III. vol. II, N4. ISSN: 1856-8327, p. 65-74.
2. Anderson T., (2002). Specification and Misspecification in Reduced Rank Regression. The Indian Journal of Statistics, 64,193-205.
3. Anderson T., (1999). Asymptotic Distribution of the Reduced Rank Regression estimator under general conditions. The Annals of Statistics, 27, 1141-1154.
4. Anderson T., (1951). Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. Ann. Math. Statist, 22, 327-351.
5. Davies, P. and Tso M., (1982). Procedures for reduced rank regression. Applied Statistic 31, 244-255.
6. Eckart, C. and Young, G., (1936). The Approximation de One Matrix for Another of Lower Rank. Psychometrika 1, 211-218.
7. Gabriel, K., (1971). The Biplot-graphic display of matrices with applications to principal component analysis. Biometrika, 58, 453-467.
8. Graybill, F.A. (1969). Introduction to Matrices with Statistical Applications. Belmont, California. Wadsworth.
9. Heinen A. and Rengifo E., (2004). Multivariate Reduced Rank Regression in non-Gaussian Contexts, Using Copulas. Center for Operations Research and Econometrics Catholic University of Louvain Belgium. Roma.
10. Hervé Abdi, (2003). Partial Least Squares (PLS) Regression. Eds. Encyclopedia of Social Sciences Research Methods
11. Höskuldsson, A., (1988). PLS regression methods. Journal of Chemometrics, 2, 211- 228.
12. Izenman, A. J., (1975). Reduced rank regression for the multivariate linear model. Journal of Multivariate Analysis 5, 248-264.
13. Johnson, R. and Wichern, (1998). Applied Multivariate Statistical Analysis. Prentice Hall, Upper Saddle River, New Jersey.
14. Jolliffe I.T. (1986) Principal Component Analysis. Springer - Verlag
15. Mardia K.V., Kent J.T. and Bibby J.M. (1979). Multivariate Analysis, London Academic Press.
16. Martens H. and NæT. (1989). Multivariate Calibration. John Wiley and Sons.
17. Montgomery D. and Peck E. (1992). Introduction to Linear Regression Analysis. Second Edition. John Wiley and Sons.
18. Peña, D., (2002). Análisis de Datos Multivariantes. Mc Graw Hill.
19. Rencher A. (2002). Methods of Multivariate Analysis. Wiley-Interscience Publication
20. Tenenhaus, M. (1998). La régression pls, théorie et pratique. Éditions Technip. Paris.
21. Ter Braak, C. and Looman, C. (1994). Biplots in reduced rank regression. Biometrical Journal 8, 983-1003.
22. Ter Braak, C. (1990). Interpreting canonical correlation analysis through biplots of structural correlations and weights. Psychometrika, 55, 519-531
23. Tso, M. K-S (1981). Reduced-rank regression and canonical analysis. J.R. Statist. Soc. B, 43, 183-189.
24. Valencia J., Díaz-Llanos F. and Sainz-Calleja (2004). Métodos de Predicción en Situaciones Límite. Editorial La Muralla, S.A. Madrid Espa.
25. Van Der Wollenberg, A.L. (1977). Redundancy Analysis. An Alternative for Canonical Correlation Analysis. Psychometrika, 42:207-219.
26. Vásquez, M. (1995). Aportaciones al Análisis Biplot: Un enfoque algebraico. Thesis Doctoral. Universidad de Salamanca.
27. Velu Raja P., Reinsel Gregory C. and Wichern Dean W. (1986). Reduced rank models for multiple time series. Biometrika Vol. 73.1, pp.105-18.
28. Wold H. (1966). Estimation of principal components and related models by iterative least squares. In Multivariate analysis (ed. P.R. Krishnaiah), Academic Press, New York.
29. Wold S., Martens H. and Wold H. (1983), in Lecture Notes in Mathematics: Matrix Pencils, ed. by A. Ruhe and B. Kagström, pp. 286-293. Springer, Heidelberg
30. Qing-Song Xu, Yi-Zeng Liang and Hai-Lin Shen (2001). Generalized PLS regression. Journal of Chemometric, Vol. 15, 135-148.