Chapter Six- Selecting the Best Innovation Model by Using Multiple Regression

6.1 Introduction

In the previous chapter, the detailed results of FA were presented and discussed. As a result, fourteen factors underlying innovation capability were identified. The main goal of this chapter is to use Multivariate Analysis (e.g. Multiple Regression) in order to select the Best Innovation Model (BIM) based on which the key drivers of innovation can be determined. These key drivers provide a plausible answer for the research question which is "What are the key drivers of innovation within firms in Malaysia from a process and product innovation perspective?"

Multiple Regression (MR) is the main statistical analysis used in this chapter. It is also the most widely used method for conducting multivariate analysis particularly when more than three variables are involved (Bryman & Cramer, 2005, p. 300). MR is not just one technique but a family of techniques that can be used to explore the relationship between one continuous dependent variable (innovation performance) and a number of independent variables or predictors (innovation capability). MR is a more sophisticated extension of correlation and is used when the researcher wants to explore the predictive ability of a set of IVs on one continuous dependent measure. This makes it ideal for the investigation of more complex real-life, rather than laboratory-based, research questions.

In this study, all the independent variables introduced in the theoretical framework (Chapter Two-Section 2.5 Theoretical Framework), along with the factor combinations which are the results of FA in the previous chapter will be employed. Therefore, fourteen independent factors[1] and eight IVs will be used for MR analysis. The dependent variable (DV), firm's innovation performance, which is the key component and of primary interest in this study will be predicted by these fourteen factors and eight IVs.

As MR enabled the researcher to achieve the main objective of the study, i.e. to predict scores on DV from the scores of the fourteen IVs, it is the statistical analysis which best suits this study.

This chapter, therefore, is classified into Six sections including this introduction. Section 6.2 presents the theoretical assumptions of MR. Section 6.3 shows the results of assumption testing. Section 6.4 discusses the process of model selection and presents the Best Innovation Model. Section 6.5 provides the answer to the research question; and finally section 6.6 presents the conclusion.

6.2 Theoretical Assumptions of Multiple Regression

A number of assumptions underpin the use of MR. These assumptions are: Ratio of cases to independent variables (sample size), Multicollinearity and singularity, Outliers, Normality, Linearity, Homoscedascity and independence of residuals.

Sample size is the first assumption to consider. According to Coakes and Steed (2007, p. 136), the minimum requirement for the number of cases (sample size) is to have at least five times more cases than independent variables.

[1] Strategic Human Resource Management, Leadership and Strategy, Technology Management, External Networking, First-Mover Advantage, Work Place Leaning, Competitor Analysis, Customer Focus, Supplier Intelligence, Internal Networking, Business Reengineering, Research and Development, Employee Performance Appraisal, and Industry Experience

Multicollinearity which refers to the relationship among the independent variables is another issue in assumption testing. According to Pallant (2001) multicollinearity exists if the independent variables are highly correlated (r=.9 and above). Furthermore, SPSS also performs 'collinearity diagnostics' on independent variables as part of the MR procedure. This can also pick up on the problems with multicollinearity that may not be evident in the correlation matrix. The results are presented in table labeled 'Coefficients'. Two values are given: Tolerance and VIF. Tolerance is an indicator of how much of the variability of the specified independent variable is not explained by the other independent variables in the model and is calculated using the formula $1–R^2$ for each variable. If this value is very small (less than .10), it indicates that the multiple correlation with other variables is high, suggesting the possibility of multicollinearity. The other value given is the VIF (Variance inflation factor), which is just the inverse of the Tolerance value (1 divided by Tolerance). VIF values above 10 would be a concern here, indicating multicollinearity[2].

Outliers is another assumption to check for. MR is very sensitive to outliers. Outliers are very high or very low scores. However, this assumption has been discussed in detail in Chapter 4, Section Pre-Analysis Stage (see sub-section: 4.4.1.1 Reducing the Influence of Outliers).

Normality and *Linearity* refer to various aspects of the distribution of scores and the nature of the underlying relationship between the variables (Coakes & Steed, 2007, p. 143). One of the ways that these assumptions can be checked is by inspecting the residuals scatterplot and the Normal Probability Plot of the regression standardized residuals. In the Normal Probability Plot, the points will lie in a reasonably straight diagonal line from

[2] This part is based on Pallant, 2005, page 150.

bottom left to top right. This would suggest no major deviations from normality. In the Scatterplot of the standardized residuals, the residuals will be roughly rectangularly distributed, with most of the scores concentrated in the centre (along the 0 point). Therefore, a clear or systematic pattern to your residuals (e.g. curvilinear, or higher on one side than the other) is a sign of abnormal distribution. Deviations from a centralized rectangle suggest some violation of the assumptions[3].

Moreover, a number of statistics are available to test normality. These are mainly Kolmogorov-Smirnov statistic, with Lilliefors significance level and the Shapiro-Wilk statistic, Skewness, and Kurtosis. However, mentioned by several statisticians (see Hair et al., 1998; Coakes & Steed, 2007; Tabachnick & Fidell, 2007) normality can be assessed to some extent by obtaining skewness and kurtosis values. Therefore, the results of skewness and kurtosis will be taken as the major statistics determinant of the normal distribution. In fact, skewness value provides an indication of the *symmetry* of the distribution, while Kurtosis provides information about the *peakedness* of the distribution. Hair et al. (1998) asserts that the distribution is perfectly normal if both values for skewness and kurtosis are *zero* which is rather an uncommon occurrence in the social sciences (see Hair et al., 1998; Tabachnick & Fidell, 2007, p. 79). Therefore, a measure of skewness of ± 1 is usually regarded as a strong deviation from normality.

For *Linearity*, the residuals should have a straight-line relationship with predicted DV scores.

[3] This part is prepared based on Pallant, 2005, pages 150-15.

6.3 Multiple Regression Assumption Testing

Before testing the assumptions, it is very important to note that fourteen (14) independent factors (result of FA) along with eight independent variables (from the original theoretical framework) entered the regression analysis. However, only five IVs and two factors were chosen for the final model selected to be analyzed[4].

Sample size. The total number of independent factors and IVs amounts to seven; and seven times five equals thirty five (7*5=35) which falls well below 85, the sample size. Therefore, following Coakes and Steed (2007, p. 136), the sample size is adequate. However, "the issue at stake here is generalizability (Pallant, 2005, p.142)". Hence, this sample size does not have the power to generate strong generalizability.

Multicollinearity. Table 6.1(Correlations**)** provides the correlations between the variables. As it shows, none of the five IVs and two independent factors correlates highly with one another. Therefore, there is no sign of Multicollinearity among our variables. Moreover, according to Table 6.4 (Coefficients), the lowest Tolerance value is 0.345 and the highest VIF value is 2.897. Thus, none of the Tolerance values is less than 0.1 and none of the VIF values is above 10. Therefore, no single violation was observed.

[4] These independent variables and factors are:IV1= Leadership and Strategy, IV2=Employee Competence, IV3= Information and Organizational Intelligence, IV4= Culture and Climate, IV5= Customer and Market Orientation, F4= External Networking, F10= Internal Networking. For more detail, see the section on "Process of Model Selection" in the present chapter.

Table 6.1: Correlations

		DV	IV1	IV2	IV3	IV4	IV5	F4	F10
Pearson Correlation	DV	1.000	.208	.357	.450	.526	.542	.442	.199
	IV1		1.000	.460	.467	.467	.550	.426	.533
	IV2			1.000	.316	.335	.429	.318	.412
	IV3				1.000	.550	.595	.693	.687
	IV4					1.000	.637	.528	.494
	IV5						1.000	.445	.566
	F4							1.000	.540
	F10								1.000
Sig. (1-tailed)	DV	.	.028	.000	.000	.000	.000	.000	.034
	IV1		.	.000	.000	.000	.000	.000	.000
	IV2			.	.002	.001	.000	.002	.000
	IV3				.	.000	.000	.000	.000
	IV4					.	.000	.000	.000
	IV5						.	.000	.000
	F4							.	.000
	F10								.
	DV	85	85	85	85	85	85	85	85
	IV1		85	85	85	85	85	85	85
	IV2			85	85	85	85	85	85
	IV3				85	85	85	85	85
	IV4					85	85	85	85
	IV5						85	85	85
	F4							85	85
	F10								85

Note: DV= Innovation Performance, IV1= Leadership and Strategy, IV2=Employee Competence, IV3= Information and Organizational Intelligence, IV4=Culture and Climate, IV5= Market and Customer Orientation, F4= External Networking, F10= Internal Networking.

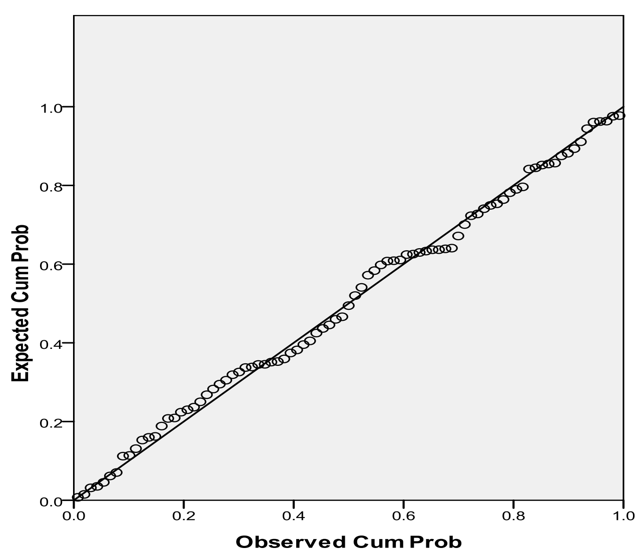
*Normality***.** Table 6.2 summarizes the results of statistics used to determine normality of the independent variables and factors. According to this table, it is evident that almost all variables are mildly skewed that is there is no perfect or ideal distribution. Table 6.2 shows that both values of skewness and kurtosis swing between -1 and +1, and all of the values are very close to zero; hence, normal distribution can be assumed and parametric tests can be used to analyze the data. This assumption is also supported by Coakes and Steed (2007, p. 58) in that "Many scales and measures used in the social sciences have scores that are skewed, either positively or negatively. This does not necessarily indicate a problem with the scale, but rather reflects the underlying nature of the construct [in this case: innovation capability] being measured".

Table 6.2: Statistics Results for Normality (N=85)

Factors	Mean	SD	Skewness	Kurtosis
IV1	36.83	5.14	-.231	-.213
IV2	22.40	2.26	-.087	-.260
IV3	93.71	16.43	-.435	.207
IV4	48.88	7.38	-.378	-.149
IV5	57.52	8.44	-.153	-.654
F4	18.30	3.28	-.113	-.136
F10	10.92	2.09	-.449	-.014
DV	23.63	5.36	.069	-.784

Note: IV1= Leadership and Strategy, IV2=Employee Competence, IV3= Information and Organizational Intelligence, IV4= Culture and Climate, IV5= Market and Customer Orientation, F4= External Networking, F10= Internal Networking; DV= Innovation Performance

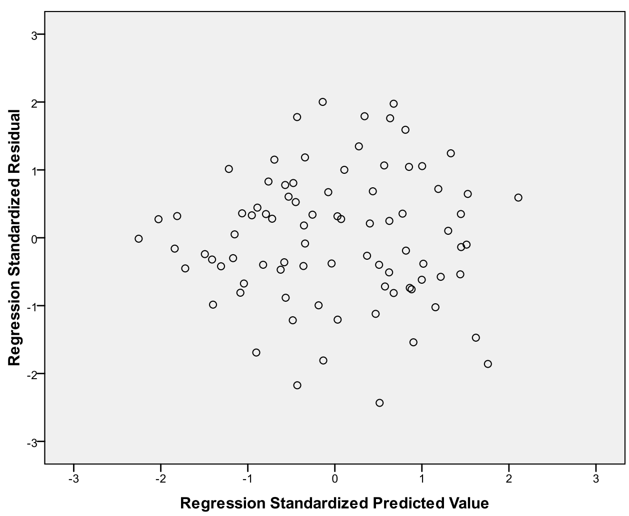
Figure 6.1: Normal P-P Plot of Regression Standard Residuals (DV: Total Innovation Performance)



Furthermore, Normal P-P Plot of each variable shows that the cases fall more or less in a straight line. This is in congruence with Coakes and Steed (2007) who claim that "in a normal probability plot, each observed value is paired with its expected value from the normal distribution; [therefore] if the sample is from a normal distribution, then the cases fall more or less in a straight line (p.34)." This also provides proof for meeting the linearity assumption.

In addition, as shown in Figure 6.2 (Scatterplot), the residuals are roughly rectangularly distributed with most of the scores concentrated in the center. Also, there is no sign of clear or systematic pattern to the residuals. Thus, normality assumption is assumed.

Figure 6.2: Scatterplot (DV: Total Innovation Performance)



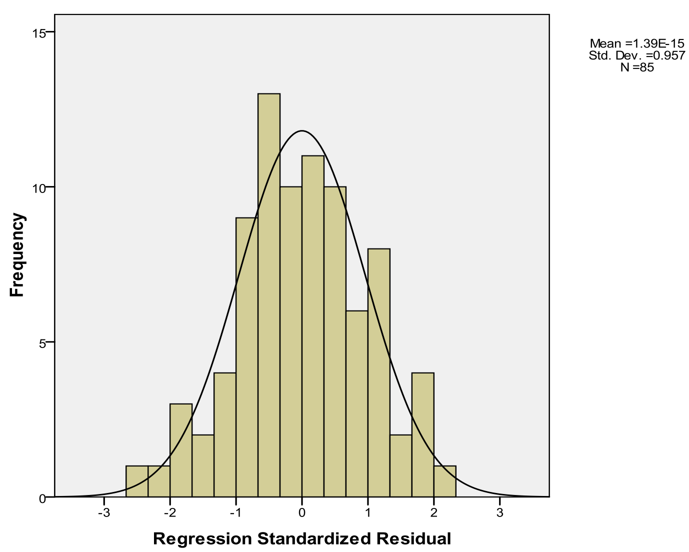
It is also very important for the DV to be normally distributed. Table 6.3 and Figure 6.3 show the results of the test of normality and the histogram for the DV of this study. According to Table 6.3, the Kolmogorov-Smirnov statistic is non-significant. In fact, a non-significant result (Sig. value of more than .05) indicates normality (Pallant, 2005, p. 57). Also, Figure 6.3 demonstrates an almost symmetrical, bell-shaped curve which has the greatest frequency of scores in the middle, with smaller frequencies towards the extremes[5]. Thus, Normality assumption for this variable is also very well met.

Table 6.3: Test of Normality (DV: Total Innovation Performance)

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
DV	.069	85	.200[*]	.975	85	.091

Notes: a. Lilliefors Significance Correction
*. This is a lower bound of the true significance.

Figure 6.3: Histogram (DV: Total Innovation Performance)



[5] See Gravetter and Wallnau, 2000, p. 52

6.4 Process of Model Selection

Eight independent variables and fourteen independent factors are used for multiple regression procedure. However, it was not proper to accept the first model generated as SPSS output for three of the IVs and twelve of the factors were not significant at 0.1(10 per cent) level of significance. Therefore, based on the chosen cut-off point for significance level, 0.1, it was necessary to eliminate the insignificant IVs and factors. This process, in this study, is called: Process of Elimination and it uses the largest value of insignificance as the starting point. Therefore, each time, the IV or factor which has the largest value of insignificance should be eliminated from the model, and the regression procedure should be repeated for the remaining IVs and factors. As a result, this procedure was repeated for sixteen (16) times and consequently three IVs and twelve factors were eliminated. The elimination process resulted in the Best Innovation Model as the outcome of this study (see Table 6.4 Model Summary).

Table 6.4: Model Summary[b]

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate	Change Statistics				
					R^2 Change	F Change	df1	df2	Sig. F Change
1	.710[a]	.504	.459	3.94472	.504	11.182	7	77	.000

Notes: a. Predictors: (Constant), IV1= Leadership and Strategy, IV2=Employee Competence, IV3= Information and Organizational Intelligence, IV4=Culture and Climate, IV5= Market and Customer Orientation, F4= External Networking, F10= Internal Networking.
b. Dependent Variable: Innovation Performance

Table 6.4 demonstrates the results of BIM. By examining the model and checking the value provided under the heading R^2, R-Square, it can be seen how much of the variance in DV (innovation performance) is explained by this model which includes five IVs and two factors (IV1= Leadership and Strategy, IV2=Employee Competence, IV3= Information and

Organizational Intelligence, IV4=Culture and Climate, IV5= Market and Customer Orientation, F4= External Networking, F10= Internal Networking) (see Table 6.6). In case of this study, the value is 0.504. This value is expressed as percentage, meaning that this model explains 50.8 per cent of the variance in innovation performance. This is a very acceptable result particularly if this percentage is compare with other results such as the one from the study conducted by Terziovski and Samson (2007) the model of which was able to explain only 42 per cent of the variance in innovation performance.

According to a few statisticians (as cited in Pallant, 2005, p. 153) when a small sample size is involved, the R^2 value in the sample tends to be rather optimistic overestimation of the true value in the population (see Tabachnick & Fidell, 2001, p. 147). Therefore, the adjusted R^2 statistic 'corrects' this value to provide a better estimate of the true population value. However, this is just a second opinion, yet if this value is considered, 0.459, still around 46 per cent of the variance in innovation performance is explained by the model. This result is also stronger than the one reported by Terziovski and Samson (2007).

These results are statistically significant and valid as shown in Table 6.5 (ANOVA). *F* ratio is a useful statistical test which is related to R^2. *F* ratio is produced by SPSS based on the multiple correlation (*R*). Multiple correlation expresses the correlation between the DV (innovation performance) and all of the IVS and factors collectively. According to Table 6.6, the multiple *R* for the multiple regression analysis is 0.710. The *F* ratio test allows us to test the null hypothesis that the multiple correlation is zero in the Malaysian population from which the sample of this study is drawn. Therefore, as the *F*= 11.182 (see Table 6.5) and the significance level is 0.000 which is really way below 0.0005, it can be concluded that it is extremely improbable that *R* in the population is zero. This means that the statistics provided here are relevant, valid.

Table 6.5: ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1217.964	7	173.995	11.182	.000[a]
	Residual	1198.185	77	15.561		
	Total	2416.148	84			

Notes: a. Predictors: (Constant), IV1= Leadership and Strategy, IV2=Employee Competence, IV3= Information and Organizational Intelligence, IV4=Culture and Climate, IV5= Market and Customer Orientation, F4= External Networking, F10= Internal Networking.
b. Dependent Variable: Innovation Performance

Table 6.6 clearly shows which variables included in the BIM contributed to the prediction of innovation performance. As Table 6.6 shows, all the retained IVs and factors are significant at 0.1.

Table 6.6: Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-5.829	4.657		-1.252	.215	-15.102	3.445					
	IV1	-.227	.110	-.218	-2.061	.043	-.446	-.008	.208	-.229	-.165	.576	1.735
	IV2	.539	.224	.228	2.408	.018	.093	.984	.357	.265	.193	.720	1.388
	IV3	.078	.045	.240	1.755	.083	-.011	.167	.450	.196	.141	.345	2.897
	IV4	.191	.082	.263	2.338	.022	.028	.353	.526	.257	.188	.510	1.959
	IV5	.250	.077	.394	3.253	.002	.097	.403	.542	.348	.261	.439	2.279
	F4	.328	.191	.201	1.723	.089	-.051	.708	.442	.193	.138	.474	2.111
	F10	-1.034	.308	-.404	-3.359	.001	-1.647	-.421	.199	-.357	-.270	.444	2.250

Note: a. DV= Innovation Performance; IV1= Leadership and Strategy, IV2=Employee Competence, IV3= Information and Organizational Intelligence, IV4=Culture and Climate, IV5= Market and Customer Orientation, F4= External Networking, F10= Internal Networking.

Based on Table 6.6, and by examining the Beta values under the column headed 'Standardized Coefficients' it is possible to compare the contribution of each IV and factor. The largest beta coefficient is 0.404[6] which is *Internal Networking* (F10). This means that Internal Networking makes the strongest unique contribution to explaining the innovation performance (DV) in the context of Malaysia, when the variance explained by all other IVS and factors in the BIM is controlled for. The second largest beta value is 0.394 which belongs to *Market and Customer Orientation* (IV5). These two Variables are significant at 0.001 and 0.002 respectively (sig. $p < 0.005$). This really testifies their unique contribution to the prediction of innovation performance.

The next two largest beta values which are 0.263 and 0.228 which belong to *Culture and Climate* (IV4) and *Employee Competence* (IV2). These beta values are significant at $p < 0.05$.

Leadership and Strategy (IV1) is another important and unique contributing variable as its beta value is 0.218 which is significant at 0.043, $p < 0.05$.

The last two Variables are *Information and Organizational Intelligence* (IV3) as well as *External Networking* (F4). These two have the beta values of 0.240 and 0.201 which are significant at 0.083 and 0.089 respectively, $p < 0.1$. As these two variables are significant at 0.1, it can be inferred that they are not making a very unique contribution to the prediction of innovation performance. However, as the level of significant for this analysis is set at 0.1, they can be still taken as significant variables but in comparison to other variables they contribute to a lesser extent. Table 6.7 presents the original items under each of these significant, contributing variables and factors.

[6] Negative signs should be ignored for the comparison purpose (see Pallant, 2005, p. 153)

Table 6.7: Items held by the Retained IVs and Factors

IV/ Factor	No. of Items	Items
IV1	9	The word 'innovation' appears in our mission statement, Our operations strategy is aligned with our innovation strategy, Senior managers actively encourage change, Senior managers implement a culture of innovation, There is a high degree of unity of purpose throughout our organization, We have eliminated barriers between departments, Senior managers show a sense of urgency relating to opportunities for innovation, We adopt an emergent (bottom up) strategy, Entrepreneurship is widely supported at middle management level.
IV2	6	Our human resource plan is clearly focused on the recruitment of creative people, Knowledge is freely shared in our organization, We have an organization-wide people development process, We have effective "top down"& "bottom up" communication processes, Employee satisfaction is measured regularly, Multi-tasking is actively used to build innovation capability.
IV3	26	Routine gathering of opinions from clients, Explicit tracking of competitor tactics, Forecasting sales, customer preferences, Market research studies, Trade magazines, government publications, Gathering of information from suppliers, Gathering of information through *Strategic Intelligence*, Licensing technologies, Patent disclosures, Publications, Informal networks with other organizations, Hired skilled employees, Reverse engineering, Independent R&D (in house or external), Networks with other organizations, Lead customers, Suppliers Consultants, Board members, Product, Services, Relative Cost Position, Operating Process, Technology, Quality Procedures, Customer Service.
IV4	12	Our organization has aligned employee behaviours with stated organizational values, Our strategic decisions are based on quantitative analysis of data, Our major operating decisions are detailed in formal written reports, We rely principally on experience-based intuition when making major operating and strategic decisions, Our major operating and strategic decisions are much more affected by industry experience, Our culture sees 'failure' as an opportunity to learn, Total Quality Management is embedded in our culture, The 'learning organization' concept is practiced in our organization, Uses hiring procedures that focus on who will best 'fit in' with the organization's culture, Promotes employees based on merit, Regularly conducts formal performance appraisal of employees, Rewards employees based on how well they perform their job, Rewards employees based on how well their work group or team performs.
IV5	15	We continuously obtain up-to-date market knowledge, All employees strive to enhance customer value creation, We place a strong emphasis on the marketing of tried products and/ or services, Our marketing and operations units work closely, We are normally the first organization to introduce new products/services in the market, Increase operating efficiencies, Develop new process innovations, Customize products /services to fit customers' needs, Develop customer loyalty, Respond quickly to customer needs, Produce a continuous stream of state-of-the-art products/services, Is 'first to market' with new products/services, Produces products/services at a cost level lower than that of our competitors, Develop 'best in industry' products/services, Responds to early market signals concerning areas of opportunity.
F4	1	Our major operating and strategic decisions are much more affected by industry experience.
F10	3	We have eliminated barriers between departments, Hired skilled employees, Routine gathering of opinions from clients.

6.5 Answering the Research Question

Based on the results which have been presented, examined, and discussed in the previous section it is now possible to answer the question of this study which is "*What are the key drivers of innovation within firms in Malaysia from a product and process innovation perspective?*"

Based on Table 6.4 (Model Summary) which represents the BIM and Table 6.6 (Coefficients), the nature of the relationship between DV (Innovation Performance) and IVs and factors (Leadership and Strategy, Employee Competence, Information and Organizational Intelligence, Culture and Climate, Market and Customer Orientation, External Networking, and Internal Networking) can be expressed as follows:

The Generic form of Linear Multiple Regression:

$Yi = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots\ldots\ldots\ldots. + \beta_j X_{ij} + \varepsilon' i$

where Y is the DV, $X_{i1}, \ldots\ldots\ldots\ldots, X_{ij}$ are the IVs and factors, β is the constant, $\beta_1 \ldots\ldots\ldots, \beta_j$ are the regression coefficients, notation i refers to the ith case in the n sample of observations, and ε' represents an error term.

The Multiple Regression Equation (E1) of This Study:

Innovation Performance (DV) = –5.829 – 0.404 **F10 +** 0.394 **IV5 +** 0.263 **IV4 +** 0.228 **IV2 –** 0.218 **IV1** + 0.240 **IV3** + 0.201 **F4 + ε**

Thus, as the equation (E1) shows, the key drivers of innovation within firms in Malaysia are: Internal Networking, Market and Customer Orientation, Culture and Climate, Employee Competence, Leadership and Strategy, Information and Organizational Intelligence, External Networking.

6.6 Conclusion

In this chapter, Multiple Regression was used to produce and select the Best Innovation Model (BIM) based on which the key drivers of innovation could be determined. According to BIM, Internal Networking, Market and Customer Orientation, Culture and Climate, Employee Competence, Leadership and Strategy, Information and Organizational Intelligence, and External Networking best predicted the dependent variable of this study which is Innovation Performance. All seven independent variables and factors collectively explain 50.8 per cent of the variance in innovation performance. This is a very acceptable result particularly if this percentage is compare with other similar studies such as the one conducted by Terziovski and Samson (2007). Their model was able to explain only 42 per cent of the variance in innovation performance.

However, the key drivers of innovations determined in this study to answer the research question are slightly different from those determined by Terziovski and Samson (2007). *Leadership and Strategy* is the only key driver of innovation which both studies share identically. In fact, Terziovski and Samson (2007) found innovation capability as a key driver of innovation performance without breaking it down into sub-constructs or variables.

As a result of this chapter, the MR equation (E1) was also built based on the BIM and Coefficients obtained as the result of regression analysis. The following Chapter which is the final or closing chapter of this thesis discusses the overall results of this study in detail and present the concluding remarks.