A Comparison of the Mallows Cp and Principal Component Regression Criteria for Best Model Selection in Multiple Regression

Randall E. Schumaker
University of North Texas

A cross validation comparison of the Mallows Cp subset model selection criteria using randomly generated data sets indicated that different subset models may be identified. The principal component regression method using Type II sum of squares with orthogonal principal component variables indicated a slightly different set of "best" variables. The two methods in the presence of multicollinearity can yield different subset models. It is recommended that researchers base regression models on substantive theory, model validation, and effect sizes for proper model testing and interpretation.

Multiple regression permits model testing wherein a set of independent variables are hypothesized to predict a dependent variable. Often when the set of variables selected does not significantly predict, the researcher searches for a "subset" of variables that provides the best prediction model. The statistical packages provide several stepwise methods for this purpose.

A review of the literature, however, indicates that most researchers misuse stepwise methods in determining the best predictor set or interpreting the importance of predictor variables (Huberty, 1989; Snyder, 1991; Thompson, 1989; Thompson, Smith, Miller, & Thomson, 1991; Welge, 1990). Tracz, Brown, and Kopriva (1991) summarized much of the literature to indicate that the results of stepwise procedures do not yield a "best" equation because different criteria can be used in the selection of different sets of variables; that when variables are intercorrelated, there is no satisfactory way to determine the relative contribution of the variables to R-squared because various subsets of variables could yield a similar R-squared value; that stepwise methods inflate Type I error rates by not using the correct degrees of freedom in calculating the change in R-squared; and that the order of variable entry is incorrectly interpreted as defining the importance of the variable or "best set" of predictors.

Current research literature indicates that the all possible subset approach is preferred over the stepwise methods for determining the best model (Berk, 1977; Cummings, 1982; Thayer, 1986; Davidson, 1988; Henderson & Denison, 1989; Welge, 1990; Thayer, 1990; Tracz, Brown, & Kopriva, 1991). Several criteria, however, are available for selecting the best subset model when using the all possible subset approach: R-squared, adjusted R-squared, mean squared error, Mallow's C_p, or a principal component regression. Constas and Francis (1992) presented a graphical method for selecting the best subset regression model using R-squared and adjusted R-squared. They plotted R-squared and adjusted R-squared against the number of predictors in the model. The maximum number of predictors for best subset model was determined at the point where the R-squared and/or the adjusted R-squared values descended.

The Mallows Cp criteria has also been recommended for selecting the best subset of predictor variables in contrast to the stepwise methods using a sample data set (Tracz, Brown, & Kopriva, 1991; Zuccaro, 1992). The Cp statistic measures the effect of underfitting (important predictors left out of the model) or overfitting (include predictors that make no contribution or are marginal). Mallows (1966; 1973) has suggested that the selection of the best subset model with the lowest bias is indicated by the smallest Mallows Cp criteria, especially in the presence of multicollinearity. The SAS package (Freund & Littell, 1991) currently prints the Mallows Cp value and a variance inflation factor (VIF) which can be used to determine which variables may be involved in the multicollinearity. Pohlmann (1983) had previously noted that multicollinearity among a set of predictor variables didn't affect the Type I error rate, but did affect the Type II error rate and width of the confidence interval. His findings suggest that sample size and model validity could compensate for multicollinearity effects, especially when certain research questions require models with highly correlated predictors, for example, $Y = \beta_1 X_1 + \beta_2 X^2_1 + e$.

The principal component regression (PCR) has also been proposed as a criteria for selecting the best predictor model. This method appears to be useful when predicting values in one sample based upon estimates from another sample and when multicollinearity exists among a set of variables (Morrison, 1976). The indication for using a PCR approach is when the mean squared error of a biased estimate is smaller than the variance of an unbiased estimate. The PCR method, however, is not appropriate for multiple regression subset models containing interactions (Aiken & West, 1993). Since the PCR method creates a set of new variables called principal components, which are uncorrelated or orthogonal, it should not be used when models depict nonlinear, correlated predicter sets.

In summary, the all possible subset approach is recommended as an alternative over stepwise methods for selecting the best set of predictor variables. The Mallows Cp criteria or a principal components regression approach is advocated for determining the best subset model over the use of R-squared, especially when the predictors are correlated. The principal component regression method, which determines the best model for prediction by creating orthogonal variables, appears more useful when estimates from one sample are used to predict in another sample or when multicollinearity exists among the predictors.

How do these criteria compare when selecting the best subset model? When might a researcher choose one criteria over another for selecting the best model? A comparison of the Mallows Cp selection criteria upon cross validation and a comparison of the parameter estimates and standard errors between the multiple regression and the PCR approach should shed further light on their usefulness for subset model selection. An applied example will further elaborate the comparison of the two criteria.

Simulation

A SAS program was used to generate a heuristic population (n = 10,000 observations) with a dependent variable and ten correlated predictor variables. The program then randomly sampled the population data set for n = 200 observations. This data set was then randomly divided to create two separate data sets of equal size ($n_1 = n_2 = 100$ observations). The SAS programs used in this simulation are available from the author.

The population correlation matrix, variable means and standard deviations are in Table 1. The correlation matrix, variable means and standard deviations for the sample data set used to compute the parameter estimates are in Table 2. The correlation matrix, variable means and standard deviations for the cross validation data set are in Table 3. Parameter estimates, computed using the ordinary least squares criterion from the first data set, were used with the second data set to calculate R^2 and the Mallows Cp values, and to determine the best variable subset models.

Table 1 Population Correlation Matrix, Means, and Standard Deviations (n = 10,000)

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	.44										
X2	.25	.10									
X3	.34	.13	.10								
X4	.43	.19	.10	.15							
X5	.42	.19	.11	.13	.19						
X6	.30	.13	.09	.11	.13	.12					
X7	.24	.11	.07	.06	.10	.08	.07				
X8	.50	.22	.13	.17	.21	.21	.16	.11			
X9	.28	.12	.08	.10	.12	.11	.09	.07	.15		
X10	.26	.11	.05	.07	.11	.12	.06	.08	.14	.08	
Mean	9.99	17.92	16.12	18.94	21.96	28.05	25.97	38.90	42.05	33.97	12.05
S.D.	2.00	4.44	8.21	6.00	4.66	4.95	6.61	8.61	4.12	6.95	8.12

Note. All values have been rounded to two decimal places.

Table 2 Sample Correlation Matrix, Means, and Standard Deviations for Estimation Sample (n_1 = 100)

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	.41										
X2	.28	.02									
X3	.41	.05	.23								
X4	.38	.23	.01	.15							
X5	.24	-.01	.04	.02	.16						
X6	.33	.02	.16	.09	.08	.08					
X7	.25	.16	.08	.03	.01	.01	.10				
X8	.39	.22	.13	-.04	.19	.06	.21	.01			
X9	.33	.19	.07	.04	.24	-.15	.03	.22	.21		
X10	.46	.23	.08	.24	.21	.03	.10	.17	.11	.17	
Mean	10.18	18.40	15.37	20.49	22.76	28.41	25.88	39.55	41.89	34.27	11.04
S.D.	1.80	4.61	8.88	5.94	4.30	4.99	6.79	7.81	4.13	6.80	8.13

Note. All values have been rounded to two decimal places.

Table 3 Sample Correlation Matrix, Means, and Standard Deviations for Cross Validation Sample (n_2 = 100)

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	.39										
X2	.28	.14									
X3	.34	-.05	-.08								
X4	.52	.03	.13	.20							
X5	.54	.17	.20	.28	.37						
X6	.26	.01	.01	.07	.18	.19					
X7	.14	.03	.05	.08	.07	.01	-.03				
X8	.55	.27	.11	.26	.26	.21	.06	.02			
X9	.32	.26	.18	-.09	.20	.07	.11	.09	.09		
X10	.31	.26	.07	.11	.12	.21	.11	.19	.09	.24	
Mean	9.94	17.91	16.55	19.26	21.37	28.40	25.34	39.23	41.92	33.93	10.38
S.D.	1.99	4.86	8.57	6.13	5.35	4.75	6.82	9.43	4.27	6.73	7.78

Note. All values have been rounded to two decimal places.

Table 4 R^2 and Cp Values for $Sample_1$ And $Sample_2$ Best Variable Subset Models ($n_1 = n_2 = 100$)

Subset Size	Variables in Subset Model	$Sample_1$ R^2	$Sample_1$ Cp
1	(10)	.21	102.92
2	(3),(8)	.33	74.44
3	(3),(8),(10)	.44	49.79
4	(1),(3),(8),(10)	.50	36.13
5	(1),(3),(6),(8),(10)	.54	27.42
6	(1),(3),(5),(8),(9),(10)	.58	19.74
7	(1),(3),(5),(6),(8),(9),(10)	.62	12.26
8	(1),(2),(3),(5),(6),(8),(9),(10)	.63	11.85
9	(1),(2),(3),(4),(5),(6),(8),(9),(10)	.64	11.27
10	(1),(2),(3),(4),(5),(6),(7),(8),(9),(10)	.65	11.00

		$Sample_2$ R^2	$Sample_2$ Cp
1	(8)	.30	101.79
2	(5),(8)	.49	50.44
3	(4),(5),(8)	.55	33.41
4	(1),(4),(5),(8)	.61	21.34
5	(1),(4),(5),(8),(9)	.63	17.05
6	(1),(3),(4),(5),(8),(9)	.65	13.37
7	(1),(3),(4),(5),(6),(8),(9)	.66	11.58
8	(1),(2),(3),(4),(5),(6),(8),(9)	.67	9.79
9	(1),(2),(3),(4),(5),(6),(7),(8),(9)	.68	9.96
10	(1),(2),(3),(4),(5),(6),(7),(8),(9),(10)	.68	11.00

Table 5 Cross Validation Comparison of R^2 and Cp Values: $Sample_1$ to $Sample_2$ for Best Variable Subset Models ($n_1 = n_2 = 100$)

Subset Size	Variables in Subset Model	$Sample_1$ Cp	$Sample_1$ R^2	$Sample_2$ Cp	$Sample_2$ R^2
1	(10)	.21	102.92	.15	159.53
2	(3),(8)	.33	74.44	.36	92.08
3	(3),(8),(10)	.44	49.79	.40	77.64
4	(1),(3),(8),(10)	.50	36.13	.45	65.44
5	(1),(3),(6),(8),(10)	.54	27.42	.47	55.78
6	(1),(3),(5),(8),(9),(10)	.58	19.74	.59	26.35
7	(1),(3),(5),(6),(8),(9),(10)	.62	12.26	.61	23.38
8	(1),(2),(3),(5),(6),(8),(9),(10)	.63	11.85	.62	20.82
9	(1),(2),(3),(4),(5),(6),(8),(9),(10)	.64	11.27	.63	10.34
10	(1),(2),(3),(4),(5),(6),(7),(8),(9),(10)	.65	11.00	.66	11.00

Table 4 indicates the model subset selection for each sample data set. Table 5 indicates a comparison between the R^2 and Mallows Cp values from the estimation sample data set to the cross validation sample data set using parameter estimates from the estimation sample. The Mallows Cp values were inflated because the parameter estimates applied to the second data set altered the residual sums of squares used in the formula to calculate them. Although the relative ordering of Cp values were the same, these values did not indicate the same single best variable subset model in the second data set.

Table 6 compares the parameter estimates using the Mallows Cp and the principal components regression method for each best variable subset model. The R^2 values will be the same regardless of which method is used. The real difference is seen when comparing the relative significance of the parameter estimates. The Mallows Cp method with correlated predictors indicated that all the parameter estimates were significant. This was not the case in the principal components regression approach. An applied example will further illustrate this distinction between the two methods.

Applied Example

Subjects

Participants in the study were a cohort of students accepted into the Texas Academy of Mathematics and Science (TAMS) at the University of North Texas in Fall, 1993. TAMS is an early college entrance program in which students earn approximately 60 hours of college credit by taking University of North Texas courses. Students enter TAMS at the beginning of their 11th year in high school. They live on campus in a special residence hall and take regular university courses in mathematics, science and the humanities. After two years, participants receive a special high school diploma and have amassed at least 60 hours of college credit. Each year approximately 200 high school sophomores, who have met the selection criteria and completed the 10th grade, are accepted into the Texas Academy of Mathematics and Science.

In the study year, TAMS accepted 204 students. Of these, 156 students attended an August orientation, which occurred a week prior to their first semester of college coursework, and completed the LASSI. There were 80 females and 76 males who participated in the study. The students who took the LASSI were similar in demographic background and academic ability as previous classes because of the academy's consistent admission requirements and pool of applicants. The participants' SAT-M and SAT-V means and standard deviations, respectively, were: M=651, SD=57; and M=530, SD=75.

Instrument

The LASSI is an English language assessment tool designed to measure college students' use of learning and study strategies. It was designed to provide assessment and pre-post achievement measures for students participating in a learning strategies and study skills project. A high-school version is available, but it was not recommended for use with accelerated students in these programs (Eldredge, 1990). The LASSI can be administered in a group setting in approximately 30 minutes. The carbonless test format allows participants to score their own assessment and take a copy of the results with them from the testing session.

The ten LASSI subscales focus on thoughts and behaviors related to successful learning. The ten subscales are (1) Attitude, (2) Motivation, (3) Time Management, (4) Anxiety, (5) Concentration, (6) Information Processing, (7) Selecting the Main Ideas, (8) Study Aids, (9) Self-testing, and (10) Test Strategies (for more details see Weinstein, 1987). Reliability studies reported Cronbach alpha internal consistency values ranging from .70 to .86 and test-retest reliabilities from .70 to .85. Validity studies have also reported normative data for high school and college students with different instruments for each group (Weinstein, Palmer, & Schulte, 1987). Students respond to individual items on each subscale using a five-point scale: (5) very typical of me; (4) fairly typical of me; (3) somewhat typical of me; (2) not very typical of me; and (1) not at all typical of me. Some item values are reverse keyed before being added to obtain a subscale score. The subscale scores are compared by graphing them onto a normal curve equivalent percentile chart.

According to the LASSI user's manual (Weinstein, 1987), students scoring above the 75th percentile do not need to improve that specific skill or strategy. Students scoring between the 75th percentile and the 50th percentile should consider improvement. Students scoring below the 50th percentile on a subscale need assistance to improve that skill or strategy. For example, students scoring below the 50th percentile on the anxiety subscale would be considered anxious about being in college. Likewise, students scoring below the 50th percentile on the motivation subscale lack appropriate motivation to do college level work effectively.

Research Question

The research question of interest was whether the ten LASSI subscales could predict a student's college grade point average after one semester of college coursework. A related question pertained to whether a "subset" of the ten LASSI subscales could better predict college grade point average for this sample of students. Students not maintaining at least a 2.50 grade point average after one semester of college coursework were dismissed from the Academy. Knowledge of which subscales are best predictors of college grade point average would aid staff in identifying potential at-risk students upon entering the Academy.

Data Analysis

The data were analyzed using a SAS statistical program. The student's college grade point average was predicted by the ten LASSI subscales using PROC REG with the SELECTION statement requesting the best subset model criteria. The PROC PRINCOMP procedure was used to create ten orthogonal principal component variables. The principal component variable parameter estimates were then computed using the

Table 6 Mallows Cp and Principal Components Regression Comparison (n_1 = 100)

Best Variable Subset Model	Mallows Cp				Principal Components				
	ß	$SE_ß$	t	p	ß	$SE_ß$	t	p	R^2
X10	.10	.02	5.00	.0001	.82	.16	5.13	.0001	.21
X3	.13	.03	4.33	.0001	.02	.15	.13	.90	.33
X8	.18	.04	4.50	.0001	1.05	.15	7.00	.0001	
X3	.10	.02	5.00	.0001	.98	.12	8.17	.0001	.44
X8	.16	.03	5.33	.0001	.42	.14	3.00	.0024	
X10	.07	.02	3.50	.0001	.21	.16	1.31	.1951	
X1	.10	.03	3.33	.0009	1.04	.11	9.45	.0001	.50
X3	.10	.02	5.00	.0001	.07	.12	.58	.59	
X8	.14	.03	4.67	.0001	.28	.15	1.87	.07	
X10	.06	.02	3.00	.0004	.14	.16	.88	.39	
X1	.11	.03	3.67	.0004	1.06	.10	.60	.0001	.54
X3	.10	.02	5.00	.0001	.11	.12	.92	.35	
X6	.06	.02	3.00	.0004	.07	.13	.54	.55	
X8	.12	.03	4.00	.0001	.19	.15	1.27	.20	
X10	.06	.02	3.00	.0004	-.02	.15	-.13	.90	
X1	.09	.03	3.00	.0004	.97	.10	9.70	.0001	.58
X3	.10	.02	5.00	.0001	.42	.11	3.92	.0004	
X5	.09	.02	4.50	.0001	.31	.12	2.58	.01	
X8	.12	.03	4.00	.0001	.22	.14	1.57	.11	
X9	.06	.02	3.00	.0004	-.11	.14	-.79	.43	
X10	.06	.02	3.00	.0004	.17	.15	1.13	.26	
X1	.10	.03	3.33	.0004	1.02	.09	11.33	.0001	.62
X3	.09	.02	4.50	.0001	.41	.11	3.73	.0002	
X5	.08	.02	4.00	.0001	-.10	.11	-.91	.37	
X6	.05	.02	2.50	.03	.09	.12	.75	.45	
X8	.10	.03	3.33	.0004	.16	.13	1.23	.24	
X9	.06	.02	3.00	.0004	.20	.14	1.43	.16	
X10	.05	.02	2.50	.03	.11	.14	.79	.44	

PROC REG procedure. The number of significant principal component parameter estimates were subsequently identified. These procedures are outlined in the *SAS System for Regression* manual (Freund & Littell, 1991).

Results

The correlation matrix, means and standard deviations of the ten LASSI subscales are in Table 7. The intercorrelations among the subscales indicated that Anxiety was not significantly correlated with Time Management, Information Processing, Support Techniques/Materials, and Self-Testing. The lowest subscale mean was on Selecting Main Ideas.

Mallows Cp

The Mallows Cp statistic is calculated as: Cp = (SSEp/MSE) - (n - 2p) + 1 (Freund & Littell, 1991) or Cp = [1/^{2} (RSSp) - n + 2p] (Mallows, 1973); where RSSp is the residual sum of squares from the best variable subset model, MSE and/or ^{2} is the mean square error from the full model with all predictor variables, n = sample size, and p = number of predictors.

The procedure for finding the optimum subset of all possible subset sizes requires computing 2^m equations. The ten subscale predictors in the model yielded 1024 regression equations (2^{10}) with associated selection criteria statistics {Note: the determination of the number of subset equations generated for p predictor variables from an m variable full model is: m!/[p!(m-p)!]. For example, the number of 2 variable subset equations generated from a 10 variable model would be 45}. Only the single best variable subset models of each size are reported.

Table 6 (cont.) Mallows Cp and Principal Components Regression Comparison (n_1 = 100)

Best Variable Subset Model	Mallows Cp				Principal Components				
	ß	$SE_ß$	t	p	ß	$SE_ß$	t	p	R^2
X1	.10	.03	3.33	.0004	1.03	.09	11.44	.0001	.63
X2	.02	.01	2.00	.05	.18	.10	1.80	.09	
X3	.09	.02	4.50	.0001	.03	.11	.27	.77	
X5	.08	.02	4.00	.0001	.30	.11	2.72	.01	
X6	.05	.02	2.50	.03	.01	.13	.08	.92	
X8	.09	.03	3.00	.0004	.12	.13	.92	.36	
X9	.05	.02	2.50	.03	.25	.14	1.78	.09	
X10	.05	.02	2.50	.03	-.05	.14	-.36	.75	
X1	.09	.03	3.00	.0004	.99	.08	12.38	.0001	.64
X2	.02	.01	2.00	.05	.24	.10	2.40	.02	
X3	.08	.02	4.00	.0001	.03	.11	.27	.77	
X4	.05	.03	1.67	.10	.10	.11	.91	.36	
X5	.07	.02	3.50	.0004	-.08	.13	-.62	.52	
X6	.05	.02	2.50	.03	.08	.13	.62	.52	
X8	.09	.03	3.00	.0004	.02	.14	.14	.91	
X9	.05	.02	2.50	.03	-.001	.14	.007	.99	
X10	.05	.02	2.50	.03	.33	.15	2.20	.04	
X1	.09	.03	3.00	.0004	.97	.08	12.13	.0001	.65
X2	.02	.01	2.00	.05	.27	.10	2.70	.008	
X3	.08	.02	4.00	.0001	.05	.10	.50	.60	
X4	.05	.03	1.67	.10	-.09	.11	-.82	.42	
X5	.07	.02	3.50	.0004	.06	.11	.55	.59	
X6	.05	.02	2.50	.03	.06	.12	.50	.60	
X7	.02	.02	1.00	.25	-.07	.12	.58	.57	
X8	.09	.03	3.00	.0004	.01	.14	.07	.94	
X9	.04	.02	2.00	.05	.23	.15	1.53	.12	
X10	.04	.02	2.00	.05	.19	.15	1.27	.21	

Note. Regression parameters have been rounded to two decimal places unless otherwise noted. The t value = ß / $SE_ß$.

Table 7 LASSI Subscale Inter-Correlations, Means, and Standard Deviations (n = 156)

LASSI Subscale	1	2	3	4	5	6	7	8	9	10
1 Attention										
2 Motivation	.59									
3 Time Mngmnt	.39	.60								
4 Anxiety/Worry	.32	.15	.09							
5 Concentration	.57	.62	.62	.33						
6 Information	.20	.15	.39	.03	.26					
7 Select Ideas	.25	.36	.31	.37	.47	.30				
8 Support	.24	.40	.47	.05	.38	.45	.40			
9 Class Prep.	.38	.50	.63	.06	.55	.56	.39	.64		
10 Test Strategy	.54	.47	.33	.50	.66	.20	.60	.21	.34	
Mean	34.33	33.12	24.91	28.38	28.56	28.94	18.32	26.03	27.36	31.46
SD	4.17	4.73	6.18	5.92	4.93	5.24	3.51	5.96	5.84	4.58

Note. The values have been rounded to two decimal places.

Table 8 Best Model Selection Criteria by Subset Size

Subset Size	Variables in Subset Model	R^2	Cp
1	(2)	.09	10.88
2	(2),(8)	.11	8.01
3	(2),(6),(8)	.14	5.16
4	(2),(4),(8),(9)	.17	2.72
5	(2),(4),(6),(8),(9)	.18	2.93
6	(2),(4),(6),(7),(8),(9)	.18	3.68
7	(1),(2),(4),(6),(7),(8),(9)	.19	5.10
8	(1),(2),(4),(6),(7),(8),(9),(10)	.19	7.05
9	(1),(2),(3),(4),(5),(6),(8),(9),(10)	.19	10.04
10	(1),(2),(3),(4),(5),(6),(7),(8),(9),(10)	.19	11.00

Note. The four variable subset model according to the Cp criteria would be selected as the best model.

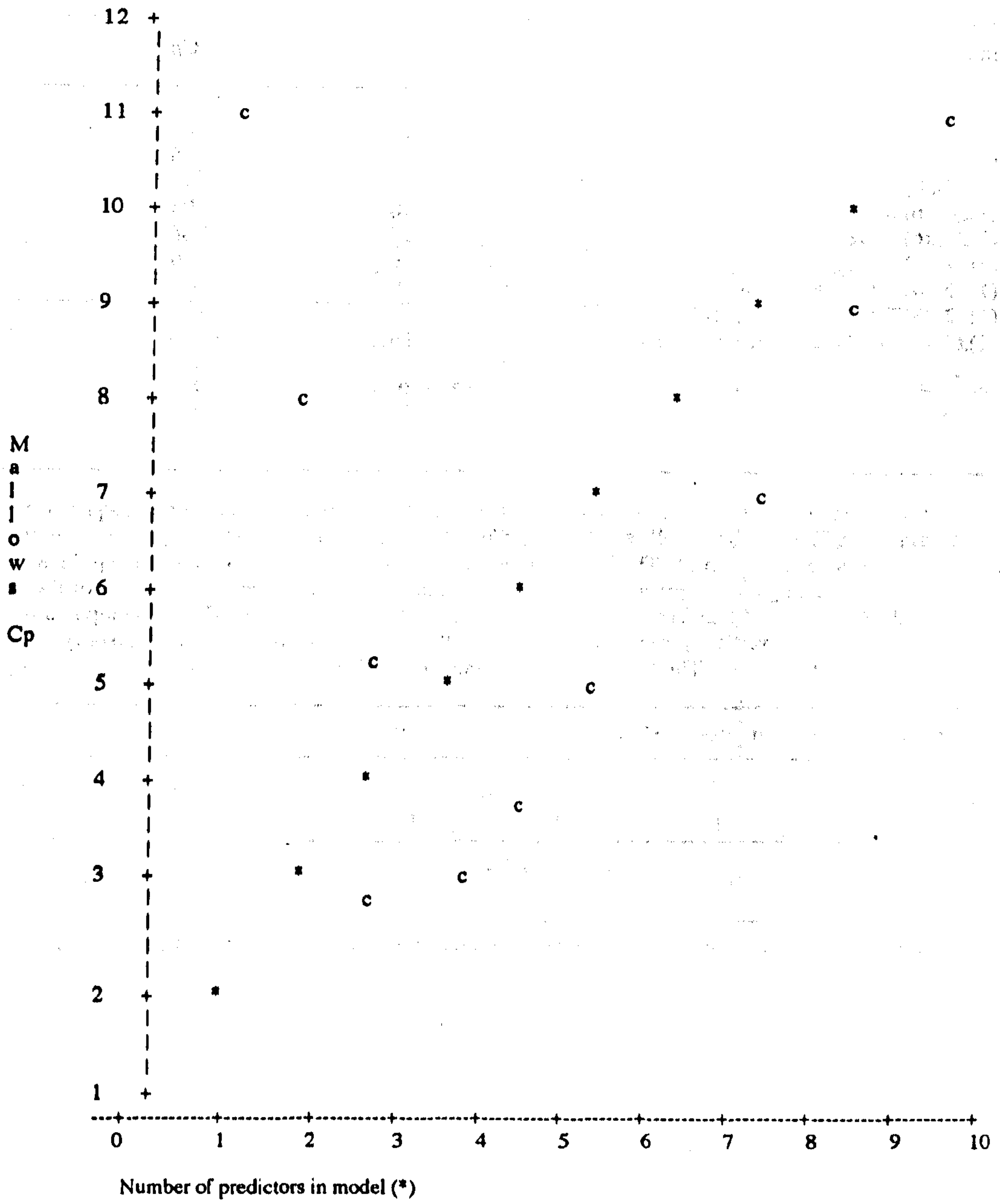
The best subset model for each subset size with the corresponding criteria are in Table 8. The Mallows Cp of 2.72 indicated a four variable subset model. The four variable subset model for predicting college grade point was Anxiety/ (4), Study Aids (8), and Self Testing (9). The Cp criteria also indicated the overfitting caused by having too many variables in the model. The large Cp values indicated equations with larger mean square error. If Cp > (p + 1), for any subset size p, then bias was present. If Cp < (p + 1), for any subset size p, then the model contained too many variables. A plot of the Cp values against the number of predictors, compared to a plot of the (p + 1) values, visually displays this phenomenon in Figure 1.

Table 9 Principal Component Regression

Model	Type II SS	df	MS	F	p.	R^2
Regression	10.76	10	1.08	3.35	.001	.19
Model Components						
(1)	4.16	1				
(2)	.99	1				
(3)	1.13	1				
(4)	1.93	1				
(5)	.09	1				
(6)	.23	1				
(7)	.58	1				
(8)	1.33	1				
(9)	.29	1				
(10)	.03	1				
Error	46.58	145	.32			
Total	57.34	155				

Note. Adj. R^2 = .13, PCR $R^2_{1,4,8}$ = 69 % (7.42/10.76).

Figure 1 Overlay Plot of Cp and (p + 1) Values



The present pattern of Cp values for the various subsets of size p are typical when multicollinearity is present. The Cp values initially become smaller, but then start to increase. The plot of Cp values is similar to a "scree" plot in factor analysis and as such a multiple regression method might also be useful in determining the number of variables to retain (Zoski & Jurs, 1993). The best subset model is indicated when the Cp values begin to increase and cross the (p + 1) values (Figure 1).

Principal Components Regression

Principal components are obtained by computing eigenvalues from the correlation matrix. The correlation matrix is used so that variables are not affected by the scale of measurement as in the use of a variance-covariance matrix. Since eigenvalues are the variances of the principal component variables, the sum of the eigenvalues equal the number of variables in the full model, just as the sum of standardized variable variances would equal the number of variables. This sum is the measure of the total variation in the data set. A wide variation in the eigenvalues would suggest the presence of multicollinearity among the variables. The number of eigenvalues greater than unity, as in factor analysis, would indicate the number of variables from the full model that would explain most of the variance in the data set. The eigenvectors, in contrast, contain the coefficients for each principal component variable. These coefficients are used to create the observed values of the original variables. These observed values are then used in multiple regression as orthogonal predictor values with no multicollinearity present.

Preliminary inspection of the model components (Type II SS) in Table 9 indicated three principal component variables (1, 4, and 8) that accounted for 69% of the variance in predicting college grade point average (7.42/10.76). The first model component alone explained 39 % of the variance (4.16/10.76).

A comparison of the full model parameter estimates in Table 10 between the original correlated predictors and the principal component regression variables sheds better insight into the best variable subset model selection criteria. The multiple regression analysis with correlated predictors identified Motivation (2) and Support (8) while the principal component method identified Attention (1), Anxiety/worry (4), and Support (8).

Summary

The Cp criteria identified a four variable predictor model as best: Motivation (2), Anxiety/worry (4), Support (8), and Class Preparation (9). This four variable subset model was further verified by examining where the plot of Cp values against the (p + 1) values crossed. The Cp criteria selected the smallest variable subset model in the presence of variable multicollinearity. The principal components approach identified Attention (1), Anxiety(4), and Study Aids (8). In examining the parameter estimates in the multiple regression analysis, only Motivation (2) and Study Aids (8) were significant relative to the other predictors in the model. The Mallows Cp and PCR criteria indicated slightly different sets of predictor variables depending upon whether the independent variables were correlated.

In using multiple regression it is important to have a theoretical basis for the regression model and to consider model validation. A common misconception in multiple regression is that the model with all the significant predictors included is the best model. This isn't always the case. The problem is that the beta values and R-squared values are data dependent due to the least squares criterion being applied to a specific sample of data. A different sample will usually result in different parameter estimates and variance explained. Although the standard errors of the beta values do provide the researcher with some indication of the amount of change expected from sample to sample, the fact remains that the estimates obtained from one sample may predict poorly when applied to a new set of sample data. The primary method to assess any change in estimates is to replicate the regression model using other sample data. The Mallows Cp criteria was similarly suspect because values were inflated upon cross validation and the best variable subset model in one sample was not identified in the other sample. Obviously, if the mean square error estimates and the residual sums of squares fluctuate, then model selection will be erroneous (see Mallows Cp formula).

The rationale behind a regression model is to estimate σ^2 (the true model's mean square error variance). Since σ^2 is not generally known, a researcher must estimate it from a knowledge of prior research ($\sigma^2 = \sigma^2_{y.x}$), obtain estimates from a model containing all theoretically relevant predictors, replicate the study, or use bootstrapping, jacknifing, and cross-validation methods. In this regard, effect size considerations, as recommended by Thompson et al. (1991), become important to consider in evaluating a regression model.

Table 10 Multiple Regression and Principal Component Parameter Estimate Comparisons

	Mallows Cp				Principal Components			
Variable	ß	$SE_ß$	t	p	ß	$SE_ß$	t	p
1	.01	.02	.68	.50	.08	.02	3.60	.001
2	.03	.02	2.29	.02	.06	.04	1.76	.081
3	.002	.01	.19	.84	-.08	.04	-1.88	.062
4	.02	.01	1.84	.07	.14	.06	2.45	.015
5	-.003	.02	-.17	.87	-.03	.06	-.53	.600
6	.01	.01	1.30	.20	.05	.06	.84	.404
7	-.02	.02	-1.02	.31	.10	.08	1.34	.182
8	-.03	.01	-2.82	.005	-.18	.09	-2.03	.044
9	.02	.01	1.28	.20	.09	.09	.95	.341
10	.005	.02	.27	.79	-.03	.10	-.31	.758

References

Aiken, L.S. & West, S.G. (1993). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: SAGE

Berk, K.N. (1977). Tolerance and condition in regression computations. *Journal of the American Statistical Association, 72*, 863-866.

Constas, M.A. & Francis, J.D. (1992). A graphical method for selecting the best subset regression model. *Multiple Linear Regression Viewpoints, 19*(1), 16-25.

Cummings, Corenna, C. (1982, March). *Estimates of multiple correlation coefficient shrinkage.* Paper presented at the American Educational Research Association annual meeting. New York, NY.

Davidson, Betty, M. (1988, November). *The case against using stepwise research methods.* Paper presented at the Mid-South Educational Research Association annual meeting. Louisville, KY.

Eldredge, J.L. (1990). Learning and study strategies inventory: a high school version (test review). *Journal of Reading, 34*, 146-149.

Freund, R.J. & Littell, R.C. (1991). SAS System for Regression (2nd Ed.) SAS Institute: Cary, NC.

Henderson, Douglas, A. & Denison, Daniel R. (1989). Stepwise regression in social and psychological research. *Psychological Reports, 64*(1), 251-257.

Huberty, C.J. (1989). Problems with stepwise methods--better alternatives. In B. Thompson (Ed.), *Advances in Social Science Methodology, 1*, 43-70. Greenwich, CT: JAI Press.

Mallows, C.L. (1966). *Choosing a subset regression.* Paper presented at the Joint Statistical Meetings. Los Angeles, CA.

Mallows, C.L. (1973). *Some comments on C_p.* Technometrics, 15, 661-675.

Morrison, D.F. (1976). *Multivariate Statistical Methods* (2nd Ed.), New York: McGraw-Hill.

Pohlmann, J. (1983, April). *A perspective on multicollinearity.* Paper presented at the American Educational Research Association annual meeting. Montreal, Canada.

Snyder, P. (1991). Three reasons why stepwise regression methods should not be used by researchers. In B. Thompson (Ed.), *Advances in Educational Research: Substantive findings, methodological developments, 1*, 99-105. Greenwich, CT: JAI Press.

Thayer, Jerome, D. (1986, April). *Testing different model building procedures using multiple regression.* Paper presented at the American Educational Research Association annual meeting. San Francisco, CA.

Thayer, Jerome, D. (1990, April). *Implementing variable selection techniques in regression.* Paper presented at the American Educational Research Association annual meeting. Boston, MA.

Thompson, B. (1989). Why won't stepwise methods die? *Measurement and Evaluation in Counseling and Development, 21(4)*, 146-148.

Thompson, B., Smith, Q.W., Miller, L.M., & Thomson, W.A.. (1991, January) *Stepwise methods lead to bad interpretations: better alternatives.* Paper presented at the Southwest Educational Research Association annual meeting. San Antonio, TX.

Tracz, S., Brown, R., & Kopriva, R. (1991). Considerations, issues, and comparisons in variable selection and interpretation in multiple regression. *Multiple Linear Regression Viewpoints, 18*(1), 55-66.

Weinstein, C.E. (1987). *LASSI user's manual.* Clearwater, FL: H&H Publishing Company, Inc.

Weinstein, , C.E., Palmer, D.R., Schulte, A.C. (1987). *Learning and Study Strategies Inventory.* Florida: H&H Publishing.

Welge, Patricia (1990, January). *Three reasons why stepwise regression methods should not be used by researchers.* Paper presented at the Southwest Educational Research Association annual meeting. Austin,TX.

Zoski, K.K. & Jurs, S.G. (1993). Using multiple regression to determine the number of factors to retain in factor analysis. *Multiple Linear Regression Viewpoints, 20*(1), 5-9.

Zuccaro, Cataldo (1992). Mallows' C_p statistic and model selection in multiple linear regression. *Journal of the Market Research Society,34*(2), 163-172.

The author wishes to acknowledge Dr. Panu Sittiwong in the Academic Computing Center at the University of North Texas for his assistance in coding the SAS programs for this study.