CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The main objective of this study is to find the effect of credit risk management on the financial performance of commercial banks listed in the Malaysian Stock Exchange. The purpose of this analysis is based on the previous research that linked the effect of better or poor credit risk management to the performance of commercial banks (Serwadda, 2018; Ekinci & Poyraz, 2019; Isanzu, 2017). In order to obtain the results, this chapter will explain the sample, variables, methodology, and models used to lead to the findings.

3.2 Variables

In this study, the dependent variable is the stock price of the commercial banks listed on the Malaysian Stock Exchange. In contrast, the dependent variables are the net charge off ratio (NCR), loan loss reserve (LRR), non-performing loan (NPL) ratio and loan loss provision ratio (LLP). The stock price exhibits the performance of the banks (Hauptmann, 2017; Zhou et al., 2021; Miralles-Quirós et al., 2019), while the four ratios capture the banks' credit risk management measurement (Joshua et al., 2020; Ephias Munangi et al., 2020; Taofeek Sola Afolabi et al., 2020). In addition, control variables regulate the variation of the stock price or market volatility and banks size. The control variables are banks' total asset (TA) and Chicago Board Options Exchange (CBOE) volatility index, also known as VIX. The variables are summarised in Table 3.1

3.2.1 Dependent Variable

As mentioned previously, the dependent variable of this study is the bank's stock price. The stock price is used as a parameter to measure the listed banks' performance (AlZou'bi et al., 2021; Köster & Pelster, 2017; Smriti & Das, 2018). Data collections process via annual reports and Fitch Connect has revealed ten commercial banks with complete data availability. The data is the quarterly stock price for nine years from 2012 until 2020. The banks are CIMB Group Holdings Berhad, Alliance Bank Malaysia Berhad, Affin Bank Berhad, BIMB Holdings Berhad, Hong Leong Bank Berhad, Hong Leong Financial Group Berhad, Malayan Banking Berhad, Public Bank Berhad, RHB Bank Berhad and AMMB Holdings Berhad.

3.2.2 Independent Variables

The first independent variable is the net charge-off ratio. A net charge-off ratio (NCR) is the dollar amount representing the difference between gross charge-offs and any subsequent recoveries of a delinquent debt (Handorf, 2018). Net charge-offs refer to the debt owed to a company that is unlikely to be recovered by that company. This "bad debt" is often written off and classified as gross charge-offs. In conclusion, Net charge-off is the final amount that the company cannot collect from its debtors. Previous researchers identified NCR as one of the bank's credit risk exposures (Munangi, 2020; Sornalatha & Aishwariya, 2020; Kemp, 2017). Hence, observing the NCR might contribute to proper credit risk management. In other words, a low NPL ratio indicates better credit risk management. The researcher, in its finding, stated that the net charge-off includes the amount of gross loans charged off in terms of bad debt expense less the recoveries made to a similar charge-off belonging to the previous period. Therefore, a reduction in net charge-off shows improving conditions of a bank's loan quality.

Furthermore, the relationship between net charge-off and net income before deducting/netting off loan loss provision shows that a lower ratio serves the bank's purpose well (Akram & Rahman, 2018). Global Portfolio Credit Risk Management, created by Siarka, P. (2021), addressed the credit risk for multi-product, where worldwide loan portfolios are acknowledged in this work. In the findings, Charge-offs and outstanding values were used for charge-off rate time series creation. The charge-off balance is the value of loans removed from the books and charged against loss reserves. Quarterly charge-off rates were calculated as a ratio of charge-offs and average portfolio outstanding for a given quarter. Hence, the charge-off rate is a risk measure similar in its nature to the default rate. However, it needs to be underlined that not every default event is automatically treated as a charge-off. Ultimately, both rates are strongly correlated and reflect the actual risk of the portfolio. This research concluded that the lower NCR is better for the bank and impacts better credit risk management.

The second independent variable is the loan loss reserve ratio (LLR). LLR is a balance sheet account representing a bank's best estimate of future loan losses (Ozili & Outa, 2017). LLR is described as the ratio used in the bank to illustrate the reserve that the company has in percentage terms to cover the estimated losses that they would have suffered due to defaulted loans. Due to its linkage with defaulted loans, the observation on LLR is one of the bank's credit risk management parameters (Oduro et al., 2019; Oketch et al., 2018; Onang'o, 2017). A Low LLR level may indicate the bank has a low reserve in covering future estimated loss due to the defaulted loans. Hence, low LLR may signify higher credit risk exposure. Ugwoke, M. J. A. (2017): Evaluation of Credit Management and Its Effects on Bank Profitability is the title of this paper. The purpose of this comparative study of First Bank PLC and Fidelity Bank PLC is to look into the credit management process in the banks as well as the impact of non-performing loans on the profitability of Nigerian banks. The findings reveal that loan loss reserves have a detrimental influence on earnings. As a result, the larger the credit risk, the higher the profit.

In another study, Al-Rdaydeh, Matar & Alghzwai (2017) analysing the effect of credit and liquidity risks on the profitability of conventional banks. In this research, the LLR is a reserve for any expected loan losses, which is the lower LLR, represent the better credit risk. LLR is a reserve for any predicted losses from loans. The lower this ratio suggests that the quality of the loan portfolio is very high and the less problematic the loans and vice versa. On the other hand, higher LLR leads to higher credit risk, which indicates poor quality of loans that increase the provisioning costs for banks and could reduce bank returns (Alta'ani & Dali, 2020).

The third independent variable is the non-performing loan ratio (NPL). NPL is a loan in which the borrower defaulted and hasn't made any principal or interest payments for some time (Koeppl & MacGee, 2017). In banking, commercial loans are considered non-performing if the borrower is 90 days or three months past due. NPL ratio is regarded as one of the credit risk measurements since the banks are exposed to losses due to defaulted consumers (Dahal & Bhaskar, 2020; Lian, 2020; Rafiq & Siddiqui, 2018). A low NPL ratio may imply that the banks have proper loan management hence reducing credit risk exposure. Akram & Rahman's (2018) findings also show that conventional banks should establish healthy loan portfolios while lowering NPL and impaired loan reserves.

Furthermore, Annor & Obeng (2017) underlined that performing loans are a charge on a bank's revenue statement. As a result, it shows the ratio of defaulted loans to loans that are truly performing. According to other empirical investigations, it is used to determine whether or not banks have effective credit risk management systems. In this instance, the restructured debts may be considered as fresh, solvent loans. Alternatively, banks might simply lend more money to grow overall loans and reduce the NPL ratio. This makes sense because the NPL ratio is derived by dividing the total loans by the amount of non-performing loans (Ranchman, Kadarusman, Anggriono & Setiadi, 2018). As a result, a rise in credit supply is accompanied by a decrease in the NPL ratio. Thus, indirectly it shows that the lower NPL effect to better credit risk management.

The fourth independent variable is the loan loss provision ratio (LLP). LLP is an income statement expense that is set aside to allow for uncollected loans and loan payments (Dahal & Bhaskar, 2020). Banks are required to account for potential loan defaults and costs to ensure they are presenting an accurate assessment of their overall financial health. Previous researchers also identified that LLP is a crucial variable to count for possible credit risk exposure (Ferrari et al., 2021; Bülbül et al., 2019; Perić et al., 2018). Due to that, proper observation of the LLP ratio induced better credit risk management. Banks who are practising optimal LLP may avoid higher credit risk exposure during unforeseen circumstances. The impact of credit risk management on the profitability of selected commercial banks listed on the Ghana stock exchange created by Annor & Obeng (2017) resulted in that loan loss provision ratio, LLP represents the amount of loans advance with respect to a total asset of the bank. Every bank takes a percentage of its deposits or liquid assets and uses them to grant or advance loans to clients. Therefore, it becomes imprudent and highly risky for a bank to advance loans equivalent to its total asset because any default or delay would be able to wreck the profitability and, by extension, the existence of the bank. Other's study, Ozili, P. K. (2018). Bank loan loss provisions, investor protection and themacroeconomy. International Journal of Emerging Markets. Larger banks in financially developed African countries have fewer loan loss provisions, while the increase in banklending leads to fewer bank provisions in countries with strong investor protection. Finally, higher bank lending is associated with higher bank provisions during an economic boom. The notion of each variable to the better credit management practice can be summarised as in Table 3.2.

There are another two independent variables that are used as the control variables. The first control variable is a total asset (TA). TA refers to the total assets owned by a bank (Sofyan, M., 2019). TA is usually recorded in the accounting records and appears in the balance sheet of the business. In addition, TA is often used to indicate a bank size (Schildbach, Schneider, & AG, 2017). Hence, for this study, TA will be used to capture the variation of the bank's size.

Another control variable for this study is the Chicago Board Options Exchange (CBOE) Volatility Index, also known as VIX. CBOE is an essential index in trading and investment activities because it quantifies market risk and investors' sentiments (Gurdgiev & O'Loughlin, 2020). In addition, CBOE is a financial benchmark designed to capture the expected market volatility (Smales, 2020). Previous researchers have used this index as a control variable to control market volatilities that would variate the stock price performance (Malkiel et al., 2018; Saini, 2019).

Table 3.1: Variables and the Formula

VARIABLE	FORMULA	REFERENCE
Dependent Variable:		
Bank's Performance	*Quarterly stock price*	Hauptmann, (2017), Zhou et al (2021), Miralles-Quirós et al. (2019)
Independent variables:		
Net Charge-off Ratio (NCR)	$\frac{loans\ charge\ off - subsequent\ recovery\ amount}{Average\ loan\ outstanding}$	Handorf, (2018),Balasubramnia et al. (2019), Haughwout, et al. (2019)
Loan Loss Reserve Ratio (LLR)	$\frac{loan\ loss\ reserve}{gross\ loan}$	Ozili, & Outa, (2017), Krüger,, Rösch, & Scheule (2018), Salo (2021)
Non-performing Loan Ratio (NPL)	$\frac{non-performing\ loan}{total\ loan}$	Koeppl & MacGee (2017), Acharya (2017), Sichinga (2020),
Loan Loss Provision (LLP)	$\frac{loan\ loss\ provision}{non-performing\ loan}$	Dahal & Bhaskar (2020), Karlsson & Strömqvist, (2019)

Control variables:

Total Asset (TA)	$non - current\ assets + current\ assets$	Sofyan (2019), Atsango.(2018), Sengupta & Vardhan (2017)
Chicago Board Options Exchange Volatility Index (CBOE)	$midpoint\ of\ realtime\ S\&P\ 500\ Index\ (SPX)\ option\ bid\ /ask\ quotes$	Gurdgiev & O'Loughlin (2020), Malkiel et al. (2018), Saini (2019).

Table 3.2: Variables Related to the Credit Risk Management.

DEPENDENT VARIABLE	NOTION TO BETTER CREDIT MANAGEMENT PRACTICE
Net Charge-off Ratio (NCR)	The lower, the better
Loan Loss Reserve Ratio (LLR)	Optimal or high
Non-performing Loan Ratio (NPL)	The lower, the better
Loan Loss Provision (LLP)	Optimal or high

3.3 Sample

This study focuses on commercial banks that are listed in the Malaysian Stock Exchange. The focus is on Malaysia since no previous research has analysed banks' possibility of credit risk management practice to influence their performance. This study also focuses only on commercial banks since credit creation is the primary activity of commercial banks. Due to that, credit risk management is an important element to the banks' sustainability. With that fact, better credit risk management may improve the listed banks' performance. As mentioned previously, all ten commercial banks listed were chosen as the study sample which are: CIMB Group Holdings Berhad, Alliance Bank Malaysia Berhad, Affin Bank Berhad, BIMB Holdings Berhad, Hong Leong Bank Berhad, Hong Leong Financial Group Berhad, Malayan Banking Berhad, Public Bank Berhad, RHB Bank Berhad and AMMB Holdings Berhad.

3.4 Methodology

Since this study has ten banks, with one dependent variable, four independent variables and two control variables, this study will use panel data analysis to produce the result. Beforehand, this study will go through two analyses: descriptive analysis and inferential analysis, to identify each variable's connectivity and characteristics. This study use Statistics and data (STATA) analysis software to generate the results for each analysis.

3.4.1 Descriptive Analysis

Firstly, descriptive analysis is used to analyse the data characteristics of the banks. Via this analysis, few statistical parameters can be identified, such as the mean, median, minimum, maximum, and standard deviation values of each bank can be compared. Descriptive analysis is essential to relate the findings of this study with the banks' specific attributes. Secondly, this model will test for high correlation independent variables and multicollinearity problem. For correlation problem, this study uses Pearson R statistical test. This test measures the strength between the different variables and their relationship. The Pearson correlation value is between -1 to 1 with 0 indicates no correlation, 1 is total positive correlation and -1 is total negative correlation. If the value of the Pearson coefficient is 0.5 and above, the variables are considered to have moderate to high correlation. The hypothesis statement for Pearson correlation test is as follows:

H_0 = There is no correlation between the independent variables

H_1 = There is correlation between the independent variables

Thus, if the Pearson test shows p-value less than 5%, then the variables are significantly correlated. The magnitude and strength of the correlation depends on the coefficient of the Pearson test.

The high correlation variables may cause multicollinearity problem to a model. This is because the independent variables should be independent otherwise, the change in one independent variable may shift the value of the other independent variable also. Hence, the model become less fit to estimate the causal relationship of independent variable with the dependent variable. In fact, it is more difficult to estimate the dependent variables if the independent variables tend to change in unison. Therefore, in order to test the possible multicollinearity problem, this study uses variance inflation factors (VIF) test. In this test, the existence of multicollinearity can be detected in the model if the VIF is more than 4. While VIF more than 10 indicates serious multicollinearity problem, VIF more than 4 may needs further investigation on the variables. These steps are crucial in order to avoid future problems to the model.

3.4.2 Panel Regression Model

As mentioned previously, this research uses the panel regression method to analyse the data. Panel data regression has multiple entities, each of which has repeated measurements at different periods. Panel data may have individual (group) effects, time effect, or both, which are analysed by fixed effect or random effect models (Balgati,2011). The benefit of the panel data regression method is that this method allows more variability and permits the analysis to explore more issues than cross-sectional or time-series data alone (Kennedy,2008).

This analysis process is expected to produce five models to answers objective 1 to objective 5. Model 1 is used to study the impact of credit risk management variables: NCR, LRR, NPL and LLP with two control variables: TA, CBOE on the bank's performance: stock price. On the other hand, models 2, 3, 4 and 5 are used to study the second objective of this study: to find the effect of individual credit risk management variable in Malaysia's listed commercial banks' performance.

The models are listed as follow:

Model 1

$$SP_{it} = \beta_0 + \beta_1 NCR_{it} + \beta_2 LLR_{it} + \beta_3 NPL_{it} + \beta_4 LLP_{it} + \beta_5 TA_{it} + \beta_6 CBOE_{it} + \mu_{it} \tag{3.1}$$

Where;

SP_{it} = stock price of the bank *i* at time *t*

NCR_{it} = net charge ratio of the bank *i* at the time *t*

LLR_{it} = loan loss reserve ratio of the bank *i* at the time *t*

NPL_{it} = non-performing loan ratio of the bank *i* at the time *t*

LLP_{it} = loan loss provision ratio of the bank *i* at the time *t*

TA_{it} = total loan of the bank *i* at the time *t*

$CBOE_{it}$ = CBOE of the bank *i* at the time *t*

$\beta_0, B_1, B_2, B_3, B_4, B_5, B_6$ = coefficient of the regression

μ_{it} =the error terms of bank *i* at the time *t*

Model 2

$$SP_{it} = \beta_0 + \beta_1 NCR_{it} + \beta_2 TA_{it} + B_3 CBOE_{it} + \mu_{it}$$

(3.2)

Where;

SP_{it} = stock price of the bank i at time t

β_0, B_1, B_2, B_3 = coefficient of the regression

NCR_{it} = net charge ratio of the bank i at the time t

TA_{it} = total loan of the bank i at the time t

$CBOE_{it}$ = CBOE of the bank i at the time t

μ_{it} =the error terms of bank i at the time t

Model 3

$$SP_{it} = \beta_0 + \beta_1 LLR_{it} + \beta_2 TA_{it} + B_3 CBOE_{it} + \mu_{it}$$

(3.3)

Where;

SP_{it} = stock price of the bank i at time t

β_0, B_1, B_2, B_3 = coefficient of the regression

LLR_{it} = loan loss reserve of the bank i at the time t

TA_{it} = total loan of the bank i at the time t

$CBOE_{it}$ = CBOE of the bank i at the time t

μ_{it} =the error terms of bank i at the time t

Model 4

$$SP_{it} = \beta_0 + \beta_1 NPL_{it} + \beta_2 TA_{it} + B_3 CBOE_{it} + \mu_{it} \tag{3.4}$$

Where;

SP_{it} = stock price of the bank i at time t

β_0, B_1, B_2, B_3 = coefficient of the regression

NPL_{it} = non-performing loan of the bank i at the time t

TA_{it} = total loan of the bank i at the time t

$CBOE_{it}$ = CBOE of the bank i at the time t

μ_{it} =the error terms of bank i at the time t

Model 5

$$SP_{it} = \beta_0 + \beta_1 LLP_{it} + \beta_2 TA_{it} + B_3 CBOE_{it} + \mu_{it} \tag{3.5}$$

Where;

SP_{it} = stock price of the bank i at time t

$\beta_0, B_1, B_2. B_3$ = coefficient of the regression

LLP_{it} = loan loss provision of the bank i at the time t

TA_{it} = total loan of the bank i at the time t

$CBOE_{it}$ = CBOE of the bank i at the time t

μ_{it} =the error terms of bank i at the time t

3.4.3 Panel Data Model Analysis

In panel data analysis, three models could be possibly matched with the dataset. The models are pooled OLS regression model, random effect model and fixed-effect model. The pooled OLS regression model has a constant coefficient referring to the intercept and slopes of the model (Hsiao, 2004). However, if this model is used to analyse panel data, all individual cross-sectional effects will be ignored. Compared to the pooled OLS model, fixed effect model is a model that considers the 'individual' of each cross-sectional unit. It allows the intercept of each bank to vary but still assume that the slope coefficients are constant across firms (Baltagi, 2005). Furthermore, the random effect model is a model that treats intercept among individuals differently from the fixed-effects model. Instead of treating intercept as fixed, random effect approach assumes the mean value of the intercept. The process contends that the banks included as sample have a common mean value of the intercept, and the individual differences in the intercept values of each bank are reflected in the error term (Gujarati & Porter, 2010). The general regression estimation of each model is as below:

1. Pooled OLS Regression Model

Pooled OLS regression model neglects the cross section and time series nature of data. This model does not distinguish between the various banks data, instead, combines the ten banks by pooling. Hence, this model denies the heterogeneity or individuality that may exist among banks. The general regression estimation for the pooled OLS model is similar as in Equation 3.1 which is

$$Y_{it} = a_0 + a_1 X_{it} + e_{it} \tag{3.6}$$

Where:

Y_{it} = dependent i variable t

a_0 = common intercept.

a_1 = the slope coefficient

X_{it} = the dependent variable i at time t

e_{it} = the error terms i at time t

The pooled model specifies constant coefficients which is the usual assumption for cross-sectional analysis.

2. Fixed Effect Model

As compared to pooled OLS model, the fixed effect model allows for heterogeneity or individuality among the banks. Each bank will have its own intercept value. The term fixed effect is due to the fact that although the intercept may differ across banks, but intercept does not vary over time, that is it is time invariant. The general regression estimation for the fixed-effect model has additional cross-sectional effect as follow:

$$Y_{it} = \sum_{i=1}^{N} D_{it} a_{0i} + a_1 X_{it} + e_{it} \tag{3.7}$$

Where;

Y_{it} = panel independent variable

$\sum_{i=1}^{N} \quad D_{it}a_{0i}$= intercept bank i at time t

a_1 = the slope coefficient

X_{it} = the dependent variable i at time t

u_{it} = the error terms

e_{it} = the error terms i at time t

Gujarati and Porter (2007) highlighted that the subscript i on the intercept suggests that the intercepts of the individual bank may be different. They further elaborate that the differences could be due to the unique features of each bank, such as managerial style, policies, and strategies. However, although the intercept across individual firms may differ, each individual's intercept does not vary over time (time-invariant), which signifies by a_{0i}

3. Random Effects Model

In the random effect model, the ten banks have a common mean value for the intercept. The general regression estimation for the random effect model is:

$$Y_{it} = a_0 + a_1X_{it} + v_i + e_{it} \tag{3.8}$$

Where;

Y_{it} = dependent i variable t

a_0 = common intercept.

a_1 = the slope coefficient

X_{it} = the dependent variable i at time t

v_i = random error term of bank i

e_{it} = the error terms i at time t

Previously, the fixed-effects model has its own (fixed) intercept a_{0i} value. In contrast, for the random effects model, the intercept a_0 represents the mean value of all the (cross-sectional) intercepts, and the error component v_i represents the (random) deviation of individual intercept from this mean value.

3.4.3.1 Panel Data Model Test

In order to test the most suitable model to match the panel dataset, few tests need to be done. Firstly, to choose between pooled OLS model with the random effect model, this study runs a test to identify the existence of the individual characteristics on the cross-sectional entities. Here, the hypothesis testing is as follows:

H_0 = Pooled OLS model is better than random effect model

H_1 = Random effect model is better than pooled OLS model

The study uses the Breusch Pagan LM test to verify whether the pooled OLS model or random effect model is the most suitable model for the panel dataset. If the p-value is less than 5%, then we can reject the null hypothesis, H_0. This means random effect model is better than the fixed effect model. Next, this study needs to test whether the fixed-effect model is more suitable than the random effect model or the other way around. In order to do that, this study uses the Hausman test to obtain the most appropriate model among the two (Rahim, Hassan, & Zakaria, 2012). Hausman test reveals the hypothesis testing as follows:

H_0 = No correlation between the independent variable and error term.

H_1 = There is correlation between independent variable and error term.

In here, if the p-value is less than 5%, then we can reject the null hypothesis, H_0 which means the fixed effect model is appropriate. After the most suitable table was identified, then, the significance of each credit risk management variable to affect the bank's performance will be tested. The hypothesis statements for the significance test are as follows

Hypothesis 1:

H_0: There is no significant effect between NCR and bank's performance.

H_1: There is significant effect between NCR and bank's performance.

Hypothesis 2:

H_0: There is no significant effect between NPL and bank's performance.

H_1: There is significant effect between NPL and bank's performance.

Hypothesis 3:

H_0: There is no significant effect between LLP and bank's performance.

H_1: There is significant effect between LLP and bank's performance.

Hypothesis 4:

H_0: There is no significant effect between LLR and bank's performance.

H_1: There is significant effect between LLR and bank's performance.

3.4.4 Diagnostic Check and Inferential Analysis

A diagnostic check is used to ensure that the estimated model is not biased (Hair et al.2007). Specific tests need to be done during the checking process, such as multicollinearity test, autocorrelation test, and heteroskedasticity test. The results of the tests are critical to ensure that the coefficient of the model is consistent and efficient. The inferential analysis will use the R^2result to determine the percentage of the variation in the dependent variable caused by the independent variable. Larger R^2 indicates better model regression fit (Morck, Yeung, and Yu, 2000). The formula for R^2 is as follows:

$$R^2 = 1 - \frac{RSS}{TSS}$$

Where:

- R^2 = coefficient of determination.
- RSS = sum of squares of residuals
- TSS = total sum of squares

The next test is the F-test, which identifies the significant collective impact of the independent variables on the dependent variable. The result will indicate the fitness of the data in the model (Park,2010). The formula for F test statistics as follows:

$$F = \frac{explained\ variance}{unexplained\ variance} = \frac{\frac{RSS_1 - RSS_2}{k-1}}{\frac{RSS_2}{n-k}}$$

Where:

- Explained variance also known as between-group variability.
- Unexplained variance also known as within-group variability.
- RSS = sum of squares of residuals
- n = total number of values
- k = independent variables

The hypothesis statement for F-test is:

H_0 = Independent variables are jointly statistically insignificant to bank's stock price

H_1 = Independent variables are jointly statistically significant to bank's stock price

If the F-test result produces less than a 5% significance level, then we can reject the null hypothesis, H_0. This means the variables are jointly significant to bank's stock price. Hence, the linear regression model better fits the data than a model containing no independent variables.

The other diagnostic tests are the multicollinearity test to test the multicollinearity problem among the variables. Multicollinearity is when two or more variables can be highly linearly related (Gujarati & Porter, 2010). Hence, a multicollinearity test observes the existence of a perfect linear relationship among regression variables. Serious multicollinearity problems can be detected using the centred Variance Inflation Factor (VIF). A VIF of one means no profound correlation among the variables, while a VIF value exceeding four is the sign of severe multicollinearity, and the models require correction. This analysis is essential to uphold the idea that each independent variable may uniquely influence the dependent variable. In other words, each variable is not highly correlated to identify the dependant variable. By doing this analysis, the suitability of the variables to be included in the model can be first determined.

This study also uses autocorrelation test as the model's diagnostic test. Autocorrelation is defined as a condition where residuals are related to each other, and it can be confirmed from the Profitability of Chi-Square of Obs*R-squared statistic. The purpose of the autocorrelation test is to investigate whether there is serial independence for the error term. It determines the correlation between two different time series data or space for cross-sectional data (Gujarati & Porter, 2010). Autocorrelation problems can be diagnosed using the Durbin-Watson test and Wooldridge test. The hypothesis testing for the autocorrelation test is as follows:

H_0 = There model has no autocorrelation.

H_1 = The model has autocorrelation

The null hypothesis that the model has no autocorrelation will be rejected if the p-value is less than 5%. Otherwise, if the p-value is higher than 5% then, there is no autocorrelation problem in the model.

The third test is heteroskedasticity test. Heteroskedasticity is refers to situations where the variance of the residuals is unequal over a range of measured values. When running a regression analysis, heteroskedasticity results in an unequal scatter of the residuals (also known as the error term) (Bollerslev, 2008). A heteroskedasticity test will identify whether the variance of the regression errors depends on the value of the independent variables. The detection of heteroskedasticity problems is an important assumption for a statistical test because they are sensitive to dissimilarities. Uneven variances in samples result in biased and skewed test results. In this study, the Wald test is used to detect the heteroskedasticity problem. The hypothesis testing for the heteroskedasticity test is as follows:

H_0 = The model is homoscedastic

H_1 = The model is heteroskedastic

In here, the null hypothesis, H_0 will be rejected if the p-value is less than 5%. The rejection indicates that the model suffers heteroskedasticity problem. In addition, this study also uses Pesaran's CD (2004) test to check the cross-sectional dependency of the models. The hypothesis statement for Pesaran's CD is as follows:

H_0 = The model has no cross-sectional dependence

H_1 = The model has cross-sectional dependence

If the results show p-value less than 5%, then the null hypothesis can be rejected. This indicates the dataset has cross-sectional dependency. According to Pesaran (2004), ignoring cross-sectional dependence in estimation can have serious consequences such as residual dependence which can resulted to estimator inefficiency and invalid test statistics.

Diagnostics tests elaborate few conditions to ensure the obtained model does not experience any problem such as dependency, multicollinearity, serial correlation and heteroskedasticity problems that may disrupt the efficiency of the result. If so, the model needs to be altered in order. In order to overcome these problems, this study will use Feasible Generalized Least Square (FGLS) method. This method suits well with the panel dataset of this study since the time series is larger than the cross-sectional data. FGLS is a 2 stages method that first run the OLS and take square of the residues, then the next stage will rewrite square of residuals as a product of sample variance of dependent variable with the exponential function of the independent variables. After that, the model will take logarithm to turn it into linear model. The equation is as follows:

Step 1: $y = \alpha + X\beta + \epsilon$

Step 2: $ln\, ln\, (\epsilon^2) = ln\, ln\, (\alpha^2) + \gamma + X\delta$

Step 3: $y = \alpha + X\beta + \sigma(e^{\{\gamma+X\delta\}})^{\frac{1}{2}}$

3.4.5 Flow Chart of the Methodology.

For a better understanding of the methodology, the summary of the methods' flow is illustrated as in Figure 3.3 below:

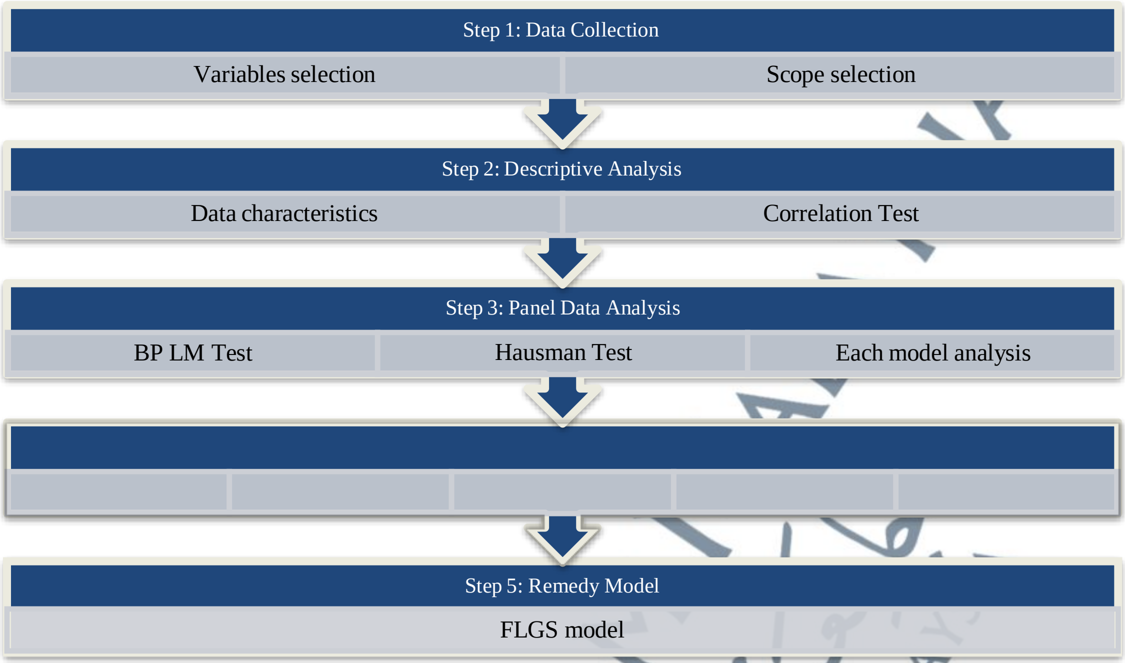

Figure 3.3: Methodology Flow