investigate the performance of indices, the long run relationship, and the causality between the indices. The second part of the study had not been studied so far in any country except in Malaysia. However, those studies were focusing on variables influencing the Islamic index compared to the non-Islamic index or on volatility. This study uses different local variables plus one external variable. In addition, this study investigates the long-term relationship as well as the causality between each index and the selected macroeconomic variables. The third part of the study focuses on whether screened and non-screened firms differ in their returns and which firm's specific variable explains returns. The following chapter discusses the variables used in each part, data, the methodology, and the hypothesis of the study.

CHAPTER FOUR

DATA AND METHODOLOGY

The methodology of this study is divided into three parts. The first section is concerned with the answering the first part questions which are investigating the performance of KLSI vs. KLCI using t-test and risk adjusted ratios. It is extended into investigating the stationarity and the long and short-term relationships between them. The second section is to answer the second part questions. It deals with the macroeconomic variables and their influence on both KLSI and KLCI, utilizes the same time series techniques used in the first part (i.e. stationarity, long and short run relationship). Since, the first and second part of the methodology is overlapping, the time series technique is explained once for both parts. The third section utilizes panel data techniques to determine whether returns of both Syariah and non-Syariah firms differ to answer the third part questions.

4.1 Part1: Comparison and Relationship between Syariah and Non-Syariah Stock Market Indices Returns

This part focuses on the performance of two indices in the Bursa Malaysia namely Syariah index (KLSI) and Kuala Lumpur Composite Index (KLCI). The methodology is divided into four parts; first, using three measurements of risk adjusted return, second, unit root analysis, third, bivariate Granger causality between KLSI and KLCI, and lastly, Vector Autoregression analysis and impulse response. The study uses secondary data of both indices (KLSI and KLCI) of the Kuala Lumpur Stock exchange and the Kuala Lumpur inter-bank offer rate (KLIBOR) as a measurement of risk free asset. Daily data of the closing prices of both indices from April 1999 up to December 2005 will be employed in this study. The data were collected from Bank Negara website, perfect analysis software, and Bloomberg database in Bursa Malaysia. The programs used in performing these tests are E-views and Microsoft Excel.

4.1.1 Definition of the Variables

4.1.1.1 Kuala Lumpur Composite Index (KLCI)

This study uses the Kuala Lumpur Stock Exchange Composite Index (KLCI), which was constructed in 1986 with the objective of effectively reflecting the performance of the companies listed on the stock exchange. KLCI is used as the non-screened index benchmark. This is because it is generally sensitive to the investors' expectations, indicative of the impact of the government policy change, and reasonably responsive to the underlying structural changes in different sectors of the economy. The criteria used in selecting the stock component in KLCI are as follows[26]:

- Companies whose annual volume and/or market capitalization fall within the first three quartiles of the Main Board companies' volume and market capitalization will be considered for inclusion.
- Companies whose annual volume and/or market capitalization fall within the last quartile of the Main Board companies' volume and market capitalization will be considered for exclusion.
- Newly listed companies will only be considered for inclusion after a minimum period of three (3) months in order to minimize any distortion of the index.
- Companies that are more than 50 percent owned by any KLCI component company and which in fact are defined as subsidiaries by the Malaysian Companies Act are excluded.
- This criterion is used to minimize, and as far as possible, avoid double counting or weight distortion in the index.
- Companies may be considered for inclusion/exclusion to better represent the objectives of the KLCI.

The index is calculated using 1977 as the base year and the method of weighting is market capitalization the formula is as follows:

Index=100*(Current aggregate Market Capitalization / Base Aggregate Market Capitalization).

[26] http://www.bursamalaysia.com/website/marketinfo/indexcomp_guidelines.htm

4.1.1.2 Kuala Lumpur Syariah Index (KLSI)

The vast majority of Syariah scholars are in agreement that investment portfolio in stocks is allowed, provided it meets certain criteria designed to minimize non-Islamic activities. this study analyzes the Kuala Lumpur Stock Exchange Syariah Index (KLSI) that was launched in April 1999, a weighted-average index with its components comprising the securities of Main Board companies which have been designated as Syariah Approved Securities by the Syariah Advisory Council (SAC) of the Securities Commission (SC). The screening criteria implemented by SAC are as follow[27]:

- Companies will be excluded if they deal with Riba, indulge in gambling, manufacturing, or selling islamically forbidden products and involve an uncertainty (*Gharar*) element in their transactions. In addition, Companies dealing in conventional insurance and Syariah non-approved securities will also be excluded.

However, companies with both islamically permissible and non-permissible activities will be scrutinized as follows[28]:

- The core activity of the company must be Islamically permissible
- The public perception of the company must be good
- The element of non-permissibility is small and involves things such as common plight, and custom and the company in general serves the benefit (*Maslaha*) of the Muslim community.

For companies with mixed activities of permissible and non-permissible the benchmarks of tolerance are used. If the contributions in turnover or profit before tax from non-permissible

[27] http://www.klse.com.my/website/education/smb_syariahindex.htm
[28] ibid

activities of a company exceed the benchmark, the securities of the company are classified as non-Syariah securities. The benchmarks are:

- The five-percent benchmark applied to assess the level of mixed contributions from the activities that are clearly prohibited such as Riba (interest-based companies like conventional banks), gambling, liquor, and pork.
- The 10-percent benchmark applied to assess the level of mixed contributions from the activities that involve the element of '*umum balwa*", which is a prohibited element affecting most people and difficult to avoid. An example is interest income from fixed deposits in conventional banks.
- The 25-percent benchmark to assess the level of mixed contributions from the activities that are generally permissible according to Syariah and have an element of public interest (*Maslahah*), but there are other elements that may affect the Syariah status of these activities. Examples include hotel, and resort operations, share trading etc., as these activities may involve other activities that are deemed non-permissible according to the Syariah.

KLSI is used to represent the screened index in Bursa Malaysia. The index is calculated using 1998 as the base year and the method of weighting is market capitalization. The formula is as follow:

Index = 100*(current aggregate market capitalization / base aggregate market capitalization).

4.1.1.3 EMAS Index

EMAS Index[29] is the abbreviation of Exchange Main Board All-Shares Index. EMAS Index is weighted by market capitalization, with the base date on 1 January 1994 and an assigned index value of 100, and 269 companies listed on the base date. It has 646 companies listed at the end of 2005. The inclusion of this index is to facilitate as a market benchmark in the calculation of the risk performance measures. Most of the performance studies in Malaysia used KLCI as a benchmark. However, Leong and Aw (1997) and Low (2007) indicated that EMAS index yielded higher R^2 compared to KLCI. This indicates that EMAS index explains the variation in returns better than KLCI.

The index is calculated using 1994 as the base year and the method of weighting is market capitalization the formula is as follows:

Index = 100*(Current aggregate market Capitalization / base aggregate market Capitalization).

4.1.2 Series Characteristics

This part deals with the descriptive statistics of both returns and prices. The usual statistical properties are included in this part. The important descriptive statistics are mean, standard deviation, skewness, kurtosis and normality. The mean of the return indicates the average return of each index. Standard deviation is a measurement of the risk of each index. Skewness and kurtosis indicate the shape of the series distribution. On the other hand, the Jaqure-Bera test of normality determines whether the series is normally distributed by not rejecting the null hypothesis of not normally distributed data. In addition, the simple correlation is calculated to show the strength of the relationship between the indices.

---

[29] FTSE Bursa Malaysia EMAS Index replaces EMAS index in 26 June 2006 (http://www.klse.com.my/website/bm/media_centre/media_releases.html?/website/bm/media_centre/media_releases/2006_resources/20060626.html).

4.1.3 Risk Adjusted Performance

The return is measured by the difference of the prices between period t and t-1. In other words, return is calculated using the following formula log (P_t /P_{t-1})[30]. Four measurement techniques are used to calculate the return.

First, the risk and return of each index and the correlation between them are measured. The Sharpe ratio (1994) (SR) is a ratio developed by Nobel Laureate Sharpe to measure risk-adjusted performance. It is calculated by subtracting the risk-free rate from the rate of return for a portfolio and dividing the result by the standard deviation of the portfolio returns[31]:

$$SR = \frac{R_i - R_m}{\sigma_i} \quad 4.1$$

Where R_i represents return on index, which in this case will be KLSI, and KLCI, R_m is the EMAS index which is benchmark in this case, and σ_i is standard deviation of the index. Generally, higher Sharpe ratio indicates higher or superior performance, while the opposite is true.

Second, The Treynor index performance measure (TI) is a ratio developed by Jack Treynor that measures returns earned in excess of that which could have been earned on a risk less investment portfolio per each unit of market risk.

The Treynor ratio is calculated as:

[30] Hussein et al. 2005, p.109.

[31] The original model is $S = \frac{\bar{D}}{\sigma_D}$ where $\bar{D}$ is the average value of the monthly differences in returns between portfolio and benchmark, and σ_D is the standard deviation of the portfolio as indicated by Sharpe (1994), pp.50.

$$TI = \frac{R_i - R_m}{\beta_i} \qquad 4.2$$

Where R_i and R_m are as defined above and β_i is beta of the index, which is calculated using CAPM. The higher the value of Treynor index the more return gained per unit of risk.

Third, the adjusted Jensen’s alpha index performance (AJAI) is a risk-adjusted performance measure that represents the average return on a portfolio over and above that predicted by the CAPM, given the portfolio's beta and the average market return. This is the portfolio's alpha. It is calculated as follow:

$$\alpha = R_i - \left[R_f + \beta_i\left(R_m - R_f\right)\right] \qquad 4.3$$

Where R_i, R_f, $ß_i$ are as defined above and R_m is the EMAS index which is the benchmark in this case. Alpha evaluates returns that the fund has generated against the returns actually expected out of the fund given the level of its systematic risk.

The difference between TI and SR is that the former deals with the systematic risk or beta[32] for the indices while the latter uses the standard deviation as a measure of risk.

Statman (1987) developed the last measurement in this part that is eSDAR. In another article, Statman (2000) and Bello (2005) indicated a modified formula of Sharpe ratio. The measurement is called excess standard deviation adjusted return, and is abbreviated as “eSDAR.” It is the excess return of the studied index (Syariah and Composite) over the

[32] Beta for SI is found by regressing the past return of the index against market returns using the following model: $R_i = a + ßR_m + u_i$ where R_m is proxied by BURSA MALAYSIA KLCI.

return of the benchmark (EMAS index) where the index is leveraged to have the benchmark's standard deviation. The mathematical expression is,

$$eSDAR = R_f + \left[\frac{R_i - R_f}{SD_i}\right] SD_{EMAS} - R_{EMAS} \tag{4.4}$$

Where R_F is the daily KLIBOR rate, R_i and SD_i are the index return and standard deviation respectively and SD_{con} and R_{con} are standard deviation and returns for the EMAS index which is the benchmark. The resultant value will indicate the return adjusted to the standard deviation of the benchmark. The higher the value of eSDAR, the greater the returns and vice versa.

Table 4.1 below shows brief summary of the variables used in this part of the thesis and their definitions.

Table 4.1 Summary of the variables

Variable	Variable name	Definition
KLCI	Kuala Lumpur Composite Index	An index of the largest 100 companies in the main board
KLSI	Kuala Lumpur Syariah Index	An index of companies that passes Shariah supervisory board screening.
EMAS	Exchange Main Board All-Shares Index	An index for the companies in the main board
KLIBOR	Kuala Lumpur Inter-bank offer rate	average of interbank deposit rates at the Interbank Money Market in Kuala Lumpur,

4.2 Part 2: The Relationship between Macroeconomic Variables and Stock Market Index Returns

This part investigates whether both indices are influenced by the same macroeconomic variables and examines studying both short and long-term dynamics between each index and the macroeconomic variables. Research conducted on Islamic stock market and its macroeconomic determinants is rather minimal. This is because it is assumed that the Islamic stock index would react to macroeconomic variables the same way a conventional index would and that the relationship will be almost the same. Although in the conventional index perspective the reaction is not exactly the same for all countries, it is not far from the realm of being a generalized fact that in developed countries the stock index reacts similarly as in developing countries. However, there is a need for studying the real factors, whether micro or macroeconomic variables behind reactions. The justification for this might be because of the screening process implemented in the Syariah compliant stock or the ethical investment portfolio for that matter.

Rudd (1981), Teper (1991), Sauer (1997), and Johnson and Neave (1996) indicated that the screening criteria would theoretically have a negative effect on the stock returns. This is because of less diversification, higher transaction cost, and higher monitoring cost. Another reason could be the implied[33] different behavior of the Muslims towards investing in Islamic stocks (Rosly, 2005). In addition, Hickman et al. (1999) indicated that the lower the correlations of returns between securities, the higher the reduction of risk. Hickman et al. (1999:74) suggested, “after all, the social performance of such investment portfolios will not be systematically changed by, for example, a recession, so there’s no reason for social investors to alter their holdings.” In addition, Hickman et al. (1999) asserted that the correlation between the overall market and screened investment portfolio tends to be lower than between the overall market and non-screened investment portfolio.

[33] The word implied here is referring to the fact the one of the objective of Syariah is to protect one’s money and therefore, it is implied that the Muslim investor might have different factors affecting his decision.

Various studies used different models to determine what variables influence stock returns. Ibrahim (2001), Yousof et. al. (2007) and others used Gordon stock valuation, Arbitrage pricing theory, or aggregate demand and supply models. The model that used in this study is the stock valuation model. Valuation in the stock market refers to a process in which an investor determines the worth of a security using the risk and return concepts, which can be applied to any asset that produces a stream of cash flow. In order to set up the value of an asset, investors must determine certain variables that affect the amount of future cash flows, the timing of the cash flows and the rate required on the investment portfolio.

Kettell (2001), to clarify the classic method of calculating intrinsic value, applied present value analysis. The present value process involves the discounting of future cash flows. The intrinsic value of a security is said to be equal to the discounted or present value of the future stream of the cash flows that investors expect to receive from the asset. This is illustrated as:

$$Estimated\ value\ of\ security = \sum_{t=1}^{n} \frac{Cash\ flow}{(1+k)^t} \qquad 4.5$$

Where *k* is the appropriate discount rate or required rate of return and t is the time interval. Since there is no specific model for valuation of screened securities, the stock valuation model is used. The study examines the impact as well as the short run and long run dynamics of selected macroeconomic variables, both real and monetary, on the KLSI as well as KLCI in Bursa Malaysia. The choice and justification of the variables are explained below. The variables used here are real activity or GDP, Kuala Lumpur Composite Index (KLCI), price level, money supply and the oil price against Kuala Lumpur Syariah Index KLSI. Of course, the choice of the variables is based on previous studies on Malaysia and other countries and intuitive financial relationships. Ibrahim (2001) asserted that a large number of variables should be included in order to disentangle the relationships and dynamics among variables. However, this might lead to a loss in the degree of freedom and inclusion of irrelevant variables. Accordingly, since this is study has a short interval of data, only five explanatory variables that influence the stock market in Malaysia which are derived from previous studies.

4.2.1 Dependent Variables

The first dependent variable in this study is Kuala Lumpur Syariah Index (KLSI). The vast majority of Syariah scholars agree that investment portfolio in stocks is allowed provided they meet certain criteria designed to minimize un-Islamic activities, which means that the Syariah index must be based on Syariah principles. this study analyzes the Kuala Lumpur Syariah Index (KLSI) that was launched in April 1999, a weighted-average index with its components comprising the securities of Main Board companies which have been designated as Syariah Approved Securities by the Syariah Advisory Council (SAC) of the Securities Commission (SC).

The second dependent variable to be included here is Kuala Lumpur Composite Index (KLCI). This study uses KLCI, which was constructed in 1986 with the objective of effectively reflecting the performance of the companies listed on the stock exchange. This is because it is generally sensitive to the investors’ expectations, indicative of the impact of the government policy change, and reasonably responsive to the underlying structural changes in different sectors of the economy.

The main reason for using KLCI is that it is considered as the one of the easily accessible references to the health of the Malaysian economy. Some studies use regional or international indices in order to examine the integration between local index and foreign index.

4.2.2 Explanatory Variables

The first explanatory variable is the real activity or the GDP of Malaysia. The relation between GDP and the stock market has been extensively studied by Chen et. al. (1986), Fama (1990), Lee (1992) and Mukherjee et. al. (1995), to name a few who proved that economic activity is positively related to stock market. The stock valuation model suggests that the stock prices are a function of the present value of expected cash flows. The cash flow depends on the performance of the company and therefore, on the performance of the economy. The GDP will positively affect the company through an increase in the expected cash flow and therefore their prices. If the economic activity may increase the performance of companies, their profit and dividends will increase too. Therefore, it is hypothesized that the relation between stock prices and real GDP is positive.

Secondly, the general price level in the economy. According to Fisher effect, which states that the nominal interest rate will anticipate the expected inflation rate, the variable affecting real interest rate are real factors like real activity and therefore, inflation has no effect whatsoever on real interest rate as well as the stock prices. This is because the interest rate or rate of return will reflect the expected inflation in the economy and investors will revalue their assets in order to hedge against expected inflation.

Madura (2003) listed two points to prove that nominal interest rate reflect expected inflation. First, nominal interest rate will compensate lenders for the reduction in purchasing power and second, it compensates lenders for forgoing current consumption.

Nevertheless, this variable relation to the stock market is generally theorized to be negative. Among others, Chen et. al. (1986), Handroyiannis et. al. (2001), Mukherjee et. al. (1995) and Maysami et. al. (2000) found that inflation has negatively affected the stock prices. If prices increase while output remains the same, this causes higher prices for goods since their input prices increase too. This therefore increases the cost of production of goods and reduces the profit of the producer causing their stock prices to decrease. In other words, Fama (1981) suggested that due to the positive correlation between real activity and stock prices, when inflation occurs, this causes the real activity to slow down and the stock prices to decrease. This is called the proxy hypothesis. Geske and Roll (1983) suggested another explanation based on the positive relationship between declining real activity and government deficit. When the economy is slowing, this decreases the government revenue; therefore, the government runs into deficit. In counter cyclical policy if the government deficit is monetized, then this triggers inflation causing stock prices to decrease, this is the reverse causality.

However, there is another opinion that stock prices and inflation have a positive relation. Maysami et. al. (2000) suggested based on the Fisher effect that one reason of holding different assets is ultimately to hedge against inflation. In the case of Malaysia, Ibrahim (2003), Ibrahim (2001), Ibrahim et. al. (2003), Wongbangpo et. al. (2002), Khil et. al. (2000) and Ibrahim et. al. (2001) found that stock prices are positively related to inflation. Subsequently, the relationship expected between inflation and stock prices is positive.

Thirdly, money supply in the narrow definition is used (i.e. M1). Money supply and inflation are closely related, since the change in money supply affects the price level in the economy. There are few ways of explaining the relationship between stock prices and the monetary policy. One way is the quantity theory of money that equates money supply to the total output of the economy. Therefore, if money supply increases the output increases for the equation to hold. However, if money supply increases and the quantity of goods and services in the economy remains the same, then the effect is transferred to an increase in the price level. The choice of the definition of money that measures the impact of money supply is debatable. Some studies choose the narrowest definition of money supply, i.e. M1, such as, Ibrahim (2003), and Wongbangpo et. al. (2002), while Ibrahim (2000), and Ibrahim (2001) used broad definition of money i.e. M2. Ibrahim (1999) and Habibullah et. al. (1996) used both M1 and M2 in investigating the relationship between stock prices and macroeconomic variables. Tan and Cheng (1995) and Tan and Baharumshah (1999) in explaining the causal relationship among prices, output, money supply, and interest rate used the three definitions of money. They concluded that M1 and M3 are the most important monetary instruments affecting prices and output respectively. Ibrahim (2003), following Tan and Baharumshah (1999), used M1 as one of the explanatory variables for prices in Malaysia. Subsequently, the narrow definition of money is used i. e. M1 following Ibrahim (2003). The hypothesized sign of M1 with stock prices is to be positive.

The fourth explanatory variable in this study is the crude oil prices. Few studies include the effect of oil prices on the stock market. In Malaysia, there are no studies available that describes the impact as well as the direction of causality between stock market and oil prices. This might be because it is almost agreeable that oil prices influence the stock market either positively or negatively depending on the nature of the country whether it is an oil exporting or oil importing country. However, the significant effect on the economy as a whole and the stock market in specific is less documented in Malaysia. Cheung et. al. (1998), Hondroyiannis et. al. (2001) and Papapetrou (2001) among others documented a negative relationship between oil prices and stock markets in industrialized countries. They found that oil prices influence the real activity and therefore the movement of stock prices. Subsequently, oil prices are hypothesized to be negatively related to stock prices.

The macroeconomic variables to be included in this study are the end of the month values reported in the bank Negara Malaysia monthly bulletin plus the closing prices of KLCI and KLSI. Real activity or GDP is be represented by Industrial production index since GDP is not available on a monthly basis. Inflation is represented by the oft-quoted bank Negara Consumer price index i.e. CPI. Money supply is represented by the narrow definition of money from bank Negara Malaysia. KLCI and KLSI are the end of the month closing prices of both indices. Lastly, the crude oil prices are represented by Oil. The data run from April 1999 to April 2006. This is because the KLSI was first initiated in 19 April 1999. Table 4.2 below shows brief summary of the macroeconomic variables used in this part of the thesis and how they were defined.

Table 4.2 Summary of the Variables

Variable	Variable name	Definition
GDP	Gross Domestic Product represented by the industrial production	An indicator that shows the production output from industrial activities, such as agriculture, forestry, logging and fishing; mining and quarrying; manufacturing; and water, gas and electricity.
M1	Narrow definition of money supply	Currency in Circulation refers to the notes and coins issued by Bank Negara Malaysia less the amount held by the commercial banks and Islamic banks.
CPI	Consumer price index (general price level)	measures the average rate of change in prices of a fixed basket of goods and services which represents the expenditure pattern of all households in Malaysia
KLCI	Kuala Lumpur Composite Index	An index of the largest 100 companies in the main board
OIL	Oil price	World prices for crude oil.
KLSI	Kuala Lumpur Syariah Index	An index of companies that passes Shariah supervisory board screening.

4.3 Part 3: Firm Specific Variables and Stock Returns

In this part, the factors or determinants of the firm's returns are discussed. The pioneers in this realm are Fama and French who did several studies on the determinant of stock returns using time series data on mostly developing countries and one on emerging markets. Some of the papers that discussed the Malaysian stock market determinant among other countries are Chui and Wei (1998), Drew et al. (2003) and Drew and Veeraraghavan (2002). One of the studies that solely discussed the Malaysian stock market determinants is Pandey (2001). They used panel data of more than 240 companies for 8 years. The variables used in these papers ranged from size-related variables to performance-related variables. The variables used in this study are based on the studies done previously in either developing or developed market. Studies discussing the determinants of the Malaysian stock market are few. This could be because it is assumed that the Malaysian stock market will react almost to the same factors as any other developing market. The other reason is the difficulty in getting the right information about the companies studied and the unavailability of the data.

This part of the study is going to use two sets of stocks, screened stocks, which is listed under the Kuala Lumpur Syariah index and the other stocks are the remaining space of stock that are not in compliance with Syariah criteria.

The stocks included in the Islamic stocks are those who have been consistently listed in KLSI since 2000 up to 2006. Therefore, at the end of the process, two sets of samples are produced one with Islamic stocks and the other is with the non-Islamic stocks.

4.3.1 Matching Process

Since this part is focused on the difference in returns for screened and non-screened firms, the need arise to match firms based on certain criteria. The most widely used matching criteria are by industry. However, in our case it is difficult to match firms based on industries. This is because some industries are not allowed into the Syariah index based on their products or activities. Therefore, it is impossible to match hotels from both indices as well as finance companies. The next criteria used for matching is size. To match firms based on size either the beginning point of the period or the study of the ending point chosen. The ending point is chosen for matching during this period, which means 2006. The matching is done as follow, starting with all the listed firms in Syariah index by 2006 then move back toward 2000 in order to determine the number of firms that is representing the non-screened firms. After doing that, the matching by size for both firms based on fiscal year end 2006 is done.

4.3.2 Dependent Variable

4.3.2.1 Stock returns

The stock returns are calculated on annual bases using the closing prices at year-end for each firm. Since the fiscal year is not consistent among firms, the fiscal year end provided by DataStream database is followed to find the stock returns closing prices.

4.3.3 Independent Variables

The selected explanatory variables that are concerned with this study are five main variables. They are variables that have received much emphasis in the past two decades. Since the model used in this study is the stock valuation model and it is difficult to include all the variables that influence stock returns, this study focuses on the most important variables that are included in previous studies. In addition, the availability of the data might be another reason for choosing these five variables. These variables are, Size, book to market ratio, market risk, price earnings ratio, and total debt.

4.3.3.1 Firm Size (MC)

The first independent variable is the firm size. Size is one of the most commonly used variables in investigating the stock returns determinants. Fama and French (1992, 1993, and 1998), Claessens et al. (1995) and Chui and Wei (1998) to name few are among those who used size as an independent variable with stock returns. They used market capitalization as a proxy to size. According to the above mentioned studies and other studies in developed markets, the relationship between market capitalization and stock returns is found to be negative. There are few explanations for the negative relationship between stock returns and size. First, according to Fama and French (1993, 1995, and 1996) size captures an independent source of systematic risk. This is because in the APT model Beta or market risk is not the only variable explaining the variation in the stock returns. Small companies face distress risk and this will cause these small firms to be more risky and therefore yield higher returns. Secondly, Lakonishok et al. (1994) suggested markets are inefficient and investors are not rational in evaluating stocks.

They further elaborated by saying that investors naively extrapolate firms past performance in the future, therefore expect the poor performance to continue in the future and when the opposite happens, heavy investment portfolio occurs in these stocks causing their returns to increase. Thirdly, Perez-Quiros and Timmermann (2000) explained that small firms are affected by tight credit market conditions. They face difficulties in financing their investment portfolios using debt. Therefore, small firms tend to use more expensive financing methods, causing their stocks to be more risky and to yield higher returns. There are in fact other explanations from the opponent of APT or the multi factor model regarding the size and book to market ratio such as data snooping, seasonality (where the effect is dominant in certain periods) and delisting bias. Therefore, from the above discussion it is expected that the a priori sign for size to be negative. Fama and French (1992, 1993, 1996, and 1998), Chui and Wei (1998), Pandey (2001) and Drew and Veeraraghavan (2002) used market capitalization as a proxy for size. They calculated it by multiplying closing price at time t by the number of outstanding shares at time t. In this study, the same method of calculating size of the firm is followed.

4.3.3.2 Book to Market Ratio (BTM or MTB)

Book to market ratio (hereafter BTM) is another variable that is used in the three factor model introduced by Fama and French (1992). Book to market ratio is the division of the book value per share over the market value per share. The relationship between BTM and stock returns is positive indicating that the higher the BM ratio or the greater the numerator (i.e. value firms), the higher the returns, and the opposite are true. The justification for this relationship is equal to those of the size. According to Fama, French (1993, 1995, and 1996), and Liew, and Vassalou (2000) book to market ratio captures an independent source of systematic risk. This is because in the APT model beta or market risk is not the only variable explaining the variation in the stock returns.

> "The relation between book-to-market equity and earnings suggests that relative profitability is the source of a common risk factor in returns that might explain the positive relation between BE/ME and average return" (Fama and French, 1993: 8)

Hence, value companies face distress risk in that their cost will be high forcing them to find highly profitable investment portfolio and this will cause these value or glamour firms to be more risky and therefore yield higher returns. Lakonishok et al. (1994) suggested that markets are inefficient and investors are not rational in evaluating stocks. They further elaborated by saying that investors naively extrapolate firms past performance in the future, therefore expecting the poor performance to continue in the future and when the opposite happens, heavy investment portfolio occurs in these stocks causing their returns to increase. Perez-Quiros and Timmermann (2000) advocated that small firms be affected by tight credit market conditions. They face difficulties in financing their investment portfolio using debt. Therefore, they tend to use more expensive financing methods causing them to be more risky to yield higher returns. Therefore, the expected sign of book to market ratio is positive. Fama and French (1992, 1993, 1996, and 1998), Chui and Wei (1998), Pandey (2001) and Drew and Veeraraghavan (2002) and others, calculated book to market as the division of the book value per share of a firm by the market value per share. This study uses the same technique to obtain book to market ratio.

4.3.3.3 Market Risk (BETA)

Market risk is the third variable used in this study. It measures the systematic risk in any security that cannot be eliminated by diversification. The measurement of this systematic risk is the beta coefficient in the capital Asset pricing model introduced by Sharpe (1964), Linter (1965) and Mossin (1966). The beta coefficient, according to Brigham and Houston (2004: 189), is "a measure of market risk which is the extent to which returns on a given stock move with the stock market." Market risk or beta has received a lot of debate regarding its explanation of the risk faced by investors. Almost all the studies reviewed in the previous chapter included beta as a variable to measure the reaction of stock returns to market risk. The established relationship between market risk and returns is positive. This means that when the market risk is high, the return will be high too. Kim (1997), Fama and French (1993, 1996, and1998), L'Her, et al. (2004) Daniel et al. (2001) and Rouwenhorst (1999) found that beta is positively related to stock returns.

Therefore, the expected sign of Beta is positive. Although some researchers use the market model in estimating beta, others use the Capital Asset Pricing Model. Following Pandey (2001), Drew and Veeraraghavan (2002) drew et al. (2003) and Elfakhani et. al. (1998) methodology in calculating beta, this paper implements the same technique. The CAPM is used to calculate Beta for each year for each firm based on weekly closing prices for all the variables in CAPM. The weekly closing prices of each company in the sample, the weekly closing of FTSE EMAS index, and the weekly KLIBOR (Kuala Lumpur Inter Bank Rate) will be used in estimating beta. The model will be as follows:

$$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + \varepsilon_{it} \qquad 4.30$$

where R_{it} is returns of company i for t, R_{ft} is the KLIBOR for t, R_{mt} is market index returns or FTSE EMAS index returns where i=1,2,3…..150 and t=2000,2001,…2006, ε_{it} is an error term, and α and β are the regression coefficients.

The estimation is done by regressing the individual returns on the market for the first year to get the beta for that year and repeat this process seven times to get the beta for the seven years.

4.3.3.4 Price Earnings Ratio (PER)

Price-earnings ratio (PER) is determined by dividing the closing market price by the company's most recent earnings per share. Investment professionals use PER as a tool to identify "good" investment portfolio opportunities. Analysts and researchers have studied the question of whether high PER periods are followed by lower stock returns. A number of investment portfolio professionals believe that a high PER indicates that the firm has growth opportunities and will translate into high future earnings, whereas another set of investors believe that a low PER indicates undervalued stocks and is a form of sound investment portfolio. Campbell and Shiller have written a series of papers on this topic since the mid 1980s. Campbell and Shiller (1987, 1988) found that future dividends could be forecasted by moving average of earnings. They also found that PER are powerful predictors of long-term stock returns. Campbell and Shiller (1998) tested to see whether the PER revert to their long-term averages. Analyzing historical data, they found that higher PER are followed by lower growth. Campbell and Shiller (1998, 2001) predicted that based on the very high PER, the future stock prices would significantly drop. In their 2001 paper, they concluded that PER and dividend-price ratios are poor predictors of future dividend growth, future earnings growth or productive. Instead, these ratios are good predictors of changes in future stock prices. Fama and French (1989) showed that the dividend yield at the beginning of the period predict a significant proportion of four-year returns, but is not a good predictor of short term return. Park (2000), on the other hand, advised that an investor should not take a high PER by itself as an alarming sign. He found that PER is explained fairly well by future earnings and interest rates and stock markets foresee a distant future of about eight years. Therefore, the PER has little use as a valuation measure.

Fisher and Statman (2000) investigated the relationship between PER, dividend yields and future returns. They concluded that PER and dividend yields are not good indicators of future stock prices, especially when looking at returns over short periods (1-2 years). However, PER and dividend yields provided much better forecasts when they were used to estimate stock returns over longer periods of time (10 years).

4.3.3.5 Total Debt (DEBT)

This variable is used to test whether screened investment portfolios differ from non-screened investment portfolio. Although the variable leverage is used in most of the previous studies, total debt, which is defined as both short and long term debt, is used in this part of the study. Dow Jones and FTSE use debt as one of their benchmarks in the exclusion criteria of stocks. Therefore, the inclusion of this variable might shed some light on whether screened investment portfolios will restrict their debt financing. Pandey (2001) used leverage which consists of total debt and found it to be significant in explaining returns. Spiess et al (1999) found that the higher the debt offerings by a firm, the lower its returns.

Table 4.3 below shows summary of the firm specific determinants used in this part of the thesis and their definitions.

Table 4.3 Summary of the variables

Variable	Variable name	Definition
R	Firms return	Calculated using compounded returns formula
MC	Market capitalization	Closing price X number of common shares
PER	Price earnings ratio	Market value per share / earnings per share
BETA	CAPM market risk	The beta coefficient in CAPM formula
DEBT	Total debt	Long term + short term debt
MTB	Market to book	Market value of the firm / book value of the firm

4.4 Time Series Techniques

4.4.1 Unit Root

In time series data such as the stock prices to do further analysis one must first test the series for non-stationarity problem. Non-stationarity or unit root is a problem where the series mean is not constant over time, the variance is not constant over time and the correlation between a variable and its lags depends on other variables.[34] The problem with non-stationarity series is that it will lead to spurious regression or nonsense regression.[35] Stock prices are usually assumed to follow a random walk process whereby price at time t equal the price at time t-1 plus shock or error term. Maddala (2003) simplified it as follows, if there is a variable Y_t which refer to stock price at time t, Y_t is said to follow a random walk process if

$$Y_t = Y_{t-1} + u_t \qquad 4.6$$

Where u_t is a purely random series with mean μ and variance σ^2. Then if Y_0=zero it means that Y_1=u_1, therefore $Y_2 = Y_1 + u_2$, hence$Y_t = \sum_{i=1}^{t} u_i$, with mean E (Y_t) = tμ and variance var (Y_t) = σ^2, since the mean and variance change with time (t) the process is not stationary.

[34] Studenmund, 2001, pp. 425.
[35] Gujarati, 2003, pp. 806.

However, it will be stationary if it is first difference. This indicates that if this process is followed by stock prices, then prices are following purely random walk process. Following the same model, if there is an increase in u_t by £ atY_t, then $Y_t$and its proceeding periods will increase by £ permanently. Then the change between two periods will equal £*u_t. However, if the model is as follow

$$Y_t = \alpha Y_{t-1} + u_t \qquad 4.7$$

Where α<1, then the effect of the shock will fade away with time. On the other hand, since investors expect their investment portfolio to appreciate over time, the above model seems to be unrealistic. Therefore, the need for a drift arises therefore the following model,

$$Y_t = \alpha + \beta Y_{t-1} + u_t \qquad 4.8$$

Therefore,

$$\Delta Y_t = \alpha + \rho Y_{t-1} + u_t \qquad 4.9$$

Where ρ=β-1

The time series will be stationary if -1 <β<1 this is equivalent to-2< ρ<0. That is, if β equals one, then ρ =zero and therefore, the series is not stationary. However, if β is between 1 and -1, then the series is stationary.

Three random walk models are tested in order to determine whether the unit root problem exist in both indices. The first model is simply the random walk model, while the second is random walk with a drift and a trend, and the last is with a drift only.

They are as follows[36]

[36] Chan et al. 1997, pp.805.

$$Y_t = \alpha Y_{t-1} + u_t \tag{4.10}$$

$$Y_t = \gamma + \beta_{t-N} + \alpha Y_{t-1} + u_t \tag{4.11}$$

$$Y_t = \delta + \alpha Y_{t-1} + u_t \tag{4.12}$$

Where, $Y_t$is the series representing the prices of composite or Syariah index, N is the number of observation, and $u_t$is the error term. The main idea is to test whether the α in each model equal to one as the null hypothesis of unit root. In other words, if α=1, then the null hypothesis cannot be rejected and this will indicate unit root problem or non-stationarity. Consequently, if the null hypothesis is not rejected, this indicates that the series under consideration is following a random walk process in any of the forms mentioned above. Hence, if the series is non-stationary, this means that the market is weakly efficient. Hakim et al. (2003), Chan et al. (1997) and Chan et al. (1992) asserted that if a series is found to be non-stationary, then it is interpreted as a sign of market efficiency, specifically weak form of efficiency. There are few types of tests to investigate the problem of unit root or non-stationarity. This study uses the Augmented Dickey Fuller test (ADF) and Phillips-Perron unit root tests (PP) to ensure the stationarity of the variables. In the ADF, the null and alternative hypothesis of unit root tests can be written as follows,

H_0: X_t there is unit root

H_1: X_t there is no unit root

The unit root hypothesis of the Augmented Dickey Fuller test (ADF) can be rejected if the t-test statistic is less than (lies to the left of) the critical value, meaning that the variable to be estimated is stationary. If the null hypothesis cannot reject, an underlying principle is that the time series has unit root or are non-stationary in the levels. However, it might be stationary in the first differences as asserted in many studies such as Hakim et. al. (2002).

The Phillips-Peron (PP) unit root tests, on the other hand, use a nonparametric statistic method to take care of the serial correlation in the error terms without adding lagged difference terms. The asymptotic distribution of the PP test is the same as the ADF test statistics (Gujarati, 2003 and Vogelvang, 2005). To guarantee that the variables are stationary, both Augmented Dickey Fuller test (ADF) and Phillips-Perron unit root tests (PP) is employed in the study.

ADF test:

$$\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \sum_{k=1}^{k} \beta_k Y_{t-k} + \varepsilon_t \quad 4.13$$

PP test:

$$Y_t = \delta_0 + \delta_1 Y_{t-1} + v_t \quad 4.14$$

where Y_t represents the stock index price for each index, $\varepsilon_t$and v_t are white noise error term and Δ is the first difference operator. The null hypothesis to be tested in both cases are for ADF α_0=0 and for PP β_0=1, which will indicate that there is unit root if not rejected. The series of logged prices is used to ensure robust results and consistency.

4.4.2 Cointegration

One of the issues that many practitioners face in time series analysis is the problem of unit root. Cointegration refers to a linear combination of non-stationarity variables. As mentioned above, if a time series is regressed against another time series without taking into consideration the unit root problem, it leads to spurious regression. Spurious regression will tend to lead to misleading conclusions and may be misguided forecasting. This can be avoided by checking if the residual of the estimated regression is stationary. Put differently, if the residual of two time series regressed appears to be stationary; it means that the series are cointegrated. Therefore, if two variables are cointegrated, they have a long-term relationship, or equilibrium between them. The theory of cointegration developed by Granger (1986) and explained by Engle and Granger (1987) deals with the subject of integrating short term dynamic with long-term relationship[37]. Generally, if two time series are integrated of any degree beyond zero I(q), it indicates that they are drifting together at approximately the same rate. Hence, they are said to be cointegrated. Cointegration can be shown in the following model

$$Y_t = \alpha + \beta X_t + u_t \quad 4.15$$

$$u_t = Y_t - \alpha - \beta X_t \quad 4.16$$

where Y_t and X_t are any two time series variables, β and α are coefficient and u_t is white noise error term, running unit root test it is concluded that both Y_t and X_t are non-stationary and hence they are stationary at the first difference or integrated of degree one I(1). Consequently, unit root test is employed on u_t if it is found stationary or integrated of degree zero or I(0), then it is concluded that although Y_t and X_t are individually integrated of degree 1, their linear combination in model 4.12 is integrated of degree zero. Gujarati (2003) mentioned that equation 4.11 is the cointegrating regression while the coefficient of the independent variable is the cointegrating parameter. Two main types of tests are used to determine the existence of cointegration between variables. The first is Engle-Granger test, where the null hypothesis of unit root of residual will be tested. Engle-Granger test is based

[37] Madala, 2003.

on evaluating whether single equation residual appears to be stationary[38]. If the null hypothesis can be rejected, then the variables are cointegrated. However, Engle-Granger test, according to Koop (2004), suffer from the same symptoms as Dickey-Fuller test, which are low power and misleading if structural breaks occur. The second test is Johansen (1988) maximum likelihood, which is based on Vector Autoregressive (VAR) approach. It commences with a general form of VAR, which is parameterized as a system of error correction mechanism, therefore it will consist of differenced lags as well as level lags of the series. According to Eun et al. (1999), the Johansen cointegration test following augmented Dickey-Fuller test in VAR system will be as follows:

$$Y_t = \alpha + \sum AkYt - k + \varepsilon_t \qquad 4.17$$

Where Y_t is an n*1 vector. By subtracting Y_{t-1} from both sides, the following model will be obtained:

$$\Delta Y_t = \alpha + \sum_{k=1}^{p-1} \Gamma k \Delta Y_{t-k} + \Gamma Y_{t-k} + \varepsilon_t \qquad 4.18$$

Where

$$\Gamma = \sum_{k=1}^{p} A_{k-I} \; , \Gamma_{\mathrm{k}} = -\sum_{\mathrm{j}=1}^{\mathrm{k}} \mathrm{A}_{\mathrm{j-I}} \qquad 4.19$$

The Johansen test of cointegration is focused on the rank of estimated matrix Γ and its characteristic roots. If rank Γ =0, then there is no cointegration or no long-term relationship between variables. However, if Γ = r when r < n, where n represents the number of

[38] Greene, 2000.

variables in the system, there exist r cointegrating vectors. Therefore, this will imply that ΓY_{t-k} is the error correction term, which reflects the equilibrium relationship. Subsequently, if there is only one cointegration equation, it means that there is only one long run equilibrium in the system.

There are five cases on how cointegration is estimated according to Johansen (1995). They are summarized below,[39]

- The level of the data has no deterministic trends and the cointegrating equations do not have intercepts.
- The level of the data has no deterministic trends and the cointegrating equations have intercepts.
- The level of the data has linear trends but the cointegrating equations have only intercepts.
- The level of the data and the cointegrating equations have linear trends
- The level of the data has quadratic trends and the cointegrating equations have linear trends.

In this study, the third case is the most probable, where the series are not trending. Both series are moving simultaneously up and down in a non-predictable manner.

The null hypothesis to be tested is that there is no cointegration between the two series. The rejection of the null hypothesis will depend on the results of Trace statistics and the maximum Eigen value, which will be obtained by E-views along with their critical values. Generally, if the statistics obtained are higher than the critical values, then this will result in

[39] Eviews 4.1 help manual

the rejection of the null hypothesis and subsequently the conclusion of cointegration between series. Granger (1986) asserted that if two series are to be cointegrated, i.e. have long-term relationship, this will indicate that there is market inefficiency since the error term of one series can be used to predict the movement of the other. Granger (1986), Chan et al. (1997) and Chan et al. (1992) support this. Nevertheless, if the series are found to be not exhibiting long-term relationship then the market is said to be efficient. To perform the Johansen cointegration test a lag length needs to be specified prior to it. The criterion that is used in this test is used in the Granger causality in the following section.

4.4.3 Vector Error Correction Model (VECM):

The issue of long-term relationship or equilibrium led to the application of the error correction mechanism, which is discussed here. Cointegration necessitates coefficient restrictions in a VAR model. The restrictions are necessary to ensure that the variables are cointegrated of degree one CI (1, 1) also guarantees that an error-correction model exists. The Granger representation theorem states that for any set of I (1) variables, error correction and cointegration are equivalent representations. As indicated earlier, the cointegration of time series will suggest that the series will be trending together. The long-term multiplier can be determined by regressing the series in the level form. Nonetheless, the short-term multiplier will be determined by the Error correction mechanism given that the series are cointegrated. To elaborate further, let us take the following example where there are two cointegrated series, namely Y and X which will be represented below,

$$\Delta Y_t = \alpha + \beta \Delta X_t + \lambda e_{t-1} + u_t \qquad 4.20$$

Where ΔY_t ΔX_t are any cointegrated time series in differenced form, e_{t-1} is the error obtained from regressing Y_t and X_t in level form, and u_t is the disturbance error and $\alpha\beta$ *and* λ, are the coefficients.

It is known in the basic regression that in the above model a change in X_t causes Y_t to change. However, in this case ΔY depends on changes in e_{t-1}. The change in e_{t-1} or the equilibrium error will be the correction process to get ΔY back to the equilibrium in the current period. Put differently, if λ e_{t-1} is negative, so will be Y_{t-1}, which is above equilibrium of $\alpha + \beta X_{t-1}$. Therefore, the change in Y_{t-1} starts falling in the next period and the corrected equilibrium is achieved. The opposite is also true. If λe_{t-1} is positive, which leads Y_{t-1} to rise in the next period to correct the error and be in equilibrium again. This explanation is applicable if there is a valid model to be tested, which is not the case here. Here two indices may be trending together and causing each other from both directions. Thus, testing them against each other in what is known to be Vector Error Correction Model or VECM. The idea behind it is to run both cointegrated variables on each other plus the error correction term in both cases. The models that is tested are as follows,

$$\Delta X_t = \alpha_0 + \alpha_1 L\Delta X_{t-1} + \alpha_2 L\Delta KLSI_{t-1} + \alpha_3 ec_{t-1} + \varepsilon_{1t} \qquad 4.21$$

$$\Delta KLSI_t = \beta_0 + \beta_1 L\Delta KLSI_{t-1} + \beta_2 L\Delta X_{t-1} + \beta_3 ec_{t-1} + \varepsilon_{2t} \qquad 4.22$$

where L is the lag operator ($LX = X_{t-1}$) , KLSI is the Syariah index prices in the natural log form, the X is all the explanatory variables in the model in natural log form, ec_{t-1} is error-correction, ε_t is disturbance terms and Δ denotes first-differences required to induce stationarity for the corresponding variables. These models shed light on the Granger causality of each variable on the other. Hence, here the causality is investigated in addition to the short-term correction mechanism. Through the error correction mechanism, alternative channels of causality will unfold. “It could be exposed by the statistically significant coefficient, first, the error correcting term included in the VECM, second, the joint significance of the lags of each explanatory variable, and third, joint test of all terms indicated previously” (Yusof, 2003:59). Hence, if the VECM is to be applied, then the clearer causality between variables will appear. When variables are cointegrated, in the short run, deviations from the long-term equilibrium feeds back on the changes in the dependent variable in order to force movements towards long-term equilibrium. That is, if the model has a statistically significant error-correcting coefficient, it means it is responding to this feedback. However, if this is not true, then the index as a dependent variable is responding to short-term shocks in the system.

4.4.4 Granger Causality

Granger Causality test was introduced by C.W. Granger in his article back in 1969 in Econometrica, which aimed to answer the question of what causes what using Vector Autoregressive VAR. The term VAR originated due to the appearance of the lagged values of the dependent variable on the right hand side of the equation. This is similar to the equation in Granger causality estimation where both X and KLSI appear on the right hand side of its own equation. “According to Sims, if there is true simultaneity among a set of variables they should be all treated on an equal footing; there should not be any a priori distinction between endogenous and exogenous variables” (Gujarati, 2003: 848). If both X and KLSI are found to be causing each other, this type of situation is perfect for the application of VAR. Below is a replication of the Granger causality model that is the same as the VAR model to be used in this study. Therefore, the following set of equations are tested

$$Y_t = \sum_{i=1}^{n} \alpha_i X_{t-i} + \sum_{j=1}^{n} \beta_j Y_{t-j} + u_{1t} \quad 4.23$$

$$X_t = \sum_{i=1}^{n} \lambda_i X_{t-i} + \sum_{j=1}^{n} \delta_j Y_{t-j} + u_{2t} \quad 4.24$$

where the X and Y are any two related variables, t is the time factor, u_{1t} and u_{2t} are the error term, which are not correlated $\alpha\beta\lambda \; and \; \delta$ and are coefficients. The model regresses each variable on its lag and other variables. In this study the model of Granger causality will be as follows,

$$KLSI_t = \sum_{i=1}^{n} \alpha_i X_{t-i} + \sum_{j=1}^{n} \beta_j KLSI_{t-j} + u_{1t} \quad 4.25$$

$$X_t = \sum_{i=1}^{n} \lambda_i X_{t-i} + \sum_{j=1}^{n} \delta_j KLSI_{t-j} + u_{2t} \quad 4.26$$

where X is the explanatory variables in the system i.e. IP, oil prices, CPI, M1, and KLCI, t is the time factor, u_{1t} and u_{2t} are the error term that are not correlated and $\alpha, \beta, \lambda \; and \; \delta$ are coefficients.

The causation could be unidirectional from KLSI to X if $\delta_j$is statistically significant but $\alpha_i$is not significant, or from X to KLSI if λ_i is statistically significant but $\beta_j$is not significant. On the other hand, if δ_j and α_i are statistically significant, it is feedback or bilateral causality. However, if none of the coefficients is significant, it indicates that there is no causation in the system. The hypotheses to be tested are that KLSI does not cause X or X does not cause KLSI. Lag requirement is needed to perform the regressions. The lag choice is chosen using Schwartz Information Criteria. The rejection of the null hypotheses indicates that there is causation from one variable to the other.

On the other hand, when the variables are found to be cointegrated the same process is applied but with adding the long-term equilibrium. The test used for such causality is the Wald test that is based on the restriction of the coefficients of the variables causing the other variable plus a restriction of the long-term equilibrium.

Therefore, the equation will be as follows,

$$KLSI_t = \sum_{i=1}^{n} \alpha_i X_{t-i} + \sum_{j=1}^{n} \beta_j KLSI_{t-j} + \vartheta e_{t-i} + u_{1t} \qquad 4.27$$

$$X_t = \sum_{i=1}^{n} \lambda_i X_{t-i} + \sum_{j=1}^{n} \delta_j KLSI_{t-j} + + \vartheta e_{t-i} + u_{2t} \qquad 4.28$$

where X is the explanatory variables in the system i.e. IP, oil prices, CPI and M1, t is the time factor, u_{1t} and u_{2t} are the error term that are not correlated and α, β, λ and δ are coefficients., KLSI is Kuala Lumpur Syariah Index and u_i is the error term are as above and e_{ti} is the cointegration equation of the system..

Since this part of the study is comparing both indices, the same equation testing the causality between macroeconomic variables and KLSI is applied to KLCI. Therefore, the following equation is used,

$$KLCI_t = \sum_{i=1}^{n} \alpha_i X_{t-i} + \sum_{j=1}^{n} \beta_j KLCI_{t-j} + \vartheta e_{t-i} + u_{1t} \qquad 4.29$$

$$X_t = \sum_{i=1}^{n} \lambda_i X_{t-i} + \sum_{j=1}^{n} \delta_j KLCI_{t-j} + +\vartheta e_{t-i} + u_{2t} \qquad 4.30$$

where X is the explanatory variables in the system i.e. IP, oil prices, CPI and M1, t is the time factor, u_{1t} and u_{2t} are the error term that are not correlated and α, β, λ *and* δ are coefficients, KLSI is Kuala Lumpur Syariah Index and e_{ti} is the cointegration equation of the system.

Accordingly, if the causality of one of the variables is to be tested, its coefficient and the coefficient of the cointegration equation are included in the restriction. Therefore if the restriction result is not significant it means that there is no causality and the opposite is true. On other words, it is testing the null hypothesis of no causation against when there is causation for that specific variable. The test statistics will use chi-square rather than F distribution since chi-square is asymptotic

4.4.5 Impulse Responses and Variance Decomposition

Ibrahim (2001) and Ibrahim et. al (2001) asserted that the Granger causality F-test only captures the direct linkage between two variables, omitting the influences of other variables in the system that might affect the variable at the same time. Put differently, a shock to the one variable not only directly affects itself, but it is also transmitted to all of the other endogenous variables through the dynamic lag structure of the VECM system of equation. An impulse response function traces the effect of a one-time shock to one of the innovations on current and future values of the endogenous variables. Impulse responses are used to confirm the direction of causality and forecast the future impact of one variable on the other. Therefore, the test will be done to check using generalized impulse to avoid the difference of responses due to different ordering.

While impulse response functions trace the direction of the causation of a shock to one endogenous variable on to the other variables in the VECM, variance decomposition separates the variation in an endogenous variable into the component shocks to the VECM. Thus, the variance decomposition provides information about the relative importance of each random innovation in affecting the variables in the VECM system. Thus, both techniques capture the direct and indirect influence of innovations on the variables measured and permit complete evaluation of its dynamics linkages and interactions with other variables. However, in the case of variance decomposition the ordering influences the results.

According to Ibrahim (2001) and Ibrahim et. al. (2001), the ordering in the variance decomposition should be starting with the goods market followed by money market and ending with the asset market. Nonetheless, Ibrahim (2003), in investigating the integration as well as the dynamics between macroeconomic variables and international stock markets suggested that the ordering should be different. He suggested that international asset prices (the US and Japan) comes in first followed by money supply inflation, industrial production, exchange rate and lastly by the domestic asset market. He justified that by indicating that international stock market (i.e. the US and Japan) are the most exogenous to the Malaysian market and that the domestic stock market is the least endogenous. Masih et. al. (1996), in determining the dynamic relation in macroeconomic activities for Malaysia and Thailand, explained that variables that respond most to current fluctuations should be placed the last in ordering. Masih et. al. (1995) and Ibrahim (2001) found that in case of Malaysia the money supply is the most exogenous variable. Therefore following Ibrahim (2001) and Ibrahim et al. (2001) industrial production is included as first in the ordering followed by oil prices, Money supply, inflation, KLSI. The second ordering is the same except that KLSI is replaced with KLCI.

4.5 Panel Data Techniques

Studies written on this kind of issue are vast and voluminous. However, using panel data and testing two submarkets, to the author's knowledge, are very few, if any. This study examines the influence of certain market factors based on previous literature on companies' returns in the Bursa Malaysia. The study chooses mainly two types of markets: screened and non-screened. The trading is done in the same market, but the criteria for listing in the screened index are the main difference, which might cause differences in the factors influencing companies. The period covered is from 2000 to 2006 because the Syariah index, which represent the screened index, has been established in 1999. The reason behind using panel data to estimate the determinants is because the individual or company's reaction to these selected firm specific variables rather than the whole market is important to this study.

Baltagi (2001) discussed the advantages of panel data. The first benefit is that since panel data is mixed between cross sectional and time series, it allows us to investigate the individual heterogeneity of the cross section, or, in other words, the failure to account for unobserved factors. That is, heterogeneity takes into consideration the effect of any omitted variables in the model. Secondly, panel data provide more observations, different individual cross section, more variability, less collinearity and most information about the units of studies. For example, income of different individuals, states and companies for an interval of time is more informative than a single individual, company and state over time (i.e. time series), or many individuals, companies, and states at one point of time (i.e. cross section). Thirdly, panel data allow us to better study dynamics of adjustment. Since it mixes both time series and cross section, it can help us study the long term adjustment of different units. For example, the prediction of the unemployment spells over from one period to the other. Fourthly, panel data can determine inter and intra-unit effect. Fifthly, panel data allow researchers to test highly complicated behavioral models. Finally, panel data deal with micro units, which allow for more accuracy and minimal bias. When dealing with micro data it is without doubt that panel data are one of the best methods.

Panel data, also called longitudinal data or time series cross-sectional data (TSCS), are data where multiple cases observed over two or more periods. While it is possible to use ordinary multiple regression techniques on panel data, they may not be optimal. The estimates of coefficients from ordinary regression may be subject to omitted variable bias. This problem arises when there is some unknown variable or a variable that cannot be controlled and that affects the dependent variable. In other words, there is an unobserved effect on the dependent variable that is not accounted for in the ordinary regression. In ordinary estimation, whether it is cross sectional or time series, whatever is not observed or omitted is considered to be included in the error or disturbance term of the model. However, in panel data this is not the case, because both cross section and time are varying.

This means that whatever is not observed or omitted could be varying over time or over individual units or both. Therefore, this unobserved effect or latent variable has to be treated with caution in order to find the best estimate to fit the model. Wooldridge (2002) pointed out that in methodological as well as application discussion about panel data, the argument is whether the unobserved effect is be treated as a fixed or random effect. That is, whether the unobserved component is viewed as random variable or a parameter to be estimated. In other words, it is necessary to know that the latent component is a random variable that has no correlation with the observed explanatory variables in the model (random effect), or it is correlated to the explanatory variable (fixed effect). Therefore, the fixed effect will show the unique unit, whether a firm or an individual, specific effect while the random effect shows the common effect. The unit specific deviation is reflected in the error component. Therefore, in the realm of panel data realm there are two main methods of estimation, fixed effect, and random effect.

In the cross sectional area the model will be as given below

$$Y_i = \beta X_i + c + u_i, \qquad \mathrm{i} = 1, 2, \ldots\ldots \mathrm{N} \qquad 4.32$$

where Y is the dependent variable, X is the vector of independent variables, c is the unobserved variable or the intercept, u is the error term and B is the coefficient for X. The Ordinary Least Square (OLS) estimator can estimate this model, which is Best Linear Unbiased Estimator (BLUE) if all the classical assumptions are fulfilled. In this case, the time is fixed but the units are varying.

Moreover the time series model will be as given below

$$Y_t = \beta X_t + c + u_t, \qquad \mathrm{t} = 1, 2, \ldots\ldots \mathrm{T} \qquad 4.33$$

Again, the OLS estimator will be BLUE if all the classical assumptions are fulfilled. In this model, the time is varying while the unit or cross section is fixed.

Therefore, in both cross section and time series either time or unit is held fixed to estimate the model. However, in the panel or longitudinal data both time and cross section are varying causing the pooled OLS estimator to be biased because it assumes the intercept and the slope of all cross section across time to be the identical. Gujarati (2003) stated that assuming intercept and slope to be identical is a restricted assumption and it might distort the true relationship between dependent and independent variables.

The pooled model of panel data is as given below

$$Y_{it} = \beta X_{it} + c + u_{it}, \qquad \mathrm{t} = 1, 2 \ldots \mathrm{T} \text{ and } \mathrm{i} = 1, 2 \ldots \mathrm{N} \tag{4.34}$$

Where

$$v_{it} = c_i + u_{it} \tag{4.35}$$

This estimation will be consistent if there is no correlation between the independent variables and the error term and the unobserved variables. That is

$$E(X_{it}u_{it}) = 0 \tag{4.36}$$

and

$$E(X_{it}c_{it}) = 0 \tag{4.37}$$

Wooldridge (2002) pointed out that even if the above assumption about the absence of correlation exists, the error term (v_{it}) or the composite error will be serially correlated due to the presence of c_i in each time period. He added that the inference using pooled OLS entails that the variance matrix and test statistics are robust as the error is serially correlated. Therefore, although pooled OLS is an easier way to estimate panel data, it is not robust and might be inconsistent and inefficient.

4.5.1 Fixed Effect

The general fixed effect model is given below:

$$Y_{it} = \beta X_{it} + v_{it}, \qquad t = 1, 2 \ldots T \text{ and } i = 1, 2 \ldots N \qquad 4.38$$

$$v_{it} = c_i + u_{it} \qquad 4.39$$

where y is the dependent variable, X is the vector of independent variables, c is the unobserved variable, u is the error term which is independent and identically distributed IID $(0,\sigma^2)$ and B is the coefficient for X.

Baltagi (2001) said that the fixed effect method assumes that the c_i is a fixed parameter that can be estimated for each firm and the u_{it} is IID with zero mean and constant variance. In addition, the X_{it} are assumed to be independent of u_{it} $E(X_{it}u_{it}) = 0$ for all t and i while $E(X_{it}c_{it}) \neq 0$. Wooldridge (2002) pointed out that the idea of fixed effect estimator of B is to transform the equations in order to eliminate the unobserved effect c_i. The transformation is obtained by first averaging equation 4.31 over time to find the cross equation

$$\overline{Y_\iota} = \overline{X_\iota}\beta + c_i + \overline{u_\iota} \qquad 4.40$$

and then subtracting equation 4.31 from equation 4.35

$$Y_{it} - \bar{Y}_i = (X_{it} - \bar{X}_i)\beta + u_{it} + \bar{u}_i \qquad 4.41$$

With c_i eliminated, pooled OLS can be applied to estimate equation 4.35, which is the term fixed effect or within estimator.

The fixed effect allows the correlation between the observed explanatory variable and the unobserved variables. According to Gujarati (2003), there are five possibilities of varying mean value for each unit. However, this study is focusing on one main possibility, namely when the intercept is varying across individual firms. The reason for this is that the time interval is short i.e. only 7 years. If this model is run regardless of the unobserved variable, then pooled model or the restricted model is estimated. The idea of LSDV is to include a variable for each firm or unit in the model. The estimated model will have m-1 dummy variables, where m is the number of the firms or N. Therefore, the result of the model will show how each individual firm differs from the others by the difference between each firm effect and the main intercept.

The advantage of the fixed effects (FE) specification according to Hsiao (2007) is that it allows the Individual, as well as time specific effects to be correlated with independent variables. In addition, it does not require an investigation in the correlation pattern. The disadvantage of this specification, however, is twofold. First, the number of unknown parameters increases with the number of sample observations. Second, fixed effect estimator does not allow the estimation of the coefficients that are time-invariant. Therefore, the second method that is the random effect estimator is explained.

4.5.2 Random Effect

The main problem with the fixed effect according to Baltagi (2001) and Gujarati (2003) is that there are too many parameters estimated which will jeopardize the degree of freedom.

Baltagi (2001) indicated that this problem can be avoided by assuming that c_i is random. That is, c_i is IID (0, σ^2), u_i is IID (0, σ^2_v), and c_i is independent of the u_i. Moreover, the independent variables X_{it} is independent from c_i and u_i. The random effect model is suitable if the individuals or cross sections are randomly drawn from a larger population. Therefore, the sample should be representative of the population.

The random effect model will be as given below

$$Y_{it} = \beta X_{it} + c_i + u_{it}, \qquad t = 1, 2 \ldots T \text{ and } i = 1, 2 \ldots N \qquad 4.42$$

Where Y is the dependent variable, X is the vector of independent variables, c is the unobserved variable or the intercept, u is the error term, and β is the coefficient for X. Now instead of treating c_i as fixed it assumed that it is a randomly drawn variable with mean of c. Hence, it can be expressed as

$$c_i = c + \varepsilon_i \qquad 4.43$$

Where ε_i is IID (0,σ^2_ε). This equation suggests that the cross section units are drawn from a larger population where they have a common intercept value = c and any individual differences between units is totally reflected in the error term ε_i.

Substituting 3.12 in 3.11 gives us

$$Y_{it} = \beta X_{it} + c + \varepsilon_i + u_{it} \qquad 4.44$$

Or

$$Y_{it} = \beta X_{it} + c + \varpi_{it} \qquad 4.45$$

The composite error term ϖ_{it} is comprised of the two error components; one is $\varepsilon_{i,}$ which is individual specific and the other is, $u_{it,}$ which is a combination of time series and cross section. Gujarati (2003) summarized the main assumptions that random effect model is built on as:

$\varepsilon_i \sim N(0,\sigma^2_\varepsilon)$

$u_{it} \sim N(0,\sigma^2_u)$

$E(\varepsilon_i u_{it}) = 0 \quad E(\varepsilon_i \varepsilon_j) = 0 \ (i \neq j)$

$E(u_{it} u_{is}) = E(u_{it} u_{jt}) = E(uit\ ujs) = 0 \ (i \neq j, t \neq s)$

That is, all individual error components are not correlated with each other and not correlated across both cross section and time series.

Hsiao (2007) discussed the advantages of random effect. First, the number of parameters remains constant when N increases. Second, it allows the derivation of efficient estimators that make use of both within and between (group) variations. Third, it allows the estimation of the impact of time-invariant variables.

4.5.3 Hausman's Specification Test

Dougherty (2006) indicated it that the choice of the model to best fit the data has to be done in a proper way. One of the ways to decide which estimator should be used is to turn to Hausman's specification test. Baltagi (2001) indicated that given the assumptions of the random effect model where there is no correlation between the independent variables and the error term, it is understood that if this assumption is not fulfilled, the coefficients estimated by random effect will be biased and inconsistent. However, the fixed effect of within estimator eliminates the ci as mentioned earlier. Therefore, the fixed effect estimator of the coefficients is considered unbiased and consistent. The Hausman test, is used to test the null hypothesis whether the correlation between the independent variables and the error term is zero or the random effect model is true or E (u_{it} X_{it}) =0. In other words, the null hypothesis is that the fixed effect and random effect do not differ substantially. Put differently, the fixed effect is consistent under both the null and the alternative while random effect is consistent under the null and inconsistent under the alternative. Therefore, rejection of the null hypothesis indicates that the fixed effect estimator is consistent and unbiased.

4.5.4 Unit Root

One of the main problems in time series data are the non-stationarity of the variables. Since panel data are composed of both time series and cross sectional data, the problem of non-stationarity might be persistent in the panel data. Therefore, it is necessary to test the variables for the unit root problem. Unit root will cause the estimation to be biased. Recent literature suggests that panel-based unit root tests have higher power than unit root tests based on individual time series. One unit root tests is used that is the Levin, Lin and Chu (2002).While this test is commonly termed panel unit root tests, theoretically, they are simply multiple-series unit root test that have been applied to panel data structures.

Levin, Lin and Chu (LLC) test assumes that there is a common unit root process, which is identical across cross-sections. The idea of common unit root is that $\delta_i = \delta$ for all of i. “this means that the coefficient of the lagged dependent variable is assumed to be homogenous across all cross-section units in panel” (Baltagi 2001: 236). Levin, Lin, and Chu (2002) developed a panel data unit root test where they test the null hypothesis whether there is a unit root against the alternative where there is no unit root. There are three main models to be tested. They are as follow, Model 1 which does not consist of any intercept or time trend

$$\Delta Y_{it} = \delta Y_{it-1} + \varepsilon_{it} \qquad 4.46$$

Model 2 which include an intercept

$$\Delta Y_{it} = \alpha_{0i} + \delta Y_{it-1} + \varepsilon_{it} \qquad 4.47$$

Model 3 which include intercept and time trend

$$\Delta Y_{it} = \alpha_{0i} + \delta Y_{it-1} + \alpha_{1,it} + \varepsilon_{it} \qquad 4.48$$

where $-2< \delta \leq 0$ for i=1, 2 …N, and ε_{it} is IID $(0, \sigma^2_{\varepsilon})$.

The first model tests the null hypothesis whether the coefficient of y_{it-1} is significant or not. In other words, H_0: $\delta = 0$, against the alternative H_a: $\delta < 0$. The second model test the null hypothesis whether H0: $\delta =0$ and $\alpha_{0i} =0$ However, in the last model the test procedure evaluates the null hypothesis whether H0: $\delta =0$ and $\alpha_{1i} =0$.

4.6 Conclusion

In short, two main methodologies are used in answering the research questions. They are time series techniques and panel data techniques. The first part addresses the difference in returns using risk and returns analysis as well as time series analysis. The second part addresses the difference in returns and whether returns are affected by the same macroeconomic variables. Again, the use of the time series techniques used in the first part are implemented to answer the questions of this part. The last part of the thesis addresses the impact of firm specific variables on the returns of screened and non-screened firms. In this part panel data techniques are used to answer the questions of this part. The next chapter discusses the development of hypotheses for each part separately.