Selected Model Systematic Sequence via Variance Inflationary Factor

Zainodin H. J.[1*], Khuneswari G.[2], Noraini A.[1], Haider F. A. A.[1]
[1] Universiti Malaysia Sabah, Faculty of Science and Natural Resources, 88400 Kota Kinabalu, Sabah, Malaysia.
[2] University of Glasgow, School of Mathematics and Statistics, G12 8QQ Glasgow, United Kingdom.

* Corresponding author. email: zainodin@gmail.com
Manuscript submitted January 3, 2015; accepted March 3, 2015.
doi: 10.17706/ijapm.2015.5.2.105-114

Abstract: Literature reviews revealed that multicollinearity always exists when model a deals with several independent variables. This phenomenon can cause the t statistic and the related probability-value to give a misleading impression of the importance of the independent variables. There are two approaches in tackling this issue. The common approach is correlation-coefficient based and the other is variance-based. Many softwares in the market have highlighted this phenomenon and offer options in minimising the effect. Currently, the variance-based approach is widely available in the software market. This is because it does not depend on the type of dependent variables. This variance-based approach via Variance Inflation Factor (VIF) quantifies the severity of multicollinearity in an ordinary least squares regression analysis. It provides an index that measures how much the variance (the square of the estimate's standard deviation) of an estimated regression coefficient is increased because of collinearity. Thus, here, a novel approach is revealed in detailing the procedures to remove several variables due to multicollinearity effects. Ultimately, the insignificant variables are eliminated. It is found that when a very stringent criterion is set for multicollinearity, the process of elimination of variables becomes smooth and easy besides shortening the number of iteration.

Key words: Hierarchically multiple regression models, insignificant effects, multicollinearity effects, selected model, variance inflation factor.

1. Introduction

Model building is an art and science. Different individual may not always agree on the best multiple regression models.

Consider the following general model:

$$Y = \Omega_0 + \Omega_1 W_1 + \Omega_2 W_2 + \ldots + \Omega_{k-1} W_{k-1} + \Omega_k W_k + u \quad (1)$$

where Y is a dependent variable (include quantitative variable, qualitative variable, dummy variable and categorical variable), W_j is the *j-th* independent variable (include single independent variable, interaction variable, generated variable, transformed variable, dummy variable), u is the random error of the model, Ω_0 is a constant of the model and Ω_j is the *j-th* parameter of the model for j =1, 2,..., k. According to Jaccard [1] such model can be classified as hierarchically multiple regression models or Cristol [2] defined a hierarchically well formulated model is an unconstraint model where it has the largest possible family of curves.

The correlation coefficient values between independent variables are usually highly correlated and the interaction variables (of any order) derived from a single variable are often highly correlated with one or more independent variables making up the interaction. Detail discussion can be found in Jaccard [1], Kutner *et al.* [3], [4] and Zainodin and Khuneswari [5]. One important problem in the application of multiple regression analysis involves the possible collinearity of the independent variables. This condition refers to situations in which two or more of the independent variables are highly correlated with each other. In such situations, there exist collinear of such variables on the dependent variable. When collinearity exists, the values of the regression coefficients for the correlated variables may fluctuate drastically, depending on which independent variables are included in the model.

2. Concept and Methodology

In dealing with the VIF issue, Bowerman and O'Connell [6] had defined as when the model expressed an independent variable as a function of the remaining independent variables. Dan & Vallant [7] stated also that correlation among the independent variables can lead to coefficient estimation with large variances. Strong multicollinearity between independent variables can cause the t statistic to be small (and the related probability-value to be large). This would give the impression that independent variable is not important (even if it really is). One method of measuring collinearity is determining the VIF value for each independent variable in a model.

$$VIF(W_j) = 1/(1-R^2(W_j))$$

where

W_j is the independent variable used as dependent variable

$R^2(W_j)$ is the coefficient of determination and $(1 - R^2(W_j))$ is the tolerance.

$R^2(W_j)$ is the coefficient of multiple determination for regression model, using independent variables W_j as the dependent variable and all other W_j variables as independent variables.

Coefficient of Multiple Regression, $R^2(W_j)$ =SSR/SST =1 –SSE/SST,

where

SSE: error sum of squares

SSR: regression sum of squares

SST: total sum of squares.

In identifying $VIF > 5$: there are 3 possible cases

There is NONE with $VIF > 5$.

Proceed to the next stage.

There is ONE variable with $VIF > 5$.

Remove that variable.

RERUN the reduced model.

There are two or more variables with $VIF > 5$.

Remove a variable with the highest VIF and

RERUN the reduced model.

Consider 3 distinct situations as follows:

Situation 1: DV: Y

IV: X_1, X_2

Find R^2 with each independent variable as dependent variable (DV) and the remaining as independent variables (IV). If there are only two independent variables, $R^2(X_1)$ is the coefficient of multiple determinations between X_1 and X_2. It is identical to $R^2(X_2)$, which is the coefficient of multiple determination between X_2 and X_1.

DV	IV	Coefficient of determination	
X_1	$f(X_2)$	$R^2(X_1)$	$VIF(X_1)$
X_2	$f(X_1)$	$R^2(X_2)$	$VIF(X_2)$

In this situation, $VIF(X_1) = VIF(X_2)$. If there exists VIF >5, a variable must be removed where a dependent variable with the lower standard error is preferred. Thus, the reduced model with all VIF values less than 5 is a selected model.

Situation 2: DV: Y

IV: X_1, X_2, X_3.

If there are 3 independent variables, then find R^2 with each independent variable as dependent variable and the remaining as independent variables.

DV	IV	Coefficient of determination	
X_1	$f(X_2, X_3)$	$R^2(X_1)$	$VIF(X_1)$
X_2	$f(X_1, X_3)$	$R^2(X_2)$	$VIF(X_2)$
X_3	$f(X_1, X_2)$	$R^2(X_3)$	$VIF(X_3)$

$R^2(X_1)$ is the coefficient of multiple determination of X_1 with X_2 and X_3;

$R^2(X_2)$ is the coefficient of multiple determination of X_2 with X_1 and X_3;

$R^2(X_3)$ is the coefficient of multiple determination of X_3 with X_1 and X_2.

The 3 VIF values are distinct. If there exist one VIF >5, the corresponding variables is removed and rerun the reduced model. Thus situation 1 is repeated.

Situation 3: DV: Y

IV: X_1, X_2, X_3, X_1*X_2, X_1*X_3, X_2*X_3

Find R^2 with each independent variable as dependent variable and the remaining as independent variable. So, there are 6 possible values:

DV	IV	Coefficient of determination	
X_1	$f(X_2, X_3, X_1*X_2, X_1*X_3, X_2*X_3)$	$R^2(X_1)$	$VIF(X_1)$
X_2	$f(X_1, X_3, X_1*X_2, X_1*X_3, X_2*X_3)$	$R^2(X_2)$	$VIF(X_2)$
X_3	$f(X_1, X_2, X_1*X_2, X_1*X_3, X_2*X_3)$	$R^2(X_3)$	$VIF(X_3)$
X_1*X_2	$f(X_1, X_2, X_3, X_1*X_3, X_2*X_3)$	$R^2(X_{12})$	$VIF(X_{12})$
X_1*X_3	$f(X_1, X_2, X_3, X_1*X_2, X_2*X_3)$	$R^2(X_{13})$	$VIF(X_{13})$
X_2*X_3	$f(X_1, X_2, X_3, X_1*X_2, X_1*X_3)$	$R^2(X_{23})$	$VIF(X_{23})$

Then, search for max {VIF(W_j)} where X_1*X_2 is the first-order interaction variable between X_1 and X_2. When VIF(X_{12}) is greater than 5, then remove the corresponding independent variable (X_{12}) from the model and the reduced model is f(X_1, X_2, X_3, X_1*X_3, X_2*X_3). Rerun or repeat similar procedure for the reduced model.

DV: Y

IV: X_1, X_2, X_3, X_1*X_3, X_2*X_3

Find R^2 with each independent variable as dependent variable and the remaining as independent variables. So, the result for rerun for the reduced model is as follows:

DV	IV	Coefficient of determination	
X_1	f(X_2, X_3, X_1*X_3, X_2*X_3)	$R^2(X_1)$	VIF(X_1)
X_2	f(X_1, X_3, X_1*X_3, X_2*X_3)	$R^2(X_2)$	VIF(X_2)
X_3	f(X_1, X_2, X_1*X_3, X_2*X_3)	$R^2(X_3)$	VIF(X_3)
X_1*X_3	f(X_1, X_2, X_3, X_2*X_3)	$R^2(X_{13})$	VIF(X_{13})
X_2*X_3	f(X_1, X_2, X_3, X_1*X_3)	$R^2(X_{23})$	VIF(X_{23})

Remove X_2 when corresponding VIF(X_2) >5 (and all other VIF values are less than 5). Then repeat similar procedure for the reduced model. Removal procedure is repeated until there are no more VIF values > 5. In case where the reduced model is FREE from collinearity then proceed to eliminate insignificant variable procedure until there is no more insignificant variable in the final model (model FREE from insignificant variable).

If a set of independent variables is uncorrelated, each VIF(W_j) is equal to 1. If the set is highly correlated then a VIF(W_j) might even exceed 10 (with W_j as the corresponding temporary dependent variable). Marquardt [8] had suggested that when VIF(W_j) was greater than 10, there were too many correlation between the variable W_j and the other independent variables. However, other statisticians had also suggested a more conservative criterion. Since then, Snee [9] had recommended using alternatives to least-squares regression if the maximum VIF(W_j) exceeded 5.

Since independent variables contain overlapping information, it is advisable to avoid interpreting the regression coefficient estimates separately because there is no way to accurately estimate the individual effects of the independent variables. One solution to the problem is to delete the variable with the largest VIF value (>5). The reduced model is often free of collinearity problems.

Details of the steps involved in Model Building Procedure are as follows:

Step 1: List all possible models (that is by compiling a list of all possible independent variables under consideration defined in equation (1)).

Step 2: Obtain selected models

For a regression model, determine VIF value for each independent variable involved.

Three possible results can occur:

1) None of the independent variables has VIF>5;

In this case proceed to Step 2.2

2) One of the independent variables has VIF >5;

In this case, remove that independent variables and

Proceed to step 2.2.

3) More than one independent variables have VIF >5;

Remove the independent variables with the highest independent variables value and

Repeat step 2.1.

(If a tie occurs, remove variable with the higher standard error)

Perform the elimination procedure (that is, eliminating one insignificant variable: p-value > 0.05).

Repeat elimination procedure until there is no more insignificant variable to be eliminated.

Go to step 2.1 for the next model otherwise go to step 3.

Obtain a BEST model using 8SC (or other criteria such as C_p statistics developed by Mallows).

Carry out the Goodness-of-fit (or any other validation tests), like randomness test and normality test.

Note: If an independent variable does not make a statistically significant contribution (p-value > α), the variable should not be included in the model. This process is termed as elimination.

Here are several possible phenomena in such removal or elimination procedure:

1) Deleting independent variables changes the regression coefficients
2) The reduction sum of squares associated with removal or eliminated independent variables vary, depending on which other independent variables are already present in the model.
3) The estimated standard deviations of the regression coefficients become large when the independent variable in the regression model is highly correlated with each other (or the corresponding variance is also large).
4) The estimated regression coefficients individually may not be statistically significant (p-value > α) even though a definite statistical relation exists between the dependent variable and the set of independent variables.

3. Numerical Illustration

Consider a data set with 3 single quantitative variables (X_1, X_3, X_6) and one quantitative dependent variable, Y. According to Zainodin and Khuneswari [5] there are N = 12 possible models involve in this situation (i.e. Models M1, M2, ..., M11, M12). For illustration, consider a model M12, with $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_6 X_6 + \beta_{13} X_{13} + \beta_{16} X_{16} + \beta_{36} X_{36} + \beta_{136} X_{136} + u$ where u is the random error variable. To complete the discussion, Zainodin *et al.* [10] suggested that there are 8 number of parameters involved in the model (where g =3, h =0 and v =2). With the suggested equation

$$\text{NP} = \sum_{j=1}^{3} {}^{3}C_j + \{(3+1)0+1\} = 3 + 3 + 1 + 1 = 8 \text{ parameters}$$

Considering all the 7 independent variables in the model M12, VIF value is obtained for each possible combination of independent variable as dependent variable and the remaining as independent variables. The corresponding possible combination values are listed in the Table 1 that follows:

Table 1. Summary for Model M12: f(X_1, X_3, X_6, X_{13}, X_{16}, X_{36}, X_{136})

DV	IV	R^2	VIF	
X_1	X_3, X_6, X_{13}, X_{16}, X_{36}, X_{136}	0.99982001	5555.9557	
X_3	X_1, X_6, X_{13}, X_{16}, X_{36}, X_{136}	0.99999803	507601.0439	
X_6	X_1, X_3, X_{13}, X_{16}, X_{36}, X_{136}	0.99999013	101339.9208	
X_{13}	X_1, X_3, X_6, X_{16}, X_{36}, X_{136}	0.99999776	446963.8245	
X_{16}	X_1, X_3, X_6, X_{13}, X_{36}, X_{136}	0.99998798	83201.9982	
X_{36}	X_1, X_3, X_6, X_{13}, X_{16}, X_{136}	0.99999925	1327622.3523	X_{36} is removed
X_{136}	X_1, X_3, X_6, X_13, X_{16}, X_{36}	0.99999920	1254269.8641	

The highest R^2 value (or VIF value >5) comes from X_{36} as dependent variable. Thus, variable X_{36} is removed and the reduced model, M12.1 is rerun. The resulting summary of the reduced model is shown in the following Table 2.

Table 2. Summary for Model M12.1: f(X_1, X_3, X_6, X_{13}, X_{16}, X_{136})

DV	IV	R^2	VIF	
X_1	X_3, X_6, X_{13}, X_{16}, X_{136}	0.996947	327.5331	
X_3	X_1, X_6, X_{13}, X_{16}, X_{136}	0.999949	19683.0352	X_3 is removed
X_6	X_1, X_3, X_{13}, X_{16}, X_{136}	0.999932	14784.6816	
X_{13}	X_1, X_3, X_6, X_{16}, X_{136}	0.999937	15873.2785	
X_{16}	X_1, X_3, X_6, X_{13}, X_{136}	0.999911	11186.4269	
X_{136}	X_1, X_3, X_6, X_{13}, X_{16}	0.994775	191.3886	

The highest R^2 value occurs when X_3 is a dependent variable. Thus, X_3 is removed. The reduced model, M12.2 is rerun and the resulting summary is shown in the following Table 3.

Table 3. Summary for Model M12.2: f(X_1, X_6, X_{13}, X_{16}, X_{136})

DV	IV	R^2	VIF	
X_1	X_6, X_{13}, X_{16}, X_{136}	0.996557	290.4218	
X_6	X_1, X_{13}, X_{16}, X_{136}	0.999826	5732.7809	X_6 is removed
X_{13}	X_1, X_6, X_{16}, X_{136}	0.986634	74.8193	
X_{16}	X_1, X_6, X_{13}, X_{136}	0.999751	4008.1418	
X_{136}	X_1, X_6, X_{13}, X_{16}	0.994355	177.1587	

The variable X_6 is removed and the reduced model is M12.3 and the result of rerun is as in Table 4.

Table 4. Summary for Model M12.3: f(X_1, X_{13}, X_{16}, X_{136})

DV	IV	R^2	VIF	
X_1	X_{13}, X_{16}, X_{136}	0.531586	2.1349	
X_{13}	X_1, X_{16}, X_{136}	0.972396	36.2270	
X_{16}	X_1, X_{13}, X_{136}	0.910609	11.1868	
X_{136}	X_1, X_{13}, X_{16}	0.983791	61.6936	X_{136} is removed

The variable X_{136} has the highest R^2 value (with VIF =61.6936) which led to the removal of X136. The reduced model, M12.4 is rerun and the result is in the following Table 5.

Table 5. Summary for Model M12.4: f(X_1, X_{13}, X_{16})

DV	IV	R^2	VIF	
X_1	X_{13}, X_{16}	0.5136	2.0561	
X_{13}	X_1, X_{16}	0.8610	7.1921	X_{13} is removed
X_{16}	X_1, X_{13}	0.8030	5.0752	

If the VIF >5 criterion is set, the model M12.4 should be the ultimate model (free from Multicollinearity) otherwise with VIF >5, variable X_{13} is removed and the reduced model produce the following summary Table 6.

Thus, using VIF> 5 criterion, the removing process stops at this stage. The final model, M12.5 is said to be free from Multicollinearity. The resulting reduced model, M12.5 is ready to go through the coefficient test. Coincidentally, the resulting coefficient values are listed in the following Table 7.

Table 6. Summary for Model M12.5: $f(X_1, X_{16})$

DV	IV	R^2	VIF
X_1	X_{16}	0.12947	1.1487
X_{16}	X_1	0.12947	1.1487

Table 7. Coefficient Test for Model M12.5

M12.5.0	Coefficients	Standard Error	t Stat	P-value
Intercept	9.04095	0.24871	36.351	9.20661E-57
X_1	-83.50315	2.09297	-39.897	2.68522E-60
X_{16}	2.23420	0.83593	2.673	0.008885824

As can be seen from Table 7, all the *p*-values that correspond to each variable is less than 5%, no elimination process takes place. Thus, the resulting model which is free from multicollinearity and non-contributing variable is, $Y = \beta_0 + \beta_1 X_1 + \beta_{16} X_{16} + u$ (or with $\hat{Y} = 9.04095 - 83.50315 * X_1 + 2.23420 * X_1 * X_6$). This final model results after undergoing two processes in sequence: first to remove the multicollinearity variable and second to eliminate non-contributing variable. The first process has been illustrated. However, the second process had not taken place since all the *p*-values < 5%. The final model obtained (model M12.5.0) is a selected model. Thus, readers are advised to refer to the detailed procedure of eliminating variables in Zainodin and Khuneswari [11] where procedures in eliminating non-contributing variables had been shown step by step until a reduced model was obtained with the corresponding *p*-values were less than $\alpha = 5\%$.

Another entity is checked: that is the number of parameters left after the process of discarding (removing and eliminating process) variables. The resulting model is M12.5.0 where $b = 5$ variables have been removed due to Multicollinearity and no variable is eliminated ($c = 0$) due to insignificant variable. Thus, it is symbolic with $(k+1) = NP - b - c = 8 - 5 - 0 = 3$. It is confirmed from Table 7 that there are 3 parameters left (the intercept, coefficient for X_1 and coefficient for X_{16}). Details in calculating $(k+1)$ number of parameters left after discarding the variable can be found in Zainodin *et al.* [12].

4. Discussion and Conclusion

The existence of serious Multicollinearity is given in the following diagnostics:

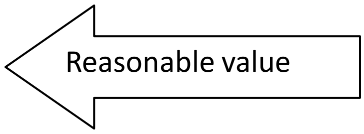
1) Large changes in estimated regression coefficients when a independent variables is added or deleted or when an observation is altered or deleted.
2) Non-significant results in individual tests on the regression coefficients for important independent variables.
3) Estimated regression coefficients with an algebraic sign that is the opposite of what is expected from theoretical considerations or prior experience.
4) Large coefficients of simple correlation between pairs of independent variables in the correlation matrix.
5) Wide confidence intervals for the regression coefficients representing important independent variables.

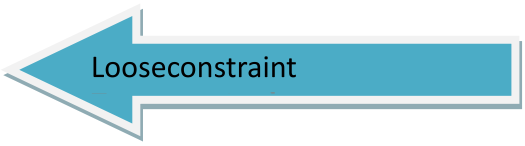
Kutner *et al.* (2008) "The largest VIF value among all independent variables is often used as an indicator of the severity of multicollinearity. A maximum VIF value in excess of 10 is frequently taken as an indication of Multicollinearity unduly influencing the least squares estimate."

Therefore, it was found that when VIF >5, many variables will be removed where $R^2 > 0.8$. However, when VIF >10, reasonable number of variables will be removed where $R^2 > 0.9$. Hence in practice VIF >10 is more reasonable to be employed in any statistical analysis because premature removal of variables can be avoided. This will then progress neatly into the process of eliminating variable due to insignificant variable before a final selected model is obtained.

Table 8. Various VIF Values for Different Coefficient of Determination Values (Meaning of Several VIF Values)

R^2	$1- R^2$	$1/(1- R^2)$	VIF
0.7	0.3	3.333333	3.33
0.71	0.29	3.448276	3.45
0.72	0.28	3.571429	3.57
0.73	0.27	3.703704	3.70
0.74	0.26	3.846154	3.85
0.75	0.25	4	4.00
0.76	0.24	4.166667	4.17
0.77	0.23	4.347826	4.35
0.78	0.22	4.545455	4.55
0.79	0.21	4.761905	4.76
0.8	0.2	5	5.00
0.81	0.19	5.263158	5.26
0.82	0.18	5.555556	5.56
0.83	0.17	5.882353	5.88
0.84	0.16	6.25	6.25
0.85	0.15	6.666667	6.67
0.86	0.14	7.142857	7.14
0.87	0.13	7.692308	7.69
0.88	0.12	8.333333	8.33
0.89	0.11	9.090909	9.09
0.9	0.1	10	10.00
0.91	0.09	11.11111	11.11
0.92	0.08	12.5	12.50
0.93	0.07	14.28571	14.29
0.94	0.06	16.66667	16.67
0.95	0.05	20	20.00
0.96	0.04	25	25.00
0.97	0.03	33.33333	33.33
0.98	0.02	50	50.00
0.99	0.01	100	100.00





References

[1] Jaccard, J. (2001). *Interaction Effect in Logistics Regression*. Thousand Oaks, California: SAGE Publication.

[2] Cristol, S. M. (2004). *Hierarchically Well Formulated Models In Epidemiology*. Doctoral Dissertation. Gorgia: Emory University.

[3] Kutner, M. C., Nachtsheim, C., Neter, J., & Li, L. (2005). *Applied Linear Statistical Models* (5th ed.). New York: McGraw-Hill/Irwin.

[4] Kutner, M. C., Nachtsheim, C., & Neter, J. (2008). *Applied Linear Regression Models* (4th ed.). New York:

McGraw Hill.

[5] Zainodin, H. J., & Khuneswari, G. (2010). The model-building approach in multiple binary logit model for coronory heart disease. *Malaysian Journal of Science and Mathematics* (MJMS), *4(1)*, 109-136.

[6] Bowerman, B. L., & O'Connell, R. T. (1990). *Linear Statistical Models: An Applied Approach* (2nd ed.). Boston: PWS-Kent Publishing Company.

[7] Dan, L. & Vallant, R. (2012). Variance inflation factors in the analysis of complex survey data. Survey Methodology, *38(1)*, 53-62.

[8] Marquardt, D. W. (1980). You should standardise the predictor variables in your regression models discussion of a critique of some ridge regression methods. In G. Smith & F. Campbell (Eds.), *Journal of the American Statistical Association*, *75*, 87-91.

[9] Snee, R. D. (1973). Some aspects of non-orthogonal data analysis, part 1: Developing prediction equations. *Journal Quality Technology*, *5*, 67-79.

[10] Zainodin, H. J., Noraini, A., & Yap, S. J. (2014). Number of parameters counting in a hierarchically multiple regression model*. Science International Journal, 2(2)*, 37-43.

[11] Zainodin, H. J., & Khuneswari, G. (2007). Justification of omitting independent variables in selecting best model. *Proceedings of National Seminar on Application of Science & Mathematics 2007* (pp. 343-356). Universiti Tun Hussein Onn Malaysia.

[12] Zainodin, H. J., Yap, S. J., & Noraini A. (2014). Body weight equation obtained from anthropometric measurements. *International Journal on Recent Researches in Science, Engineering & Technology*, *2(11)*.

H. J. Zainodin was born on February 8th, 1946 in Malaysia. He obtained his BSc in mathematics from Bandung Institute of Technology, Indonesia in 1974. Then, he got an MSc in industrial mathematics from Aston University United Kingdom in 1976. Finally, he obtained a PhD in statistics from Nottingham University, United Kingdom in 1984. His major field of study is statistics.

Currently, he's a professor in mathematics and statistics at Universiti Malaysia Sabah (UMS). He has the rare distinction of excelling in 3 crucial areas: his in-depth and up-to-date analysis of his own statistical field, the ability to write clearly, quickly and concisely, and the vocal ability to speak and present well. Also, he is one member of editorial board in international journals & is in the scientific organizing committee in international conferences. In UMS, he is the internal examiner, chairman of viva-voce session, programme coordinator & consultant and supervisor to many students. He is also the chief editor of Borneo Science Journal. In the past, he had been entrusted with many positions of responsibility His current research interests are multiple regression models, multiple binary logistics models, multinomial logistics regressions, parameter estimation and dendrochronology.

Prof. Zainodin is a member/fellow of many professional bodies: Japan Mathematical Society, Royal Statistical Society (UK), Malaysia Institute of Statistics, Majlis Profesor Negara (National Professorship Body) and Pakistan Institute of Statistics. He received many awards, such as ORS (Overseas Research Scholarship), Christopher Cox Memorial Fund, etc. He has many publications to his credit. He has been published more than 100 titles in various journals, 80 titles in seminar/conferences proceedings and 10 books titles. Above all, Professor Dr Zainodin is still actively involved in analyzing and editing papers for publication. This also explains his many awards of travel grants obtained.

G. Khuneswari was born on November 13th, 1984 in Malaysia. She obtained her BSc in mathematics with economics in 2007 and the MSc in statistics in 2009 from Universiti Malaysia Sabah (UMS). Currently, she is studying her PhD in Statistics at the University of Glasgow, United Kingdom. She is a tutor in the Faculty of Science, Technology and Human Development, University Tun Hussein Onn Malaysia (UTHM) and her PhD is funded by UTHM and Ministry of Education (Malaysia). Her major field of study is statistical modelling. Her current research interest and PhD work are related to model selection and model averaging. G. Khuneswari is a fellow of the Royal Statistical Society (United Kingdom). Even as a MSc student, she has written several joint papers & presented nationally and internationally. She has proven herself as a serious & hardworking researcher. Her youth, energy, enthusiasm and drive will blaze the trail for more exciting research collaboration and results in future.

A. Noraini was born on 3rd December, 1959 in Malaysia. Noraini Abdullah is a senior lecturer in the mathematics and economics programme in the Faculty of Science and Natural Resources, Universiti Malaysia Sabah. She obtained her bachelor of science degree from Universiti Kebangsaan Malaysia in 1999, the master of science in industrial mathematical modelling from Loughborough University of Technology, United Kingdom in 2003 and her PhD degree in mathematical economics from Universiti Malaysia Sabah in 2013. Currently, she is a member of PERSAMA (Persatuan Matematik Malaysia) and ISM (Institut Statistik Malaysia). On top of that, she is a journal reviewer, programme consultant, research consultant and supervisor to many students. At present, she is the vice-chairman of the Mathematical and Statistical Application Research Group (MASA) of Universiti Malaysia Sabah. Her current research interests are in multiple regression models, multiple binary logistics models, multinomial logistics regressions, parameter estimation and mathematics with economics.

Haider F. Abdul Amir obtained his master and PhD degrees in engineering physics (with concentration on semiconductor devices, instrumentations and nuclear), from the Faculty of Engineering, University of Gadjah Mada, Indonesia, Secondment of Osaka University Japan. He had his first degree in the Faculty of Sciences, University of Baghdad at Iraq. He had taught at many universities and Institutions, in Iraq and Indonesia. He is currently working as an associate professor in Universiti Malaysia Sabah (UMS), Malaysia since 2006.

His current research interests are nanotechnology in electronic, semiconductor materials, nuclear physics and green technology. He has a track record of fundamental research on these topics which is documented by numerous publications. He has been reviewer for numerous journals and indexed papers, such as Borneo Science, IEEE, ScienceDirect.

He is a life member of Iraqi Society of Mathematics and Physics, the Energy Conservation Society in Indonesia, MS Student Evaluation Committee in Yogyakarta Universities and Malaysian Physicists Organization. He is also a senior member of IACSIT (International Association of Computer Science and Information Technology (IACSIT).