

## PERFORMANCE COMPARISONS OF MODEL SELECTION CRITERIA: AIC, BIC, ICOMP AND WOLD'S FOR PLSR

Özlem GÜRÜNLÜ ALMA\*

### ABSTRACT

*Partial least squares regression (PLSR) is a statistical method of modeling relationships between  $Y_{N \times M}$  response variable and  $X_{N \times K}$  explanatory variables which is particularly well suited to analyzing when explanatory variables are highly correlated. In partial least square part, some model selection criteria are used to obtain the latent variables which are the most relevant variables describing the response variables. In typical approach to select the numbers of latent variables are Akaike information criterion (AIC) and Wold's R criterion.*

*In this study, we are interested in the performance of Bayesian Information Criterion (BIC) and Information Complexity Criterion (ICOMP) criteria besides the traditional methods AIC and Wold's R criteria as the model selection criteria for partial least squares regression when the number of observations are higher than predictor variables. Performances of AIC, BIC, ICOMP and Wold's R criteria were compared by real life data and simulation study. Simulation results were obtained from different sample sizes, different number of predictor variables and different number of response variables. The simulation results demonstrate that the BIC and ICOMP model selection methods are more effective than AIC and Wold's R criteria selecting of latent variables for known PLSR models.*

**Keywords:** AIC, BIC and ICOMP information criteria, K-fold cross-validation, Model selection, Partial least squares regression, Wold's R criterion.

### 1. INTRODUCTION

The partial least squares regression is a generalization of multiple linear regression analysis. It was developed by Herman Wold (1966) as an econometric technique but became popular as a tool to analyze data from chemical applications. PLSR is also used in multivariate statistical data analysis (Geladi and Kowalski, 1986; Wold, 1982). It is useful when the predictor variables are highly correlated and/or the number of dependent variables is greater than or equal to the number of observations (Wold, 1982). This success has led to the development of extensions methods of PLSR with objectives other than simple multivariate linear regression. A statistical overview of PLSR can be found in Geladi and Kowalski (1986), Wold et al. (2001), and Abdi and Salkind (2007).

The PLSR's goal is to predict or analyze a set of response variables from a set of independent variables or predictors. This prediction is achieved by extracting from the predictors a set of orthogonal factors called latent variables which have the best predictive power. Associations are established with latent factors extracted from predictor variables that maximize the explained variance in the response variables. These latent factors are defined as linear combinations constructed between predictor

---

\*Assistant Professor Dr., Department of Statistics, Faculty of Sciences, Muğla Sıtkı Koçman University, 48000 Muğla, Turkey, e-mail: [ozlem.gurunlu@gmail.com](mailto:ozlem.gurunlu@gmail.com)

and response variables, such that the original multidimensionality is reduced to a lower number of orthogonal factors to detect the structure in the relationships between predictor variables and between these latent factors and the response variables (Abdi and Salkind, 2007; Helland, 1990; Wold, 1982). Although as many latent variables as  $\min(N,K)$  can be calculated, where  $N$  is the sample size and  $K$  is number of explanatory variables, it is conjectured that the lower order latent variables are associated with process noise and should be excluded from the model. Therefore to remove the noise, a criterion is required for selecting the number of latent variables to include in the PLSR model (Li et al., 2002).

Various approaches have been proposed in the literature for model order selection methods, including Final Prediction Error criterion (FPE), Multiple Correlation Coefficient ( $R^2$ ), Adjusted Multiple Correlation Coefficient ( $R_a^2$ ), Normalized Residuals Sum of Squares (NRSS), Mallow's Statistics ( $C_p$ ), Predicted Error of Sum of Squares (PRESS), Wold's R criterion, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Information Complexity Criterion (ICOMP). A review of these model order selection criteria can be found in Haber and Unbenhauen (1990); Bozdoğan (2000); Li et al. (2002); Clark and Troskie (2006). To evaluate the performance of the different criteria, simulated models allow the underlying structures of the models to be known (Bedrick and Tsai, 1994; Eastment and Krzanowski, 1982). Practical case studies as described in Bozdogan (2000), Myung (2000), Li et al. (2002), and Clark and Troskie (2006).

The PLSR creates latent variables for both explanatory and response variables using different algorithms. As well as the standard NIPALS, SIMPLS and Kernel algorithms, many different algorithms have been proposed to compute PLSR parameters such as IVS-PLS, PoLiSh, UVE-PLS, GA-PLS, and etc (Jouan-Rimbaud Bouveresse and Rutledge, 2009).

In this paper, the subset of latent variables that best fit the data is sequentially determined. Firstly, the latent variables are extracted using partial least squares algorithm, secondly, the number of latent variables can be consistently estimated using information criterion. The performance of information criterion is considered with the generation of experimental data. We have shown the behaviour of AIC, BIC, ICOMP and Wold's R criteria for different sample sizes and different dimension of PLSR models by simulation study.

The article is organized as follows. Section 2 includes the PLSR algorithm and describes how to obtain the latent variables. Section 3 gives summary information about the model selection criteria which are AIC, BIC, ICOMP and Wold's R criteria. In Section 4, real life data and simulation models are described, and the simulation results are given. This section focuses on the empirical results which show the performance of information criteria for various configurations of data sets. Finally a summary of simulation results and conclusions are given in section 5.

## **2. PARTIAL LEAST SQUARES REGRESSION MODEL**

The objective of all linear PLSR algorithm is to project the data down onto a number of latent variables ( $t_a$  and  $u_a$ ), and then to develop a regression model between latent

variables. It uses both the variation of  $\mathbf{X}$  and  $\mathbf{Y}$  to construct latent variables. The intension of PLSR is to form components that capture most of the information in the  $\mathbf{X}$  variables, which is useful for predicting response variables, while reducing the dimensionality of the regression problem by using fewer components than the number of  $\mathbf{X}$  variables (Garthwaite, 1994).

$\mathbf{X}_{N \times K}$  represents the data matrix of  $N$  observation units on  $K$  explanatory variables and  $\mathbf{Y}_{N \times M}$  the data matrix of  $N$  observation units on  $M$  response variables.  $t_a$  and  $u_a$  ( $a=1, \dots, A$ ) are latent variables, where  $A$  is the number of the latent variables, and then a regression model between latent variables is written as follows:

$$\mathbf{u}_a = \mathbf{b}_a \mathbf{t}_a + \mathbf{e}_a, \quad a=1, \dots, A \quad (1)$$

where  $\mathbf{e}_a$  is vector of errors and  $\mathbf{b}_a$  is an unknown parameter estimated by  $\hat{\mathbf{b}}_a = (\mathbf{t}_a' \mathbf{t}_a)^{-1} \mathbf{t}_a' \mathbf{u}_a$ . The latent variables are computed by  $t_a = \mathbf{X}_a \mathbf{w}_a$  and  $u_a = \mathbf{Y}_a \mathbf{q}_a$ , where both  $\mathbf{w}_a$  and  $\mathbf{q}_a$  have unit length and are determined by maximizing the covariance between  $t_a$  and  $u_a$ .

$\mathbf{X}_{a+1} = \mathbf{X}_a - \mathbf{t}_a \mathbf{p}_a'$  where  $\mathbf{X}_1 = \mathbf{X}$  and  $\mathbf{p}_a = \mathbf{X}_a' \mathbf{t}_a / (\mathbf{t}_a' \mathbf{t}_a)$  and  $\mathbf{Y}_{a+1} = \mathbf{Y}_a - \mathbf{b}_a \mathbf{t}_a \mathbf{q}_a'$  where  $\mathbf{Y}_1 = \mathbf{Y}$ . Letting  $\hat{u}_a = \hat{\mathbf{b}}_a \mathbf{t}_a$  be prediction of  $u_a$ , the matrices  $\mathbf{X}$  and  $\mathbf{Y}$  can be decomposed as the following (Li et al., 2002):

$$\mathbf{X} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}_a' + \mathbf{E}, \quad \text{and} \quad \mathbf{Y} = \sum_{a=1}^A \hat{\mathbf{u}}_a \mathbf{q}_a' + \mathbf{F}, \quad (2)$$

where  $\mathbf{E}$  and  $\mathbf{F}$  are the residuals of  $\mathbf{X}$  and  $\mathbf{Y}$  after extracting the first "a" pairs of latent variables.

### 3. ESTIMATING NUMBER OF LATENT VARIABLES USING INFORMATION CRITERION

The problem of estimating the true error of hypothesis using different adjustable parameters in order to choose the best one is known as model selection (Hastie et al., 2001). The necessity of introducing the concept of model evaluation has been recognized as one of the important technical areas, and the problem is posed on the choice of the best approximating model among a class of competing models by a suitable model evaluation criterion given a data set. Model evaluation criteria are defined as figures of merit, or performance measures, for competing models (Bozdoğan, 2000). In this section a number of criteria for PLSR model selection can be briefly summarized for multivariate regression models. In PLSR model, the information criteria used to find the number of latent variables and  $\mathbf{T}_a, \mathbf{Y}, \mathbf{M}, V(a, \mathbf{T}_a), a$  was used instead of  $\mathbf{X}, \mathbf{Y}, \mathbf{M}, \Sigma, K$ . Let  $\mathbf{T}_a$  be a matrix of latent variables, and  $V(a, \mathbf{T}_a)$  is the sum of squared residuals.

$$V(a, \mathbf{T}_a) = \min_{\mathbf{T}_a} \frac{1}{MN} \sum_{k=1}^M \sum_{i=1}^N (Y_{ik} - \hat{Y}_{ik})^2 \quad (3)$$

$$V(a, T_a) = \min_{T_a} \frac{1}{MN} \sum_{k=1}^M \sum_{i=1}^N (Y_{ik} - T_{ia} b_{ak})^2 \quad (4)$$

Selection of the number of latent variables to build a representative model is an important issue in PLSR. The main goal of model selection is to approximate the true model using candidate models and then retain the model that entails a minimum loss of information. A metric frequently used by chemometricians for the determination of the number of latent variables is that of Wold's R criterion, whilst more recently a number of statisticians have advocated the use of AIC (Li et al., 2002). Generally a good model has small residuals and few parameters, and then it is preferred, chosen with the smallest value of information criterion. However, it is well known that different information theoretic criteria with proper choice of penalty function can be used to choose the correct model (Kundu and Murali, 1996). Bedrick and Tsai (1994) modified the AIC criterion which is corrected version of the multivariate AIC for the small sample case (Bedrick and Tsai, 1994).

AIC and BIC are the two penalized criteria that are based on two different model selection approaches. AIC is aimed at finding the best approximating model to the unknown data generating process whilst BIC is designed to identify the true model. AIC does not depend directly on sample size. Bozdoğan (1987) noted that because of this, AIC lacks certain properties of asymptotic consistency. Although BIC takes a similar form like AIC, it is derived within a Bayesian framework, reflects sample size and have properties of asymptotic consistency. For reasonable sample sizes, BIC apply a larger penalty than AIC, thus other factors being equal it tend to select simpler models than does AIC. From a Bayesian view point this motivates the adoption of the Bayesian information criteria. AIC and BIC have been compared theoretically and empirically (Kuha, 2004; Weakliem, 2004) and examined empirically with respect to theselection of stock-recruitment relationships (Wang and Liu, 2006; Henry de-Graf, 2010). Although, AIC, BIC, and Bozdogan information criteria compared theoretically and empirically in many areas, there has been no empirical comparison for their relative performance in PLSR modeling context.

### 3.1 The Akaike Information Criterion

The Akaike information criterion was developed by Akaike (Akaike, 1974). AIC has played a significant role in solving problems in a wide variety of fields for analyzing actual data. The AIC is defined as,

$$AIC = -2\log L(\hat{\theta}) + 2K, \quad (5)$$

where  $\hat{\theta}$  is the maximum likelihood estimator of the parameter  $\theta$  for an approximating statistical model  $Y$ ,  $L(\hat{\theta})$  is the maximized likelihood function, and  $K$  is the number of free parameters in  $Y$ . The multivariate version of AIC was given by Bedrick and Tsai (1994),

$$MAIC = N(\log \left| \hat{\Sigma} \right| + M) + 2d[MK + M(M+1)/2], \quad (6)$$

where  $d = N/[N - (K + M + 1)]$  and  $\hat{\Sigma}$  is the maximum likelihood estimator of  $\Sigma$ . This is the corrected version of the multivariate AIC for the small sample case. When the sample size is large,  $d$  value can be equal to one, thus equation (6) may be further simplified. Also, Bozdoğan (2000) derived a score information theoretic criteria under the multivariate normal assumption for the multivariate regression model which are given as follows,

$$AIC = NM \log(2\pi) + N \log |\hat{\Sigma}| + NM + 2[MK + M(M + 1)/2] \quad (7)$$

Since 1974, AIC has been modified in many ways. For example, many model selection criteria including CAIC, CAICF (Bozdoğan, 1987), GAC (Torr, 1998), GAIC (Kanatani, 2002) and MAIC (Boyer et al., 1994) are derived from AIC.

### 3.2 The Bayesian Information Criterion

The Bayesian Information Criterion is an information criterion based on Bayesian method proposed by Schwarz (1978), has recently been applied to the selection of models. BIC is shown as,

$$BIC = -2 \log L(\hat{\theta}) + K \log(N), \quad (8)$$

where  $\hat{\theta}$  is the maximum likelihood estimator of the parameter  $\theta$  for an approximating model  $M$ ,  $L(\hat{\theta})$  is the maximized likelihood function, and  $K$  is the number of the estimated parameters. The multivariate version of BIC was given by Bedrick and Tsai (1994). It is shown as follows,

$$BIC = N \log |\hat{\Sigma}| + \left[ \frac{MK + M(M + 1)}{2} \right] \log(N). \quad (9)$$

BIC favourizes more parsimonious models than AIC due to its penalization. AIC, but not BIC, is biased in the following sense: if the true model belongs to the family  $M_i$ , the probability that AIC chooses the true model does not tend to one when the number of observations goes to infinity. AIC and BIC have similar formulas but originates from different theories and there is no rationale to use simultaneously AIC and BIC: AIC is an approximation of the Kullback-Leibler divergence between the true model and the estimated one, while BIC comes from a bayesian choice based on the maximisation of the posterior probability of the model, given the data (Saporta, 2008).

### 3.3 The Information Complexity Criterion

The development of ICOMP has been motivated partly by AIC, and partly by information complexity concepts and indices. In contrast to AIC, the new procedure ICOMP is based on the structural complexity of an element or set of random vectors via a generalization of the information based covariance complexity index. ICOMP inverse Fisher information matrix (ICOMP(IFIM)) is shown as for multiple regression (Bozdoğan, 2000; Bozdoğan, 2004),

$$\text{ICOMP(IFIM)} = N \log(2\pi) + N \log(\hat{\sigma}^2) + N + C_1(\hat{F}^{-1}(\hat{\theta})), \quad (10)$$

where  $\hat{\sigma}^2$  is the estimated variance of regression model, and K explanatory variables in regression model. Bozdoğan (2000) introduced ICOMP (IFIM) information theoretic criterion for the multivariate regression model, and it is also used when there is multicollinearity in regression model. It is shown as follows,

$$\text{ICOMP} = NM \log(2\pi) + N \log|\hat{\Sigma}| + NM + 2C_1(\hat{F}^{-1}(\hat{\theta})). \quad (11)$$

The complexity measure  $C_1(\hat{F}^{-1}(\hat{\theta}))$  is given by,

$$C_1(\hat{F}^{-1}(\hat{\theta})) = \frac{M(M+K)}{2} \log \left[ \frac{\text{tr}(\hat{\Sigma})\text{tr}(X'X)^{-1} + \frac{1}{2N} \left[ \text{tr}(\hat{\Sigma}^2) + (\text{tr}\hat{\Sigma})^2 + 2 \sum_j \hat{\sigma}_{jj} \right]}{M(M+K)} \right] - \frac{1}{2}(M+K+1) \log|\hat{\Sigma}| - \frac{M}{2} \log|(X'X)^{-1}| - \frac{M}{2} \log(2). \quad (12)$$

### 3.4 Wold's R Criterion

Wold's R criterion is based on cross validation which can be calculated from the Predicted Residual Sum of Squares (PRESS) values, and it can be explained as follows:

$$R = \frac{\text{PRESS}(a+1)}{\text{PRESS}(a)}, \quad (13)$$

where PRESS(a) denotes the PRESS after including the first a latent variables. Wold's R criterion terminates when R is greater than unity or a given threshold and hence  $A=a$  (Li et al., 2002). PRESS is a measure of how well the use of the fitted values for a subset model can predict the observed responses of a dependent variable, and its value for the  $i^{\text{th}}$  observation is calculated as follows:

$$\text{PRESS} = \sum_{i=1}^n (y_i - \hat{y}_{i(i)})^2, \quad (14)$$

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  $i^{\text{th}}$  case and by the second subscript (i), it is shown that  $i^{\text{th}}$  case is omitted when the regression function was fitted. The smaller PRESS value shows that it is the best model to predict. In some situations PRESS should reach a minimum and start to rise again. To avoid building a model that is either overfit or underfit, the number of components where the PRESS value reaches a minimum would be the obvious choice for the best model. While the minimum of the PRESS may be the best choice for predicting the particular set of samples, most likely it is not the optimum choice for predicting all unknown samples in the future. That is, the optimum number of factors was determined rather than the selection of the model, which yields a minimum in PRESS; the model selected is the one with the fewest number of factors such that

PRESS for that model is not significantly greater than the minimum PRESS (Niazi and Azizi, 2008). A solution to this problem has been suggested in which the PRESS values for all previous factors are compared to the PRESS value at the minimum.

#### 4. REAL LIFE DATA EXAMPLE, DESIGN OF SIMULATION STUDY AND RESULTS

In this paper, real life data are used and a simulation study is conducted to gain a better understanding of AIC, BIC, ICOMP, and Wold's R criteria performances for PLSR model selection; in fact it is a designed experimental simulation study for choosing the true latent variables. The experiment has various characteristics of the simulation models, in order to quantify the expected performance of information criteria. In the next subsection, the steps of data generation and performance results of criteria to PLSR model selection are shown by means of a simulation study. Additionally, in order to select model number of components to be retained in the final model, k-fold cross validation in kernel PLSR algorithm is used (Kohavi, 1995).

##### 4.1 Real Life Data Example

Performances of AIC, BIC, ICOMP and Wold's R criteria have been tested considering a real life dataset: the Body Fat Measurement. This data set has been used by Bozdoğan (2004) for subset selection of best predictors using Genetic Algorithms. In this data set, it is determined that the best subset predictors of  $y$ =Percent body fat from Siria (1956) equation, using  $k=13$  predictors are  $x_1$ =Age (years),  $x_2$ =Weight(lbs),  $x_3$ =Height (inches),  $x_4$ =Neck circumference (cm),  $x_5$ =Chest circumference(cm),  $x_6$ =Abdomen 2 circumference (cm),  $x_7$ =Hip circumference (cm),  $x_8$ =Thigh circumference (cm),  $x_9$ =Knee circumference (cm),  $x_{10}$ =Ankle circumference (cm),  $x_{11}$ =Biceps (extended) circumference (cm),  $x_{12}$  = Forearm circumference (cm),  $x_{13}$  =Wrist circumference (cm). The data contain the estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements for  $n = 252$  men. This is a good example to illustrate the versatility and utility of our approach using multiple regression analysis with GA. This data set is maintained by Dr. Roger W. Johnson of the Department of Mathematics & Computer Science at South Dakota School of Mines and Technology<sup>1</sup>. A variety of popular health books suggest that the readers assess their health, at least in part, by estimating their percentage of body fat. In Bailey (1994), for instance, the reader can estimate body fat from tables using their age and various skin-fold measurements obtained by using a caliper. Other texts give predictive equations for body fat using body circumference measurements (e.g. abdominal circumference) and/or skin-fold measurements. See, for instance, Behnke and Wilmore (1974); Wilmore (1976); or Katch and Mc Ardle (1977).Percentage of body fat for an individual can be estimated once body density has been determined. Siria (1956) assumes that the body consists of two components-lean body tissue and fat tissue.

Letting,

$$D = \text{Body Density (gm/cm}^3\text{)}$$

$$A = \text{proportion of lean body tissue}$$

$$B = \text{proportion of fat tissue (A+B=1)}$$

$$D = 1/[(A/a) + (B/b)]$$

$$B = (1/D)*[ab/(a-b)] - [b/(a-b)].$$

<sup>1</sup>E-mail: [rwjohnso@silver.sdsmt.edu](mailto:rwjohnso@silver.sdsmt.edu), and web address: <http://silver.sdsmt.edu/~rwjohnso>



a = density of lean body tissue (gm/cm<sup>3</sup>)  
 b = density of fat tissue (gm/cm<sup>3</sup>)

Using the estimates a=1.10gm/cm<sup>3</sup> and b=0.90 gm/cm<sup>3</sup> (Katch and McArdle, 1977) or Wilmore (1976), we come up with Siri's equation:

$$\text{Percentage of Body Fat (i.e. } 100*B) = 495/D - 450.$$

Volume, and hence body density, can be accurately measured by a variety of ways. The technique of underwater weighing computes body volume as the difference between body weight measured in air and weight measured during water submersion. In other words, body volume is equal to the loss of weight inwater with the appropriate temperature correction for the water's density (Katch and McArdle, 1977). Using this technique,

$$\text{Body Density} = \text{WA}/[(\text{WA}-\text{WW})/\text{c.f.} - \text{LV}]$$

where, WA=Weight in air (kg), WW=Weight in water (kg), c.f.=Water correction factor (=1 at 39.2 deg F as one-gram of water occupies exactly one cm<sup>3</sup> at this temperature, =.997 at 76-78 deg F), LV=Residual Lung Volume (liters) (Katch and McArdle, 1977). Other methods of determining body volume are given in Behnke and Wilmore (1974).

For this data set, PLSR model is established using Minitab package program tool, and validation technique is selected as k-fold cross validation, k=5. Then, the results of model selection and validation are as follows:

**Table 1. Model selection and validation for percent body fat**

Number of latentvariables	Relativecumulativevariance of components	Sum of squareerror	R-Square	PRESS	R-Sq (pred)
1	0.59	0.039	0.56	0.041	0.54
2	0.70	0.012	0.86	0.015	0.83
3	0.75	0.004	0.95	0.005	0.93
4	0.81	0.002	0.96	0.003	0.96
5	0.84	0.002	0.97	0.002	0.97
6	0.87	0.002	0.97	0.002	0.97
7		0.002	0.97	0.002	0.97
8		0.002	0.97	0.002	0.97
9		0.002	0.97	0.002	0.97
10		0.002	0.97	0.002	0.97

As seen from the results in Table 1, the number of latent variable is 6, and Figure 1 shows the model selection plot.



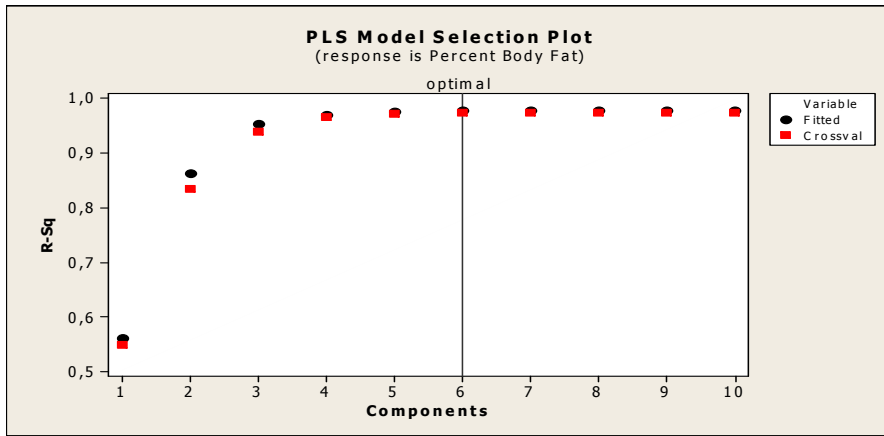


Figure 1. Partial least squares regression model selection plot

This data set is used to compare the performance of model selection criteria AIC, BIC, ICOMP and Wold's R. For this data set, AIC, BIC and ICOMP find the optimal number of components (6) whereas Wold's R finds it as 7.

#### 4.2 Design of Simulation Study

In this subsection, simulation experiments are performed to evaluate the performance comparisons of AIC, BIC, ICOMP, and Wold's R criteria. The framework for the simulation models are based on the study of Naes and Martens (1985), Li and et al. (2002). It is extended in this paper to the situation where there exist multiple response variables and different number of explanatory variables. In the simulation study, the multivariate regression models are first developed from which data are generated, and then model selection criteria are applied. The resulting models are then compared with the true models and finally an evaluation of the different criteria for PLSR model selection is made through a comparison of the success rate as to which the true model is selected (Bedrick and Tsai, 1994; Li et al., 2002).

The  $\mathbf{X}$  and  $\mathbf{Y}$  block data, with sample size  $N$ , are generated as:

$$\mathbf{X} = \sum_{i=1}^{A^*} \mathbf{r}_i \xi_i' + \tilde{\mathbf{E}}, \quad (15)$$

$$\mathbf{Y} = \sum_{i=1}^{A^*} \mathbf{z}_i \eta_{A^*i}' + \boldsymbol{\psi} = \sum_{i=1}^{A^*} \mathbf{r}_i \eta_{A^*i}' + \tilde{\mathbf{F}}_{A^*}, \quad (16)$$

where  $\tilde{\mathbf{E}}$  and  $\mathbf{r}_i$  are generated from mutually independent normal variables. It is noted that  $\text{var}(\mathbf{r}_i) + \text{var}(\mathbf{e}_j)$  is the largest eigenvalue of  $\text{cov}(\mathbf{X})$ .  $\boldsymbol{\psi}$  is generated from a multivariate normal distribution,  $\tilde{\mathbf{F}}$  is a noise matrix, and  $\mathbf{Z}$  is constructed as  $\mathbf{z}_i = \mathbf{r}_i + \mathbf{f}_i$ ,  $\mathbf{f}_i$  are generated as independent normal variables.  $\{\xi_i\}$  and  $\{\eta_{A^*i}\}$  are normalized orthogonal vector series, and  $\mathbf{r}_i$  are mutually independent random variables.

Comparing equation (2), with equations (15) and (16), it can be calculated that latent variable  $\mathbf{t}_i$ , loading vectors  $\mathbf{p}_i$  and  $\mathbf{q}_i$  obtained from PLSR algorithm are approximately

equal to  $\mathbf{r}_i$ ,  $\{\xi_i\}$  and  $\{\eta_{A^*i}\}$   $i=(1,\dots,A^*)$ , respectively. The  $\mathbf{Y}$ -block data,  $\mathbf{Y}$  of the response variables then essentially depends on  $\mathbf{r}_i$ ,  $i=(1,\dots,A^*)$ , plus noise. This means that the theoretical value of the number of latent variable is equal to  $A^*$  (Li et. al. 2002).

To carry out simulations run, it is proceeded on different simulation models and a fixed number of blocks for  $k$ -fold cross-validation in kernel algorithm;  $k$  is selected as 5. The dimensions of explanatory variables are extended as  $N \times 5$ ,  $N \times 8$ ,  $N \times 10$ ,  $N \times 15$ , and  $N \times 20$ . The dimensions of response variables matrix,  $\mathbf{Y}$ , are chosen as  $N \times 3$ ,  $N \times 4$ ,  $N \times 5$ , and sample sizes are selected as  $N=50, 100, 250, 500, 1000$ . For each of the combinations of parameters in Table 2, 10000 data sets are generated taking into account the dimension of partial least squares regression models and sample sizes, so that  $25 \times 10000$  data sets are generated.

**Table 2. The elements of experimental data sets**

Number of latent variables	The dimension of response variables matrix	The dimensions of explanatory variables	Sample sizes
5*3 <sup>1</sup>	$Y_{N \times 3}$	$X_{N \times 5}$	$N=\{50, 100, 250, 500, 1000\}$
8*4	$Y_{N \times 4}$	$X_{N \times 8}$	$N=\{50, 100, 250, 500, 1000\}$
10*4	$Y_{N \times 4}$	$X_{N \times 10}$	$N=\{50, 100, 250, 500, 1000\}$
15*5	$Y_{N \times 5}$	$X_{N \times 15}$	$N=\{50, 100, 250, 500, 1000\}$
20*5	$Y_{N \times 5}$	$X_{N \times 20}$	$N=\{50, 100, 250, 500, 1000\}$

<sup>1</sup>5\*3 shows that the number of predictor variables is 5, and these variables are reduced to number 3 for the number of latent variables.

Then these data sets are applied to AIC, BIC, ICOMP, and Wold’s R criteria. Explanatory data matrix,  $\mathbf{X}$ , is generated from equation (15), and  $\mathbf{Y}$ , is generated from equation (16). Generation of  $\mathbf{X}$  and  $\mathbf{Y}$  data matrices are just explained for 5\*3 which is shown in Table 3, and the other data matrices are given in Appendix. The components of  $\mathbf{X}$  and  $\mathbf{Y}$  data matrices are given in Table 3 ( $i=[1,\dots,A^*]$ ,  $A^*=3$ ).

**Table 3. The components values of X and Y matrices for 5\*3**

$\tilde{\mathbf{E}}$	$\tilde{\mathbf{E}}=[e_1,\dots,e_3] \sim N(0,0.02)$
$\mathbf{r}_i$	$r_1 \sim N(0,15), r_2 \sim N(0,7.5), r_3 \sim N(0,3)$
$\xi_i$	$\xi_1 = [0.6247 \quad 0.5635 \quad 0.4472 \quad 0.2871 \quad 0.0989]'$ $\xi_2 = [0.5635 \quad 0.0989 \quad -0.4472 \quad -0.6247 \quad -0.2871]'$ $\xi_3 = [0.4472 \quad -0.4472 \quad -0.4472 \quad 0.4472 \quad 0.4472]'$
$\mathbf{f}_i$	$f_1 \sim N(0,0.5), f_2 \sim N(0,0.25), f_3 \sim N(0,0.1)$
$\eta_{A^*i}$	$\eta_{31} = [0.7887 \quad 0.5774 \quad 0.2113]'$ $\eta_{32} = [0.5774 \quad -0.5774 \quad -0.5774]'$ $\eta_{33} = [0.2113 \quad -0.5774 \quad 0.7887]'$
$\Psi$	$\Psi = \begin{pmatrix} 0.00010 & 0.00006 & 0.00006 \\ 0.00006 & 0.00010 & 0.00006 \\ 0.00006 & 0.00006 & 0.00010 \end{pmatrix}$

The variance inflation factor (VIF) values for 5\*3 design matrix shows that there is multicollinearity (Table 4). The VIF values are calculated by Minitab package program. Table 5 shows the relative cumulative variances by the five latent variables for the X and Y blocks, averaged over 10000 simulation experiments. It can be seen from the first two rows of Table 5 that on average, for  $A^* = 3$ , first three latent variables capture 100% and 98% of the variances in the X and Y data sets, respectively. This verifies the theoretical value of the number of latent variables  $A^* = 3$ .

**Table 4. The VIF values for 5\*3. (N=100, k=5)**

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
VIF	67.6	238.0	142.0	72.3	42.8

**Table 5. Relative cumulative variances of X and Y for 5\*3. (N=100, k=5)**

True Model	Blocks	Number of Latent Variables				
		1	2	3	4	5
$A^*=3$	X-block	0.55	0.89	1.00	1.00	1.00
	Y-block	0.71	0.84	0.98	0.98	0.98

### 4.3 Results of Simulation Study and Performance Comparison of Model Selection Criteria

In this study, we compare the performance of model selection criteria by using the percentage of success which shows the precision in finding the number of latent variables by model selection criteria. We also compute performances of all criteria for Li et. al. (2002) data when dimension to reduction PLSR model is 6\*4, and the results are shown in Table 6. As seen from the results, AIC, BIC, ICOMP methods provide the best selection of the number of latent variables.

**Table 6. Comparison of the percentages of the selected number of latent variables for 6\*4**

N	AIC	BIC	ICOMP	MAIC	Wold's R
100	100	100	100	84.0*	47.6*
1000	100	100	100	75.8*	49.0*

\* MAIC and Wold's R results are taken from Li et al. study (2002).

All results of the simulations for various sample sizes and dimensions are given in Table 7 and these are obtained by 10000 replications. It illustrates the ability of AIC, BIC, ICOMP, and Wold's R criteria in selecting latent variables for all situations.

**Table 7. Percentages of performance comparison for each criterion**

N	AIC					BIC					ICOMP					Wold's R				
	50	100	250	500	1000	50	100	250	500	1000	50	100	250	500	1000	50	100	250	500	1000
5*3	100	100	100	99	100	100	100	100	99	100	100	100	100	99	100	60	58	53	50	49
8*4	100	100	100	100	100	100	100	100	100	100	99	100	100	100	100	29	20	11	23	13
10*4	59	88	87	73	51	57	88	91	81	64	61	88	91	81	67	47	41	54	43	32
15*5	45	35	24	24	23	37	35	27	34	27	50	35	26	30	26	47	45	47	43	41
20*5	63	59	58	55	43	59	59	64	63	58	67	62	64	63	58	46	47	45	43	42

As it can be seen from Table 7, for all experimental data sets, almost AIC, BIC, and ICOMP criteria have similar performances except Wold's R criterion. It has the lowest success rate compared to other criteria. For  $(p^*a) \leq (8^*4)$ , AIC, BIC, and ICOMP criteria perform better than the Wold's R criterion. When  $(p^*a) > (8^*4)$  the success rates of AIC, BIC, ICOMP and Wold's R criteria decrease as the sample sizes and the dimensions of models increase. Overall ICOMP criterion provides the best selection of the number of latent variables among AIC, BIC, and Wold's R criteria.

These comparisons of performances are graphically presented in Figure 2(a)-(e) and Figure 3(a)-(f).

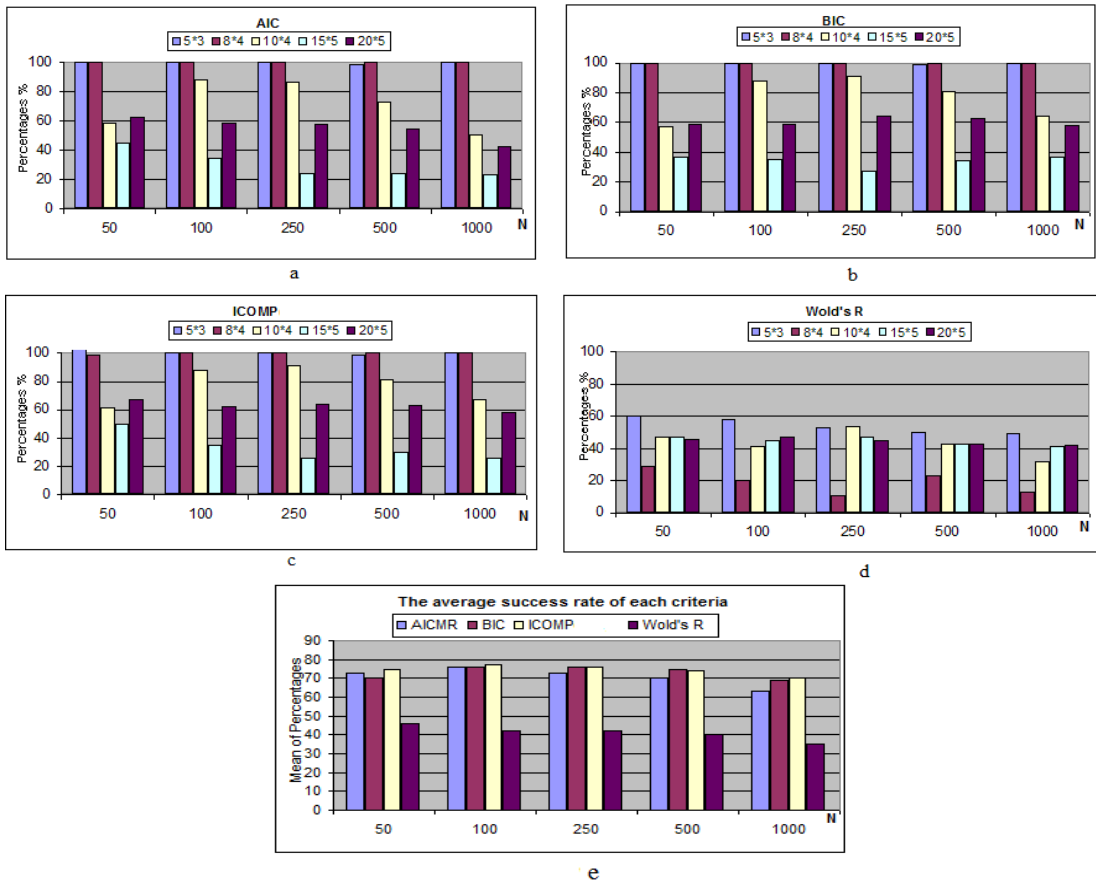


Figure 2. (a)-(d) Performance comparison of each model selection criterion for various dimension of models and sample sizes. (e) The average success rate of model selection criterion for various sample sizes

As can be seen from Figure 2 (a)-(f), there is dependency among dimension of the PLSR models, sample sizes and model selection criteria. AIC, BIC, and ICOMP truly estimate the latent variables of the underlying known PLSR models for the dimension  $5^*3$  and the dimension  $8^*4$ . When the sample sizes increase and the dimension of PLSR models is constant, these criteria have a slight tendency to over-fit their PLSR models. The simulation results show that BIC and ICOMP criteria achieved selecting the true number of latent variables with such a rate of approximately eighty percent for all design matrices. Generally it can be said that, when N and the dimension of PLSR

models increases, PLS creates a model with a high number of latent variables, which is statistically significant.

While the dimensions of the PLSR model change and the sample size is constant, variation in the criteria of performances is shown in Figure 3 (a)-(f).

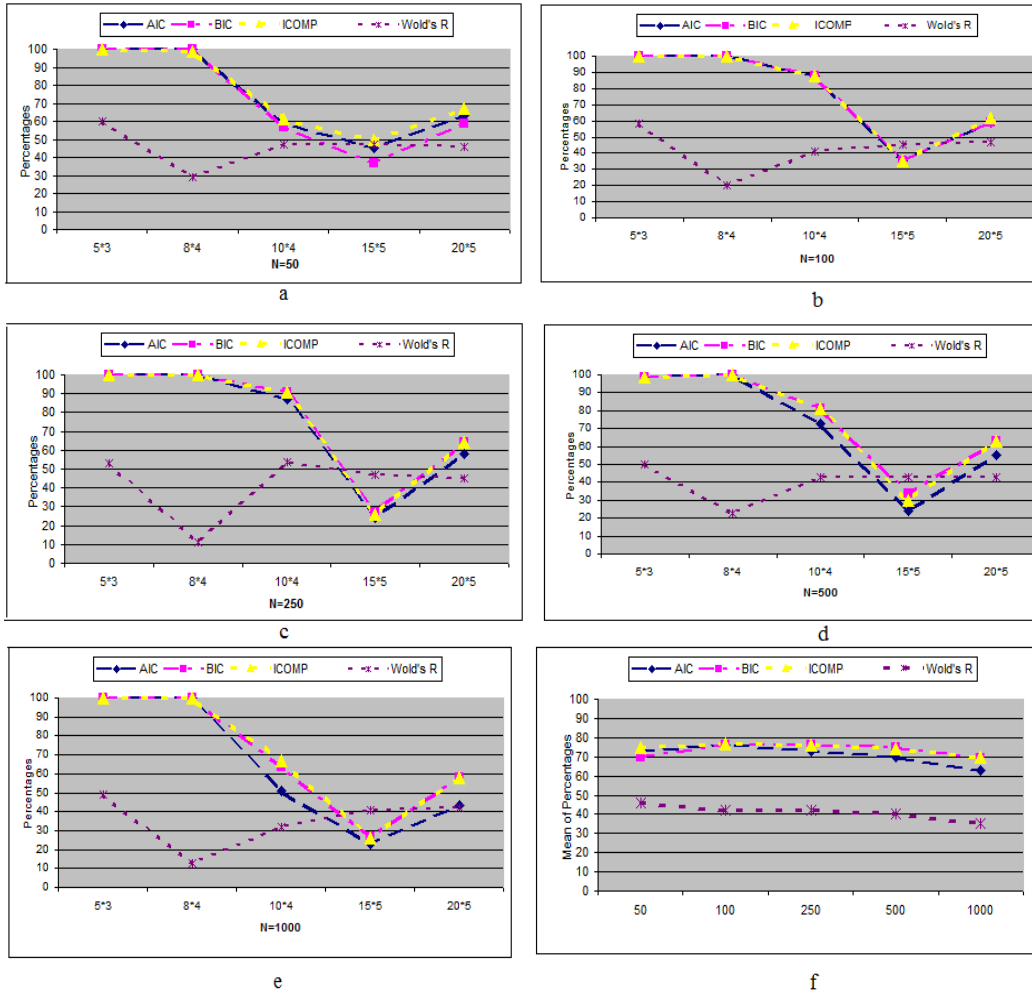


Figure 3. (a)-(e) Performance comparison of each model selection criterion for various dimensions of models for each sample size. (f) The average success rate of model selection criterion for various sample sizes

In each experiment all model selection criteria is applied to test how well they can identify the true known PLSR model. Figure 3(a)-(f) show the success rate of each criterion in identifying the true model. Since the performance of every criterion can be affected by the sample sizes and the dimension of PLSR models, the performances of AIC, BIC, and ICOMP criteria in general show a similar characteristic. Especially, when the dimension of models is smaller than  $(p * a) \leq (10 * 4)$ , AIC, BIC, and ICOMP criteria have acceptable performance and almost more accurately select the latent variables than the Wold's R criterion for all dimensions. As shown in Figure 3 (a)-(e), Wold's R criterion does not work well for any sample size and dimension. BIC and ICOMP criteria perform quite well, and in general select the true number of latent variables for known PLSR models.

Results depicted in Figure 3(a)-(e) clearly show that there is a significant reduction in the performances of AIC, BIC, and ICOMP criteria as the dimensions of the PLSR model increase. However, the performance of the Wold's R criterion almost stays the same but it is never satisfactory. In order to provide an overall measure of success, the average success rate is calculated and shown in Figure 2(f) and Figure 3(f) for various sample sizes. It can be seen from this figure that on the average AIC, BIC, and ICOMP criteria outperform Wold's R criterion, and BIC and ICOMP criteria success rates are higher than the AIC and Wold's R criteria.

## 5. CONCLUSION

The major contribution of this paper is that this study evaluating the performances of AIC, BIC, ICOMP, and Wold's R criteria for model selection in PLSR, where the number of observations is typically much larger than the number of predictor variables. The aim of the analysis is to extract latent variables with respect to their partial contribution to total variance to build representative models of PLSR. Given this ranking to the latent variables, AIC, BIC, ICOMP, and Wold's R criteria are used to determine a consistent estimate of the dimension of the model. In a real life data set AIC, BIC, ICOMP criteria truly find the number of latent variables. It seems that AIC, BIC and ICOMP criteria are considerably better in choosing the right model when these are applied to our data set.

A simulation study is undertaken to compare the performances of AIC, BIC, ICOMP, and Wold's R criteria. Synthetic data are generated for different number of sample sizes and different dimensions of PLSR models. The simulation studies results clearly show much improved performances of BIC and ICOMP criteria in comparison to AIC and Wold's R criteria methods. The AIC, BIC, and ICOMP criteria properly find latent variables for  $(p \cdot a) < (10 \cdot 4)$ , for all sample sizes except the Wold's R criterion. It is seen from Table 7 that there is a dependency between dimension and sample size of the PLSR models and the success rates of the model selection criteria except for the Wold's R criterion.

In conclusion, these experiments showed that BIC and ICOMP criteria are considerably better than the traditional methods (AIC and Wold's R criteria) in choosing the right model when it is applied to our experimental set of synthetic data. An important point to make is that there is a big difference between the performances of AIC, BIC, ICOMP for different number of observations and the dimensions of PLSR models. Thus, one should select the sample size and the dimensions of the experiment in advance depending on the success rates of the criteria given in this study.

## 6. REFERENCES

Abdi, H., Salkind N. 2007. Encyclopedia of Measurement and Statistics. Thousand Oaks (CA): Sage.

Akaike, H., 1974. A new look at the statistical model identification, IEEE Transaction on Automatic Control 19, 716-723.

- Bailey, C., 1994. *Smart Exercise: Burning Fat, Getting Fit*. Houghton-Mifflin Co., Boston, pp: 179-186.
- Bedrick, E. J., Tsai, C. L., 1994. Model selection for multivariate regression in small samples. *Biometrics* 50, 226-231.
- Behnke, A. R., J.H. Wilmore, 1974. *Evaluation and Regulation of Body Build and Composition*. Prentice-Hall, Englewood Cliffs, N. J., Pages: 236.
- Boyer, K. L., Mirza, M. J., Ganguly, G., 1994. The Robust Sequential Estimator: A General Approach and its Application to Surface Organization in Range Data, *IEEE PAMI* 16, 987-1001.
- Bozdoğan, H., 1987. Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions, *Psychometrica* 52, 345-370.
- Bozdoğan, H., 2000. Akaike's Information Criterion and Recent Developments in Information Complexity. *Journal of Mathematical Psychology* 44, 62-91.
- Bozdoğan, H., 2004. *Statistical Data Mining and Knowledge Discovery*. Chapman and Hall/CRC, USA.
- Bozdoğan, H., 2004. *Intelligent statistical data mining with Information Complexity and Genetic Algorithms in Statistical Data Mining and Knowledge Discovery*. Chapman and Hall/CRC, USA.
- Clark, A. E., Troskie, C. G., 2006. Regression and ICOMP: A Simulation Study. *Communications in Statistics Simulation and Computation* 35, 591-603.
- Eastment H. T., Krzanowski W. J. 1982. Cross-validators choice of the number of components from a principal component analysis. *Technometrics* 24, 73-77.
- Garthwaite, P. H., 1994. An interpretation of partial least squares, *Journal of the American Statistical Association* 89, 122-127.
- Geladi, P., Kowalski, B. R., 1986. Partial least-squares regression a tutorial. *Anal. Chim. Acta.* 185, 1-17.
- Haber, R., Unbenhauen, H., 1990. Structure identification of nonlinear dynamic systems-a survey on input/output approaches. *Automatica* 26 (4), 651-677.
- Hastie, T., Tibshirani, R., Friedman J., 2001. *The elements of statistical learning: data mining, inference, and prediction*. New York, Springer.
- Helland, I. S. 1990. Partial Least Squares Regression and Statistical Models, *Scandinavian Journal of Statistics* 17(2), 97-114.
- Henry de-Graft, A. 2010. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics* 2(1): 001-006.



- Jouan-Rimbaud Bouveresse, D., Rutledge, D. N., 2009. Two new extensions of principal component transform to compute a PLS2 model between two wide matrices: PCT-PLS2 and segmented PCT-PLS2. *Analytica Chimica Acta*. 642 (1-2), 37-44.
- Katch, F., W. McArdle, 1977. *Nutrition, Weight Control and Exercise*. Houghton-MifflinCo., Boston.
- Kanatani, K., 2002. Model Selection for Geometric Inference, The 5th Asian Conference on Computer Vision, Melbourne, Australia, pp. xxi-xxxii, January.
- Kohavi, R., 1995. A study of cross-validation and boots trap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, San Francisco, CA, USA, pp. 1137–1143. Morgan Kaufmann Publishers Inc.
- Kuha, J. 2004. AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods and Research*. (33)2: 188-229.
- Kundu, D., Murali G., 1996. Model selection in linear regression, *Computational Statistics and Data Analysis* 22 (5), 461-469(9).
- Li, B., Morris, J., Martin E. B., 2002. Model Selection for Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems* 64, 79-89.
- Myung, I. J., 2000. The Importance of Complexity in Model Selection. *Journal of Mathematical Psychology* 44, 190-204.
- Naes, T., Martens, H., 1985. Comparison of prediction methods for collinear data. *Communication in Statistics Simulation and Computation* 14, 545-576.
- Niazi, A., Azizi, A., 2008. Orthogonal Signal Correction-Partial Least Squares Method for Simultaneous Spectrophotometric Determination of Nickel, Cobalt and Zinc. *Turkish Journal of Chemistry* 32, 217-228.
- Saporta, G., 2008. Models for Understanding versus Models for Prediction. In *Compstat 2008, Part IX*, 315-322.
- Schwarz, G., 1978. Estimating the Dimension of a Model. *The Annals of Statistics* 6 (2), 461-464.
- Siria, W. E., 1956. Gross Composition of the Body. In *Advance in Biological and Medical Physics*, Lawrence J.H. and C.A. Tobias (Eds.). Academic Press, New York.
- Torr, P. H. S., 1998. Model Selection for Two View Geometry: A Review, Model Selection for Two View Geometry: A Review, Microsoft Research, USA, Microsoft Research, USA.

Wang Y, Liu, Q., 2006. Comparison of Akaike information criteria (AIC) and Bayesian information criteria (BIC) in selection of stock recruitment relationships. Fisheries Research 77(2): 220-225.

Weakliem, L. D., 2004. Introduction to the Special Issue on Model Selection. Sociological Methods and Research. 33(2): 167-186.

Wilmore, J., 1976. Athletic Training and Physical Fitness: Physiological Principles of the Conditioning Process. Allynand Bacon, Inc., Boston.

Wold, H., 1966. Estimation of principal components and related models by iterative least squares. In P.R. Krishnaiah (Ed.). Multivariate Analysis. (pp.391-420) New York: Academic Press.

Wold, H., 1982. Soft Modelling, The basic design and some extensions, in: K.- G. Jöreskog, H. Wold (Eds.), Systems Under Indirect Observation. Vols.I and II, North-Holland, Amsterdam.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory 58, 109-130.

## KEKKR İÇİN MODEL SEÇME KRİTERLERİNİN PERFORMANS KARŞILAŞTIRMALARI: AIC, BIC, ICOMP ve WOLD'S R

### ÖZET

*Kısmi en küçük kareler regresyonu (KEKKR), çoklu bağlantının olduğu durumlarda, yanıt değişkeni  $Y_{N \times M}$  ile açıklayıcı değişkenler  $X_{N \times K}$  arasında modelleme yapabilen istatistiksel bir yöntemdir. Kısmi en küçük kareler bölümünde, yanıt değişkenini en iyi açıklayabilecek gizli (latent) değişkenlerin elde edilmesi için bazı model seçme kriterleri uygulanır. Gizli değişkenlerin seçiminde kullanılan genel yaklaşımlar Akaike bilgi kriteri (AIC) ve Wold's R kriteridir.*

*Bu çalışmada, gözlem sayısının açıklayıcı değişken sayısından fazla olduğu durumlarda, geleneksel yöntemler AIC ve Wold's R'a ek olarak Bayes bilgi kriteri (BIC) ve Bilgi karmaşıklık kriteri de (ICOMP) KEKKR için model seçme kriterleri olarak incelenmiştir. AIC, BIC, ICOMP ve Wold's R model seçme kriterlerinin performansları gerçek veri örneği ve benzetim çalışması yoluyla karşılaştırılmıştır. Benzetim çalışması sonuçları, farklı örneklem büyüklükleri, farklı sayıda açıklayıcı değişken ve yanıt değişkeninin olduğu durumlarda elde edilmiştir. Yapılan benzetim çalışması sonuçları BIC ve ICOMP model seçme kriterlerinin KEKKR modelleri için, gizli değişkenin seçiminde diğer model seçme kriterlerinden (AIC ve Wold's R) çok daha etkili olduklarını ve daha doğru sayıda gizli değişken seçimi yaptıklarını göstermiştir.*

**Anahtar Kelimeler:** AIC, BIC ve ICOMP bilgi kriterleri, K çapraz doğrulama, Kısmi en küçük kareler regresyonu, Model seçimi, Wold's R kriteri.

APPENDIX

Table 1. The components of X and Y matrix for 8\*4

$\tilde{\mathbf{E}}$	$\tilde{\mathbf{E}}=[e_1, \dots, e_8] \sim N(0, 0.01)$
$\mathbf{r}_i$	$r_1 \sim N(0, 10), r_2 \sim N(0, 5), r_3 \sim N(0, 2), r_4 \sim N(0, 0.5)$
$\xi_i$	$\xi_1 = [0.1612 \ 0.3030 \ 0.4082 \ 0.4642 \ 0.4642 \ 0.4082 \ 0.3030 \ 0.1612]'$ $\xi_2 = [0.3030 \ 0.4642 \ 0.4082 \ 0.1612 \ -0.1612 \ -0.4082 \ -0.4642 \ -0.3030]'$ $\xi_3 = [0.4082 \ 0.4082 \ 0.0000 \ -0.4082 \ -0.4082 \ -0.0000 \ 0.4082 \ 0.4082]'$ $\xi_4 = [0.4642 \ 0.1612 \ -0.4082 \ -0.3030 \ 0.3030 \ 0.4082 \ -0.1612 \ -0.4642]'$
$\mathbf{f}_i$	$f_1 \sim N(0, 0.25), f_2 \sim N(0, 0.125), f_3 \sim N(0, 0.05), f_4 \sim N(0, 0.0125)$
	$\eta_{41} = [0.2887 \ 0.5000 \ 0.5774 \ 0.5000]'$ $\eta_{42} = [0.5000 \ 0.5000 \ 0.0000 \ -0.5000]'$ $\eta_{43} = [0.5774 \ 0.0000 \ -0.5774 \ -0.0000]'$ $\eta_{44} = [0.5000 \ -0.5000 \ -0.0000 \ 0.5000]'$
$\Psi$	$\Psi = \begin{pmatrix} 0.00010 & 0.00006 & 0.00006 & 0.00006 \\ 0.00006 & 0.00010 & 0.00006 & 0.00006 \\ 0.00006 & 0.00006 & 0.00010 & 0.00006 \\ 0.00006 & 0.00006 & 0.00006 & 0.00010 \end{pmatrix}$

Table 2. The components of X and Y matrix for 10\*4

$\tilde{\mathbf{E}}$	$\tilde{\mathbf{E}}=[e_1, \dots, e_{10}] \sim N(0, 0.02)$
$\mathbf{r}_i$	$r_1 \sim N(0, 20), r_2 \sim N(0, 10), r_3 \sim N(0, 4), r_4 \sim N(0, 1)$
$\xi_i$	$\xi_1 = [0.4458 \ 0.4349 \ 0.4132 \ 0.3813 \ 0.3401 \ 0.2904 \ 0.2337 \ 0.1711 \ 0.1044 \ 0.0351]'$ $\xi_2 = [0.4349 \ 0.3401 \ 0.1711 \ -0.0351 \ -0.2337 \ -0.3813 \ -0.4458 \ -0.4132 \ -0.2904 \ -0.1044]'$ $\xi_3 = [0.4132 \ 0.1711 \ -0.1711 \ -0.4132 \ -0.4132 \ -0.1711 \ 0.1711 \ 0.4132 \ 0.4132 \ 0.1711]'$ $\xi_4 = [0.3813 \ -0.0351 \ -0.4132 \ -0.3401 \ 0.1044 \ 0.4349 \ 0.2904 \ -0.1711 \ -0.4458 \ -0.2337]'$
$\mathbf{f}_i$	$f_1 \sim N(0, 0.5), f_2 \sim N(0, 0.25), f_3 \sim N(0, 0.1), f_4 \sim N(0, 0.025)$
$\eta_{A_i}$	$\eta_{41} = [0.6935 \ 0.5879 \ 0.3928 \ 0.1379]'$ $\eta_{42} = [0.5879 \ -0.1379 \ -0.6935 \ -0.3928]'$ $\eta_{43} = [0.3928 \ -0.6935 \ 0.1379 \ 0.5879]'$ $\eta_{44} = [0.1379 \ -0.3928 \ 0.5879 \ -0.6935]'$
$\Psi$	$\Psi = \begin{pmatrix} 0.00010 & 0.00006 & 0.00006 & 0.00006 \\ 0.00006 & 0.00010 & 0.00006 & 0.00006 \\ 0.00006 & 0.00006 & 0.00010 & 0.00006 \\ 0.00006 & 0.00006 & 0.00006 & 0.00010 \end{pmatrix}$

Table 3. The components of X and Y matrix for 15\*5

$\tilde{\mathbf{E}}$	$\tilde{\mathbf{E}}=[e_1, \dots, e_{15}] \sim N(0, 0.02)$
$\mathbf{r}_i$	$r_1 \sim N(0, 20), r_2 \sim N(0, 10), r_3 \sim N(0, 5), r_4 \sim N(0, 3.5), r_5 \sim N(0, 0.8)$
$\xi_i$	$\xi_{1(1, \dots, 10)} = [0.3646 \ 0.3607 \ 0.3527 \ 0.3409 \ 0.3253 \ 0.3062 \ 0.2838 \ 0.2582 \ 0.2298 \ 0.1989]'$ $\xi_{1(11, \dots, 15)} = [0.1658 \ 0.1309 \ 0.0945 \ 0.0571 \ 0.0191]'$ $\xi_{2(1, \dots, 10)} = [0.3607 \ 0.3253 \ 0.2582 \ 0.1658 \ 0.0571 \ -0.0571 \ -0.1658 \ -0.2582 \ -0.3253 \ -0.3607]'$ $\xi_{2(11, \dots, 15)} = [-0.3607 \ -0.3253 \ -0.2582 \ -0.1658 \ -0.0571]'$ $\xi_{3(1, \dots, 10)} = [0.3527 \ 0.2582 \ 0.0945 \ -0.0945 \ -0.2582 \ -0.3527 \ -0.3527 \ -0.2582 \ -0.0945 \ 0.0945]'$ $\xi_{3(11, \dots, 15)} = [0.2582 \ 0.3527 \ 0.3527 \ 0.2582 \ 0.0945]'$ $\xi_{4(1, \dots, 10)} = [0.3409 \ 0.1658 \ -0.0945 \ -0.3062 \ -0.3607 \ -0.2298 \ 0.0191 \ 0.2582 \ 0.3646 \ 0.2838]'$ $\xi_{4(11, \dots, 15)} = [0.0571 \ -0.1989 \ -0.3527 \ -0.3253 \ -0.1309]'$ $\xi_{5(1, \dots, 10)} = [0.3253 \ 0.0571 \ -0.2582 \ -0.3607 \ -0.1658 \ 0.1658 \ 0.3607 \ 0.2582 \ -0.0571 \ -0.3253]'$ $\xi_{5(11, \dots, 15)} = [-0.3253 \ -0.0571 \ 0.2582 \ 0.3607 \ 0.1658]'$
$\mathbf{f}_i$	$f_1 \sim N(0, 0.05), f_2 \sim N(0, 0.025), f_3 \sim N(0, 0.0125), f_4 \sim N(0, 0.05), f_5 \sim N(0, 0.0125)$
$\eta_{A^i}$	$\eta_{s1} = [0.6247 \ 0.5635 \ 0.4472 \ 0.2871 \ 0.0989]'$ $\eta_{s2} = [0.5635 \ 0.0989 \ -0.4472 \ -0.6247 \ -0.2871]'$ $\eta_{s3} = [0.4472 \ -0.4472 \ -0.4472 \ 0.4472 \ 0.4472]'$ $\eta_{s4} = [0.2871 \ -0.6247 \ 0.4472 \ 0.0989 \ -0.5635]'$ $\eta_{s5} = [0.0989 \ -0.2871 \ 0.4472 \ -0.5635 \ 0.6247]'$
$\Psi$	$\Psi = \begin{pmatrix} 0.00010 & 0.00006 & 0.00006 & 0.00006 & 0.00006 \\ 0.00006 & 0.00010 & 0.00006 & 0.00006 & 0.00006 \\ 0.00006 & 0.00006 & 0.00010 & 0.00006 & 0.00006 \\ 0.00006 & 0.00006 & 0.00006 & 0.00010 & 0.00006 \\ 0.00006 & 0.00006 & 0.00006 & 0.00006 & 0.00010 \end{pmatrix}$

Table 4. The components of X and Y matrix for 20\*5

$\tilde{\mathbf{E}}$	$\tilde{\mathbf{E}}=[e_1, \dots, e_{20}] \sim N(0, 0.02)$
$\mathbf{r}_i$	$r_1 \sim N(0, 30), r_2 \sim N(0, 20), r_3 \sim N(0, 10), r_4 \sim N(0, 6), r_5 \sim N(0, 3)$
$\xi_i$	$\xi_{1(1, \dots, 10)} = [0.3160 \ 0.3140 \ 0.3102 \ 0.3044 \ 0.2967 \ 0.2872 \ 0.2759 \ 0.2629 \ 0.2483 \ 0.2322]$ $\xi_{1(11, \dots, 20)} = [0.2147 \ 0.1958 \ 0.1757 \ 0.1545 \ 0.1324 \ 0.1095 \ 0.0858 \ 0.0617 \ 0.0372 \ 0.0124]$ $\xi_{2(1, \dots, 10)} = [0.3140 \ 0.2967 \ 0.2629 \ 0.2147 \ 0.1545 \ 0.0858 \ 0.0124 \ -0.0617 \ -0.1324 \ -0.1958]$ $\xi_{2(11, \dots, 20)} = [-0.2483 \ -0.2872 \ -0.3102 \ -0.3160 \ -0.3044 \ -0.2759 \ -0.2322 \ -0.1757 \ -0.1095 \ -0.0372]$ $\xi_{3(1, \dots, 10)} = [0.3102 \ 0.2629 \ 0.1757 \ 0.0617 \ -0.0617 \ -0.1757 \ -0.2629 \ -0.3102 \ -0.3102 \ -0.2629]$ $\xi_{3(11, \dots, 20)} = [-0.1757 \ -0.0617 \ 0.0617 \ 0.1757 \ 0.2629 \ 0.3102 \ 0.3102 \ 0.2629 \ 0.1757 \ 0.0372]$ $\xi_{4(1, \dots, 10)} = [0.3044 \ 0.2147 \ 0.0617 \ -0.1095 \ -0.2483 \ -0.3140 \ -0.2872 \ -0.1757 \ -0.0124 \ 0.1545]$ $\xi_{4(11, \dots, 20)} = [0.2759 \ 0.3160 \ 0.2629 \ 0.1324 \ -0.0372 \ -0.1958 \ -0.2967 \ -0.3102 \ -0.2322 \ -0.0617]$ $\xi_{5(1, \dots, 10)} = [0.2872 \ 0.0858 \ -0.1757 \ -0.3140 \ -0.2322 \ 0.0124 \ 0.2483 \ 0.3102 \ 0.1545 \ -0.1095]$ $\xi_{5(11, \dots, 20)} = [-0.2967 \ -0.2759 \ -0.0617 \ 0.1958 \ 0.3160 \ 0.2147 \ -0.0372 \ -0.2629 \ -0.3044 \ -0.0124]$
$\mathbf{f}_i$	$f_1 \sim N(0, 0.4), f_2 \sim N(0, 0.1), f_3 \sim N(0, 0.5), f_4 \sim N(0, 0.02), f_5 \sim N(0, 0.00125)$
$\boldsymbol{\eta}_{A_i}$	$\eta_{s1} = [0.4472 \ 0.4472 \ 0.4472 \ 0.4472 \ 0.4472]'$ $\eta_{s2} = [0.4472 \ 0.5635 \ -0.0989 \ -0.6247 \ -0.2871]'$ $\eta_{s3} = [0.4472 \ -0.0989 \ -0.2871 \ 0.5635 \ -0.6247]'$ $\eta_{s4} = [0.4472 \ -0.6247 \ 0.5635 \ -0.2871 \ -0.0989]'$ $\eta_{s5} = [0.4472 \ -0.2871 \ -0.6247 \ -0.0989 \ 0.5635]'$
$\Psi$	$\Psi = \begin{pmatrix} 0.00010 & 0.00006 & 0.00006 & 0.00006 & 0.00006 \\ 0.00006 & 0.00010 & 0.00006 & 0.00006 & 0.00006 \\ 0.00006 & 0.00006 & 0.00010 & 0.00006 & 0.00006 \\ 0.00006 & 0.00006 & 0.00006 & 0.00010 & 0.00006 \\ 0.00006 & 0.00006 & 0.00006 & 0.00006 & 0.00010 \end{pmatrix}$