Rangsit Journal of Arts and Sciences, July-December 2013 RJAS Vol. 3 No. 2, pp. 147-154

ISSN 2229-063X (Print)/ISSN 2392-554X (Online)
DOI : 10.14456/rjas.2013.3

A variable selection method in multiple linear regression models based on Tabu Search

Kannat Na Bangchang

Faculty of Science and Technology, Valaya Alongkorn Rajabhat University, Patumthani 13180, Thailand

Email: kn_matu@msn.com

Submitted 9 January 2013; accepted in final form 17 October 2013

Abstract

This research has proposed a variable selection method based on the Tabu Search for multiple linear regression models. In this study two objective functions used in the Tabu Search are mean square error (MSE) and the mean absolute error (MAE). The results of the Tabu Search are compared with the results obtained by the stepwise regression method based on the hit percentage criterion. The simulations cover both cases, without and with multicollinearity problems. Without multicollinearity problem, the hit percentages of the stepwise regression method and the Tabu Search using the objective function of MSE are almost the same but slightly higher than the Tabu Search using the objective function of the MAE. But with multicollinearity problem the hit percentages of the Tabu Search using the objective function of MSE or the MAE are higher than the hit percentage of the stepwise regression method. Additionally, the correlation coefficients between the independent variables X_1 and X_4 are higher; yielding hit percentages that are lower.

Keywords**:** *stepwise regression, Tabu Search, variable selection*

บทคัดย่อ

งานวิจัยนี้ได้เสนอวิธีการคัดเลือกตัวแปรในตัวแบบการถดถอยเชิงเส้นพหุ โดยใช้วิธีการค้นหาแบบต้องห้าม ในการศึกษานี้มีฟังก์ชันเป้าหมาย 2 ฟังก์ชันที่ใช้ในการค้นหาแบบต้องห้าม คือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย แล้วนำผลมาเปรียบเทียบกับวิธีการถดถอยแบบขั้นตอน เกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ร้อยละของจำนวนครั้งที่คัดเลือกตัวแปรในตัวแบบได้ถูกต้อง การศึกษาได้ทำในกรณีที่ไม่มีและมีปัญหาสหสัมพันธ์เชิงเส้นพหุ โดยใช้วิธีการจำลอง กรณีไม่มีปัญหาสหสัมพันธ์เชิงเส้นพหุ ความสามารถในการคัดเลือกตัวแปรของวิธีการถดถอยแบบขั้นตอนและการค้นหาแบบต้องห้ามที่มีฟังก์ชันเป้าหมายเป็นค่าความคลาดเคลื่อนกำลังสองเฉลี่ยมีร้อยละของความถูกต้องใกล้เคียงกัน แต่มากกว่าการใช้วิธีการค้นหาแบบต้องห้ามที่ใช้ฟังก์ชันเป้าหมายเป็นค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเล็กน้อย กรณีที่มีปัญหาสหสัมพันธ์เชิงเส้นพหุ วิธีการค้นหาแบบต้องห้ามที่มีฟังก์ชันเป้าหมายเป็นเป็นค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยมีร้อยละของความถูกต้องในการคัดเลือกตัวแปรสูงกว่าวิธีการถดถอยแบบขั้นตอน นอกจากนี้พบว่าเมื่อสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระ X_1 และ X_4 เพิ่มขึ้น ทำให้ร้อยละของความถูกต้องในการคัดเลือกตัวแปรเข้าสู่ตัวแบบลดลง

คำสำคัญ*: การถดถอยแบบขั้นตอน, การค้นหาแบบต้องห้าม, การคัดเลือกตัวแปร*

1. Introduction

At the present time, in many branches of research the use of statistical methods is a necessity if one is to arrive at an overall picture. Inferential statistics are used in the analysis of data in order to summarize the overall features of a population, while at the same time locating the interrelationships among dependent and independent variables. However, there are many methods, depending on the objectives of the research. One popular approach is that of "regression analysis" which is a study of the relationships between independent and dependent variables, in order to explain or predict the values of dependent variables on the basis of independent variables (Montgomery, Peck, & Vining, 2006).

In the literature, there are many examples of the Tabu Search being used and modified for specific result criteria. The Tabu Search is applied widely such as when Sacchi and Armentano (2011) used compute parametric Tabu Search for solving 0-1 mixed integer programming problems by solving a series of linear programming as weighted terms in the objective function (Sacchi & Armentano, 2011). Pacheco, Casado, & Nunez (2009) analyzed for selecting variables that are subsequently used in logistic regression models and found that yields the

greatest percentage of hits in logistic regression (Pacheco et al., 2009). Shen, Shi, & Kong (2010) modified the Tabu Search approach for variable selection in quantitative structure-activity relationship demonstrated that the modified Tabu Search is a tool for variable selection which converges quickly towards the optimal position (Shen et al., 2010). Finally, Turajlić and Dragović (2012) used Tabu Search for selecting the set of services with the best overall quality of service by solving the web service selection problem (Turajlić & Dragović, 2012). Consequently, they used Tabu Search in many ways and always used in variable selection.

The validity of applying a linear regression model depends on an appropriate choice of model, requiring a sufficient number of independent variables to explain the dependent variables. But if more independent variables are selected than necessary for application to our model, a large error in prediction will result (Seenoi, 2010). Currently there are many methods for selecting variables, but one method that is extensively used is "stepwise regression". This is a method of choosing independent variables for a regression model. In addition to the technique of selecting variables for use in the method of stepwise regression, there are "combinatorial" techniques for finding optimal values, which can likewise be used for selecting suitable independent variables, for instance the methods of "Tabu Search" (Glover, 1990) and "genetic programming" (Holland, 1975). In this study the Tabu Search is applied to selecting independent variables, comparing the results with that of stepwise regression.

2. Objectives

The purpose of this study is to compare the selection of independent variables by the Tabu Search using mean square error (MSE) and the mean absolute error (MAE) with that of stepwise regression, in the multiple linear regression model, by using data that simulates sample sizes of 25 and 100, for data without and with multicollinearity.

3. Theory

In this study, the multiple linear regression model is of the form $y = X\beta + \varepsilon$, where y is the dependent variable vector, of size $n \times 1$, X is the matrix of independent variables of size $n \times (k + 1)$, β is the vector of model parameters, of size $(k + 1) \times 1$, ε is the error vector of size $n \times 1$, subject to the condition $\varepsilon \sim N_n(0, \sigma^2 I_n)$, where I_n is an identity matrix of size $n \times n$, n is the sample size, and k is the number of independent variables in the model. Since the essential objective of a regression analysis (Bruce, David, & Richard, 1990) is to arrive at a suitable regression model, which depends on the criteria for selecting variables to be entered in the model, one method that is used extensively is that of stepwise regression.

Stepwise regression is a method for selecting independent variables to be entered into the regression model, by selecting each independent variable that is most closely related to a dependent variable in the regression model. The independent variable entered at each step of the model is checked for its significance. If at any step an independent variable entered into the model is insignificant, it is removed. An independent variable entered into the regression model must be tested for its part in explaining variation in the dependent variable. In the case that there are other independent variables in the model (Efroymson, 1960), any independent variable selected for inclusion in the regression model may later be removed if it is found that the variable has no significance. As to specifying the level of significance at which to select an independent variable for inclusion in the model (α_{entry}), with that at which to remove a variable from the model (α_{stay}), these should be specified as of equal value (Smith & Richard, 1981)

Tabu Search is one method of finding a metaheuristic which is used to find a solution close to the optimum value (Glover, 1990) and which is popularly applied to decision problems, such as the example of the "travelling salesman problem" (Knox, 1989) and "job scheduling". The stages in this procedure are not complicated and the result is highly efficient.

Further, a study has been made of methods for finding optimal values of a function for which the variables are continuous, using a Tabu Search for the solution (Siary & Berthiau, 1997). The Tabu Search for a solution is applied in the selection of the independent variables for a multiple linear regression analysis, finding the highest among the adjusted R^2) R_a^2 ((Salhi, Drezner, & Marcoulides, 1999). The Tabu Search for a solution can take into consideration more objective functions of varying forms. Though, this method of selecting variables to be entered into the model is different from other methods. Apart from this, studies have been made of the selection of appropriate variables for linear

regression analysis by using the methods of searching for on "optimal" value, "mathematical programming", "Lagrangian relaxation" and the "greedy randomized adaptive search procedure" (Eksioglu, Demirer, & Capar, 2005), while there have been studies of the selection of appropriate models for logistic regression analysis, using the Tabu Search for a solution (Pacheco et al., 2009).

4. Methods

This research studies the case of independent variables having a uniform distribution and the error having a normal distribution. There are 6 independent variables in the "full model" and there are 4 independent variables in the model used to find values of the dependent variables, that is the variables X_1, X_2, X_3, and X_4, specifying a population size N = 100,000. We set up the independent variables as $X_1 \sim U(15,80)$, $X_2 \sim U(20,150)$, $X_3 \sim U(10,100)$, $X_4 \sim U(50,50)$, $X_5 \sim U(-100,100)$, and $X_6 \sim U(-80,400)$. The error is set up to have a normal distribution $\varepsilon \sim N0,2500)$. We specify the coefficient of regression, which is, $\beta' = (100,24,15,-8,5,0,0)$. The sample sizes used are 25 and 100. Apart from this, we specify the correlation coefficient between the independent variables X_1 and X_4 equal to 0.999. After that, calculation of the correlation coefficients between X_1 and X_4 within each sample are obtained.

Consider the division into groups of randomly obtained samples in accordance with the correlation coefficients between X_1 and X_4, as follows:

$0.495 \leq r \leq 0.504$ Classify them into groups with equal correlation coefficients 0.50.

$0.945 \leq r \leq 0.954$ Classify them into groups with equal correlation coefficients 0.95.

$0.955 \leq r \leq 0.964$ Classify them into groups with equal correlation coefficients 0.96.

$0.965 \leq r \leq 0.974$ Classify them into groups with equal correlation coefficients 0.97.

$0.975 \leq r \leq 0.984$ Classify them into groups with equal correlation coefficients 0.98.

$0.985 \leq r \leq 0.994$ Classify them into groups with equal correlation coefficients 0.99.

Repeat this process 500 times in each time of correlation coefficient group and sample size.

Later, the selection of a suitable model is undertaken by the method of stepwise regression by specifying the level of significance at which to select an independent variable for entry into the model (α_{entry}) together with that at which to remove a variable from the model (α_{stay}), to be equal in value at 0.05. For a Tabu Search the researchers specified the "Tabu tenure list" as equal to 10, and considered an objective function

$$\text{MSE} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n-k-1} \quad (1)$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|}{n} \quad (2)$$

In consideration of finding the parameter estimators for a regression analysis that provided a minimum value for both objective functions. After that the average MSE for 500 repetitions of the equation for selection of the variables were calculated. Besides, the criteria in this study was to calculate the percentage of the number of times the selection of the correct in variable selection as concluded in Tables 1 and 2.

5. Results

It is found from Table 1, that when the independent variables X_1 and X_4 have a correlation coefficient of less than 0.1, the method of Tabu Search, using as objective functions the MSE and MAE, correctly selects the independent variables to go into the model at a rate of 88.60% and 85.60% respectively. This is similar to the method of stepwise regression, which correctly selects the independent variables to go into the model at a rate of 89.40%

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.495 to 0.504, the Tabu Search, using as objective functions the MSE and MAE, correctly selects independent variables to go into the model at a rate of 79.40% and 79.20% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 45.20%.

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.945 to 0.954, the Tabu Search, using as objective functions the MSE and MAE, correctly selects independent variables to go into the model at a rate of 71.40% and 70.20% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 15.60%.

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.955 to 0.964, the Tabu Search, using as objective functions the MSE and MAE, correctly selects independent variables to go into the model at a rate of 70.80% and 70.00% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 12.40%.

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.965 to 0.974, the Tabu Search, using as objective functions the MSE and MAE, correctly selects independent variables to go into the model at a rate of 71.00% and 70.40% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 10.20%.

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.975 to 0.984, the Tabu Search, using as objective functions the MSE and MAE, correctly selects independent variables to go into the model at a rate of 71.20% and 69.40% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 9.20%.

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.985 to 0.994, the Tabu Search, using as objective functions the MSE and MAE, correctly selects independent variables to go into the model at a rate of 70.60% and 69.60% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 7.40%.

Table 1 The mean of regression coefficient and mean square error (MSE) by the correlation coefficient[1] and the method of variable selection When ε_i ~N(0,2500) , n = 25

	b_0	b_1	b_2	b_3	b_4	b_5	b_6	MSE	The hit percentage
r < 0.1									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	99.7028	26.2503	15.3315	-10.3452	7.5213	0.0414	-0.0223	2509.13	**88.60**
Tabu_MAE	98.9161	26.1159	15.3436	-10.2703	7.6571	-0.0817	-0.0248	6002.32	**85.60**
Stepwise	100.9034	23.9646	15.0186	-8.0125	5.0032	0.0606	0.0799	2513.25	**89.40**
0.495 ≤ r ≤0.504									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	104.3217	22.8864	15.4021	-7.0276	4.1275	-0.0898	0.0314	2511.32	**79.40**
Tabu_MAE	104.2415	22.9238	15.4273	-6.9981	4.0321	0.0731	0.0206	6179.11	**79.20**
Stepwise	59.1386	26.7712	15.0089	-8.0095	9.4512	-0.0054	0.0201	2517.24	**45.20**
0.945 ≤ r ≤0.954									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	105.7548	20.9375	15.4692	-6.9312	3.8820	-0.1098	0.0654	2521.52	**71.40**
Tabu_MAE	105.7425	20.9416	15.4937	-6.7558	3.6995	0.1123	0.0456	6879.11	**70.20**
Stepwise	-34.1497	28.8845	15.0109	-8.0175	12.4643	-0.0181	0.0322	2523.11	**15.60**
0.955 ≤ r ≤0.964									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	104.8070	20.9193	15.7283	-6.8772	3.6059	-0.2167	0.0568	2521.54	**70.80**
Tabu_MAE	104.7409	20.9776	15.4020	-6.7680	3.4288	0.2389	0.0901	6880.34	**70.00**
Stepwise	-36.2087	29.8401	15.0093	-8.0199	15.2276	-0.1658	0.0327	2524.17	**12.40**
0.965 ≤ r ≤0.974									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	103.7689	21.0725	15.2612	-6.7661	3.7082	-0.1176	0.0567	2521.24	**71.00**
Tabu_MAE	103.6011	21.2781	15.3523	-6.9222	3.6174	0.0875	0.0885	6787.11	**70.40**
Stepwise	-48.8832	30.6709	15.0092	-8.0156	16.9987	-0.1844	0.0717	2522.25	**10.20**
0.975 ≤ r ≤0.984									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	104.4026	21.0057	15.7771	-6.7475	3.6870	-0.0812	0.0765	2520.52	**71.20**
Tabu_MAE	107.8179	21.4606	15.5949	-7.0149	3.7354	0.0487	0.0875	6809.35	**69.40**
Stepwise	-58.6637	30.6233	15.0132	-8.0128	20.1427	-0.0774	0.0621	2524.11	**9.20**
0.985 ≤ r ≤0.994									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	104.6102	21.2412	15.7270	-6.8583	3.8569	-0.0776	0.0342	2523.54	**70.60**
Tabu_MAE	103.8257	21.3493	15.6940	-6.5812	3.7460	0.0648	0.0432	6802.34	**69.60**
Stepwise	-79.1431	31.6635	15.0129	-8.0062	20.4422	-0.0755	0.0548	2526.12	**7.40**

[1] The regression coefficient for simulation

[2] The correlation coefficient between X_1 and X_4

Table 2 The mean of regression coefficient and mean square error (MSE) by the correlation coefficient[1] and the method of variable selection When ε_i ~N(0,2500) , n = 100

	b_0	b_1	b_2	b_3	b_4	b_5	b_6	MSE	The hit percentage
r < 0.1									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	100.1283	25.8183	15.1431	-9.5979	7.5622	0.0130	-0.0158	2496.23	**90.00**
Tabu_MAE	99.2788	26.0340	15.5058	-10.0237	7.9466	0.0149	0.1054	5987.98	**87.20**
Stepwise	99.7283	24.0044	15.0022	-8.0078	4.9939	0.0377	0.0142	2499.15	**91.00**
0.495 ≤ r ≤0.504									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	105.3216	23.0012	15.2156	-7.2897	4.4173	-0.0529	0.0219	2507.62	**82.40**
Tabu_MAE	104.9913	23.2075	15.3071	-7.1125	4.2794	0.0698	0.0235	6109.22	**81.20**
Stepwise	67.3206	26.3129	15.0066	-8.0071	9.0315	-0.0034	0.0191	2515.13	**54.80**
0.945 ≤ r ≤0.954									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	106.2804	20.5615	15.1781	-6.7096	3.9536	-0.0678	0.0567	2497.16	**74.40**
Tabu_MAE	104.7179	21.1603	15.2607	-6.6858	3.8422	0.0789	0.0812	6780.35	**73.20**
Stepwise	-14.8497	27.3814	15.0214	-8.0178	12.4569	-0.0482	0.0315	2512.33	**24.60**
0.955 ≤ r ≤0.964									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	106.9925	20.9708	14.7114	-6.8436	3.7251	-0.0235	0.0876	2497.01	**74.60**
Tabu_MAE	104.9351	21.0167	15.4016	-6.7386	3.7644	0.0767	0.0689	6770.51	**73.20**
Stepwise	-19.2647	28.1305	15.0221	-8.0192	14.7835	-0.0859	0.0388	2502.12	**22.20**
0.965 ≤ r ≤0.974									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	107.3514	20.9038	15.0048	-6.8593	3.7441	-0.0589	0.0812	2496.23	**74.20**
Tabu_MAE	104.0410	21.0909	15.4366	-6.9650	3.6830	0.0598	0.0865	6712.11	**73.40**
Stepwise	-20.8481	30.0913	15.0251	-8.0180	15.6083	-0.0647	0.0449	2503.17	**19.80**
0.975 ≤ r ≤0.984									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	107.5602	21.2171	15.0951	-6.7796	3.7327	-0.0768	0.0435	2495.16	**75.00**
Tabu_MAE	104.9340	20.9719	15.5013	-6.8817	3.8298	0.0241	0.0245	6687.35	**74.60**
Stepwise	-28.3969	30.5877	15.0241	-8.0197	18.8786	-0.0565	0.0313	2501.23	**16.20**
0.985 ≤ r ≤0.994									
Defined[2]	100	24	15	-8	5	0	0		
Tabu_MSE	107.1490	21.5488	14.9392	-7.1208	3.8325	-0.0543	0.0612	2498.34	**74.80**
Tabu_MAE	105.8325	21.3514	15.3534	-7.1485	3.7613	0.0439	0.0786	6788.51	**74.00**
Stepwise	-32.8049	32.0402	15.0209	-8.0156	19.1203	-0.0371	0.0409	2503.12	**12.40**

[1] The regression coefficient for simulation
[2] The correlation coefficient between X_1 and X_4

It is found from Table 2 that, when the independent variables X_1 and X_4 have a correlation coefficient of less than 0.1, the method of Tabu Search, using as objective functions the MSE and MAE, correctly selects the independent variables to go into the model at a rate of 90.00% and 87.20% respectively, which is similar to the method of stepwise regression, which correctly selects the independent variables to go into the model at a rate of 91.00%

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.495 to 0.504, the Tabu Search, using as objective functions the MSE and MAE, correctly selects independent variables to go into the model at a rate of 82.40% and 81.20% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 54.80%.

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.945 to 0.954, the Tabu Search, using as objective functions the MSE and MAE, correctly selects independent variables to go into the model at a rate of 74.40% and 73.20% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 24.60%.

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.955 to 0.964, the Tabu Search, using as objective functions the MSE and MAE, correctly selects

independent variables to go into the model at a rate of 74.60% and 73.20% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 22.20%.

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.965 to 0.974, the Tabu Search, using as objective functions the MSE and MAE, correctly selects independent variables to go into the model at a rate of 74.20% and 73.40% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 19.80%.

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.975 to 0.984, the Tabu Search, using as objective functions the MSE and MAE, correctly selects independent variables to go into the model at a rate of 75.00% and 74.60% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 16.20%.

When the independent variables X_1 and X_4 have a correlation coefficient in the range 0.985 to 0.994, the Tabu Search, using as objective functions the MSE and MAE, correctly selects independent variables to go into the model at a rate of 74.80% and 74.00% respectively, which is higher than the method of stepwise regression, which correctly selects independent variables to go into the model at a rate of 12.40%.

6. Discussion

When comparing the efficiency of selection of models, as the sample size increases, in the selection of independent variables by the method of Tabu Search using either of the two objective functions and by the method of stepwise regression, the percentage of models correctly selected increases. This occurs in cases where the data have or do not have multicollinearity.

When comparing the efficiency of selection of independent variables as to whether the data have multicollinearity, it is found that when the correlation coefficient among the independent variables X_1 and X_4 increases, the method of selection by Tabu Search, with either of the two objective functions, the percentage of models correctly selected is similar. In the case of stepwise regression the percentage of models correctly selected is lower.

With the estimation of values by the method of least squares (OLS), when the independent variables have a multicollinearity problem, it is found that the estimated value for the regression coefficient for the related independent variables trends away from the value specified in the model and is higher than the value of the correlation coefficient for the independent variables obtained by Tabu Search, whether the MSE or the MAE is used as the objective function. This is because Tabu Search method does not use $(X'X)^{-1}$ for an estimate parameter leading to error in estimation (Piriyakul, 1986). However, there is one caveat to this method. The standard error in the estimated value, in the case that the independent variables show a multicollinearity problem, is higher than in the case where they do not, and the value of the standard error as obtained from the Tabu Search is of a similar character to that from the method of stepwise regression, in consequence of using the OLS formula for calculation. However the MSE value obtained with either of the two objective functions causes the calculated value to be lower than it should be. Thus it will make it possible to report that, in the case where the data have a multicollinearity problem, the Tabu Search will enable the researcher to estimate values for the regression coefficient closer to the model parameters than does the OLS method. Apart from this, it can be better used in the estimation of the regression coefficient than can the method of "ridge egression," since it is not necessary to use judgment when deciding on the "ridge trace". Moreover, this is like a robust form of redundancy analysis, seeking directions in the factor space that are associated with high variation in the responses but biasing them toward directions that are accurately predicted (Gujarati, 2009).

7. Conclusion

We define terms as follows:

For the independent variables X_1 and X_4 to have a low level multicollinearity means that the correlation coefficient among independent variables X_1 and X_4 has a value less than 0.1.

For the independent variables X_1 and X_4 to have a medium level multicollinearity means that the correlation coefficient among independent variables X_1 and X_4 has a value of 0.5.

For the independent variables X_1 and X_4 to have a high level multicollinearity means that the correlation coefficient among independent variables X_1 and X_4 has a value of 0.95, 0.96, 0.97, 0.98 or 0.99.

In the case that the independent variables X_1 and X_4 have a low level multicollinearity, the method of selecting independent variables by stepwise regression yields a value of the regression coefficient for every independent variable similar to the value specified in the model, and the highest percentage of models that are correctly selected. In the case of the Tabu Search using either of the two objective functions a value of the regression coefficient is obtained for the independent variables $X_1(\beta_1)$, $X_3(\beta_3)$ and $X_4(\beta_4)$ higher than that specified in the model and the percentage of models correctly selected is a little lower than with the method of stepwise regression. Apart from this, it is found that the percentage of models correctly selected increases when the sample size increases both with the Tabu Search and the method of stepwise regression.

In the case that the independent variables X_1 and X_4 have a medium or high level multicollinearity, the selection of independent variables by the method of Tabu Search, using either of the two objective functions yields a value of the regression coefficient for every independent variable similar to the value specified in the model in all circumstances, and the percentage of models correctly selected is higher than with the method of stepwise regression. Apart from this, it is found that the percentage of models correctly selected increases when the sample size increases and is similar to that when the value of the correlation coefficient among independent variables X_1 and X_4 changed. When the independent variables are selected by the method of stepwise regression, this gives a regression coefficient for the independent variables (β_0) that indicated a difference from that specified in the model. Apart from this, it is found that the value of the regression coefficient for the independent variables $X_1(\beta_1)$ and $X_4(\beta_4)$ has a value higher than that specified in the model in all circumstances. In addition, it is found that the percentage of models correctly selected increases when the sample size increases, but the value of the correlation coefficient among independent variables X_1 and X_4 decreases.

8. Acknowledgements

I would like to express my special thanks to Associate Professor Jirawan Jitthavech and Associate Professor Vichit Lorchirachoonkul who gave good advice and guidance for this study.

9. References

Bruce, L. B., David, A. D., & Richard, T. O. (1990). *Linear Statistical models: an applied approach* (2nd ed.). Boston, USA: Duxbury Press.

Efroymson, M. A. (1960). Multiple regression analysis. In A. Ralston & H.S. Wilf (Eds.), *Mathematical methods for digital computers* (pp. 191-203). New York, USA: Wiley.

Eksioglu, B., Demirer, R., & Capar, I. (2005). Subset selection in multiple linear regression: a new mathematical programming approach. *Computer & Industrial Engineering*, *49*(1), 155-167.

Glover, F. (1990). Tabu search: A tutorial. *Interfaces*, *20*(4), 74-94.

Gujarati, D.N. (2009). *Basic Econometrics* (4th ed.). New York, USA: McGraw-Hill.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MA, USA: The University of Michigan Press. Republished by the MIT Press 1992.

Knox, J. (1989). *The application of tabu search to the symmetric traveling salesman problem*.
A doctoral dissertation. University of Colorado at Boulder, CO, USA.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to linear regression analysis* (4th ed.). New Jersey, USA: John Wiley & Sons, Inc.

Pacheco, J., Casado, S., & Nunez, L. (2009). A variable selection method based on Tabu Search for logistic regression models. *European Journal of Operational Research (EJOR)*, *199*(2), 506-511.

Piriyakul, M. (1986). *Regression Analysis*. Bangkok, Thailand: Chuanpim Press. (In Thai)

Sacchi, L. H., & Armentano, V. A. (2011). A Computational study of parametric tabu search for 0-1 mixed integer programs. *Computer & Operational Research*, *38*(12), 464-473.

Salhi, S. and Drezner, Z. and Marcoulides, G. (1999) Tabu search model selection in multiple regression analysis. *Communication in Statistics: Simulation and Computation*, *28*(2), 349-367.

Seenoi, P. (2010). *Test Statistics for selecting multiple linear regression models*. (Master's thesis, National Institute of Development Administration, Bangkok, Thailand). Retrieved from http://libsearch.nida.ac.th

Shen, Q., Shi, W., & Kong, W. (2010). Modified tabu search approach for variable selection in quantitative structure-activity relationship studies of toxicity of aromatic compound. *Artificial Intelligence in Medicine*, *49*(2), 61-66.

Siary, P., & Berthiau, G. (1997). Fitting of tabu search to optimize functions of continuous variables. *International Journal for Numerical Methods in Engineering*, *40*(13), 2449-2457.

Smith, D., & Richard, N. (1981). *Applied regression analysis* (2nd ed.). New York, USA: Wiley & Sons, Inc.

Turajlić, N., & Dragović, I. (2012). A hybrid metaheuristic based on variable neighborhood search and tabu search for the web service selection problem. *Electronic Notes in Discrete Mathematics*, *39*(1), 145-152.