CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes a detailed explanation of the variables and a general overview of the research methodology to be adopted and adapted in the study. This chapter comprises of five sections. The next section represents variables definition followed by the data collection process in Section 3.3. Section 3.4 shows the general overview of research methodology and then followed by the motives behind the application of research methodology in Section 3.5. Section 3.6 shows the diagnostics test while the last section concludes the chapter.

3.2 Variables Definition

All the variables; dependent and independent, are selected based on previous evidence. This study has two dependent variables namely operational performance (ROA, ROE & NIM) and bank stability (Z-score). There is another proxy for the performance measure that is called Tobin Q (market-based measurement). The Tobin q was examined as an indicator of the firm's effectiveness from an investment perspective (Wolfe, & Sauaia, 2003). Tobin's q is defined as the sum of market value of common stock, preferred stock, book value of long-term debt and short-term debt divided by replacement costs of net plant and equipment and inventories (Wang, 2002). Tobin Q which is applied for the market-based measure (Al-Saidi, & Al-Shammari, 2013). However, based on the above discussions it is concluded that Tobin Q is market-based

measurement tools with the listed banks. The study could not use Tobin Q (market base measure) as the performance measurement since all the banks are not listed.

To explain the dependent variables, several explanatory variables such as focal variables; bank size and level of bank sizes i.e., large, medium, and small based on total assets, total deposits, and operating income, the financial intermediary role is measured by the cost to income (economies of scale) and loan to deposit (economies of scope). While non-financial intermediary roles such as non-interest expenses to non-interest income, the dummy for the mode of financing, i.e., cash or stock, control variables, namely bank-specific variables such as loan loss reserve to gross loan, liquidity, and capitalization macro-economic variables such as GDP and inflation are deployed.

This study uses focal variables such as bank size, level of bank sizes the financial intermediary role, the non-financial intermediary role and modes of financing. The levels of bank size are divided into three level according to its total assets, total deposits and operating income: namely large, medium and small. The financial intermediary is measured by the cost to income (economies of scale) and loan to deposits (economies of scope). Non-interest expense to non-interest income used to measure non-financial intermediary role. This study also included several controlled variables commonly found in the literature. These variables include bank-specific variables namely loan loss reserve to gross loan, liquidity, capitalization, and macro-economic variables such as GDP and inflation (Ibrahim & Rizvi, 2017; Salaber, Rao-Nicholson & Cao, 2016; Abbas et al., 2014; Kandil & Chowdhury, 2014; Gattoufi et al., 2014; Sufian & Habibullah, 2009; Al-Sharkas et al., 2008; Linder & Crane, 1993). Macro-economic variables like GDP and inflation are also used following Amene, & Alemu (2019) and Erel et al. (2017). Choi & Jeon (2011) confirmed previous literature results that found GDP being the most relevant factor in long-run relationships and determining the

aggregate mergers activity trend. Mohamed & Sidiropoulos (2010), however, applied GDP as a proxy of economic size.

3.2.1 Dependent Variables

This study has two dependent variables, i.e., operational performance (ROA, ROE, and NIM) and bank stability (Z-score).

3.2.1.1 Operational Performance

Accounting-based ratios, such as return on asset (ROA) & return on equity (ROE) and net interest margin (NIM) are selected as the proxies for operational performance i.e., to represent the operational performance of the bank. There are two reasons for choosing an accounting base ratio. Firstly, accounting ratios used are recorded in the financial statement and are easily tractable. The other being, the benefit or outcome resulting from M&As are said to be reflected best in ratios such as ROA, ROE and NIM (Hitt, Harrison, Ireland & Best, 1998).

3.2.1.1.1 Return on Asset (ROA)

ROA indicates how efficient the management of the banks in using their assets to maximize the outcomes at a given cost to generate income (Rao-Nicholson, Salaber, & Cao, 2016; Abdulazeez, Suleiman, & Yahaya, 2016; Rani, Yadav, & Jain, 2016; Rehan, Khan, & Khan, 2018); Kandil, & Chowdhury, 2014).

3.2.1.1.2 Return on Equity (ROE)

This ratio measures the return from shareholders' investment that reflects the effectiveness of the banks' management to produce additional shareholder earnings. The higher the ROE, the greater the total shareholders' equity growth and, consequently, shareholders benefit from higher stock prices (Abdulazeez, Suleiman, & Yahaya, 2016; Gattoufi et al., 2014; Rani, Yadav, & Jain, 2016; Rehan et al., 2018; Jallow, Masazing, & Basit, 2017).

3.2.1.1.3 Net Interest Margin (NIM)

Net interest margin is a measure of the difference between the interest income generated by banks and the amount of interest paid out to their depositors/lenders relative to the number of their assets. It is similar to the gross margin of non-financial companies (Rani, Yadav, & Jain, 2016; Khan, & Khan, M2018; Larasati, Agustina, Istanti, & Wijijayanti, 2018).

3.2.1.2 Bank Stability (Z-score)

It is expected that the bank should be financially sound after M&A. Along with the bank's operational performance, this study chose bank stability as the second dependent variable. Bank stability is proxied by the Z-score index. It combines profitability, leverage and return volatility in a single bank distance-to-risk measure Ibrahim & Rizvi, 2018; (Moutsianas, & Kosmidou, 2016**;** De Haan, & Poghosyan, 2012; (Čihák, & Hesse, 2010; Ibrahim & Rizvi, 2017; Anginer, Demirguc-Kunt, & Zhu, 2014; Nguyen et al., 2012; Wahid, & Dar, 2016; Diaconu, & Oanea, 2015; Khaddafi, Heikal, & Nandari, 2017; Li, X., Tripe, & Malone, 2017; Li et al., 2020; Karim, Al-Habshi, & Abduh, 2016). The formula for the Z-score is,

Z-Score = [{ROA + Equity to total assets)}/ standard deviation of total assets]

3.2.2 Independent Variables

In line with the previous evidence, several variables, i.e., focal variables, control variables (i.e., bank-specific and macroeconomic variables), and dummy variables are used as independent variables.

3.2.2.1 Focal Variables

Focal variables imply variables that are derived from theories that are applied in the thesis. A number of focal variables are used namely bank sizes, level of bank sizes, financial intermediary role (e.g., economies of scale, economies scope), non-financial intermediary role (e.g., non-interest cost to non-interest income) and the modes of financing (cash or stock). Bank size and level of bank sizes i.e., large, medium, and small based on the natural log of total assets, total deposits, and operating income are used and these are supported by resource dependency theory (Bhyrovabhotla, 2012; Smirnova, 2014; Morris, 2004; Nair, Trendowski & Judge, 2008). Meanwhile, the financial intermediary role is proxied by economies of scale (the cost to income), economies of scope (total loan to deposit). Non-financial intermediary role proxied by non-interest income to non-interest expenses are used based on efficiency theory and the theory of financial intermediation. A free cash flow hypothesis is applied to support modes of financing i.e., cash or stock.

3.2.2.2 Market Structure

3.2.2.2.1 The Herfindahl-Hirschman Index (HHI)

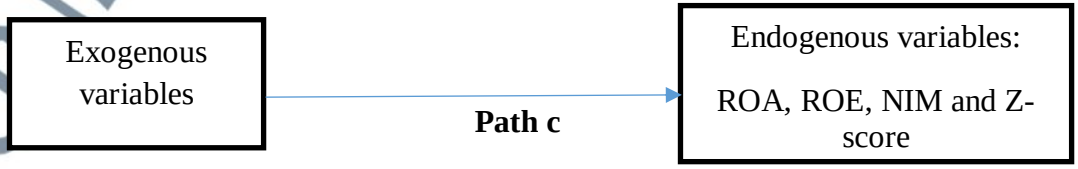
The Herfindahl-Hirschman Index (HHI) is measured by the square of a market share divided by the total number of shares. A score below 1500 shows weak concentration but strong competition, a score from 1500 to 2500 reveals moderate concentration and competition, and any score above 2500 until 10,000 points implies strong concentration but weak competition. HHI is used to check the relative impact of M&As on the market. A higher value of HHI indicates weak competition but a stronger level of concentration of the market and vice versa.

HHI is also used to monitor the potential impact of M&As on an industry. It is a quantitative measure that regulators can cite to veto any M&As transaction. Companies can include the index in their M&As proposal to indicate that the merger would not lead to a monopolistic market. The lower the HHI is, the more power consumers hold in that industry. Thus, prices are usually lower, and company margins are compressed.

3.2.2.2.2 Concentration Ratio (CR)

The concentration ratio (CR) is calculated by summing up the percentage of market share. If close to 1, the industry is highly concentrated with weak competition, however it is close to zero, the industry is less concentrated with strong competition.

Hair, Ringle, & Sarstedt (2013) and Preacher, & Hayes (2008) prove many rules representing types of mediation relationships. The following steps are given.



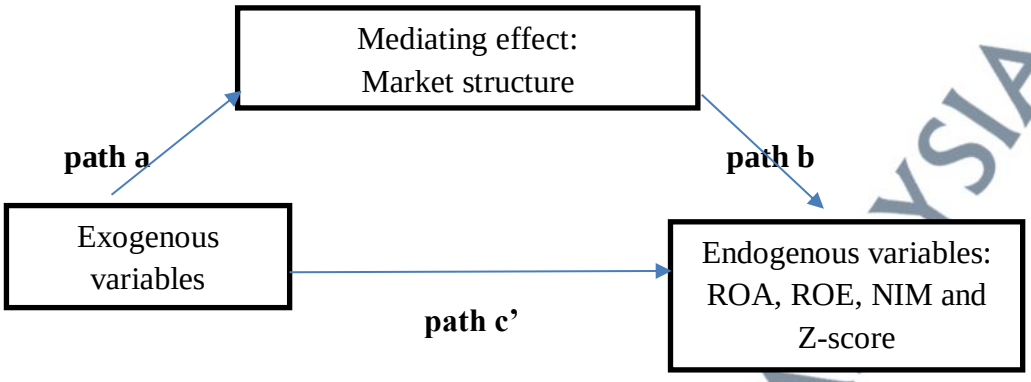


Figure 3.1: Flow of Structural Equation Modeling (SEM)

Step 1: Regress the endogenous variables by the exogenous variables (path c) ensuring that exogenous variables predict endogenous variables.

$$Y = \alpha + \beta_1 X_n + €, \text{ while } \beta_1 \text{ is significant.} \quad (3.1)$$

Where:

Y = endogenous variables,

α = constant term,

β_1 = coefficient of exogenous variables,

Xn = exogenous variables, and

€ = error term

Step 2: Regress the mediator variable by the exogenous variables to confirm that the exogenous variables are the significant predictor for the mediator.

$$M = \alpha + \beta_1 X_n + €, \text{ while } \beta_1 \text{ is significant} \quad (3.2)$$

Where:

M = mediator variable,

α = constant term,

β_1 = coefficient of exogenous variables,
Xn = exogenous variables, and
€ = error term

Step 3: Regress the endogenous variables by the mediator and exogenous variables to confirm that the mediator has a significant predictor of the endogenous variables and the previously significant coefficient of exogenous variables in step #1 is slightly reduced. Moreover, by entering a mediator variable into a model, the exogenous variables may no longer affect the endogenous variables (e.g., complete mediation i.e., indirect only mediation) or get a week (e.g., partial mediation i.e., complementary and competitive mediation), and a causal mediating model exists (Namazi, & Namazi, 2016).

$$Y = \alpha + \beta_1 X_n + \beta_2 M + €, \text{ while } \beta_2 \text{ is significant} \quad (3.3)$$

Where:
Y = dependent variable
M = mediator variable,
α = constant term,
β_1 = coefficient of exogenous variables,
Xn = exogenous variables, and
€ = error term

β_1 should be smaller in absolute value than the original mediation effect. Table 3.1 shows the typology of mediating effects.

Table 3.1: Typologies of Mediation Effect

No	Types of Mediation	Mediation Effects	Description
1	Complementary Mediation	Partial Mediation	The indirect effect (path a x b) and direct effect (path c) are both significant and the signs pointing the same direction

2	Competitive Mediation		The indirect effect (path a x b) and direct effect (path c) are both significant and the signs pointing the opposite direction
3	Indirect Only Mediation	Full Mediation	The indirect effect (path a x b) is significant, but the direct effect (path c) is not significant
4	Direct Only No-Mediation	No mediation	The indirect effect (path a x b) is not significant, but the direct effect (path c) is significant
5	No Effect; Non-Mediation		Neither the indirect or direct effect is significant

Sources: Ramli (2014); Zhao, Lynch Jr, & Chen (2010); Rucker, Preacher, Tormala, & Petty (2011); Hair et al. (2013)

3.2.2.3 Dummy Variables

In this study, several dummy variables are used. For example, the level of bank sizes, i.e., large, medium, and small and modes of financing i. e., cash or stock.

3.2.2.4 Control Variables

It is necessary to control for other factors that have been proposed in the literature as possible determinants of performance and stability (Hassan, & Bashir, 2003). Likely, control variables are the independent variables that may predict performance and stability but are not the variable of interest of the study (Musah., Abdulai, & Baffour, 2020). This study uses two types of control variables namely, bank-specific variables and macro-economic variables.

3.2.2.4.1 Bank Specific Variables

This study uses bank-specific variables as controlled variables. The credit risk ratio is represented by the loan loss reserve to the gross loan (CR), capitalization ratio (CAP), and liquidity ratio (LIDY). High capital and liquidity holding increases the banks creditworthiness and enables banks to expand their business. In addition, it also

enables banks to absorb negative shocks and reduces banks bankruptcy risk. As a result, higher capital holdings are linked to higher profitability and stability (Brissimis & Delis, 2009; Mirzaei et al., 2013; Pasiouras & Kosmidou, 2007). The banks' credit risk is considered by including the ratio of loan loss provisions to total loans. This ratio is a forward-looking measure of banks' loan quality. Higher provision reflects borrowers' inability to honour their loan obligation in a timely manner.

3.2.2.4.2 Macroeconomic Variables

To observe the effect of macro-economic variables and in line with previous studies, gross domestic product (GDP) and inflation (INF) are selected. Both are used to represent macroeconomic aspects in examining operational performance and bank stability. Sufian & Habibullah (2012), Köhler (2014), Bourkhis & Nabi (2013) and Cihák & Hesse (2007) have used these macroeconomic factors as the control variables in explaining the variations in the responding variables.

Table 3.2 provides a detailed explanation of the dependent and independent variables.

Table 3.2: Description of the Variables Investigating the Impact of M&As on the Bank's Operational Performance and Stability

Variables	Code	Definition	Features/ Description	Sources	Expected sign	References
A. Dependent Variables						
Operational Performance	ROA,	ROA and ROE measure the operational performance of the banking sector. ROA; how efficient the manager is in have better ROA by using bank assets.	ROA is defined as tax income as a percentage of total assets.	Fitch Connect database, bankscope and bank's annual report	Positive	(Rao-Nicholson, Salaber, & Cao, 2016; Abdulazeez, Suleiman, & Yahaya, 2016; Rani,Yadav, & Jain, 2016; Khan, & Khan, 2018; Jallow., Masazing, & Basit, 2017; Kandil, & Chowdhury, 2014; Gattoufi, Al-Muharrami, & Shamas, 2014; Nim Rani, Yadav, & Jain, 2016; Larasati, Agustina, Istanti, & Wijijayanti, 2018)
	ROE and	ROE implies profit generated from the money invested	ROE is defined as tax income as a percentage of total equity.			
	NIM	Net interest margin (NIM) reveals the amount of money that a bank is earning in interest on loans compared to the amount it is paying in interest on deposits	Net Interest Margin (NIM) is calculated by profit after tax/total profit			
Bank Stability	Z-score	Bank Stability	Measures the level of risk of the banking sector. It is measured by Z-score = ROA+ equity to total asset) / standard deviation of ROA).	Author calculation	Positive	(Ibrahim & Rizvi, 2018; (Moutsianas, & Kosmidou, 2016**;** De Haan, & Poghosyan, 2012; (Čihák, & Hesse, 2010; Ibrahim & Rizvi, 2018; Nguyen, Skully, & Perera, 2012; Wahid, & Dar, 2016; Diaconu, & Oanea, 2015; Khaddafi, Heikal, & Nandari, 2017; Li, X., Tripe, & Malone, 2017; Li et al., 2020; Karim, Al-Habshi, & Abduh, 2016).

Table 3.2, continued

Variables	Code	Definition	Features/ Description	Sources	Expected sign	References
B. Independent Variables						
1. Focal Variables						
Bank sizes	BSTA, BSTD & BSOI	Bank sizes are measured based on the total assets (BSTA), total deposits (BSTD), and operating income (BSOI). The units for BSTA, BSTD & BSOI are USD Million	Natural log of total assets, total deposits and operating income.	Fitch Connect database, bankscope and bank's annual report	Positive	(Ibrahim & Rizvi, 2017; Kandil & Chowdhury, 2014; Markides & Ittner, 1994; Cybo-Ottone & Murgia, 2000; Kosmidou, Pasiouras, Doumpos, & Zopounidis, 2006; Aladwan, 2015; Micco et al., 2007; Gamra, & Plihon, 2011)
Level of bank sizes	$BSTA_{LMS}$ $BSTD_{LMS}$ $BSOI_{LMS}$	Bank size total assets large, medium, and small ($BSTA_{LMS}$) Bank size total deposits large, medium, and small ($BSTD_{LMS}$) Bank size operating income large, medium, and small ($BSOI_{LMS}$)	Sorting banks measure the level of bank sizes from the lowest to the largest. i.e., there are 24 banks. all banks are arranged from the lowest to the largest. And then the first 8 banks are sorted as the small banks, the second 8 banks are sorted as medium-sized banks while the last 8 banks are sorted as the largest banks.	Author calculation	Positive	(Katib & Mathews, 2000; Kosmidou, Pasiouras, Doumpos, & Zopounidis, 2006; Aladwan, 2015; Ibrahim & Rizvi, 2018; Čihák, & Hesse, 2010; Wahid, & Dar, 2016)
Economies of scope	Escope	To measure the effect of lending activities of the bank, the loan ratio is used; loan to the total assets; lower ratio indicates excessive liquidity.	Loan ratio; it is defined as loan/total deposits (economies of scope)	FicthConnect database	Positive	(Gamra, & Plihon, 2011; Fithria, & Sholihin, 2018; Beck, Demirgüç-Kunt, & Merrouche, 2010; Amene, & Alemu, 2019; Gattoufi, Al-Muharrami, & Shamas, 2014; Smirnova, 2014; Antoniadis et al., 2014; Abbas et al., 2014; Beck, Guo & Yang, 2013; William, 2009; McClure, 2009; Pasiouras & Zopounidis, 2008; Cigola & Modesti, 2008) [cont.]

Economies of scale	Escale	The efficiency ratio implies how efficient the bank is. A lower or negative value indicates the better of the bank.	Cost/income (economies of scale)	FicthConnect database	Negative	[cont.] (Piloff & Santomero, 1998; Ibrahim & Rizvi, 2017; Sufian, 201; Piloff & Santomero, 1998; Ibrahim & Rizvi, 2017; Sufian, 2011)
Non-financial intermediary role	NFIR	Non-interest cost to non-interest income, to measure non-financial intermediary role. Lower is better.	Non-interest cost/non-interest income	FicthConnect database	Negative	(Sufian,& Habibullah, 2009; Nguyen et al., 2012); (Lepetit, Nys, Rous, & Tarazi, 2008).
Modes of financing (cash or stock)	FIN	M&As are financed by stock or cash	Payment of M&As deal	FicthConnect database	Positive	(Kwenda, Oyetade, & Dobreva, 2017; Iankova (2014; Nor & Ismail, 2006; Ramaswamy & Waegelein, 2003; Tuch & O'Sullivan, 2007)
2. Control Variables						
Credit risk	CR	Loan loss reserve to the gross loan is used to measure the credit risk of the bank.	Loan loss reserve to gross loan	FitchConnect database, bankscope and bank's annual report	Negative	(Boloupremo & Ogege (2019; Fayed (2013; Altunbaş, Y., & Marqués, D. (2008; Kandil, & Chowdhury, 2014; Gattoufi, Al-Muharrami, & Shamas, 2014; Al-Sharkas et al., 2008)
Liquidity	LIDY	To know about the liquidity position of the bank liquidity ratio is used in the study.	Liquidity ratio; it is defined as liquid asset/total deposits	FitchConnect database, bankscope and bank's annual report	Negative/ Positive	(Amene & Alemu (2019; Fayed, 2013; Altunbaş, & Marqués, 2008; Abbas et al., 2014)
Capitalization	CAP	Capitalization represents the financial condition of the banks especially the shareholder's portion.	Total equity/total assets	FitchConnect database, bankscope and bank's annual report	Positive	(Kandil, & Chowdhury, 2014; Gattoufi, Al-Muharrami, & Shamas, 2014; Rao-Nicholson & Cao, 2016; Abbas et al., 2014)
3. Macro-economic Variables						

Gross domestic product	GDP	To measure M&As performances along with bank-specific factors, macro factors were also used. Gross domestic product is used to represent the economic size of the particular country	Natural log of GDP constant term to USD	International Monetary Fund (IMF)	Positive	(Mohamed & Sidiropoulos, 2010; Sufian & Habibullah, 2012; Köhler, 2014; Bourkhis & Nabi, 2013; Cihák & Hesse, 2007; Amene, & Alemu, 2019; Erel et al., 2017; Choi & Jeon; 2011; Hassan, & Bashir, 2003)
Inflation	INF	Inflation has been measured by the Consumer Price Index (CPI)	Inflation	International Monetary Fund (IMF)	Positive/ negative	
4. Industry-specific variables (Market structure)						
Log of Herfindahl Hirschman Index	LHHI	Hirschman Herfindahl Index (HHI), a measurement of market structure (concentration) for the country based on total assets (TA), Total deposits (TD) and Gross loan (GL)	It implies the level of competition in the sector. HHI is calculated by squared root summation of the bank's market share based on total assets. Since the measurement of the other two variables, namely total deposits and gross loan are the same, I select total assets for the analysis.	Author calculation		(Rinkevičiūtė, & Martinkute-Kauliene, 2014; Galetić, & Obradović, 2018; Al-Muharrami, & Matthews, 2009; Khan, Ahmad, & Chan, 2018; Mohammed, Ismail, & Muhammad, 2016; Hakim & Chikr, 2014; Abdul Kadir, Habibullah, Siong Hook, & Mohamed, 2014; Sufian & Habibullah, 2013; Gajurel, 2010; Turk-Ariss, 2010; Abdul Majid & Fadzlan, 2007; Pawlowska, 2005; Repon and Islam, 2016; Rinkeviciute & Martinkute Kauliene, 2014; Bikker, & Haaf, 2002)
Concentration ratio	CR3	Concentration ratio (CR), a measurement of market structure (concentration) for the country based on total assets (TA), Total	It implies the level of concentration in the sector. CR3 is calculated by taking 3 largest banks based on total assets.	Author calculation		As above

	deposits (TD) and Gross loans (GL).	Since the measurement of the other two variables, total deposits and gross loan, is the same, I select total assets for the analysis.	

3.3 Data Collection

Unbalanced panel data are collected from Q1 2004 until Q4 2020, and from 24 banks consists of 14 conventional and 10 Islamic banks (see Appendix Table A2.1). These banks represent 6 countries: Saudi Arabia, United Arab Emirates, Qatar, Kuwait, Bahrain, and Pakistan.

This study used 17 years of quarterly data from 2004 until 2020. Total number of observations for Islamic banks is 200 in pre-M&As and post M&As while 280 observations for conventional banks in pre-M&As and post M&As. The six countries were selected because there were M&As aspects in the Islamic banking sector within the selected period of times. Data were collected from several secondary sources: the bank's financial statement, FitchConnect (Bankscope) database, Bloomberg and Thomson Reuters. Meanwhile, the macroeconomics variables i.e., quarterly GDP and inflation were collected from the World Bank and the International Monetary Fund (IMF) database. Two periods were selected for the analysis i.e., 5 years before and 5 years after the M&As. It has been divided into 5 years pre & post following which year M&As has occurred for that particular's banks. Notably, the year of happening M&As is not the same for all banks.

3.4 General Overview of Research Methodology

Various previous studies have applied different techniques to empirically analyze the impact of M&As. For instance, regression analysis has been applied by (Kandil & Chowdhury, 2014; Abbas Hunjra, Azam, Ijaz, & Zahid, 2014; Bernad, Fuentelsaz & Gómez, 2010; Ravichandran, Mat Nor, & Mohd-Said, 2010) whereas other studies like (Ibrahim & Rizvi, 2017; Barros & Caporale, 2012; Neto, Brandão & Cerqueira, 2010;

Mohamed & Sidiropoulos, 2010; Zou & Simpson, 2008; Amihud et al., 2002) have applied panel techniques to analyze and explain M&As activity in the banking sectors. This study applies POLS and panel data techniques (fixed effects and random effects) is conducted along with structural equation modeling (SEM) using Stata package 14.2 (Acock, 2013; Huber, 2014; Venturini, & Mehmetoglu, 2019) to examine and analyze research questions. Due to the smaller number of groups (24 banks), GMM techniques cannot be applied since number of instruments are high.

3.4.1 Panel Data Techniques

3.4.1.1 Fixed Effects (FE) and Random Effects (RE)

Panel data techniques are often considered an efficient and popular analytical method in handling econometric data. It allows the inclusion of data for N cross-sections (e.g., banks) and T periods (e.g., years). It may be a balanced panel (same time periods for all units) or an unbalanced panel (different time periods for all units). In this study, the unbalanced panel is adopted. The diagnostic test of panel data is discussed in Section 3.5. To choose the most appropriate estimator between fixed effects and random effects, a statistical test namely the Hausman test is used. The Hausman statistic tests the null hypothesis that RE is appropriate for a particular sample compared to the FE and allows us to decide which model gives the best estimation. The Hausman test allows verifying the presence of correlation between the unobservable heterogeneity and the explanatory variables (Wooldridge, 2002). The basic framework for panel data is defined as the following regression model.

$$Y_{nt} = \alpha + vx_{nt} + \varepsilon_{nt}$$

(3.4)

Where:

Y_{nt} = dependent variables; ROA, ROE, NIM and Z-score

α = constant term,

v = vector of all explanatory variables effecting M&As performance and stability

x = number of explanatory variables

n = cross-section unit

t = periods

ε = denotes error disbursement.

The econometrics specifications of the model

a. Operational performance

$$PER_{nt} = \alpha + \beta_1 BSTA_{nt} + \beta_2 BSTD_{nt} + \beta_3 BSOI_{nt} + \beta_4 Escale_{nt} + \beta_5 Escope_{nt} + \beta_6 NFIR_{nt} + \beta_7 LIDY_{nt} + \beta_8 CAP_{nt} + \beta_9 CR + \beta_{10} GDP + \beta_{11} INF + £_1 FIN_{nt} + £_2 BSTA_{LMSnt} + £_3 BSTD_{LMSnt} + £_4 BSOI_{LMSnt} + \varepsilon_{nt}$$

(3.5)

b. Bank stability (Z-score)

$$Z\text{-score}_{nt} = \alpha + \beta_1 BSTA_{nt} + \beta_2 BSTD_{nt} + \beta_3 BSOI_{nt} + \beta_4 Escale_{nt} + \beta_5 Escope_{nt} + \beta_6 NFIR_{nt} + \beta_7 LIDY_{nt} + \beta_8 CAP_{nt} + \beta_9 CR + \beta_{10} GDP + \beta_{11} INF + £_1 FIN_{nt} + £_2 BSTA_{LMSnt} + £_3 BSTD_{LMSnt} + £_4 BSOI_{LMSnt} + \varepsilon_{nt}$$

(3.6)

Where:

PER = return on asset (ROA), return on equity (ROE), net interest margin (NIM)

$Z\text{-score}_{nt}$ = bank stability (Z-score),

α = constant term

β = coefficient of exogenous variables

£ = coefficient of dummy variables

ε = error term

BSTA = bank size total assets,

BSOI = bank size operating income,

BSTD = bank size total deposits,

$BSTA_{LMS}$ = level of bank sizes i.e., large, medium, and small based on total assets,

$BSOI_{LMS}$ = level of bank sizes i.e., large, medium, and small based on operating income,

$BSTD_{LMS}$ = level of bank sizes i.e., large, medium, and small based on total deposits,

Escale = cost to income,

Escope = deposit to loan,

NFIR = non-interest expense to non-interest income,

FIN = dummy for the mode of financing; cash or stock,

LIDY = liquidity; liquid asset/total assets,

CAP = capitalization; total equity / total assets,

CR = credit risk; loan loss reserve to gross loan,

GDP = gross domestic product,

INF = inflation,

Panel data models investigate group (individual-specific) effects, time effects, or both to deal with heterogeneity or individual effect that may or may not be observed. These effects are either fixed or random effects. Moreover, the fixed and random effect is called a static model.

3.4.1.1.1 Fixed Effect (FE)

The most common panel estimators for relatively large T (T>N) are macro panel data. FE is also known as within estimator or least square dummy variable estimator or covariance estimator. Fixed effects regression is adopted to control for omitted variables that differ between cases but are constant over time. FE allows a researcher to observe the effects of omitted independent variables on the dependent variables. It imposes equality of all slop coefficients and error variance across the variance and only the intercept across units such as firms, banks, or countries are allowed to vary. Moreover, FE is more restrictive. FE expresses the following model,

$$Y_{it} = \alpha + \beta X_{it} + \acute{\omega} Z_{it} + \varepsilon_{it} \quad (3.5)$$

Xit; [1 for i^{th} individual i=2....N, 0 otherwise] and Z; [1 for t^{he} period, t=2...n, 0 otherwise]

3.4.1.1.2 Random Effect (RE)

The random effect model is the estimator if some omitted variables do not vary over time and differ across the cases. In contrast, others may be fixed between cases and vary over time. It is the less restrictive estimator. It also imposes the equality of all slop coefficients but allows error variances and intercept to differ across countries. It assumes random intercepts, the mean of which is captured by the constant term and the variance of which is captured by the variance of the error term. RE model express as the following

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it} \ (\varepsilon_{it} = \mu_i + \nu_t + \acute{\omega}_{it}) \quad (3.6)$$

Where:

Y_{it} = dependent variables,

X = explanatory variables,

α = constant term,

β = slop coefficients,

$\mathcal{E}$ = the composite error term,

μ ~N (O, $\sigma^2\upsilon$) = cross-section error component,

ν ~N (O, $\sigma^2\nu$ = time-series error component,

ώ ~N (O, $\sigma^2\acute{\omega}$) = combined error component

i = units and

t = period.

If the $\sigma^2\upsilon$ and $\sigma^2\nu$ are zero, then the RE model becomes the FE model.

If there is a doubt, the correlation between the missing cross-sectional characteristics and the explanatory variables, the generally accepted way of choosing between FE and RE is running the Hausman test. Here, the null hypothesis is no such correlation exists, i.e., RE against alternative hypothesis correlation exists, i.e., FE. If the p-value is less than 0.05, Ho is rejected. Then it is safe to use FE, meaning that the fixed effect model is appropriate for the study. Accordingly, the Chow test selects between POLS and FE while the Lim test for choosing between POLS and RE (Maulidar, & Majid, 2020; and Zulfikar, & STp, 2019).

3.4.2 Structural Equation Modelling (SEM)

SEM is a very general statistical modeling technique widely used in the behavioral sciences. SEM is a multivariate statistical analysis technique that is used to analyze structural relationships. This technique is the combination of multiple regression analysis, and it is used to analyze the structural relationship among variables. It can be viewed as a combination of factor analysis and regression or path analysis. The interest in SEM is often

on theoretical constructs, which are represented by the latent factors. Hair Jr, Sarstedt, Hopkins, & Kuppelwieser (2014) mentioned that SEM has become the dominant analytical tool for testing cause-effect relationships models with latent variables. The relationships between the theoretical constructs are represented by regression or path coefficients between the factors (Hox, & Bechger, 1998). SEM brings two advantages. Firstly, SEM makes it possible to study complex patterns of relationships among the constructs in a conceptual model in an integrative fashion. Secondly, the measurement of unobserved (latent) variables by observed fallible indicators can be modeled explicitly, and the effect of measurement error (both random and systematic) on structural relationships can be considered. Especially it is a variance-based Structural Equation Modelling approach. This study uses SEM to test the mediation part using the Stata software for pooled data set (Mayerl, & Andersen, 2019; Sattar, 2019; Gu, Cao, & Wang, 2019). Furthermore, bootstrapping t-test is analysed for further support the SEM results.

The bootstrapping techniques proposed by Preacher and Hayes (2004). Bootstrapping assigns measures of accuracy (bias, variance, confidence intervals, prediction error, etc.) to sample estimates. The test further supports the result of SEM. More specially by dint of Bootstrapping test, it can be confirmed and further support which variable showed mediation relationship.

3.4.2.1 Mediating 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 has become increasingly

vital to examine the direct and mediation effect determining bank performance. As stated earlier, mediation is a set of causal hypotheses. Technically speaking, the predictor (X) may influence the outcome variable (Y) through the mediating variable (M). In short, the inclusion of a third variable clarifies the relationship of two variables, i.e., the predictor variable (X) and the outcome variable (Y). It is also shown by indirect effects (path a and path b) and direct effects (path c). Whereas, path a state of predictor variables (X) on mediating variables (M), similarly, path b indicates the impact of mediating variable (M) on outcome variables (Y) and simultaneously path c shows predictor variable (X) on outcome variables (Y). The mediation relationship exists when a third variable plays an essential role in governing the relationship between two other variables. This relationship can be seen in Figure 3.1.

Nitzl, Roldan, & Cepeda (2016) noted that the current mediation literature discusses two different types of mediation, full and partial mediation. Partial mediation can be divided again into complementary and competitive partial mediation. The typology of mediation is shown in Table 3.1 and Appendix Table A4.

3.4.2.1.1 Full Mediation

Full mediation is indicated in the case where the direct effect c' is not significant whereas the indirect effect path (a × b) is significant, which means only the indirect effect via the mediator exists. In other words, technically speaking, full mediations means that the effect of the variable X to Y is completely transmitted with help of another variable M. It also means the condition Y completely absorbs the positive or negative effect of X. In this

way, it can completely pass an effect or completely hinder the effect in terms of another effect.

3.4.2.1.2 Partial Mediation

Partial mediation is indicated in other situations under the condition that both the direct effect c' and the indirect effect path (a × b) are significant represent partial mediation. Two types of partial mediations can be distinguished namely complementary mediation and competitive mediation. A decision tree for a typology of mediation effects is reported in Appendix Figure A4.

3.4.2.1.3 Complimentary Mediation

In a complementary partial mediation, the direct effect path c' and indirect effect path (a × b) point in the same (positive or negative) direction (Baron and Kenny, 1986). It is an often observed result that path (a × b) and path c' are significant. A × b × c' is positive, which indicates that a portion of the effect of X on Y is mediated through M. In contrast, X still explains a portion of Y that is independent of M. This complementary mediation hypothesis suggests that the intermediate variable explains, possibly confounds, or falsifies the relationships between the independent and dependent variables. Complementary mediation is often called a 'positive confounding' or a 'consistent' model (Zhao et al., 2010).

3.4.2.1.4 Competitive Mediation

In a competitive mediation, the direct effect path c' and indirect effect path (a × b) point are in a different directions. A negative a × b × c' value indicates the presence of competitive mediation. As mentioned above, this indicates that a portion of the effect of X on Y is mediated through M, whereas X still explains a portion of Y that is independent of M. In the past, researchers often focused only on complementary mediation (Zhao et al., 2010). The competitive mediation hypothesis assumes that the intermediate variable will reduce the magnitude of the relationship between the independent and dependent variables.

3.4.2.1.5 Only Direct Effect Non-Mediation

If the indirect effect path (a × b) is not significant whereas the direct path c' is, the mediator variable has no impact; this indicates that an immediate, non-meditating effect is present. In this case, the study was perhaps searching for an inappropriate mediation relationship. However, it is possible that an unrecognized mediation relationship still exists and another mediation variable is present that mediates an effect between X and Y (Shrout and Bolger, 2002). Thus, a researcher should rethink the model's theoretical basis if the expected mediation relationship cannot be found (Zhao et al., 2010). A decision tree for the typology of mediation effects is reported in Appendix Figure A4.

3.4.2.1.6 No Effect Non-Mediation

There is no effect if neither the indirect effect path (a × b) nor the direct effect path c' is significant. The total effect can still be significant. First of all, in this case, the researcher should determine whether the sample size has enough power to show an effect

when there is an effect. Putting the last two cases together – the indirect effect path (a × b) is not significant and the direct path c' is or is not – frequently indicates a problematic or flawed theoretical framework (Zhao et al., 2010). In this case, the researcher should thoroughly examine the hypothesized model. When, for example, the total effect path c is significant, it can indicate that the mediation variable should be deleted because it brings no further degree of explanation. Technically speaking, if the mediation variable M has no real effect, it only dilutes the effect of the direct variable X and should be deleted. A decision tree for the typology of mediation effects is reported in Appendix Figure A4.

3.4.2.2 Model Selection Criteria

Many researchers propose several model selection criteria for SEM. The criteria are divided into absolute fit indices and incremental fit indices. Absolute fit indices determine how well a priori model fits the sample data (McDonald, & Ho, 2002) and demonstrate which proposed model has the most suitable fit. These indices provide a fundamental indication of the model fits (Hooper, Coughlan, & Mullen, 2008). Unlike the incremental fit indices, their calculation does not rely on the comparison with the baseline model but is instead a measure of how well the model fits in comparison to no model at all. Absolute fit indices include model chi-square, Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR). While incremental fit indices are also comparative (Miles, & Shevlin, 2007) or relative fit indices (McDonald, & Ho, 2002) are a set of indices that do not use chi-square in their raw form but compare the values to a baseline model. An example of incremental fit indices is the comparative Fit Index (CFI).

3.4.2.2.1 Model Chi-Square

The chi-square value is the traditional measure for evaluating the overall model fit and assessing the magnitude of discrepancy between the sample and the fitted covariance matrices. A good model fit would provide an insignificant result at a 0.05 threshold (Barret, 2007). Thus, the chi-square statistics are often referred as either a badness of fit (Kline, 2015) or a lack of fit measure. Cut off for good fit is p-value >0.05 (Kline, 2015)

3.4.2.2.2 Root Mean Square Error of Approximation (RMSEA)

The RESEA is the second fit statistic reported in the LISREL program and was first developed by Steiger and Lind. The RMSEA signals us how well the model with unknown but optimally chosen parameter estimates would fit the population's covariance matrix. As reported by the researcher, in recent years it has been regarded as one of the most fitted indices (Diamantopoulos, Siguaw, & Siguaw, 2000). One of the greatest advantages of the RMSEA is its ability for a confidence interval to be calculated around its value. This is possible due to the kwon distribution value of the statistics and subsequently allows for the null hypothesis (poor fit) to be tested more precisely (McQuitty, 2004). The cut-off value of RMSEA should be less than 0.08 (Kline, 2015).

3.4.2.2.3 (Standardized) Root Mean Square Residual (S) RMR

The RMR and SRMR are square roots of the differences between the residuals of the sample covariance matrix and the hypothesized covariance model. The cut of values for the (S) RMR ranges from zero to one with a well-fitting model obtaining values less than 0.05

(Diamantopoulos et al., 2000). However, values more than 0.08 are deemed acceptable. SRMR values close to zero imply good fits of the model (Kline, 2015).

3.4.2.2.4 Comparative Fit Index (CFI)

The CFI is the revised form of NFI which considers the sample size and performs well even when the sample size is small (Berger, 1990). The index was introduced by Bentler in 1990 and subsequently included in the fit indices in his EQS program (Kline, 2015). A cut-off criterion of <=0.90 was initially advanced, however, a recent study has shown that a value of more than 0.90 is required to ensure that incorrect specified models are not accepted. Meanwhile, Kline (2015) stated that a value of more than 0.85 is presently recognized as a good fit.

3.5 Motives Behind Applying Panel Data Estimation Techniques

The main motive behind implementing panel techniques is data constraint. Firstly, controlling the individual heterogeneity to take risk is described as the unobserved heterogeneity. If these influence the variable of interest and are correlated with observed explanatory variables (Xs), then the estimated effects of this variable will be biased and not precise. Secondly, panel data allows for larger observations by pooling individual and time dimensions (cross-sectional units-banks*times) as well as more variability. Thirdly, the technique is also dynamic (more adjustment of nature or environmental or geographic). Fourthly, this method could identify the parameter that could not be traced in the pure cross-section or pure time series.

The research questions can be investigated through conventional regression analysis that only traces the relationship between dependent and independent variables. Regression assumes long term theoretical relationships between the variables and expresses exogenous variable and endogenous variable. It does not give a lead-lag relationship among the selected variables, emphasizing this strategic decision.

Moreover, the time series technique does not apply to this study due to data constraints. It is difficult to obtain the significance of t-ratios or F-statistics from regressions. Panel technique is used because it allows the inclusion of data for (N) cross-sections (e.g., banks) and (T) periods (e.g., years) benefited, better controlling of individual heterogeneity, more information data sets (integration of time series and cross countries, therefore more degree of freedom), suitable to study the dynamics of adjustment (how individuals adjust over time), identification of parameters that cannot be identified with pure cross-sections or pure time series and reduced multicollinearity problem.

3.6 Diagnostics Test

This study used an unbalanced panel data set of 24 banks of 10 Islamic and 14 conventional banks from six countries[2] year ranging from the 1st quarter of 2004 until the 4th quarter of 2020. Several tests such as heteroscedasticity, multicollinearity, and auto-correlation were used to test the panel data set's accuracy and avoid any bias in the estimation. The diagnostic tests are discussed below.

[2] 6 countries are Qatar, Kuwait, Bahrain, Saudi Arabia, and UAE and Pakistan

3.6.1 Heteroscedasticity

It is used to test for heteroscedasticity in a linear regression model and assumes that the error terms are normally distributed. It tests whether the variance of the errors from regression is dependent on the values of the independent variables. The test result showed that chi2 (1) = 34.02, Prob>chi2 = 0.0000. Since the probability value (p-value) is less than 0.05, the null hypothesis is rejected. Meaning that there is a problem with heteroscedasticity. To correct the problem, Whites (1980) heteroskedastic-consistence covariance matrix estimation is used throughout the regressions.

3.6.2 Multicollinearity

Multicollinearity occurs when independent variables in a regression model are correlated. This correlation is a problem because independent variables should be independent. If the degree of correlation between variables is high enough, it can cause problems. Vif value is 2.19 (Appendix Table A1.1) and since is less than 10, it implies that there is no problem with multicollinearity.

3.6.3 Auto-correlation

Autocorrelation is a characteristic of data that shows the degree of similarity between the values of the same variables over successive time intervals. According to the auto-correlation test, F(1, 18) = 29.755, Prob > F = 0.000. Since the p-value is less than 0.05, the null hypothesis of no autocorrelation is rejected. It shows that the presence of an auto-correlation problem in the study. To solve the problem, vce (robust) is used throughout the regressions.

3.7 Chapter Summary

This chapter provides the fundamental research design of the study. It outlines the variables that were explained, and data collection procedures were identified. Models and empirical research techniques are used in the analysis. Then, the motive of behind applying panel techniques and Structural Equation Modelling (SEM) was discussed in detail.