CHAPTER FIVE : METHODOLOGY

5.1 Introduction

Two methods were developed in order to reach a conclusion and reports findings by researchers, these includes qualitative and quantitative research. Qualitative research is empirical research where the data are not in the form of numbers (Punch, 1998, p. 4). Qualitative research is multimethod in focus, involving an interpretive, naturalistic approach to its subject matter. This means that qualitative researchers study things in their natural settings, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them. Denzin & Lincoln (1994, p. 2). Quantitative research gathers data in a numerical form which can be put into categories, or in rank order, or measured in units of measurement. This type of data can be used to construct graphs and tables of raw data. Quantitative researchers aim to establish general laws of behavior and phenomenon across different settings/contexts. Research is used to test a concept and ultimately support or reject it.

In statistical analysis, collection of data plays a significant part. The method of collecting information is divided into two different sections, such as primary data and secondary data. In this process, primary data is assembling of data or information for the first time, but, secondary data is the data that has been already gathered or collected by others. The most crucial variable of primary data is it's original and first-hand, whereas, secondary data is the interpretation and analysis of the primary data.

This chapter explains the data and method used in this study (i.e. quantitative and secondary data). The first section started by describing the data and its sources and concluded with methods of the data collection. This section is followed by the explanation of the method followed in deriving the LCR and NSFR series in order to conduct the empirical analysis. It also explains the data filtering technique and criteria for sample selection. The subsequent section explained the software (i.e., PLS-SEM 3.0) used to conduct the analysis and why the researcher chooses this software, although there is other software that can perform the same function. A brief explanation of the model specification was made in the next section, and it is followed by an explanation of the data before as well as after the crisis period. The section will also explain various equations used and their directions and it shows why all the variables in the study are formative rather than reflective. Later the chapter will thoroughly discuss the structural sub-model and mediation effects used in the study. The chapter closes with the summary of the multi-group analysis (MGA) methods used in this research to compare Islamic and conventional banks as well as before and after the crisis period.

5.1.1 Research Methodology

The initial raw data sample for this research includes annual financial data of over 100 (including both commercial, development, Agrcultural and Industry) Islamic and conventional Malaysian banks for the period from 2000 to 2016. The study period was chosen because of the availability of data as prior to 2000, most data are not reported (i.e., they are missing). In addition, the sample is chosen to enable the researcher to determine a population's characteristics by directly observing only a portion (or sample) of the population. In order to conduct econometric analysis, the data was subsequently narrowed down to achieve a more homogeneous sample of 42 banks from both conventional and Islamic banks operating in Malaysia (see Table 2.1 and 2.2 in chapter 2). The sample was narrowed down because of two reasons. Firstly, there is enough data quality for banks throughout the sample period. This comprises continuous observations of data series, which are essential for the investigation of dynamic effects. At the same time, some banks disclosure does not allow a consistent calculation of LCR and NSFR in all the sample periods. Nevertheless, since the study is considering only one country, the data are analogous with regards to banking culture and tradition, geography and economy, and regulatory and political environment. Thus, aspects such as system and country-specific control variables should be less needed in the analysis framework. Relatively, the limited sample may make it very difficult to draw definite conclusions for the whole Malaysia or even the South East Asia banking system. Nonetheless, it allows the researcher to analyze the research questions for significant and representative results towards the diverse angles of the Malaysian financial market.

Other applied filtering criteria (see table 4 of appendices) include bank size, the structure of the asset criteria (the financing activity of the bank must be substantial), and selected activity specializations as defined by the Fitch connect database. The second condition is specifically to exclude specialized financial institutions like finance or security companies and investment banks. As soon as the bank is chosen, additional criteria are applied to ensure LCR and NSFR financial data are developed based on the following guidelines: (1) certain and consistently available and (2) showcase satisfactory qualitative standards.

5.1.2 Variables and their Proxies

In this research, bank specific factors from the Fitch connect database is used. The database provides annual financial information for banks in over 200 countries around the world. Fitch connects databased coverage is comprehensive in many countries, whereby the data for banks cover for almost 90 percent of all bank's assets and Granularity of over 700 financial data points. Data on macroeconomic factors are acquired from various sources, including World Bank database, IMF and World Economic Outlook.

The most critical factors analyzed in this research is liquidity which was initially introduced in chapter one and two. These ratios have been defined quite recently (see Basel Committess on Banking Supervision, 2010) and the calculations require the; raw balance sheet, maturity structures of banks deposits and financings activities, and off-balance sheet data to reproduce both LCR and NSFR. Unfortunately, some data elements required for LCR and NSFR calculations were unavailable. An example is a breakdown of customer deposits into stable and less stable components and maturity profiles of bank deposits and financing activities. However, a good approximation of LCR and NSFR seems feasible. The following approach is inspired by Ötker-Robe & Pazarbasioglu (2010) who derived their NSFR time series in a comparable way and Yaakub, et al (2017) who derived their LCR and NSFR by following the guidelines issued by Islamic Financial Services Board, namely the 'Guidance Note on Quantitative Measures for Liquidity Risk Management in Institutions Offering Islamic Financial Services Excluding Islamic Insurance (Takāful) Institutions and Islamic Collective Investment Schemes' to measure LCR and NSFR of Islamic Banks in Malaysia (Islamic Financial Services Board (Islamic Financial Services Board, 2015).

The Liquidity Coverage Ratio (LCR): This ratio measures the high-quality liquid assets of the bank over the total net cash outflows. The main goal of Liquidity Coverage Ratio (LCR) is to ensure sufficient high-quality liquid assets for the banks to maintain activities up to 30 days under severe stress conditions to meet short term institution-specific and systemic withdrawal. It also protects the bank from run on its wholesale liabilities, including secured funding. In order to obtain stock of high-quality liquid assets held by a bank, the researcher multiplies the on-balance sheet accounting values by internationally set factor and consider the quality of each asset's category. Total net cash outflows are calculated by deducting total expected cash inflows from the total expected cash outflows for the next 30 days in the same period up to an aggregate of 75% of the total expected cash outflows in the specified stress scenario (outflows-min {inflows; 75 % of outflows}. Total expected cash outflows are obtained through the multiplication of the outstanding on- and off-balance sheet commitments by the rates at which they are expected to run off or drawn down over the next 30 days under stress, individual or systemic conditions. Table 1 and 2 in the appendix summarized the factors used to calculate LCR and NSFR guided by the available data in the bank's database and other researches (Yaakub, et al., 2017).

Net Stable Funding Ratio (NSFR) is a ratio of available to required stable funding (Basel Committee on Banking Supervision, 2010). The Available Stable Funding (ASF) is a weighted sum of funding sources according to their stable features. Similarly, the Required Stable Funding (RSF) is a weighted sum of uses of funding sources according to their liquidity". In order to measure the required amount of stable funding, specific RSF factors are applied to the assets in the on-balance sheet assets and liabilities and the off-balance sheet. The RSF factor denotes the percentage or amount of the exposure that should be supported by stable funding: higher liquid asset has lower RSF factor. Table 3 in the appendix provides a summary of ASF and RSF factors applied in this research. It represents a simplified version of Basel Committee on Banking Supervision (2010) rules and in line with the method applied in Ötker-Robe & Pazarbasioglu (2010), supported based on the financial data available in the database.

The bank specific factors and macroeconomic factors use for this research are categorized into eight. First is the capital adequacy ratios measured by equity to total assets (Mohd Amin, et al. 2017) and equity over net loans (Ghenimi , & Omri 2015).

Second factors are variables that show the quality of the assets held by the banks; these include; growth of gross loans and Loan Loss Provision (Dietrich, 2014). Third factors are the variables that show the management efficiency, these are, overhead over total assets (Dietrich, 2014), total expenses over total revenue (Said et, al 2008) and personnel expenses over net income. Fourth factors are the variables that measure the bank’s earnings; these are; non-interest share, interest earning ratio and operating income over average total assets (Dietrich, 2014). The fifth factors are the variables that show the credit risk which is the ratio of risk-weighted assets and off-balance-sheet activities divided by assets (Horvath et al. 2016) and the sixth factors are the variables that measure the concentration in which HHI is used to measure concentration risk as used by Horvath et al. (2016). As for the macroeconomic independent variables, the research employed factors such as inflation, GDP growth and GDP per Capital (Dietrich, 2014). These independent variables are to be measured against liquidity (LCR and NSFR) as dependent variables first, while in the second and third steps, these variables are to be measured against ROE, ROA, Net Interest Margin and Z score as dependent variables using PLS-SEM approach.

Table 5. 1: Measurement of Bank Specific Factors and Macroeconomic Factors

No.	Variables	Proxies/Items	Description (Items)	Factor
1	Capital Adequacy Ratio	1. EQNL and 2. EQTA	1. Equity over total assets (in %) and 2. Equity to net loans (in %)	Bank-Specific factor
2	Assets Quality	1. GGL, 2. RILIL and 3. CRISK	1. Annual growth of Gross loans (in %), 2. Reserves for Impaired Loans/ Impaired Loans and 3. Risk-weighted assets and off-balance-sheet activities divided by assets	Bank-Specific factor
3.	Management Efficiency	1. TE/TR, 2. PE/NI and 3. OH/TA	1.Total Expense over Total Revenue 2. Personnel Expenses over Net income and 3. Overhead over Total Assets	Bank-Specific factor
4	Earnings Quality	1.NII/GR. 2. OI/AS and 3. IER	1.Total non-interest income over total income (in %), 2. Operating income over average total assets. 3. Net Interest Income/Total Revenue.	Bank-Specific factor
5	Inflation	Inf.	Yearly average Inflation (in %)	Macro Factor
6	Gross Domestic Products	1. GDPC, and 2. GDPG	1. GDP over total population and 2. yearly real GDP growth (in %)	Macro factor
7	Concentration Risk	CON	Herfindahl-Hirschman Index (HHI)	Bank-specific factor
8	Liquidity	1. LCR and 2. NSFR	1. LCR is defined as the total of high-quality liquid assets divided by total cash outflows; the ratios are defined based on the funding weight shown in table 3 of the appendix and 2. NSFR is defined as a bank's available stable funding (ASF) divided by its required stable funding (RSF) (in %). ASF and RSF are calculated based on funding weights shown in Appendix 3	Bank-Specific factor
9	Profitability	1. ROA, 2. ROE and 3. NIM	1. Net profits over average[a] total assets (in %), 2. Net profits over average[a] total equity (in %) and 3. Net interest income divided by total average[a] assets (in %),	Profitability Factor
10	Insolvency Risk	Z score	ROA (E/A)/ Σroa	Insolvency Factor

a. Average of beginning and end of year reported value.
Note: Girardone (2004); for Islamic Banks, 'loans' are identified as financing activities, 'interest income' is called 'financing revenue' and 'interest expenses' are labeled as 'financing expenses.'

5.1.3 Sample Data

Out of the entire sample, incomplete records were excluded from the data analyzed, also non- financial institutions such as manufacturing companies, insurance companies and investment companies, which generally have different characteristics from the banks, and are not subject to any regulatory requirements (e.g., the minimum capital required) have also been excluded. The final sample consists of 13,828 (i.e. 561 banks year samples) banks year observations (from 2000-2016) which were divided into 7 subsamples. The subsamples for Islamic and conventional banks are 4,196 (222 banks year samples) and 9,632 (339 banks year samples) banks year observations, respectively. On the other and, the subsamples for Islamic banks before and after the crisis are 1,135 (49 banks year samples) and 3,061 (124 banks year samples) respectively, while subsample of conventional banks before and after the crisis are 4,780 (195 banks year samples) and 4,852 (199 banks year samples) respectively. The study has 10 constructs (Latent and Observable variables) for 21 indicators whose proxies have been indicated in table 5.1. Out of these 10 constructs, 8 are latent variables because they includes more than 1 items while 2 (i.e. inflation and insolvency risk) are consider as obserble variables because they have only 1 item measurement.

5.1.4 Data Analysis Software

A unique bank specific factors and macroeconomic factors of a bank's liquidity in the risk-return trade-off concept is often observed non-directly. The accounting proxies cannot be represented precisely for each of the factors that determine liquidity. This research employed the approach of PLS-SEM in the Smart PLS 3 software, which practically and conceptually made a combination of both multiple regressions and principal component analysis (PCA). As one of the main contribution of this study is the effect of mediation on the model, the PLS can be used to handle the model since the objective of the research is for the prediction and concept development in the liquidity, profitability and insolvency risks.

5.1.5 Methodology and Model Specification

As illustrated in the Conceptual Framework Model Figures 4.2 and 4.3 of chapter four, the research is analyzed in three steps. Step one, an investigation on how bank and macroeconomic factors (exogenous variables) affect the level of profitability and insolvency risks (endogenous variables) , which is called as “X on Y”. Step two, analyzed bank specific factors and macroeconomic factors affect on the level of liquidity (X on M). Step three analyzed the impact of liquidity on banks profitability and insolvency risks (M on Y). It should be noted that the whole mediating process was carried out simultaneously using PLS-SEM software. Finally the study analysed the mediating effects of liquidity between the relationship of bank specific factors and macroeconomic factors of profitability and insolvency risk. Based on this, the research broadly explains banks’ profitability to include scopes of traditional financial performance in addition to macro-prudential factors and insolvency risks measured through equity plus ROA divided by the standard deviation of the ROA (Vander Vennet, 1996; and De Haan & Poghosyan, 2012). Once a sample of banks is identified, the researcher analyzed descriptive statistics for any significant differences across the factors between the two groups[4]. A comprehensive explanation of the bank specific factors and macroeconomic factors considered in this study is included in table 5.1.

5.2 Partial Least Square-Structural Equation Modelling

In this section, the conceptual structure of the research is presented to serve as an outline of how and why the research was performed using Partial Least Squares (PLS, hereafter), which is a variance-based Structural Equation Modelling (SEM, hereafter) approach. An explanation of the characteristics of PLS-SEM using the

[4] Islamic and Conventional banks

formative measurement model, PLS path model assessment, sources and types of information, sampling procedure and techniques to collect and analyze data are given.

In recent years, partial least squares (PLS) path modeling has become increasingly attractive in different disciplines (Hair et al., 2012a, b; Hair et al., 2012a, Henseler et al., 2009; Lee et al., 2011; Ringle et al., 2012; Sosik et al., 2009 and Roemer E. (2016). Despite the increasing number of publications (Hair et al., 2014), relatively few researchers have used PLS in their studies so far. Some notable exceptions are, for example, Jacobs et al. (2011), Johnson et al. (2006) or Jones et al. (2002), and Ramli (2014). The small number of papers using PLS in their studies is surprising, since changes in technologies nowadays enable us to collect a vast amount of data. For example, online survey tools enable researchers to collect online data at low cost (e.g., Callegaro et al., 2014). As another example, consumers' online transaction data are continuously collected in e-commerce (Turban et al., 2015). The amount of data currently challenges many researchers and companies simultaneously creating a need for methods and tools to analyze such data (Columbus, 2015).

The need to analyze banks and companies quantitative data also holds true for phenomena in industrial management and finance that are more complex in nature and require specific research designs (Hassett & Paavilainen-Mäntymäki, 2013; Van de Ven, 1992 and Ramli, 2014). For example, (i) academician/ researchers are interested to know the impact of certain internal and external factors that drives the company's performance (Ramli 2014) (ii) industrial managers are interested in understanding and predicting the adoption of new integrated service solutions over time (Davies et al., 2006; and Ulaga & Loveland, 2014). Or, (iii), industrial managers strive after understanding and predicting the creation of value in business relationships at different stages of the relationship lifecycle (Eggert et al., 2006).

The analysis of unobservable, complex variables such as “adoption” or “relationship value” over time as well as related causal relationships requires adequate methods. In this case, researches turn toward structural equation modeling (SEM) techniques instead of regression analyses to incorporate unobservable variables (subsequently termed constructs, see Rigdon, 2012) that are indirectly measured by indicators (Hair et al., 2014; and Mooi & Sarstedt, 2011). Researchers have the option to choose between covariance-based SEM, represented by LISREL, and variance-based SEM, with PLS path modeling as the most prominent method (Henseler et al., 2009; Mooi & Sarstedt, 2011). The choice of the appropriate SEM procedure should be based on its methodological characteristics (Hair et al., 2014; Henseler et al., 2009).

In this thesis, it is argued that PLS path modeling is highly appropriate to analyze development and change in constructs, since it offers three favorable methodological characteristics. First, constructs often need to be predicted in evolutionary models (e.g., Shea & Howell, 2000). Especially PLS path modeling gives researchers and practitioners the possibility to predict constructs (Hair et al., 2011; Henseler et al., 2009). Second, model complexity quickly increases when development and change shall be analyzed in companies/banks quantitative data. This is due to the larger number of constructs that are measured at different points in time and the respective effects between those constructs (see also Johnson et al., 2006). PLS path modeling is highly suitable to deal with such complex models (e.g., Fornell, 1982; Fornell & Cha, 1994; Wold, 1985). Third, sample sizes can become quite small (see also Jones et al., 2002), for example, due to panel attrition (e.g., Frees, 2004; Laurie, 2008). This argument becomes even more severe, if the studies are conducted in areas of research in which the sample sizes are notoriously small, such as industrial market research (e.g., Slater & Narver, 2000). PLS path modeling is particularly appropriate in case of small sample sizes (Henseler et al., 2009) as well as the large sample size (Ramli. 2014). For these reasons, PLS has already been used in some papers for some research designs (e.g., Hennig-Thurau et al., 2006; Jacobs et al., 2011; Johnson et al., 2006; Jones et al., 2002; Shea & Howell, 2000; and Ramli 2014). However, the way how data are treated, how PLS models are built and how additional analyses enrich PLS path analyses varies significantly between those papers.

5.2.1 Benefits of using PLS – SEM

Structural equation modeling (SEM) includes a diverse set of mathematical models, computer algorithms, and statistical methods that fit networks of constructs to data (Haenlein & Kaplan, 2004). SEM includes confirmatory factor analysis, confirmatory composite analysis, path analysis, partial least squares (PLS) path modeling, and latent growth modeling (Kline, 2011). In a tutorial paper by Division of Statistics and Scientific Computation University of Texas at Austin in 2012; among the top reasons the researchers use SEM which deal with its own language is because of its ability to fairly use some stringent statistical assumptions. SEM has a number of attractive virtues which includes:

- Assumptions underlying the statistical analyses are clear and testable, giving the investigator full control and potentially furthering understanding of the analyses.
- Graphical interface software boosts creativity and facilitates rapid model debugging (a feature limited to selected SEM software packages).
- SEM programs provide overall tests of model fit and individual parameter estimate tests simultaneously.

- Regression coefficients, means, and variances may be compared simultaneously, even across multiple between-subjects' groups.
- Measurement and confirmatory factor analysis models can be used to purge errors, making estimated relationships among latent variables less contaminated by measurement error.
- Ability to fit non-standard models, including flexible handling the data, databases with autocorrelated error structures (time series analysis), and databases with non-normally distributed variables and incomplete data.
- This last feature of SEM is its most attractive quality. SEM provides a unifying framework under which numerous linear models may be fit using flexible, powerful software.

5.2.2 PLS-SEM vs Traditional Regression Method

Ramli (2014). examines capital structure determinants using a simultaneous causal model with interaction effects between manifest and latent variables over the period 1990 to 2010. In comparing PLS-SEM versus regression analysis, the following superiority of PLS SEM over traditional regressions were observed by the researcher. First, the standard errors of PLS-SEM coefficients are smaller than those from regression analysis. Thus, the PLS SEM results are preferable because of their smaller standard error. The reason for the superiority might be because regression analysis involves several individual regressions whereas PLS-SEM is a simultaneous approach. Iacobucci et al. (2007, p.146) state that "fitting components of models simultaneously is always statistically superior to doing so in a piece-meal fashion….", and they argue that a simultaneous approach produces more consistent estimates that are closer to the true values. Thus, the powerful simultaneous approach might have been able to reduce standard errors. Second, the significant t-values for testing the effects of mediation in PLS-SEMs appear to be consistent. These results demonstrate the power of PLS-SEM over regression analysis in determining the presence of mediation even in a simple model. The results from regression analysis show inconsistent significance estimates.

In sum, the choice between regression and PLS-SEM matters even with the simplest scenarios per item for each of the constructs. PLS-SEM is more beneficial for researchers because of its convenience to knowing all existing mediation effects. Because, one is more likely to spot whether there is mediation by looking at the model diagram rather than the regression approach that needs to separately set up four models. In addition, this study supports Iacobucci et al.’s (2007) argument that PLS-SEM estimates tend to be true and close to the known population structural characteristics and finally be more convincing and reliable with simultaneous statistical estimations.

5.2.3 Formative Measurement Construct

In PLS-SEM, there are two measurement construct called as “formative” and “refective”. In a formative construct, the indicators cause the construct, whereas in a more conventional latent variables, sometimes called reflective constructs, the indicators are caused by the latent variable. The term for the formative construct is a "cause indicator" or "composite indicator," which represent the same meaning. Formative constructis calculated through multiple regression analysis. For example, if the "Socio-Economic Status (SES)" is modeled as a formative indicator (cause), then SES can be defined by a combination of formative indicators such as income, home ownership, qualification of a professional certificate, and the prestige of occupation. If a person finishes his professional certificate in a university, SES will probably increase even if the salary, occupation and house ownership stay the same.

This differs from the reflective construct that a higher degree of SES does not imply that a person’s salary increased, gets a new house property, finishes his/her professional tertiary education and better job position all at the same time. One of these events itself may be sufficient. Thus, the concept of causal indicators can be positively correlated, negatively correlated or not be correlated. MacKenzie et al. (2005) claim: “Dropping a measure from a formative-indicator model may omit a unique part of the conceptual domain and change the meaning of the variables, because the construct is a composite of all the indicators. This means that the construct is a weighted linear combination of all its indicators measured. Removing the income indicator can eliminate an important item that contributes to SES. Thus, instead of assessing the error term at the indicator level, it would account for it on the construct itself. Besides, internal consistency and reliability evaluation are not essential if there is more than one formative indicator.

The mathematical representation of the equation is:

$$Y_1 = \gamma_1 y_1 + \gamma_2 y_2 \ldots\ldots\ldots\ldots + \gamma_n y_n + \varsigma_1 \qquad \text{(Eq. 2)}$$

where Yi denotes the LV, y1 denotes the indicators determining the LV rather than the inverse (i.e.,, y1, income; y2, homeownership; y_3, educational qualification; y_4, occupational prestige). ς1 denotes measurement errors for the Y_i (LV), and γ_{i1} denotes the coefficients that are expected to be the effect of y_1 on Y_1. The associated coefficients for this relationship in PLS-SEM are referred to as the *outer weight*. The notation assumes that Y_1 and y_s are deviation scores, COV $(y_i, \varsigma_1) = 0$ for all *i,* and E $(\varsigma_1) = 0$ (Bollen & Lennox, 1991). The formative equation is considered regression equations. Note that in equation 2, y is the explanatory variable and the LV (Y) is the dependent variable.

5.2.4 Reasons for Choosing Formative Construct in this Study

Formative indicators are an integral part of the construct, as such, it causes creation, and it can change the construct direction when any of it is removed or altered. Since the correlation between formative indicators is low, it is unnecessary to test the reliability of measurement items, and a high level of internal consistency is not required (Bollen & Lennox, 1991; Chin, 1998). However, the research uses the Variance Inflation Factor (VIF) to test the multicollinearity for all the variables under the formative measurement model.

Furthermore, according to Rodgers & Guiral, (2011), all corporate outcomes, including financial performance and customer satisfaction, are classified as formative indicators variables. Based on this, corporate performance (i.e., bank's liquidity and profitability) measurement has to comply with the formative procedures for measurement (Fornell, 1982). This study affirms that liquidity and profitability evaluation using formative measures is more appropriate that will narrow the gap between academia and practice and it gained credibility because it allows for the determination of the weight of sub-indicators and causal relationships. Accordingly, this research employed a formative indicator to assess the liquidity and performance of banks in Malaysia through mediation. It should be noted that, in this research both latent and observable variables are employed. If there are more than one item for a particular construct, then it will be latent variable (i.e. Bank profitability). But, if there is only one item for particular construct, then it will be observable variable (example, inflation).

5.2.5 The Structural Sub-Model

The structural sub-model between the path coefficients of the construct, X, M, and Y are the construct (LVs) that relate the exogenous variable (X) and the endogenous variable (Y). However, “M” can be exogenous to the endogenous “Y” and endogenous to exogenous X. The exogenous variable is used to describe the construct (LV) that can never be the dependent variable; exogenous variables are not involved with any structural path relationships pointing to them. The endogenous variable is generated within the model; the value of the variable is changed (determined) by one functional relationship in the model. In short, the construct is explained by other constructs via the structural model relationships. The structural sub-model mathematically is given in equation 3:

$$Y = \beta_0 + \beta X + \beta M + \varepsilon \qquad \text{(Eq. 3)}$$

where Y is the endogenous variable, β is the vector-matrix of regression coefficients to the vector of exogenous variables X and M, and ε is the residual for the structural equation model (inner model). For example, in this study, the path coefficient relationship "a" is between the bank specific factors and macroeconomic factors (X) and bank’s liquidity (M). The path coefficient relationship "b" is between bank liquidity (M) and profitability and insolvency risks (Y) and the path coefficient relationship "c" is between the bank specific factors and macroeconomic factors (X) and bank profitability and insolvency risks (Y). Bank liquidity (M) is assumed to be mediated by the direct relationship between the the bank specific factors and macroeconomic factors (X) and bank profitability and insolvency risks (Y). "Mediated" means the third variable (M), the bank liquidity, influences the changes in the relationships between X and Y.

5.3 Autocorrelation Test

Autocorrelation is a mathematical representation of the degree of similarity between a given time series and a lagged version of itself over successive time intervals. It is the same as calculating the correlation between two different time series, except autocorrelation uses the same time series twice: once in its original form and once lagged one or more time periods. Though as explained in section 5.2.1 above, the PLS SEM has the ability to handle autocorrelation issues, however, in order to further confirm that the data is accurate, prior to the running the analysis, the research conducted an autocorrelation to find out whether the data has suffered from autocorrelation. The Durbin Watson (DW) statistic is a test for autocorrelation in the residuals from a statistical regression analysis. The Durbin-Watson statistic will always have a value between 0 and 4. A value of 2.0 means that there is no autocorrelation detected in the sample. Values from 0 to less than 2 indicate positive autocorrelation and values from from 2 to 4 indicate negative autocorrelation. Acccording to the autocorrelation tests conducted, DW statistics shows a value of about 2 which means no autocorrelation is detected in the sample (see appendix B for autocorrelation tests at the bottom of the thesis).

5.4 Heteroskedasticity Tests

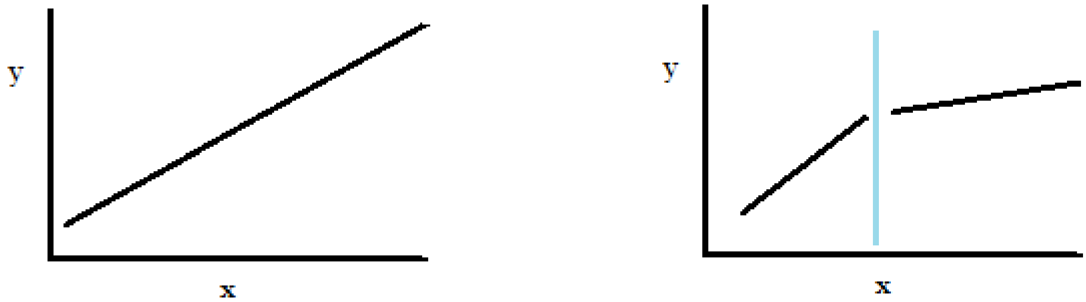
Heteroscedasticity means unequal scatter (non normal distribution). In regression analysis, the researchers talk about heteroscedasticity in the context of the residuals or error term. Specifically, heteroscedasticity is a systematic change in the spread of the residuals over the range of measured values. An important assumption assumed by the classical linear regression model is that the error term should be homogeneous in nature. Whenever that assumption is violated, then one can assume that heteroscedasticity has occurred in the data. In section 5.2.1 above, the researcher shows the SEM has the ability to handle any non normality in the data. However, in order to further confirm that the data does not suffer from heteroskedasticity issues, SPSS software is used through graph plot to examine whether the error terms are normally distributed (i.e. no heteroskedasticity is found in the sample). The plotted graph shows the error terms are homoscedastic (i.e. are normally distributed), thus, the reseach concluded that the sample did not suffered from any heteroskedasticity issues (see appendix C for heteroskedasticity tests at the bottom of the thesis).

5.5 Chow Test (Stability/Structural Break of the Data Test)

The Chow test tells if the regression coefficients are different for split data sets. Basically, it tests whether one regression line or two separate regression lines best fit a split set of data.

Sometimes data will have a break point or structural point (a period of significant or violent change), splitting a data set into two parts. For example:

- Donations given to an organization before and after a natural disaster.
- Stock market prices before and after Black Friday.
- House prices before and after a significant interest change.
- Asset prices before and after civil war.



The dataset on the left has a single regression line. The set on the right has a break point in the middle and two regression lines.

If the two parts can be represented by one single regression line, it can be said that the regression can be "pooled." Let's say the linear regression analysis of two parts of a data set (shown on the right above) resulted in the following two linear regression equations:

- First part of the data: $y_t = X_1*b_1 + \mu_1$
- Second part of the data: $y_t = X_2*b_2 + \mu_2$

The Chow test would tell if the coefficients $b_1 = b_2$ and $\mu 1 = \mu_2$. If they *are* equal, the data set can be represented with a single regression line.

Procedure for Running the Test

The null hypothesis for the test is that there is no break point (i.e. that the data set can be represented with a single regression line). The following procedure is followed in order to run the Chow test successfully:

1. Run a regression for the entire data set (the "pooled regression"). Collect the error Sum of Squares data.
2. Run separate regressions on each half of the data set. Collect the Error Sum of Squares data for the two regressions.
3. Calculate the Chow F statistic using the SSE from each subsample. The formula

$$CHOW = \frac{(RSS_p - (RSS_1 + RSS_2))/k}{(RSS_1 + RSS_2)/(N_1 + N_2 - 2k)},$$

   is:
   where:
   - RSS_p = pooled (combined) regression line.
   - RSS_1 = regression line before break.
   - RSS_2 = regression line after break.
4. Find the F-critical value from the F-table.
5. Reject the null hypothesis if the calculated F-value falls into the rejection region (i.e. if the calculated F-value is greater than the F-critical value).

5.5.1 Rationale for Pooling and Separating the Data in this Study

In order to confirm whether there is a stability or structural break in the data or not, Chow (Stability/Structural break) test for 17 years' time series data was conducted and the results shows there is no structural breaks in the data as F test is statistically significant (i.e. F test =7.65, and p. value = 0.0058). This suggests pooling the data for 17 years (the procedure followed to conduct the F tests is explained above and the Stata software was used to run the test). However, since according to the study by Sulaiman & Abdul Basit (2017), the 2008 financial crisis significantly impacts earnings per share, share price and deposit lending ratio of Malaysian banks. However, their study did not measure the other factors (i.e. liquidity and other profitability measures). This research closes this gap by separating the data between pre and post crisis (8 years pooled data) and use other factors not used by other studies. The result for the pre-crisis confirmed that between 2000-2008 there was no structural breaks for the data (suggesting the pooled data in these 8 years is okay), the F test and p value are 16.32 and 0.0001 respectively. This means we rejected he null hypothesis in favour of alternate hypothesis. On the other hand, the post crisis pooled data (i.e. 2009-2016) also shows no sign of any structural breaks in the data as F test and p value are 3.72 and 0.0546 respectively. In summary, the F test finding shows this study can either pooled the sample for 17 years and/or we can separate between pre and post crisis and pooled the data in each category.

5.5.2 Structural Break Tests for Pooling 17 year and 8 year- Period

In order to address the above issues, a Chow test was conducted. The Chow test tells if the regression coefficients are different for split data sets. Basically, it tests whether one regression line or two separate regression lines best fit a split set of data. Sometimes the data will have a break point or structural point (a period of significant or violent change), in this research, since other researches shows 2008 financial crisis negatively affected Malaysian banking sector (Sulaiman & Andulbasit 2017), splitting data set into two parts (i.e. before and after the crisis period) is feasible. After splitting the data for both 17 and 8 years, the F test (i.e. 7.65, 3.72 and 16.32) shows that there is no structural break in the data. Accordingly, this justify either pooling or separating the data is feasible in all the cases.

5.5.3 Separating the Data Between Islamic and Conventional Banks

Unlike the PLS-MGA findings for pre and post crisis period, the PLS-MGA findings between Islamic and conventional banks showed some variables has significant difference between Islamic and Conventional banks, based on that, the study separated the data in order to know the overall effect of the variables in each banking sector. Fortunately, after separating the data, the results of the Islamic banks significantly differ from that of conventional banks and accordingly, the recommendations made in the thesis for each banking category differs. Furthermore, the Chow test which is the tests developed in order to know whether pooling the data is appropriate or also separating the data will gives more information is conducted using the stata software. The results indicated that separating the data will gives more information as F test (Chow test shows an insignificant value (i.e. F test =1.47, P-value =0.2266). This means we failed to reject the null hypothesis. So, there is a structural break when the data is pooled. This also justified the reasons why the research further done an analysis for Islamic and Conventional banks separately.

5.6 Mediation Effect

Testing the causal and indirect effects such as mediation has been popular in marketing, social science, psychology, sociology, and education. However, those effects are notable exceptions in the finance field because only recently it became increasingly vital to examine the direct and indirect/mediation effects in determining the relationships (see De Jong et al., 2008; Ramadhan et al., 2012). As stated earlier, mediation is a set of causal hypotheses. The predictor (X) may influence an outcome variable (Y) through a mediating variable (M). In short, the inclusion of a third variable clarifies the effect of the relationship of two variables, i.e., the predictor (X) and the outcome variable (Y). The mediation relationship exists when a third variable plays an essential role in governing the relationships between the other two variables. The commonly asked question is, "Does the X variable predicts or cause the Y variable?". However, this simple question could move beyond to slightly more complex issues such as "how" or "why" X causes Y. This is where the third variable (mediator) plays its role. A mediator variable explains the relationship between a predictor X and an outcome (Y). For example, a larger bank size enables the banks to generate higher returns on assets and revenue because of unlimited access to the money market and other short-term instruments (liquidity). Bank size can be a proxy for predictor X, the return can be a proxy for outcome Y and the funding market (liquidity) can be a proxy for mediator variable M.

5.6.1 Typology of Mediation

Understanding the typology of mediation and non-mediation as proposed by Zhao et al. (2010) and Rucker et al. (2011), cited by Ramli (2014), can be helpful in the discussion for the mediation effects.

"A typology of mediation is as follows:

i. *Complementary mediation: the indirect effect (path: a x b) and direct effect (path c) both significant and the signs are pointing in the same direction. For example, the three path coefficients a x b x c are significant, and multiplying the three coefficients results in a positive number.*

ii. *Competitive mediation: the indirect effect (path: a x b) and direct effect (path c) both are significant, and the sign is pointing in the opposite direction. For example, the three path coefficients a x b x c are significant and multiplying the three coefficients results in a negative number.*

iii. *Indirect-only mediation: the indirect effect (path: a x b) significant, but the direct effect (path c) is not significant.*

iv. *Direct-only non-mediation: the indirect effect (path: a x b) is not significant, and the direct effect (path c) is significant;*

v. *No-effect non-mediation: neither the indirect nor the direct effect is significant."*

Zhao et al. 2010 and Rucker et al. 2011, as cited by Ramli 2014.

5.6.2 Bootstrapping

Bootstrapping is the final procedure to obtain the significant value of the t-statistics for the path relationships in the structural sub-model. This procedure is conducted after the measurement model has been estimated through the PLS-SEM algorithm's iterative procedures (if the research variables use reflective constructs). However, even though this study uses formative variables, but it still used the VIF to measure the multicollinearity. Bootstrapping is claimed to be a powerful method to test hypotheses about causal and indirect/mediating variable effects (Bollen & Stine, 1990; Hayes, 2009; Mackinnon et al., 2004; Shrout & Bolger, 2002). Bootstrapping is a process that repeatedly resamples the distribution by treating the sample size as mimicking the original sampling process to obtain the standard error for hypothesis testing. This sampling is conducted many times with replacement to a total k, which is a large number, e.g., 1000, 2000 or 5000, to construct the new sample. Then, the estimation of the significance test is done from the bootstrap sample in PLS-SEM (Hair et al., 2017;). This significance test procedure is on the sampling distribution's standard deviation for each model parameter that acts as a proxy for the empirical standard error. Bootstrapping analysis enables a statistical test to determine if a particular coefficient is zero (the null hypothesis) as opposed to the coefficient not equal to zero (alternative hypothesis). Chin (1998) proposes the test statistics for PLS: $t_{emp} = \frac{w}{se(w)}$, where: t_{emp} is the empirical t-value, w is the original PLS estimate of a certain path coefficient, and $se(w)$ is the bootstrapping value of the standard error. The significance test estimates are claimed to perfectly suit the PLS-SEM technique (Hair et al., 2017).

5.6.3 Multi-Group Analysis

A multi-group analysis is an advance in PLS for group comparison. The multi-group analysis is used to compare path coefficients estimate between two or more groups of data. It is typically applied when researchers want to test for heterogeneity in sample data. Heterogeneity occurs when two or more group of samples reveal significant differences in the model, i.e., countries, culture and gender. In PLS-SEM, Hair et al. (2017) state that the categorical variables such as gender or country are called moderator variables. This means that it is assumed that there is a categorical variable (moderator) that influences the relationships in the PLS-SEM model. Note that according to Baron & Kenny (1986), the moderator variable (a qualitative variable such as sex, class, race, group or sector, or a quantitative variable such as the level of incentive) can affect the direction or strength of the independent- or predictor- and dependent – criterion relationship (Henseler & Chin, 2010)-. In PLS, it is impossible to test a group comparison on a global criterion. Nevertheless, it is possible to compare two group path coefficients at a time. Meaning that the effects of differences between groups can be examined. In this study, the different mechanisms concerning the impact on banks' profitability and insolvency risks and liquidity can be revealed by the models' comparative results of the path estimators across groups (i.e., banks). The population parameter β in the multi-group analysis (MGA) can be hypothesized differently for two sample populations: β (1) # β (2). In this study, the sample populations are from two different bank categories operated in Malaysia, i.e., Islamic and conventional banks, and before and after the 2008 financial crisis. This study examines whether the relationship between the bank specific factors and macroeconomic factors and profitability and insolvency risks in the two populations are different in their path coefficients. The PLS-MGA approach is as follows: first, the subsamples are analyzed separately, and the bootstrap outcomes are used to test the hypotheses for group differences. Secondly, the parameter estimates for the two subsamples, i.e., a path coefficients β and β, and the standard error for both path parameters, i.e., s.e. and s.e. determine the test statistics. The test-statistics should consider whether the standard errors between the two subgroups' path estimators are equal. If they are not, Chin (2000) states that the t-test is computed as follows:

$$\mathbf{t} = \frac{path\theta_{sample\,1} - path\theta_{sample\,2}}{\frac{(n^{(1)}-1)}{(n^{(1)})} .se\,(p^{(1)})^2 + \frac{(n^{(2)}-1)}{(n^{(2)})} .se\,(p^{(2)})^2}$$

Where $\theta^{(1)}$ is the path coefficient for group one (Islamic banks or pre crisis period), ($\theta^{(2)}$) is the path coefficient for group two (conventional banks or after the crisis period). se($p^{(1)}$) is the standard error coefficient for group one (Islamic banks or before the crisis period), and se($p^{(2)}$) is the standard error coefficient for group two (conventional banks or after the crisis period). $n^{(1)}$ and $n^{(2)}$ are the number of observations in group 1 and 2. This test assumes that the variances between the subsamples are different (which is called invariance), which is based on the parametric approach from the two subsamples after the data has been split. This approach uses the re-sampling procedure for the standard errors from the two samples of the structural path coefficients obtained from the bootstrapping process that is generally evaluated by Chin's (1998) model.

5.7 Summary of the Chapter

This chapter has outlined the research methodology and model used to test all the hypotheses and to satisfy the research objectives for the study. The chapter started with the introductory explanation of qualitative and quantitative methods of research. In addition, the chapter discussed the data and the methods followed to elicit the data, in this sense the chapter explain the period for the study (i.e. 2000-2016). The chapter also explained the details 10 LVs and 21 indicators used for the analysis for data. Accordingly, the chapter revealed that PLS SEM software is suitable for the data analysis of the LVs developed in this research, however, before such an analysis is performed, an econometrics issues (i.e. autocorreltion, heteroekedasticity and Chow test for structural breaks that are not available in the PLS SEM must be conducted using other softwares like stata or SPSS) . The chapter thoroughly explained PLS SEM used for quantitative studies citing relevant journals to back how the research and how the econometrics issues are solved by using PLS SEM. Though there are quate alots of research done recently on PLS SEM, however, only few papers were found to used PLS SEM for quantitative studies. This chapter explains the justifications and benefits for using PLS SEM and the superiority of PLS SEM over traditional regressions analysis.

In PLS SEM, the LVs are either reflective or formative LVs. The chapter revealed that all the LVs variables in this research are formative, this is based on the Rodgers (2010) paper. Since the structural sub-model revealed the study analyzed the mediation effects, the chapter explained the mediation effect and typology for mediation. The bootstrapping analysis procedure is explained in the second to the section of the chapter while the explanations on the Multi-Group Analysis is done the last section.