European Journal of Applied Sciences 4 (6): 257-264, 2012
ISSN 2079-2077

DOI: 10.5829/idosi.ejas.2012.4.6.1111

A Performance Assessment of Model Selection Criteria When the Number of Objects Is Much Larger than the Number of Variables in PLSR

[1]Elif Bulut and [2]Özlem Gürünlü Alma

[1]Department of Business Administration, Ondokuz Mayıs University, Samsun 55139, Turkey
[2]Department of Statistics, Mugla University, Mugla 48000, Turkey

Abstract: Partial Least Squares Regression (PLSR) is a method for constructing predictive models when the variables are many and highly collinear. Its goal is to predict a set of response variables from a set of predictor variables. This prediction is achieved by extracting a set of orthogonal factors called latent variables from the predictor variables. This study investigated the performances of model selection criteria in selecting the true number of latent variables from PLSR models for data sets that have various observations and variable numbers. Their performances have been compared in terms of the simulation study and 5-fold cross validation. This simulation has been performed for different numbers of predictor variables and different numbers of observation units to compare the performance of two types of Multivariate Akaike Information criterion and three types of Wold's R criterion in finding the number of true latent variables. The simulation results show that all criteria achieved the true number of latent variables for a small-sized design matrix. It was noticed that when the observation numbers were increased, PLSR worked with a larger number of latent variables, except for some cases. Wold'S R_2 and Wold'S R_3 found less numbers as the number of latent variables.

Key words: K-fold Cross-validation · Multivariate Akaike Information Criteria · Partial Least Squares · WOLD'S R

INTRODUCTION

PLSR is a wide class of methods for modeling relations between sets of observed variables by means of latent variables. It is a combination of partial least squares analysis and multiple linear regression and allows modeling and prediction for multiple response variables and highly correlated or collinear multiple predictor variables. The aim of this technique is to summarize a large set of variables in terms of statistical similarities between the original variables by losing the minimal amount of information, which is obtained by the help of some algorithms. The application fields of PLSR also cover cases, in which there are more than one response variable, thus serving as an alternative to MANOVA designs. In these cases, where multiple response variables are used, PLSR creates other latent variables from the linear combination of the original response variables that act as synthetic response variables [1].

The pioneering work in PLS was done in the late sixties by H. Wold in the field of econometrics. The use of the PLS method for chemical applications was pioneered by the groups of S. Wold and H. Martens in the late seventies after an initial application by [2, 3]. [4] was offered a review of historical development of PLS. PLS regression was studied and developed from the point of view of statisticians by [5]. The book [6] used statistical concepts that began to provide a theoretical basis for PLS. The recent investigations were provided by [7-12]. In PLS, latent variables are obtained by using different algorithms which make dimension reduction by using singular or eigen value decomposition. These latent variables, which are the new predictor variables and the linear combinations of the original predictor variables and which do not have linear relationship among them, are used in the regression partition and they also ideally model the response variables. Following dimension reduction in PLS part, some model selection criteria are used to obtain the latent variables, the most relevant ones in describing the variability in the response variables. These methods are called model selection methods and play an important role in selecting the best model.

Corresponding Author: Elif Bulut, Department of Business Administration, Ondokuz Mayıs University, Samsun 55139, Turkey.

In this study, we provide a simulation study of dimension reduction methods for PLSR using kernel algorithm and 5-fold cross validation. Our goal is to carry out a simulation experiment: varying the number of predictors and observation units in order to compare the performances of M_{AKAKIE}, $M_{BEDRICK}$ and Wold's R (for three thresholds values) criteria for selecting the true number of latent variables. PLS was briefly reviewed and model selection criteria were described in Sections 2 and 3, respectively. The data simulation procedure used to study the performance of these dimension reduction methods is described in Section 4. Results of the simulation study are described in Section 5.

MATERIALS AND METHODS

PLS Model and PLS Algorithms: In PLS partition, the data matrices can be decomposed as the sum of the latent variables given in [13] as follows:

$$\mathbf{X} = \sum_{j=1}^{A} \mathbf{t_j p_j'} + \mathbf{E} \quad (1)$$

$$\mathbf{Y} = \sum_{j=1}^{A} \mathbf{u_j q_j'} + \mathbf{F} \quad (2)$$

where $\mathbf{t_j}$ and u_j are latent variables and $\mathbf{t_j}$'s are orthogonal to each other and also $\mathbf{t_j}$ is orthogonal to the subsequent $\mathbf{u_j}$.

M predictor variables are reduced to a fewer number of latent variables, A($A \leq M$), by algorithms in partial least squares. The most commonly used algorithms are NIPALS algorithm, UNIPALS algorithm, KERNEL algorithm, SAMPLS algorithm and SIMPLS algorithm. Early studies were about NIPALS algorithm. Then, the other algorithms were investigated on the basis of this algorithm. SIMPLS algorithm was studied by [14]. KERNEL algorithm was studied by [15]. Also, [16] studied the KERNEL algorithm.

Choice of algorithm depends strongly on the shape of data matrices to be studied. In some studies, the number of observations is much larger than the number of variables. This causes the algorithm to work with a matrix which is independent of the number of observations. For an opposite case, where the number of variables exceeds the number of observation units, choosing an algorithm that works with a matrix independent of the number of variables will be the best choice [17, 18]. Modified Kernel algorithm #2 was generated by [19]. This algorithm was modified from kernel algorithm, developed by [15], to provide two faster and more economical algorithms [19]. In modified kernel algorithm #2, the covariance matrix $\mathbf{X'X}$ is computed only once with the original $\mathbf{X}$. For more information, see [19].

In this study one of the Kernel algorithms, Modified Kernel algorithm #2, was used to select the number of latent variables. It was an alternative to the NIPALS algorithm for handling datasets where N>>M. This algorithm uses $\mathbf{X'YY'X}$ matrix since it is independent of the number of observations. This property provides working with a small matrix. This algorithm innovates to update $\mathbf{X'Y}$ variance-covariance matrix by the multiplication of an updating matrix ($\mathbf{I - wp'}$) without interfering with the original $\mathbf{X}$ and $\mathbf{Y}$ matrices. In ($\mathbf{I - wp'}$), I represents identity matrix.

Model Selection Criteria for Dimension Reduction in PLSR: Model selection is a critical subject in predicting the performance of regression models. Selecting the best model depends on the correct selection of variables. Hence, the model prediction error is minimized and the model is prevented from redundant variables. Such a criterion is required to ensure that the same decision process is used for both models for each simulation and in order to make a simulation study feasible. In this study two different forms of Multivariate Akaike Information criterion, Wold's R and adjusted Wold's R criteria, were used to select latent variables.

Akaike Information Criterion was developed by Hirotsugu Akaike under the name of an information criterion (AIC) in [20] and proposed by [21]. It is a measure of the goodness of fit of an estimated statistical model. It is a way of selecting a model from a set of models. Given a data set, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best. The multivariate version of AIC from [22] under the multivariate normal assumption for the multivariate regression model is given as follows:

$$\text{MAIC} = \text{Nalog}(2\pi) + N\log\left|\hat{\Sigma}\right| + Na + 2\left[aK + \frac{K(K+1)}{2}\right]. \quad (3)$$

For more than one response variable (K>1), multivariate version of AIC was given by Bedrick and Tsai (1994):

$$MAIC(a) = N\left(\log\left|\hat{S}\right| + K\right) + 2d\left[aK + \frac{K(K+1)}{2}\right], \quad (4)$$

where d = N/[N – (a + k + 1)] and $\hat{\Sigma}$ is the maximum likelihood estimator of Σ. In this paper, MAIC from [22] and MAIC from [23] have been shown as M_{AKAIKE} and $M_{BEDRICK}$, respectively.

Cross-validation is a very popular technique for model selection and validation. It is used to investigate the predictive validity of a linear regression equation. It is conceptually very simple to understand; however, it is the most computationally intensive method for optimizing a model. Besides, it is the most common approach to estimating the true accuracy of a given model and it is based on splitting the available sample into a training set and a validation set [24].

The other variable selection criterion, Wold's R, depends on the Prediction Sum of Squares (PRESS) values in cross validation. An alternative when data set is large enough is to split the data into two types. The first set, called the model building set or the training set, is used to develop the model. The second data set, called the validation or prediction set, is used to evaluate the reasonableness and predictive ability of the selected model. This validation procedure is often called cross-validation [25]. In this study, 5-fold cross-validation was used and PRESS values were determined each time deleting one of the k-fold of observations and variables. It is also possible to choose models based on leave-one-out cross validation. Nevertheless, this is computationally expensive. The PRESS value for the i^{th} observation is as follows:

$$PRESS = \sum_{i=1}^{N}\left(y_i - \hat{y}_{i(i)}\right)^2 \quad (5)$$

where the notation $\hat{y}_{i(i)}$ is used for the fitted value. By the first subscript i, it is shown that it is a predicted value for the i^{th} case and by the second subscript (i), it is shown that i^{th} case is omitted when the regression function is fitted. The smaller PRESS value shows that it is the best model to predict. In some cases, PRESS should reach a minimum and begin to rise again. Wold's R can be calculated from the PRESS values. It can be explained as follows:

$$R = \frac{PRESS(a+1)}{PRESS(a)} \quad (6)$$

where PRESS(a) denotes the PRESS value including the first a latent variables. Wold's R criterion terminates when R is greater than unity or a given threshold and, hence, A=a and adjusted Wold's R criterion terminates when R is greater than a given threshold, A=a [13].

In this paper, three threshold values, 1 for Wold's R criterion and 0.90 and 0.95 for adjusted Wold's R criterion, were used. If R value is greater than one of these threshold values, this means that no additional latent variable will be included in the PLSR model. If it provides significant predictions, this means that R value is smaller than the threshold value. Then, the latent variable (a+1) will be included in the model. Wold's R_1, Wold's R_2 and Wold's R_3 have showed the threshold values 1, 0.95 and 0.90, respectively.

Simulation Procedures: Contrary to the fixed effect model, the quality of prediction for a random model does not always increase with the number of latent variables used in the model. Typically, the quality first increases and then decreases. If the quality of the prediction decreases when the number of latent variables increases, this indicates that the model is overfitting the data. Therefore, for a random model, it is critical to determine the optimal number of latent variables to be maintained to build the model. A straightforward approach is to stop adding latent variables as soon as the PRESS decreases [12]. And, as also mentioned by [16], the choice of the number of latent variables, A, is a matrix rank problem which causes overfitting when A has been chosen too large and underfitting when A has been chosen too small. Thus, in this simulation study, the performance of the model selection criteria in PLSR models is investigated.

The framework for the simulation model was based on [13] for the problem of multiple response variables. The paper's simulation study was extended for various dimensions for predictor variables matrix and various magnitudes for observation units to compare the performances of model selection criteria. In this study, the true number of latent variables was shown with $A^*(A^* \leq A)$. The dimensions of predictor variables matrix N × M was extended with M=6, 8, 10 and 12.

The dimension of response variables matrix, Y, was chosen as **Y**, K=4 and N; and the observation units were chosen as 100, 250, 500 and 1000. Matlab codes were written for 5-fold cross validation in Modified Kernel Algorithm #2. The simulation experiments were repeated 10,000 times for all design matrices. All these data generations were based on the PLS model. Predictor data matrix, **X**, was generated as follows:

$$\mathbf{X} = \sum_{a=1}^{A^*}\mathbf{r_a}\xi_a' + \mathbf{E} \quad (7)$$

where $\mathbf{r_a}$ and **E** were generated as mutually independent normal variables and ξ_a were generated as normalized orthogonal and unit vectors. var($\mathbf{r_1}$) + var($\mathbf{e_m}$) is the largest eigenvalue of cov(**X**). The components of **X** matrix are given in Table 1.

Response variables matrices, **Y**, were generated from Equations (8) and (9). These equations are shown as follows:

Table 1: The components of X matrix

Dimension of data matrix	R	E
N × M	R = [$\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, $\mathbf{r}_4$], a = 1,2,3,4 were generated as mutually independent normal variables with mean zero and, var($\mathbf{r}_1$) = 10 var($\mathbf{r}_2$) = 5 var($\mathbf{r}_3$) = 2 var($\mathbf{r}_4$) = 0.5	E = [$\mathbf{e}_1$, $\mathbf{e}_2$, …, $\mathbf{e}_m$], m = 1,…,M were generated as mutually independent random variables with mean zero and, var($\mathbf{e}_m$) = 0.01.

Table 2: The generated values fot φ

Dimension of data matrix	φ
N × K	φ = [φ_1, φ_2, φ_3, φ_4] was genrated from multivatiate normal distribution with mean zero and following variance-covariance matrix; $\begin{bmatrix} 0.00010 & 0.00006 & 0.00006 & 0.00006 \\ 0.00006 & 0.00010 & 0.00006 & 0.00006 \\ 0.00006 & 0.00006 & 0.00010 & 0.00006 \\ 0.00006 & 0.00006 & 0.00006 & 0.00010 \end{bmatrix}$

Table 3: The generated data F for Y

Dimension of data matrix	**F**
N × K	F = [$\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$, $\mathbf{f}_4$], a = 1,2,3,4 were generated as mutually independent normal variables with mean zero and var($\mathbf{f}_1$) = 0.25 var($\mathbf{f}_2$) = 0.125 var($\mathbf{f}_3$) = 0.05 var($\mathbf{f}_4$) = 0.0125

Table 4: VIF values for N × 6

Predictors	X_1	X_2	X_3	X_4	X_5	X_6
VIF	534.4	526.7	1009.8	661.9	621.4	866.9

Table 5: Relative cummulative variances of **X** and **Y** for N × 6

		Number of Latent Variables					
True Model	Blocks	1	2	3	4	5	6
$A^* = 4$	X-block	0.53055	0.87616	0.96319	0.99984	1.00000	1.00000
	Y-block	0.94539	0.96692	0.97603	0.97634	0.97690	0.97700

$$\mathbf{Y} = \sum_{a=1}^{A^*} \mathbf{z_a} \eta'_{\mathbf{A}_a^*} + \varphi = \sum_{\mathbf{a=1}}^{\mathbf{A}^*} \mathbf{r_a} \eta'_{\mathbf{A}_a^*} + \mathbf{F}_{\mathbf{A}^*} \quad (8)$$

$$\mathbf{F} = \sum_{a=1}^{A^*} \mathbf{f_a} \eta'_{\mathbf{a}} + \varphi \quad (9)$$

where **E** and **F** are the residuals of **X** and **Y** after extracting the first A pairs of latent variables. φ was generated from a multivariate normal distribution, while $\eta_{A_a^*}$ were generated as normalized orthogonal vectors. $\mathbf{z_a} = \mathbf{r_a} + \mathbf{f_a}$ were generated as independent normal variables.

Table 1, Table 2 and Table 3 show how data matrices were generated. After generating all data sets, the variance inflation factor (VIF) was calculated for N × 6 latent variables in Minitab packet program to see whether there was multicollinearity or not. The VIF values are shown only for N × 6 in Table 4, with the high values for VIF indicating a high degree of multicollinearity.

Table 5 gives the cumulative variances that explained the variations of latent variables in both predictor and response variables for N × 6 design matrix.

The analysis showed that for the other-sized design matrices, both the VIF values were high as N × 6 and the number of latent variables was the same with N × 6. So, A* = 4 was taken as the true number of latent variables.

RESULTS AND DISCUSSION

The selecting numbers of latent variables were obtained for each simulated data set using model selection criteria from PLSR models. The performance of the five procedures, M_{AKAKIE}, $M_{BEDRICK}$, Wold's R_1, Wold's R_2 and Wold's R_3 criteria, across 10,000 simulated data sets are summarized in Figures 1 to 5. Ten thousand data sets were generated for each dimension of predictor variables, M=6,8,10,12 and each observation unit, N=100,250,500,1000. For each plot in Figure 1, the y-axis is the number of latent variables and the x-axis is the percentage of model selection criteria. The percentage values show the percentage of maximum iteration number for latent variables for each case in 10,000 repetitions. For example, the performance of M_{AKAKIE} is 53.23% ((5323/10,000)*100) when N=100 and M=6.

For each plot in Figure 1, the y-axis is the number of latent variables and the x-axis is the percentage of model selection criteria. The percentage values show the percentage of maximum iteration number for latent variables for each case in 10,000 repetitions. For example, the performance of M_{AKAKIE} is 53.23% ((5323/10,000)*100) when N=100 and M=6.

Figure 1(a-d) shows that the true number of latent variables is obtained by all criteria for small design matrix as in Li and Morris (2002), $A = A^* = 4$. When the number of predictor variables is increased, all criteria find the number of latent variables close to the number of predictor variables for N=100 except for Wold'S R_1, Wold'S R_2 and Wold'S R_3 criteria for M=10. All criteria find these numbers for latent variables with a higher percentage value than the other numbers. Only for M=10, Wold's R_1 finds the number of latent variables as seven with 50.24% and this is the figure with the closest percentage value to the value, with which the number of latent variables is found as 2 with the largest percentage. In the same case, Wold's R_2 and Wold's R_3 find the number of latent variables as ten with 80.23% and 91.83%, respectively and this is the figure with the closest percentage value to the value, with which the number of latent variables is found as 2 with the largest percentage.

Figure 2 shows that M_{AKAIKE} and $M_{BEDRICK}$ find the number of latent variables with a larger number when observation units are increased. For this case, all Wold's R criteria find the number of latent variables as the true number of latent variables, $A = A^* = 4$ with 100%. $M_{BEDRICK}$ finds the number of latent variables with higher percentages than M_{AKAIKE}. In this case, Wold's R criteria are the best because of finding the true number of latent variables.

Wold's R criteria find different numbers of latent variables as the number of latent variables when observation units increase and M=8. M_{AKAIKE} and $M_{BEDRICK}$ work with larger numbers of latent variables as the number of latent variables. They find the number of latent variables as 8 with an increasing percentage with increasing observation numbers. The percentage values for M_{AKAIKE} are 79.72%, 99.98% and 100% for the rest. The percentage values for $M_{BEDRICK}$ are 53.93%, 99.93% and 100% for the rest.

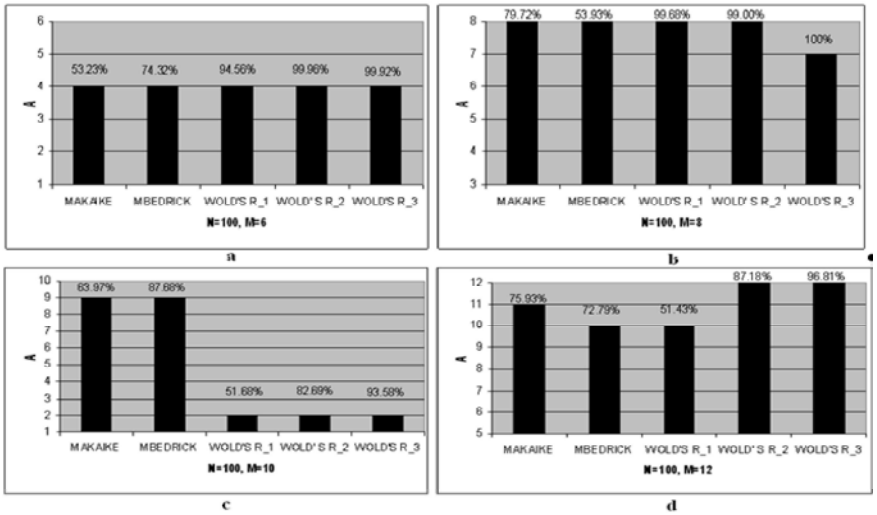

Fig. 1(a-d): All criteria for changing M and fixed N. A*=4

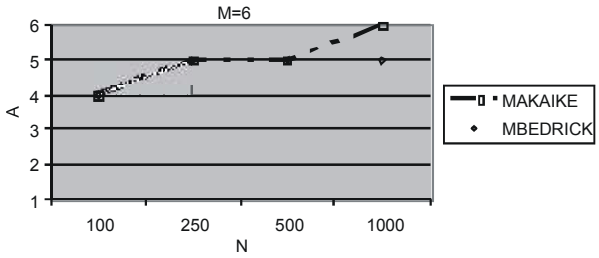

Fig. 2: Model selection criteria for M=6 and varying N. A*=4

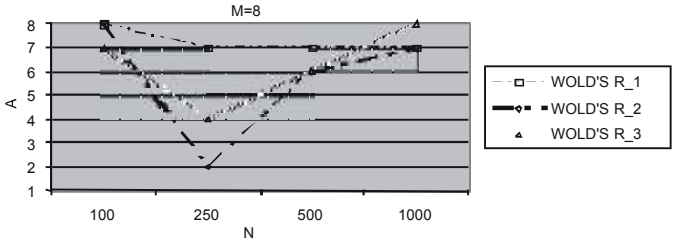

Fig. 3: Model selection criteria for M=8 and varying N. A*=4

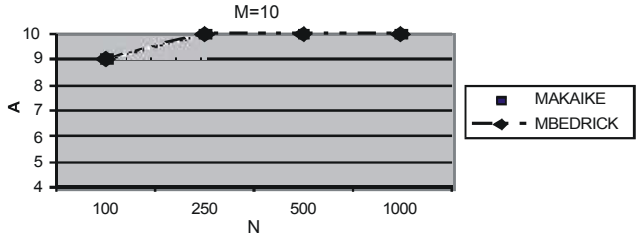

Fig. 4-a: Model selection criteria for M=10 and varying N

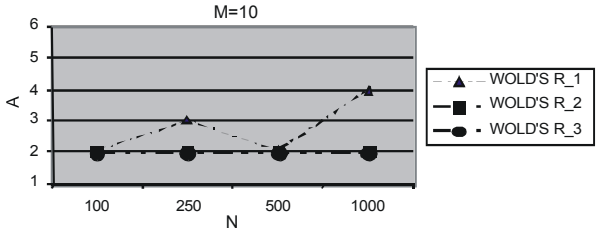

Fig. 4-b: Model selection criteria for M=10 and varying N

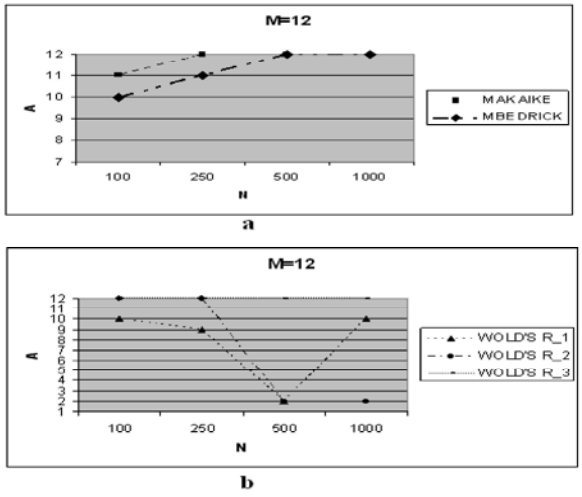

Fig. 5 (a-b) Model selection criteria for M=12 and varying N

As seen from Figure 4 (a-b), M_{AKAIKE} and $M_{BEDRICK}$ show the same tendency to find the number of latent variables. They found the number of latent variables as 9 from small observation units. When the numbers of observations get larger, these criteria work with the same number of predictor variables as the number of latent variables. Nevertheless, WOLD'S R criteria exhibit a different situation. Wold's R_2 and Wold'S R_3 work with small and the same numbers as the number of latent variables; however, Wold's R_1 shows a discrepancy for different numbers of observations. Nevertheless, all Wold's R criteria work with a small number of latent variables. Wold's R_2 and Wold's R_3 are more sensitive than Wold's R_1 in finding the number of latent variables due to low threshold values. In the low threshold value, it is more powerful to select the (a)th latent with regard to (a+1)th latent variable.

Figure 5 (a-b) shows that M_{AKAKIE} and $M_{BEDRICK}$ work with an increasing number for the number of latent variables for increasing observation number, but Wold's R criteria have variability in finding the number of latent variables for increasing observation number.

CONCLUSION

This study is based on the comparison of the performances of MAIC from Bozdogan and Bedrick and Tsai and Wold's R criterion for various dimensions of data matrices in finding the true number of latent variables. The simulation results show that all criteria achieved the true number of latent variables for small-sized design matrices. However, the results for the other-sized design matrices greatly varied and they consistently showed different numbers for the number of latent variables. Generally, it can be said that when observation numbers increase, PLS creates a model with a high number of latent variables, which is statistically significant. The simulation studies also show that Wold's R criterion is effective for a N × 6 design matrix. That is, it gave the same result as in [13] when the data were generated according to their paper. Nevertheless, when the data were generated according to the assumptions of PLSR, it seemed that Wold's R criterion did not give desirable results in higher-dimensional data. MAIC from Bozdogan and Bedrick and Tsai found almost the same results as the number of latent variables, but they could not find the true number of latent variables for higher-dimensional data. Simulation studies also showed that Wold's R criterion was almost more sensitive than MAIC in working with small numbers for the number of latent variables. Especially, Wold's R_2 and Wold's R_3 work with a less number than Wold's R_1. In conclusion, it is noted that experimental results clearly demonstrated an excellent performance of MAIC from Bozdogan and Bedrick and Tsai in N × 6 when N=100. However, for higher-dimensional data matrices, all model selection criteria have a tendency to find the number of latent variables close to the number of predictor variables. These results suggest that MAIC and Wold's R criteria are affected by variations in the dimensions of regression models and observation numbers in selecting the true model.

Notations:

- a → Index of latent variables
- A → Number of latent variables in PLS model
- i → Index of observation units
- N → Number of observations
- M → Number of predictor variables (m=1,2,…,M)
- K → Number of response variables (k=1,2,…,K)
- $\mathbf{X}$ → Matrix of predictor variables with dimension (N × M)
- $\mathbf{Y}$ → Matrix of response variables with dimension (N × K)
- $\mathbf{E}$ → Matrix of X residuals (N × M)
- $\mathbf{F}$ → Matrix of Y residuals (N × K)
- $\mathbf{t}_j$ → J_{th} latent variable for X in PLS model
- $\mathbf{u}_j$ → J_{th} latent variable for Y in PLS model
- $\mathbf{p}_j$ → X loading vector of latent variable j (M × 1) in PLS model
- $\mathbf{q}_j$ → Y loading vector of latent variable j (K × 1) in PLS model
- $\mathbf{w}_j$ → X weight of component j (M × 1) in PLS model
- $\mathbf{r}_a$ → Latent variable in simulation study, approximately equal to t_j in PLS regression model
- $\boldsymbol{\xi}_a$ → Approximately equal to p_j in PLS regression model
- $\boldsymbol{\eta}_a$ → Approximately equal to q_j in PLS regression model
- M_{AKAKIE} → MAIC from Bozdogan (2000).
- $M_{BEDRICK}$ → MAIC(a) from Bedrick and Tsai (1994).

In this paper, uppercase bold characters represent matrices and lowercase bold variables represent vectors.

REFERENCES

1. Carrascal, L.M., I. Galva´n and O. Gordo, 2009. Partial least squares regression as an alternative to current regression methods used in ecology. Oikos, 118: 681-690.
2. Kowalski, B., R. Gerlach and H. Wold, 1982. Chemical Systems under Indirect Observation, In: K. Jöreskog and H. Wold, (Eds.), Systems under Indirect Observation, North-Holland, Amsterdam, pp: 191-209.
3. Geladi, P. and R. Kowalski, 1986. Partial least squares regression: A Tutorial. Analytica Chimical Acta, 185: 1-17.
4. Geladi, P., 1988. Notes on the history and nature of partial least squares (PLS) modelling. J. Chemometrics, 2: 231-246.
5. Höskuldson, A., 1988. PLS regression methods. J. Chemometrics, 2: 211-228.
6. Martens, H. and T. Naes, 1989. Multivariate calibration. John Wiley and Sons.
7. Helland, I.S., 1990. Partial least squares regression and statistical models. Scandinavian J. Statistics, 17: 97-114.
8. Garthwaite, P.H., 1994. An interpretation of partial least squares. J. American Statistical Association, 89: 122-127.
9. Wold, S., M. Sjöström and L. Eriksson, 2001. PLS-regression:a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58: 109-130.
10. Tobias, R.D., 2003. An introduction to partial least squares regression from: www.ats.ucla.edu/ stat/ sas/ library/pls.pdf.
11. Abdi, H., 2007. Partial least square regression (PLS regression). In: N.J. Salkind, (edc), Encylopedia of Measurement and Statistics Sage.
12. Abdi, H., 2010. Partial least squares regression and projection on latent structure regression (PLS Regression). Wiley Interdisciplinary Reviews: Computational Statistic, 2(1): 97-106.
13. Li, B., J. Morris and B. Martin, 2002. Model selection for partial least squares regression. Chemometrics and Intelligent Laboratory Systems, 64: 79-89.
14. De Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. Chemometrics and Intelligent Laboratory Systems, 18: 251-263.
15. Lindgren, F., P. Geladi and S. Wold, 1993. The Kernel algorithm for PLS. J. Chemometrics, 7: 45-59.
16. Rännar, S., F. Lindgren, P. Geladi and S. Wold, 1994. A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and Algorithm. J. Chemometrics, 8: 111-125.
17. De Jong, S. and C.J.F. Ter Braak, 1994. Comments on the PLS kernel algorithm. J. Chemometrics, 8: 169-174.
18. Lindgren, F. and S. Rannar, 1998. Alternative partial least-squares (PLS) algorithm. Perspective in Drug Discovery and Design 12/13/14, pp: 105-113.
19. Dayal, B. and J. MacGregor, 1997. Improved PLS algorithms. J. Chemometrics, 11: 73-85.
20. Akaike, H., 1971. Autoregressive model fitting for control. Ann. Inst. Statistics Math., 23: 163-180.
21. Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control., 19: 716-723.
22. Bozdogan, H., 2000. Akaike's information criterion and recent developments in Information Complexity. J. Mathematical Psychology, 44: 62-91.
23. Bedrick, E.J. and C.L. Tsai, 1994. Model selection for multivariate regression in small samples. Biometrics, 50: 226-231.
24. Last, M., 2006. The uncertainty principle of cross-validation. Proceedings of IEEE-GrC. pp: 275-280, Atlanta, USA, May 10-12, 2006.
25. Neter, J., M.H. Kutner, C.J. Nachtsheim and W. Wasserman, 1996. Applied linear regression models. Times Mirror Higher Education Group, Inc.