Variable Selection in Multivariable Regression Using SAS/IML

By

Ali A. Al-Subaihi
Institute of Public Administration
Riyadh 11141, Saudi Arabia
subaihia@ipa.edu.sa

Abstract

This paper introduces a SAS/IML program to select among the multivariate model candidates based on a few well-known multivariate model selection criteria. Stepwise regression and all-possible-regression are considered. The program is user friendly and requires the user to paste or read the data at the beginning of the module, include the names of the dependent and independent variables (the **y**'s and the **x**'s), and then run the module. The program produces the multivariate candidate models based on the following criteria: Forward Selection, Forward Stepwise Regression, Backward Elimination, Mean Square Error, Coefficient of Multiple Determination, Adjusted Coefficient of Multiple Determination, Akaike's Information Criterion, the Corrected Form of Akaike's Information Criterion, Hannan and Quinn Information Criterion, the Corrected Form of Hannan and Quinn (HQ_c) Information Criterion, Schwarz's Criterion, and Mallow's $\mathbf{C_P}$. The output also constitutes detailed as well as summarized results.

Keywords: Multivariate model selection; SAS/IML module; Stepwise regression; All-possible-regression.

1. Introduction

Applications where several quantities are to be predicted using a common set of predictor variables are becoming increasingly important in various disciplines (Breiman & Friedman, 1997; Bilodeau & Brenner, 1999). For instance, in a manufacturing process one may want to predict various quality aspects of a product from the parameter setting used in the manufacturing. Or, given the mass spectra of a sample, the goal may be to predict the concentrations of several chemical constituents in the sample (Breiman & Friedman, 1997). A natural class of models that accommodate this would be generalization of a univariate multiple regression model, called multivariate multiple regression (MMR). In MMR, q dependent variables (y_1, y_2, …, y_q) are to be predicted by linear relationships with k independent variables (x_1, x_2, …, x_k).

The statistical linear model for the MMR model is

$$\boldsymbol{Y}_{n\times q} = \boldsymbol{X}_{n\times(k+1)}\mathbf{B}_{(k+1)\times q} + \mathbf{E}_{n\times q} \qquad (1.1)$$

where $\boldsymbol{Y}$ represents n (independent) observations of a q-variate normal random variate, $\boldsymbol{X}$ represents the design matrix of rank $k+1$ with its first column being the vector 1, **B** is a matrix of parameters to be estimated and **E** represents the matrix of residuals.

In practice, MMR uses include a large number of predictors where some of them might be slightly correlated with the **y**'s or they may be redundant because of high correlations with other **x**'s (Spark et al., 1985). The use of poor or redundant predictors can be harmful because the potential gain in accuracy attributable to their inclusion is outweighed by inaccuracies associated with estimating their proper contribution to the prediction (Spark et al., 1985).

The problem of determining the "best" subset of independent variables in multiple linear regression has long been of interest to applied statisticians, and it continues to receive considerable attention in recent statistical literature (McQuarrie & Tsai, 1998). Two approaches

are suggested in the statistical literature to deal with this problem. The *first* approach is to find the "best" set of predictors for each individual response variable using one (or more) of the multiple model selection criteria that are available in most of statistical packages such as S-*plus*, SAS, SPSS etc. In this approach, researchers perform model selection procedures on a univariate basis q times, where q is the number of the **y**'s in the model. This can lead to q different subset of predictors, one for each **y**. The *second* approach is to find the "best" set of predictors for all response variables simultaneously, where one subset of predictors that "best" predict all **y**'s using an analogous matrix expression of one of the univariate variable selection criteria is selected.

Sparks et al. (1985) criticized univariate model selection methodology as compared to multivariate techniques and stated two reasons for dealing with target variables jointly rather than separately. One reason is simply that it is computationally more efficient because the number of times required doing necessary computations for model selection would be reduced from q to one. A second reason is that researchers sometimes need to establish which subset of predictors can be expected to perform well for all target variables, especially if there are costs associated with sampling the predictors.

Although the second approach is becoming increasingly important in various disciplines, to-date statistical software such as SAS and SPSS cannot be utilized to implement the second approach (SAS/STAT User's guide, 1990; and SPSS Base System, 1992). In this paper, we present a SAS module to select the "best" subset of predictors that can be conveniently used to predict all **y**'s jointly utilizing popular multivariate model selection criteria. Our SAS module performs model selection using three automatic search procedures (Forward Selection, Forward Stepwise Regression, and Backward Elimination), and nine all-possible- regression procedures (**MSE, R^2, *Adj*R^2**, **AIC**, **AIC_C**, **HQ**, **QH_C**, **BIC**, and **C_p**).

Spark et al. (1983) were the first to introduce the multivariate version of variable selection using the multivariate $\mathbf{C}_p$-statistic. Later, Spark et al. (1985) presented a multivariate selection method that uses the mean squared error of prediction rather than tests of hypotheses as the basis for selection. They also discussed the relationship between these two approaches. Bedrich and Tsai (1994) developed a small-sample criterion ($\mathbf{AIC}_C$), which adjusts the Akaike information criterion (**AIC**) to be an exact unbiased estimator for the expected Kullback-Liebler information, for selecting MMR models. Another modification of **AIC** and $\mathbf{C}_p$ has been proposed by Fujikoshi and Satoh (1997); their modification of the **AIC** and $\mathbf{C}_p$ criteria were intended to reduce bias in situations where the collection of candidate models includes both underspecified and overspecified models. Recently, McQuarrie and Tsai (1998) present and compare the performance of several multivariate as well as univariate variable selection criteria for two special models and give comprehensive details on model selection.

2. Description of Model selection Methods

Stepwise regression and all-possible-regression are two types of variable selection procedures that are employed by most of the statistical software packages, and used in practice. In the former, investigators delete or add variables one at a time using a stepwise method and in the later they examine all possible subsets and choose one model based on some criteria.

Before presenting a detailed description of each procedure, we note that all variable selection criteria in MMR involve matrices and functions such as trace (tr(•)), determinant (| • |), or the largest eigenvalue (λ(•)) can be used to obtain the scalar counterpart of the univariate criteria. Our module will use only the determinant function because tr(•) deals only with diagonal elements and does not take into account the contribution of off-diagonal entries, and λ(•) uses only one root which makes it unreliable (Spark et al., 1983).

It would be helpful also to introduce a standard notation for all variables, vectors, matrices, and functions used before describing each criterion. The following table presents notations and definitions of variables and functions used in defining the criteria:

Table 1 Notations and definitions of variables and functions used.

Symbol	Definition
n	The number of observations
p	The number of parameters including the intercept
k	The number of **x**'s in the "full model"
q	The number of **y**'s
$\mathbf{Y}$	The matrix of dependent variables
$\mathbf{X}$	The matrix of all candidate independent variables with its first column being the vector **1**
$\mathbf{X_p}$	The submatrix of **X** containing the vector **1** and the columns corresponding to selected variables $\mathbf{x}_p$ in the model.
$\mathbf{J}$	The $q \times q$ matrix of ones
$\boldsymbol{\Lambda}$	The Wilks' $\boldsymbol{\Lambda}$ statistic, which is analogous to **F** random variable, defined as the ratio of two independent chi-square random variables divided by their respective

Table 1 Notations and definitions of variables and functions used.

Symbol	Definition
	degrees of freedom
$\hat{\boldsymbol{\Sigma}}$	The sum squared error for a "full-model" including the intercept
$\hat{\boldsymbol{\Sigma}}_{\mathbf{p}}$	The sum squared error for a model with *p* parameters including the intercept
ln	The natural logarithm
$\lvert\bullet\rvert$	The determinant function

2.1 Stepwise Regression

Stepwise regression consists of three procedures: Forward Selection, Forward Stepwise Regression, and Backward Elimination (Barrett & Gray, 1994; Rencher, 1995). Although the forward stepwise regression is probably the most widely used procedure (Neter et al., 1996), all three criteria will be presented and used in the module.

The usual criteria used for adding (or deleting) an **x** variable is either partial Wilks' $\boldsymbol{\Lambda}$ or partial **F** criterion. Our SAS module employs only the partial Wilks' $\boldsymbol{\Lambda}$. The Wilks' $\boldsymbol{\Lambda}$ is analogous to **F** random variable, defined as the ratio of two independent chi-square random variables divided by their respective degrees of freedom (Rencher, 1998). It is defined as

$$\Lambda(x_1, x_2, x_3, ..., x_p) = \frac{\left|\mathbf{Y}'[\,\mathbf{I} - \mathbf{X}_\mathbf{p}(\mathbf{X}'_\mathbf{p}\mathbf{X}_\mathbf{p})^{-1}\mathbf{X}'_\mathbf{p}\,]\,\mathbf{Y}\right|}{\left|\mathbf{Y}'[\,\mathbf{I} - \frac{1}{\mathbf{N}}\mathbf{J}\,]\,\mathbf{Y}\right|} \text{ which is distributed as } \Lambda_{q,1,n-p-1}.$$

A variable would be a candidate for addition when the minimum partial Wilks' $\boldsymbol{\Lambda}$ value falls below a predetermined threshold value. The variable would be a candidate for deletion when the maximum partial Wilks' $\boldsymbol{\Lambda}$ value exceeds a predetermined value.

2.1.1 Forward Selection

The forward selection technique begins with no variables in the model. For each of the independent variables, the forward method calculates partial Wilks' Λ statistics that reflect the variable's contribution to the model if it is included. The minimum value for these partial Λ statistics is compared to a predetermined threshold value. If no Λ statistic falls below the predetermined threshold value, the forward selection stops. Otherwise, the forward method adds the variable that has the lowest partial Wilks' Λ statistic to the model. The forward method then calculates partial Wilks' Λ statistics again for the variables still remaining outside the model, and the evaluation process is repeated. Thus, variables are added one by one to the model until no remaining variable produces a significant partial Wilks' Λ statistic. Once a variable is in the model, it stays. (For more details, the reader is referred to Rencher, 1995)

2.1.2 Forward Stepwise Regression

The stepwise method is a modification of the forward selection technique and differs in that variables already in the model do not necessarily stay there. As in the forward selection method, variables are added one by one to the model, and the partial Wilks' Λ statistic for a variable to be added must have an entry significant value (i.e., the minimum value of partial Λ statistics falls below a predetermined threshold value). After a variable is added, however, the stepwise method looks at all the variables already included in the model and deletes any variable that does not produce an stay significant partial Λ statistic (i.e., its partial Λ value exceeds a predetermined value). Only after this check is made and the necessary deletions accomplished can another variable be added to the model. The stepwise process ends when none of the variables outside the model has an entry significant partial Λ statistic and every variable in the model is significant to stay, or when the variable to be added to the model is the one just deleted from it.

2.1.3 Backward Elimination

The backward elimination method begins with all **x**'s included in the model and deletes one variable at a time using a partial Λ. At the first step, the partial Λ for each **x** is calculated and the variable with largest partial Λ statistic that exceeds the predetermined threshold value is deleted. At the second step, a partial Wilks' Λ is calculated for each of the q-1 remaining variables, and again the least important variable in the presence of the others is eliminated. This process continues until a step is reached at which the largest partial Λ is "significant" (i.e., does not exceed the predetermined value), indicating that the corresponding variable is apparently not redundant in the presence of the other variable in the model.

2.2 All-Possible-Regression

The all-possible-regression procedure calls for considering all possible subsets of the pool of potential predictors and identifying for detailed examination a few "good" subsets according to some criterion (Neter et al., 1996). Various criteria for comparing the regression model may be used with the all-possible-regression selection procedure. Residual mean square error (**MSE**), coefficient of multiple determination ($\mathbf{R^2}$), adjusted coefficient of multiple determination ($\boldsymbol{Ajd}\mathbf{R}^2$), Akaike's information criterion (**AIC**), Hannan and Quinn information criterion (**HQ**), Schwarz criterion (**BIC**), and Mallow's C_P are some of these procedures (see e.g., Rencher, 1995 and McQuarrie & Tsai ,1998). These techniques are included in the program because of their popularity in selecting the "best" subset of predictors.

2.2.1 Residual Mean Square Error

The residual mean square error is the variance estimator for each model and is defined by

$$\mathbf{MSE} = \left| \frac{\hat{\boldsymbol{\Sigma}}_{\mathbf{p}}}{n-p} \right| \tag{2.3}$$

where $\hat{\boldsymbol{\Sigma}}_{\mathbf{p}} = \mathbf{Y}'[\mathbf{I} - \mathbf{X}_{\mathbf{P}}(\mathbf{X}'_{\mathbf{P}}\mathbf{X}_{\mathbf{P}})^{-1}\mathbf{X}'_{\mathbf{P}}]\mathbf{Y}$ is the sum squared error for a model with p parameters including the intercept. It is often suggested that the researcher choose the model with minimal value of **MSE**.

2.2.2 $\mathbf{R}^2$ *Selection Criterion*

$\mathbf{R}^2$ is the coefficient of multiple determination and the method finds subsets of independent variables that best predict a dependent variable by linear regression in the given sample. It efficiently performs all possible subset regressions and displays the models in decreasing order of matrix ($\mathbf{R}^2$) magnitude within each subset size. The $\mathbf{R}^2$ is computed as:

$$\mathbf{R}^2 = \left| [\mathbf{Y}'(\mathbf{I} - \tfrac{1}{\mathbf{n}}\mathbf{J})\mathbf{Y}]^{-1}[\mathbf{Y}'(\mathbf{X}_{\mathbf{p}}(\mathbf{X}'_{\mathbf{p}}\mathbf{X}_{\mathbf{p}})^{-1}\mathbf{X}'_{\mathbf{p}} - \tfrac{1}{\mathbf{n}}\mathbf{J})\mathbf{Y}] \right| \tag{2.4}$$

The $\mathbf{R}^2$ method differs from the other selection methods; it always identifies the "best" model as the one with the largest $\mathbf{R}^2$ for each number of variables considered.

2.2.3 *Adjusted* $\mathbf{R}^2$ *Selection Criterion*

Since the number of parameters in the regression model is not taken into account by $\mathbf{R}^2$, as $\mathbf{R}^2$ does not decrease as p increases, the adjusted coefficient of multiple determination (***Ajd***$\mathbf{R}^2$) has been suggested as an alternative criterion. The ***Ajd***$\mathbf{R}^2$ method is similar to the $\mathbf{R}^2$ method and it finds the "best" models with the highest ***Adj***$\mathbf{R}^2$ within the range of sizes. The criterion is

$$\textbf{\textit{Adj}}\mathbf{R}^2 = 1 - \frac{(n-1)(1-\mathbf{R}^2)}{n-p} \tag{2.5}$$

2.2.4 Akaike's information criterion (AIC)

The **AIC** procedure (Akaike, 1973) is used to evaluate how well the candidate model approximates the true model by assessing the difference between the expectations of the vector *y* under the true model and the candidate model using the Kullback-Leibler (K-L) distance. The Kullback-Leibler (K-L) distance is the distance between the true density and estimated density for each model. The criterion is

$$\mathbf{AIC} = \ln |\, \hat{\mathbf{\Sigma}}_{\mathbf{p}} \,| + \frac{2pq + q(q+1)}{n} \qquad (2.6)$$

The model that best predicts the **y**'s jointly, with this procedure, is the one that has the minimum **AIC**'s value.

2.2.5 The Corrected Form of Akaike's information criterion ($\mathbf{AIC_c}$)

Bedrick & Tsai (1994) pointed out that the Akaike's information criterion might lead to overfitting in small samples. Thus, they proposed a corrected version ($\mathbf{AIC_C}$) of **AIC**, which is

$$\mathbf{AIC_C} = \ln |\, \hat{\mathbf{\Sigma}}_{\mathbf{p}} \,| + \frac{(n+p)q}{n-p-q-1} \qquad (2.7)$$

The "best" subset of **x**'s, with this procedure, is the one that has the minimum $\mathbf{AIC}_C$'s value.

2.2.6 Hannan and Quinn (HQ)

Although **HQ** information criterion introduced by Hannan and Quinn (1979) was intended for use with the autoregressive models, it also can be applied to regression models (McQuarrie & Tsai, 1998). The criterion is

$$\mathbf{HQ} = \ln |\, \hat{\mathbf{\Sigma}}_{\mathbf{p}} \,| + \frac{2\ln(\ln(n))pq}{n} \qquad (2.8)$$

The "best" model is the model that corresponds to the minimum **HQ** value.

2.2.7 The Corrected Form of Hannan and Quinn (HQ_c) Information Criterion

The Hannan and Quinn (**HQ**) Information criterion usually overfits when applied to small samples McQuarrie & Tsai (1998). Therefore, McQuarrie & Tsai (1998) proposed a corrected version of it, which is

$$\mathbf{HQ}_C = \ln|\widehat{\mathbf{\Sigma}}_p^2| + \frac{2ln(ln(n))pq}{n-p-q-1} \tag{2.9}$$

Similarly, the procedure identifies the "best" subset of the **x**'s that yields the smallest value.

2.2.8 Schwarz's Criterion (BIC)

The function computes Schwarz 's Bayesian information criterion for each model using the Kullback-Leibler (K-L) distance (Schwarz 1978; SAS/STAT User's Guide, 1990), which can be utilized to identify the "best" model. The criterion is

$$\mathbf{BIC} = \ln|\widehat{\mathbf{\Sigma}}_p^2| + \frac{ln(n)p}{n} \tag{2.10}$$

The "best" model by the procedure is the model that corresponds the minimal value.

2.2.9 Mallow's C_P

The C_P criterion was initially suggested by Mallow's (1973) for univariate regression and extended by Spark et al. (1983) to multivariate multiple regression. It evaluates the total mean squared error of the *n* fitted values for each subset regression. The criterion is obtained by using the formula

$$C_P = (n-k)\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{\Sigma}}_{\mathbf{p}} + (2p-n)I \tag{2.11}$$

where $\hat{\mathbf{\Sigma}} = \mathbf{Y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}$, and **I** is the identity matrix of size $(q \times q)$. The procedure identifies the "best" subset of the **x**'s with the one that gives both *small* $\mathbf{C_P}$ (in terms of scalar

function of matrix such as determinant) and near $p\mathbf{I}$ (Spark et al., 1983; Rencher, 1995). If $2p-n<0$, however, $|\mathbf{C_P}|$ is negative and unreliable (Spark et al., 1983). Hence, modification of $|\mathbf{C_P}|$ has been suggested by Spark et al. (1983) to remedy this problem, the quality $|\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{p}}|$ is always positive and is written in terms of $\mathbf{C_P}$ as

$$\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{p}} = \frac{\mathbf{C}_{\mathbf{p}} + (n-2p)\mathbf{I}}{n-k} \tag{2.12}$$

When the bias is 0, $\mathbf{C}_{\mathbf{p}} = p\mathbf{I}$, and (2.12) becomes

$$\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{p}} = \frac{n-p}{n-k}\mathbf{I} \tag{2.13}$$

Hence, subsets are sought that satisfy

$$\left|\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{p}}\right| \le \left(\frac{n-p}{n-k}\right)^{q} \tag{2.14}$$

3. Practical Example

Anderson and Bancroft (1952, p. 205) presented data on chemical components of 25 tobacco leaf samples. The dependent variables are

Y_1 : Rate of cigarette burn in inches per 1000 seconds

Y_2 : Percent sugar in the leaf

Y_3 : Percent nicotine

The independent variables are

X_1 : Percentage of Nitrogen

X_2 : Percentage of Chlorine

X_3 : Percentage of Potassium

X_4 : Percentage of Phosphorus

X_5 : Percentage of Calcium

X_6 : Percentage of Magnesium

Table 3-1 presents the data.

Spark et al. (1983) considered these data and found the "best" subsets of predictors based on the multivariate $\mathbf{C}_p$ criterion. We use this data set here to obtain the "best" subset(s) of predictors using all the criteria mentioned.

Three straightforward steps are needed in order to run the program properly for the data: (1) paste the data at the beginning of the program and read it using DATA statement, (2) name the dependent and the independent variables, and (3) run the program.

TABLE 3-1 The Tobacco Data

Subject ID	Dependent Variables			Independent Variables					
	Y1	**Y2**	**Y3**	**X1**	**X2**	**X3**	**X4**	**X5**	**X6**
1	1.55	20.05	1.38	2.02	2.90	2.17	0.51	3.47	0.91
2	1.63	12.58	2.64	2.62	2.78	1.72	0.50	4.57	1.25
3	1.66	18.56	1.56	2.08	2.68	2.40	0.43	3.52	0.82
4	1.52	18.56	2.22	2.20	3.17	2.06	0.52	3.69	0.97
5	1.70	14.02	2.85	2.38	2.52	2.18	0.42	4.01	1.12
6	1.68	15.64	1.24	2.03	2.56	2.57	0.44	2.79	0.82
7	1.78	14.52	2.86	2.87	2.67	2.64	0.50	3.92	1.06
8	1.57	18.52	2.18	1.88	2.58	2.22	0.49	3.58	1.01
9	1.60	17.84	1.65	1.93	2.26	2.15	0.56	3.57	0.92
10	1.52	13.38	3.28	2.57	1.74	1.64	0.51	4.38	1.22
11	1.68	17.55	1.56	1.95	2.15	2.48	0.48	3.28	0.81
12	1.74	17.97	2.00	2.03	2.00	2.38	0.50	3.31	0.98
13	1.93	14.66	2.88	2.50	2.07	2.32	0.48	3.72	1.04
14	1.77	17.31	1.36	1.72	2.24	2.25	0.52	3.10	0.78
15	1.94	14.32	2.66	2.53	1.74	2.64	0.50	3.48	0.93
16	1.83	15.05	2.43	1.90	1.46	1.97	0.46	3.48	0.90
17	2.09	15.47	2.42	2.18	0.74	2.46	0.48	3.16	0.86
18	1.72	16.85	2.16	2.16	2.84	2.36	0.49	3.68	0.95
19	1.49	17.42	2.12	2.14	3.30	2.04	0.48	3.28	1.06
20	1.52	18.55	1.87	1.98	2.90	2.16	0.48	3.56	0.84
21	1.64	18.74	2.10	1.89	2.82	2.04	0.53	3.56	1.02
22	1.40	14.79	2.21	2.07	2.79	2.15	0.52	3.49	1.04
23	1.78	18.86	2.00	2.08	3.14	2.60	0.50	3.30	0.80
24	1.93	15.62	2.26	2.21	2.81	2.18	0.44	4.16	0.92
25	1.53	18.56	2.14	2.00	3.16	2.22	0.51	3.73	1.07
Sum	**42.20**	**415.39**	**54.03**	**53.92**	**62.02**	**56.00**	**12.25**	**89.79**	**24.10**

The Stepwise and All-Possible-Regression Procedures

Table 3-2 presents the "best" subset of predictors based on all stepwise regression procedures and all-possible-regression methods. It shows that the forward selection, forward stepwise regression, and mean square error criteria selected the model with $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_4$, and $\mathbf{x}_6$, as the "best" model to predict **y**'s jointly. The backward elimination method selected the model with $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, and $\mathbf{x}_6$, as the "best" one. The adjusted R^2 selected the model with $\mathbf{x}_2$ and $\mathbf{x}_4$. The Akaike's information

criterion and its corrected form, and Hannan and Quinn information criterion and its corrected form selected the model with $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_6$. The Schwarz's Bayesian information criterion selected the model with all **x**'s in it.

The Mallow's $\mathbf{C}_p$ criterion, on the other hand, gave an array of 'best" models because the program compares the values of $\left|\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{\Sigma}}_{\mathbf{p}}\right|$ to $(\frac{n-p}{n-k})^q$ and selects the model with $\left|\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{\Sigma}}_{\mathbf{p}}\right| \leq (\frac{n-p}{n-k})^q$ as the "best" one. In using the $\mathbf{C}_p$ criterion, we seek to identify subsets of **x**'s for which (1) the $|\mathbf{C}_p|$ value is small and (2) the $\mathbf{C}_p$ value is near $p\mathbf{I}$ (i.e., $\left|\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{\Sigma}}_{\mathbf{p}}\right| \leq (\frac{n-p}{n-k})^q$). Table 3-2 shows the values of $(\frac{n-p}{n-k})^q$ at each p.

TABLE 3-2 The upper limit values for each p

p	2	3	4	5	6	7
$(\frac{n-p}{n-k})^q$	2.086248	1.825789	1.587963	1.371742	1.176097	1

TABLE 3-3 The "best" subset of predictors for all model selection criteria

Model selection Criteria		Subset of Predictors
Stepwise Regression	FORWARD	1246
	BACKWARD	1234
	STEPWISE	1246
All-Possible-Regression	MSE	1246
	ADJRSQ	24
	AIC	126
	AICC	126
	HQ	126
	HQC	126
	BIC	123456
	Cp	126
		1236
		1246
		1256
		12346
		12356
		12456

Note: FORWARD = Forward Selection; BACKWARD = Backward Elimination; STEPWISE = Forward Stepwise Regression; MSE = Mean Square Error; ADJRSQ = Adjusted R^2; AIC & AICC= Akaike's information criterion and its corrected form, respectively; HQ & HQc = Hannan and Quinn information criterion and its corrected form, respectively; BIC = Schwarz's Bayesian information criterion; and C_p = Mallow's criterion where.

4. Summary and Conclusion

A SAS/IML program has been written to locate the multivariate candidate models for several multivariate model selection criteria. Three straightforward steps need to run the program properly for a new data: (1) Paste (or read) new data in the place of the example data at the beginning of the program and read it using DATA statement, (2) change the dependent and the independent variables' names in IML procedure, and (3) run the program. The program can also be used in a univariate linear regression case.

Al-Subaihi (2002) stated that it is important for the investigator not to depend entirely on variable selection criteria because none of them works well under all conditions. A simulation study conducted by Bedrick and Tsai (1994) shows that factors such as sample size, number of dependent variables, number of independent variables, and the correlation between the $\boldsymbol{y}$'s play a role in deciding which criterion should be used. Thus, the researcher needs to include predictors based on theory(ies) of the field under study and have low correlation with each other and high correlation with all $\boldsymbol{y}$'s. He needs to utilize more than one criterion in evaluating possible subset of $\boldsymbol{x}$ variables. Finally, the researcher needs to evaluate the final *good* models using various diagnostic procedures.

Bibliography

[1] Akaike, H. (1973), "Information Theory and an Extension of The Maximum Likelihood Principle", In B.N. Petrov and F. Csaki ed., 2nd *International Symposium on Information Theory*, pp. 267-281, Akademia Kiado, Budapest.

[2] Al-Subaihi, Ali A. (2002), "Univariate Variable Selection Criteria Available In SAS or SPSS", paper presented at the American Statistical Association annual meeting- August 11-15, 2002, New York, NY.

[3] Anderson, R. L. and Bancroft, T. A. (1952), *Statistical Theory in Research*, McGraw-Hill Book Company, Inc., New York, NY.

[4] Barrett, B. E. and Gray, J. B. (1994), "A Computational Framework for Variable Selection in Multivariate Regression", *Statistics and Computing*, **4**, 203-212.

[5] Bilodeau, M. and Brenner, D. (1999), *Theory of Multivariate Statistics*, Springer-Verlag New York, Inc., New York.

[6] Breiman, L. and Friedman, J. H. (1997), "Predicting Multivariate Responses in Multiple Linear Regression", *Journal of the Royal Statistical Society*, **59** (No. 1), 3-54.

[7] Fujikoshi, Y.; and Satoh, K. (1997), " Modified AIC and Cp in Multivariate Linear Regression", *Biometrika*, **84** (3), 707-716.

[8] Hannan, E. J. and Quinn, B. G. (1979), "The Determination of The Order of an Autoregression", *Journal of the Royal Statistical Society*, B **41**, 190-195.

[9] Mallows, C. L., (1973), "Some Comments on Cp", *Technometrics*, **15** (4), 661-675.

[10] McQuarrie A. D., and Tsai, C. (1998), "*Regression and Time Series Model Selection*", World Scientific Publishing Co. Pte. Ltd., River Edge, NJ.

[11] Miller, A. J. (1990), *Subset Selection in Regression*, Chapman and Hall, New York, NY.

[12] Neter, J., Kutner, M., Nachtsheim, C., and Wasserman, W. (1996), "*Applied Linear Statistical Models*", McGraw-Hill Companies, Inc., NY.

[13] Rencher, A. C. (1995), "*Methods of Multivariate Analysis*", John Wiley & Sons Inc., New York, New York.

[14] Rencher, A. C. (1998), " *Multivariate Statistical Inference and Applications*", John Wiley & Sons Inc., New York, New York.

[15] *SAS/STAT User's Guide*, Version 6, 4th Edition, SAS Institute Inc., Cary, NC (1990).

[16] Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461 - 464.

[17] Sparks, R. S.; Coutsourides, D.; and Troskie, L. (1983), " The Multivariate Cp", *Commun. Statistis. –Theor. Meth.,* **12** (15), 1775-1793.

[18] Sparks, R. S.; Zucchini, W.; and Coutsourides, D. (1985), " On Variable Selection in Multivariate Regression" , *Commun. Statistis. –Theor. Meth.,* **14** (7), 1569-1587.