CHAPTER 1:

INTRODUCTION

1.1 Background of Study

Time series analysis involves a various method that compromised a researcher's intention in analyzing a sequence of data or observation that recorded over a range of time. As parts of the most importance usability in time series analysis, the modelling and forecasting process can be applied to predict a future situation based on past and current data. One of the assumptions that has to be satisfied in modelling and forecasting process is variance consistency among the observations (Adhikari and Agrawal, 2013). The presence of any forms of disturbance in a data series may interrupt the variance consistency and thus the modelling and forecasting a time series data becomes incorrect. Outliers and structural breaks are examples of disturbance that usually happened in a time series data.

Outliers, include an astonishing spike or dip, are unusual observations that significantly differs from the other observations. There are two types of outliers that are most commonly known namely additive outliers and innovative outliers (Fox, 1972, Denby and Martin, 1979). Additive outlier affects at a single observation while innovative outlier affects all the afterwards observations. For instance, if autocorrelation and partial autocorrelation function is used in modelling an Autoregressive-Moving Average (ARMA) models, the presence of additive or innovative outliers cause the ARMA model to be deviated (Deutsch, Richards and Swain, 1990). Hence, the prediction of the future observations using this deviated model will lead to an erroneous.

On the other hand, structural break include level shift or level breaks occurred when a variable in time series data suddenly changes at a time interval and it usually occurred in the mean or variance of a data series (Sharma, Swayne and Obimbo, 2016). For instance, at a period of time, a structural break changes from a zero mean to a non-zero mean data generating process (DGP). It causes links between variables involved to be changed or shifted as well. If a structural break occurred in a DGP with zero mean, the forecasting of the future observations based on zero mean will not be accurate. (Castle and Hendry, 2019) described that a structural break does not only happen in the mean and variance, but also in distributional and slopes of the time series data. This research will investigate the presence of structural breaks in the mean of a time series data.

In financial time series and econometrics, which is the focus of this research, outliers and breaks are example of economic shocks. Their presence may represent an economic situation such as economic boom, financial crash, sudden changes in policies or a data processing error has occurred. Most economic time series data are not stationary with no pattern and difficult to identify any outliers and structural breaks. For example, in a large daily historical data of stock index that recorded over the past 13 years (see Figure B.5 and Figure B.6), it is quite impossible to identify the exact date of the outliers and structural breaks occurrence. Hence, there is a need to define an impressive and competent method for detecting the outliers and structural breaks in a financial time series data in order to investigate the causes, effects, and plan for the recovery actions including policies. In fact, the outliers and structural breaks need to be identified and analyse for any treatments in order to have a valid conclusion and interpretation about the data modelling and forecasting.

This research attempts to investigate the performance of indicator saturation as one of the latest approaches in outliers and structural breaks detection literature. The indicator saturation approach takes each observation in the data series to be the dummy variables collectively, and their significance are being analyzed through the regressors significance testing process. This implies that the indicator saturation adopted zero-one indicators for the observations in a data series and these dummy variables will act as additional variables to the model. The main part of this approach is that, it utilizes model selection based on the general-to-specific modelling technique. Through general-to-specific technique, a general model called General Unrestricted Model (GUM), with full set of saturation indicators together with another independent variables are used in the beginning before it is simplified to the most satisfactory model that compromised the research framework. Simplification here means that producing the most meaningful model that satisfies some criteria such as having the highest adjusted R-squared value, the lowest mean square error and the smallest information criterion value. Besides, a meaningful model only has the most significant independent variables and all other insignificant variables are eliminated. On a side note, overfitting and multicollinearity might occur when too many variables included in a GUM (Goggin, 1986, Zhang, 2014) causes the retained significant variable to be biased. In this research, stationary first order autoregressive, AR(1) DGP will be used in the simulations. The reason of using this model is that, it is well known that AR(1) model depends linearly on the previous values. Thus, it is suitable with the financial time series as the model can be used to predict future values based on the past values. For example, this model helps to forecast a market stock's future price based on its previous performance. Second, as stationary AR(1) depends on the AR(1) parameter, we would like to investigate if there are any effects to the performance of indicator saturation when varying these parameter. To test this approach in a real data, it will be tested to identify the presence of outliers and structural breaks in Islamic indices and their conventional counterparts indices.

1.2 Autoregressive Model - An Overview

It is well known that AR(p) processes consist of an observation, y_t that is obtained and regressed against the previous observation (i.e. $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$) from the same time series (Adhikari and Agrawal, 2013). Then, AR(p) DGP will be:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \epsilon_t \tag{1.1}$$

where β_0 is an intercept, β_p is the AR(p) parameter, and ϵ_t is the randomness term. It is clear that p is the number of previous observations in the series that are used to forecast the value of the present time. p is also known as the order of an autoregressive process. Thus, the first-order autoregressive model, AR(1) take the form of:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t \tag{1.2}$$

Most statistical forecasting requires a time series data is in stationary form. This means, the data properties such as mean, variance, autocorrelation need to be constant over time. Consequently, stationary AR(1) process depends on the AR(1) parameter, β_1. AR(1) is stationary when $|\beta_1| < 1$ (Nason, 2006).

In R, the selection of an autoregressive model can be done using **auto.arima** function (Hyndman et al., 2019). The model selections are based on the information criterion values which are Akaike Information Criterion (AIC) (Akaike, 1973), corrected Akaike Information Criterion (AICc) (Sugiura, 1978) and Bayesian Information Criterion (BIC) (Schwarz et al., 1978). BIC has the greater penalty imposed to penalized the model compared to AIC and AICc and thus BIC is trying to find the true model among the the set of candidates. As this research trying to find a true AR(1) model among the autoregressive models, this research uses BIC to estimate the AR(1) model of the real data.

Back to the time series fundamentals, there are two types of time series that commonly known in the literature namely univariate and multivariate time series. The latter is when the series consist of more than one variable while the former is when the series have only one variable. Equation (1.2) is example of univariate AR(1) time series. For multivariate time series, vector autoregressive model of order one VAR(1) is used. The VAR(1) process with k regressors will be:

$$\begin{aligned}
y_{t,1} &= \beta_1 + \beta_{11} y_{t-1,1} + \beta_{12} y_{t-1,2} + \beta_{13} y_{t-1,3} + \ldots + \beta_{1k} y_{t-1,k} + \epsilon_{t,1} \\
y_{t,2} &= \beta_2 + \beta_{21} y_{t-1,1} + \beta_{22} y_{t-1,2} + \beta_{23} y_{t-1,3} + \ldots + \beta_{2k} y_{t-1,k} + \epsilon_{t,2} \\
y_{t,3} &= \beta_3 + \beta_{31} y_{t-1,1} + \beta_{32} y_{t-1,2} + \beta_{33} y_{t-1,3} + \ldots + \beta_{3k} y_{t-1,k} + \epsilon_{t,3} \\
\vdots \quad & \vdots \quad \vdots \qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad \vdots \\
y_{t,k} &= \beta_k + \beta_{k1} y_{t-1,1} + \beta_{k2} y_{t-1,2} + \beta_{k3} y_{t-1,3} + \ldots + \beta_{kk} y_{t-1,k} + \epsilon_{t,k}
\end{aligned} \tag{1.3}$$

where $t = 1, \ldots, T$ and ϵ is the randomness term.

1.3 Background of Case Study

An Islamic index, also known as Shariah Compliant Index is an index that complies the Shariah screening criteria stated by the Islamic scholars or institutions. Any financial activities that practicing interest (or *riba*), gambling, production or trade of non-*halal* goods and services such as alcohol, pork, and pornography are among the non-permissible criteria listed in Shariah screening criteria, by means, the market index from these activities cannot be included in order to consider an index is listed as Islamic index (El Khamlichi, Sarkar, Arouri and Teulon, 2014). Meanwhile, conventional indices do not require to follow the Shariah screening criteria. They can have a mixture of non-permissible and permissible criteria. Although the Islamic capital market products and services has been enormously grown over the past few decades, but the studies on the Islamic indices is still not widely covered. In fact, the performance of indicator saturation approach using Islamic indices as the empirical evidence has not been investigated yet.

The global financial crisis named Great Recession (Duignan, 2019) which occurred from 2007 to 2009, was the most significant worldwide economic shocks. It was the longest economic recession and affecting developed and developing countries. This economic downturn firstly affects the economies of the United States and European developed country. The effect spread around the world including developed economies in Asia like Japan and including developing economies in Southern/Southeastern Asia such as Malaysia, Indonesia and India. In general, developed economy is one of the characteristics of a developed country with high per capita income, high level of industrialization, high level of human development index, and etc. Meanwhile, developing economy is come from developing countries with low or medium per capita income, level of industrialization, level of human development index, and etc. In this research, the performance of indicator saturation is applied to the Islamic indices and its conventional indices from developed and developing economies.

According to United Nation (Nation, 2019), Malaysia, Indonesia and India are among the listed developing economies, while Japan and European are among the listed developed economies. The name of the indices used for investigation in this research are:

1. Malaysia Islamic index: FTSE Bursa Malaysia EMAS Shariah Index
2. Malaysia Conventional index: FTSE Bursa Malaysia EMAS Index
3. Indonesia Islamic index: Jakarta Islamic Index
4. Indonesia Conventional index: Jakarta Composite Index
5. India Islamic index: Nifty 50 Shariah Index
6. India Conventional index: Nifty 50 Index

7. Japan Islamic index: S&P Japan 500 Shariah Index
8. Japan Conventional index: S&P Japan 500 Index
9. European Islamic index: S&P Europe 350 Shariah Index
10. European Conventional index: S&P Europe 350 Index

All the data above were collected from Datastream spans daily from January 03, 2007 to June 28, 2019, with $T = 3068$. The daily market prices are shown in Figure B.5 and Figure B.6. However, for the analysis, this thesis used the return values in each data that were calculated from the log first difference in percentage of daily market price. The formula used is:

$$\mathbf{r}_t = \ln \frac{P_t}{P_{t-1}} \times 100, \tag{1.4}$$

where $\mathbf{r}_t$ is the return stock series with time recorded, t, P_t is the stock price at time t (i.e. the current stock price) and P_{t-1} is the stock price at time $t-1$ (i.e. the day-before stock price). The return price of indices used in this research has been approximately modelled by AR(1) model using **auto.arima** function in R under null hypothesis of no outliers and structural breaks present.

1.4 Scope and Limitation

The focus of this research is to manifest the indicator saturation approach in detecting the time location of outliers and structural breaks, that presence in autoregressive model of order one, AR(1). Hence, the main concern is *outlier and structural break dates detection*, so the effect of this approach to the modelling and forecasting; and treatment of the outliers will not be discussed in this research. This research will look into different perspective as compared to the existing literature because it uses R programming language instead of $OxMetrics$.

The performance of indicator saturation approach in R is examined in two modes which is in Monte Carlo simulations and in the real data sets. In the Monte Carlo simulations, the impulse indicators are used to identify the presence of an additive outlier and multiple outliers separately. Next, step indicators are used to identify the presence of a single as well as multiple shifts. As this research focus on the stationary AR(1) model, the AR(1) parameter, $|\beta_1|$ is varied from 0.1 to 0.9, together with the variation of significance level, α.

Stock market indices from different type of economies are used to inspect the performance of indicator saturation in the real data. The real data is modelled by AR(1) model based on null hypothesis of no presence of outliers and structural breaks. Since each stock market indices have different base dates, so the starting date for the research investigation is set at January 03, 2007 and limit until June 28, 2019. The main concern of using the real data is for *outlier and structural break dates detection that matched to the global financial crisis period and other economy-affected events*. This research will not cover the performance comparison between the indices including advance statistical analysis. In this research, the return values, $\mathbf{r}_t$ of the stock price are calculated and will be used to achieve the research' objectives, instead of the original price of the stock price. The outliers and structural break dates detection in the real data that obtained from R, and will be compared with the results from $OxMetrics$.

There are some limitations to this study. First, the observations number in the simulation experiments are less associate to the real date experiments as the largest observations number in the simulation experiment is six time less than the observation number in the real data. This is due to computer performance capacity (i.e. central processing unit, CPU) that took a long period of time of simulation experiments in order to conduct a big observation number. The larger the observation number, the longer the CPU will take to finish the simulation experiments. Undoubtedly, the longer time will be consumed for a larger number of simulations. However, the largest observation number used for Monte Carlo simulations are larger than other observation number that has been used by the previous researchers (see Table 2.3). Second limitation is that the currency recorded in daily price of the stock indices are measured in the local currency unit. This means, the currency across the stock indices are not the same and this could lead to a bias in a comparison part. The reason of using the local currency is that it reflects a real overview of a stock index in a country.

1.5 Research Objectives

Generally, the utmost objective of this research is to fill some gaps in the detection of outliers and structural breaks literature using indicator saturation approach and find some added values to this approach, by providing perceptive analysis and impressive data interpretation.

Specifically, the objectives for this research are:

- To determine the optimum significant level, effect of variation in AR(1) parameter and other properties required for saturation indicator to work well in detecting outliers and structural breaks in the stationary AR(1) model.

- To prove that the breaks detection mechanism proposed by (Doornik, Hendry and Pretis, 2013) also works using impulse indicator saturation in detecting outliers, and the mechanism also works for different form of step indicator saturation structure.

- To investigate the performance of indicator saturation approach using **gets** package in R as comparison to **Autometrics** in $OxMetrics$.

- To investigate whether the impulse and step indicator saturation capable to detect any outliers and structural breaks corresponds to the global financial crisis in Islamic and conventional counterpart indices within the considered time span.

- To determine the relationship between the Islamic indices and their conventional counterparts indices in developed and developing economies during the crisis period.

1.6 Organization of the Thesis

The direction of this thesis are as follows:

- Chapter 1: This chapter introduces the preliminaries and overviews about this research. It includes, preamble of the research, research objectives, backgrounds, and scopes and limitations.

- Chapter 2: This chapter is about the literature review on the indicator saturation field. It covers the previous studies that has been done before and the studies that related to the outliers and structural breaks detection literature.

- Chapter 3: This chapter discuss in details the methodology used for this research. It includes how indicator saturation works in general-to-specific modelling in detecting outliers and structural breaks. Moreover, it also covers the properties of the impulse and step indicator saturation in AR(1) model.

- Chapter 4: Analysis and Findings; is about the results from the Monte Carlo simulations regarding the performance of impulse indicator saturation in detecting outliers and step indicator saturation in detecting structural breaks through the concept of potency and gauge. Five cases are considered here which is the case of no presence any outliers and structural breaks, the case with the presence of single and multiple additive outliers, and the case with the presence of single and multiple structural breaks. Then, this chapter also reported the findings from the real data.

- Chapter 5: This chapter provides conclusion for this thesis and future research that can be performed to improve this research.

CHAPTER 2:

LITERATURE REVIEW

The literature of detecting outliers and structural breaks approach has been expanded from manual method to mechanical method. Manual method is conducted by inspecting the location of an outlier or a structural break through the time series plot (Clarke, Coladarci and Minium, 1999). A classical approach of manual method is observing any outliers using box plot during data exploratory process. These approaches are not practical when the data have irregular and complicated patterns, and when there are large number of observations with many independent variables. Mechanical method that will be applied throughout this research, requires certain techniques and frameworks that may works effectively and efficiently.

2.1 Outliers in Time Series

In a time series data, the presence of outliers is a common event. Before a modelling and prediction process is carried out, a data cleaning has to be done. One of the data cleaning elements is outliers detection. There has been a variety definition of outliers in the literature. Here are a few definitions that are commonly used:

- (Hawkins, 1980) defines an outlier as *an observation that deviates so much from the other observations as to arouse suspicion that it was generated by a different mechanism*

- (Johnson and Wichern, 2002) introduces the term outlier as *an observation that appear to be not consistent with the remainder of the same data set.*

- (Barnett and Lewis, 1994) suggests that an outlier is *an observation that appears to deviate markedly from other members of the sample in which is occurs.*

Even though there are different definitions of the outliers, but there are common words used that described the behaviour of the outliers which is "*deviates so much*", "*not consistent*", "*deviate markedly*". Thus, it has come to the understanding that *an outlier is an observation that are significantly different or diverge from the majority of the other observations in a data series.* In financial time series, these outliers indicate the effect of economic, political or financial events occurred in an economy that is highly unlikely to occur again (Franses and Van Dijk, 2000). (Fox, 1972) proposed that the presence of the outliers may lead to bias parameter estimation. This causing the modelling a financial time series data to end up with error-forecasting the future situation and consequently affecting the policy that has to be made regarding the financial events occurred (i.e. the causes of outliers). Therefore, it is crucial to identify the presence of the outliers in a financial time series.

2.1.1 Outlier Detection Methods

(Fox, 1972) introduces two types of outliers from autoregressive (AR) models namely additive outliers and innovative outliers. The former affects at a single observation while the latter affects all the afterwards observations. There has been a several available approaches to identify the presence of outliers in an autoregressive-based time series data.

(a) Method based on likelihood ratio test: (Chang, Tiao and Chen, 1988) and (Tsay, 1988) presented an iterative approach for outlier detection in ARMA and autoregressive integrated moving average (ARIMA) models. It is an extended version of likelihood ratio approach proposed by (Fox, 1972). The test is used to detect and identify the type of the outliers which is either it is an additive or innovative outlier. The idea of this method is to use maximum of the likelihood ratio statistics and standardized estimated errors in the observation that suggested to be an outlier.

(b) Method based on observation influence: This method uses an influence function to measure the *small change* effect of the data distribution. See (Tolvi, 1998) for an example of the influence function. The *small change* is including to assume some of the observations to be missing values and eliminate them during the estimation process. Another example is to add some weight to the observations that suspected to be the outliers. (Abraham and Chuang, 1993) uses this method to detect outlier in AR models by introducing independent and identical distributed of Bernoulli random variables as weight to the outlier.

(c) Method based on Gibbs sampling: (McCulloch and Tsay, 1994) introduced the use of Gibbs sampler in estimating AR parameter and in detecting the presence of outliers. (Justel, Peña and Tsay, 2001) then extend this study to detect outlier in patches form. This method requires conditional posterior distribution for each parameter including the outlier's magnitude in the AR process given the rest of other parameters. Then a probability is carried out based of the prior and posterior probability.

The detection of outliers using these approaches maybe unsuccessful especially when there are multiple outliers detected due to masking effects which means an outlier is undetected under the large occurrence of outliers. These approaches also tend to misspecify "good" data points as outliers. This phenomenon called swamping or smearing effect (see (Justel et al., 2001)). In addition, these approaches require a correct distribution and thus a wrong distribution chosen could lead to misclassified an outlier.

2.2 Structural Breaks in Time Series

Literally, structural breaks in a time series can be easily understood. (Castle and Hendry, 2019) proposed a formal definition of a structural break as a sudden shifting in the behaviour of a variable over time. The breaks in behaviour of a variable includes changes in distribution and parameter of a data series such as the mean and variance. (Hansen, 2001) mentioned that a structural break is occurred when at least one of the parameters - mean, variance and trend over time - of the data series has changed at a period of time. (Tsay, 1988) proposed additional types of outliers to the study introduced by (Fox, 1972) which are level shifts in means, variance changes and transient changes. However, following (Castle and Hendry, 2019), level shifts and variance changes are examples of structural breaks occurred in a time series and will be used in this research. Variance and transient changes will not be discussed in this research.

In financial time series, the structural break effects are more investigated compared to the outlier effect. This is due to the fact that the structural breaks in a financial time series indicate the consequences from an economic or a non-economic event that took over a period of time, instead of at a single time location. Examples of economic events that causes a structural break happens are emerging markets and integration of world equity markets, changes in exchange rate system and the introduction of a single currency in Europe (see (Andreou and Ghysels, 2009)); meanwhile examples of non-economic events are war, natural disaster, and political turnmoil (see (Castle and Hendry, 2019)).

The effects of the structural breaks in term of statistical analysis are parallel with the effect of the outliers, in which distortions in parameter estimation and forecasting. In economics, unaccounted the presence of the structural breaks cause the economic relationship goes astray, erroneous in forecasts and misinformation in policy suggestions (Hansen, 2001). (Pettenuzzo and Timmermann, 2011) showed that including structural breaks analysis in financial time series can improves long-horizon forecasts. (Boot and Pick, 2020) suggests that a forecast performance will enhance better by taking account the modelling of the structural breaks based on mean square forecast error loss. Hence, structural breaks detection has to be an important element in financial time series analysis.

2.2.1 Structural Break Detection Methods

There has been a several commonly used approaches to identify the structural break dates in a financial time series data:

(a) (Chow, 1960) introduced an econometric test of equality in linear regression's coefficients of two different data samples that are split at a predetermined break point. This shows that a structural break date needs to be known exogenously in order to test whether it is a true structural break date or not.

(b) Quandt (Quandt, 1960) introduced Quandt Likelihood Ratio (QLR) test which is an extension to Chow test. This test tries to avoid the exogenously select a break date by measuring the Chow test at all breakpoints. The most significant Chow test statistics from all breakpoints are selected as the Quandt statistics as it is highly likely to be a breakpoint.

(c) (Bai and Perron, 1998, 2003) developed the structural breaks' estimation as global minimizers of the sum of squared residual. This test has the capability of detecting multiple structural break dates at unknown timing.

Table 2.1 shows the advantages and disadvantages of the existing structural break detection methods mentioned above.

Table 2.1: Advantages and disadvantages of the structural break detection methods

Structural breaks detection method	Advantages	Disadvantages
Chow Test	Useful to identify a single known break date. Easy to implement because it shows F-statistics and standard distribution for comparison of the two data sets (sub-periods).	Less power to detect multiple break dates A break date needs to be known exogenously in order to work with the Chi distribution
Quadnt Likelihood Ratio Test	Useful to identify an unknown break date. Easy to detect the unknown break date from the QLR statistic plot.	High computationally cost and time consuming More extensive distribution required that depends on variable numbers and trimming data range used.
Bai and Perron Test	Useful to identify unknown as well as multiple break dates. Available in most statistical software like R	Causes severe size distortions in persistent series. (see (Prodan, 2008))

2.3 Identification of Both Outlier and Structural Break Dates

Since the effects of outliers are almost parallel to the effects of structural breaks, this has triggered researchers to carry out only one element which is either outlier detection or structural break detection in their research. Apparently, the research of both outlier and structural break dates detection in financial time series is limited. For instance, (Chatzikonstanti, 2017) studied the outliers and structural breaks analysis in the United States stock market and (Halari, Tantisantiwong, Power and Helliar, 2015) identified the presence of outliers and structural breaks in Karachi Stock Exchange.

In econometrics, the detection of both outliers and structural breaks in a financial time series is crucial in order to investigate the association between short-lived effects and long-lasting effects. In addition, (Andreou and Ghysels, 2002) and (Rodrigues and Rubia, 2006) proved that a distortion in a structural break test used in their research could arise in the presence of the extreme observations or outliers in the data series. Hence, there is a need to define an approach that can detect both outlier and structural break simultaneously.

2.4 Indicator Saturation Approach

Indicator saturation approach is one of the latest approaches in outliers and structural breaks detection literature. It is still new in the econometrics and require more studies on its performance.

This approach originally defined by (Hendry, 1999) in investigating the presence of outliers in the United States annual expenditure from the period of 1931 to 1989 using impulse dummy. Consequently, based on the research by (Hendry, 1999), (Hendry, Santos and Johansen, 2008) and (Johansen and Nielsen, 2009) had proposed the null distribution of the impulse indicator saturation approach in detecting outliers. The extension of this impulse indicator has resulted in introducing the step indicator saturation by (Doornik, Hendry and Pretis, 2013) to detect structural breaks in time series data. The analysis includes the detection of single and multiple breaks in simulations and in real data. Then, (Castle, Doornik, Hendry and Pretis, 2015) explain further this step indicator by comparing it with an alternative of structural break method called lasso and least angle regression.

These primary researches study the development of indicator saturation approach which introduce the impulse indicator in detection outliers and step indicator in detecting structural breaks. From these primary researches, we can summarise that the indicator saturation approach applies zero-one dummy variables correspond to each observation's time location in a data series. Next, by utilizing general-to-specific modelling scheme, these dummy variables which act as additional variables will be regressed for significance testing starting from a general model until a simplified model is formed. To avoid the lack of degree of freedom from general-to-specific modelling as observation number is greater than variable number, block splitting of the variables is used.

Table 2.2: The general properties of impulse and step indicator

Indicator Saturation	Function	Dummy Variable	Description
Impulse	Ourlier date detection	$I_{tj} = \{1_{\{t=j\}}\}$	The dummy variable takes value 1 if $t = j$, and zero otherwise
Step	Structural break date detection	$S_{tj} = \{1_{\{t \leq j\}}\}$	The dummy variable takes value 1 if $t \leq j$, and zero otherwise

Table 2.2 shows the preliminaries of impulse and step indicator saturation approaches described from the primary researches. The details on the properties of impulse and step indicator saturation and how it works will be discussed in Chapter 3. There is another type of saturation indicator called trend indicator saturation to detect the trend component of structural time series data (see (Pretis, Mann and Kaufmann, 2015)), but it will not be included in this study.

2.5 Existing Empirical Studies on Indicator Saturation Approach

The performance of indicator saturation approach has been investigated in various empirical studies. There are difference in research properties related to indicator saturation approach used in the existing empirical researches such as field of the real data, type of model used in data generating process, the uses of impulse, step or both indicators, the computational software used, the structure of step indicator, and the observation number used in simulations and real data. Table 2.3 shows the difference between the available existing researches in indicator saturation approach.

Table 2.3: Difference in existing empirical research

Article	Real Data Field	Model Used	Indicator Saturation	Structure of Step Indicator	Software Used	Observation Number
(Santos and Hendry, 2006)	-	AR(1)	Impulse	-	Autometrics	T={100,200,300}
(Hendry and Mizon, 2011)	US food expenditure	Static and dynamic model with 2 lags	Impulse	-	Autometrics	T = 71
(Ericsson and Reisman, 2012)	Financial Crisis Data	Vector autoregressive model	Impulse	-	Autometrics	T = 97
(Castle et al., 2012)	US Interest rate	Static Model	Impulse	-	Autometrics	T = 100
(Doornik et al., 2013)	US interest rate	Static Model	Step	$S_{tj} = \{1_{\{t \leq j\}}\}$	Autometrics	T = 250
(Marczak and Proietti, 2016)	European industrial production series	Structural Time Series	Impulse and Step	$S_{tj} = \{1_{\{t \leq j\}}\}$	Autometrics	T = 277
(Pretis et al., 2016)	Climate Temperature	Static and AR(1) model	Step	$S_{tj} = \{1_{\{t \leq j\}}\}$	Autometrics	T = 115*
(Frydman and Stillwagon, 2016)	Stock market	Dynamic model with 2 lags	Impulse and Step	$S_{tj} = \{1_{\{t \leq j\}}\}$	Autometrics	*survey data*

* T = 115 is used in 10 subsamples

All of these existing empirical researches use **Autometrics** to study the performance of indicator saturation, however, the performance of indicator saturation using another computation software like **gets** in R has not been scrutinized yet (as of the time of writing this thesis). In fact, only (Pretis et al., 2016) did mention about the time processing taken by **Autometrics** to detect outliers and structural breaks using indicator saturation. In their study, **Autometrics** took almost 5 minutes to complete one replication of $T = 1150$ (from 10 subsamples of $T = 115$).

To fill the gaps in the indicator saturation literature, the study on indicator saturation approach in detecting outliers and structural breaks is extended further in this research by using a dynamic process which is stationary AR(1) from financial time series data. In contrast to (Santos and Hendry, 2006), this research includes real data investigation and implement **gets** in R to detect outliers and structural breaks. The hopes of this research is it could give added values to the literature as it implement **gets** in R in which, has never been used by any researches (to the best of our knowledge at the time of writing this thesis). Besides, while the existing empirical researches apply the same structure of step indicators which is $S_{tj} = \{1_{\{t \leq j\}}\}$, this research apply the different structure of step indicator which is $S_{tj} = \{1_{\{t \geq j\}}\}$.

2.6 Computational software: *R* and *OxMetrics*

The advancement in computer technology has given rise to changes in statistical methodology, which can be portrayed as revolutionary (Hand, 2015). **Autometrics**, a module from $OxMetrics$ programming (Doornik, 2009*b*) has been used by previous researchers, in investigating general-to-specific modelling and the indicator saturation approach theoretically and empirically. In 2018, as the indicator saturation literature grows, (Sucarrat, 2018) introduced **gets** package in R programming (R Core Team, 2014), focussed on automatic model selection based on general-to-specific modelling and indicator saturation method to detect outliers, breaks and trends in a data series. This package is an extension to the package **AutoSEARCH** (Sucarrat, 2015a).

Table 2.4: General properties of **gets** and **Autometrics**

	gets in R	**Autometrics** in $OxMetrics$
Modus Operandi	Use coding language	Point-and-click programme
Operating system	Open source	Close source
Fees	Free	Subscription fees

Table 2.4 shows the summary general properties of **gets** in R and **Autometrics** in $OxMetrics$. The details in performance of **gets** and **Autometrics** in detecting outliers and structural breaks will be discussed throughout Chapter 3 and Chapter 4.

2.7 Outliers and Structural Breaks in Islamic Stock Market Index

The studies on the relationships between an Islamic index and its conventional counterpart index has not been widely covered. In fact, the comparison studies between both indices on the issues of the presence of outliers and structural breaks is still at a minuscule level. This means, due to its shariah compliance restrictions, there is not much evidence whether the Islamic indices robust to the global financial crisis or any economy-affected events. The current available studies only cover either outliers detection or structural breaks detection in Islamic and/or conventional indices.

Table 2.5: Studies on outliers and structural breaks in Islamic stock indices

Article	Outlier Detection Method	Structural Break Detection Method	Data Used
(Wahid, 2014)	-	Chow Test	Islamic Stock Index of 7 Countries
(Aloui, Hammoudeh and ben Hamida, 2015)	-	Bai and Perron Test	Islamic Indices of Gulf Cooperation Countries
(Majdoub, Mansour and Jouini, 2016)	-	Bai and Perron Test	Both Islamic and Conventional of Dow Jones Indices
(Abu Bakar, 2019)	Box-and-Whisker plot	-	Malaysia Islamic Stock Price
(Rejeb and Arfaoui, 2019)	-	Bai and Perron Test	Both Islamic and Conventional of Various Countries

Table 2.5 shows the available researches that study the outliers and structural breaks in the Islamic only or with its conventional counterpart indices. The investigation of handling the outliers detection together with the structural breaks detection in both Islamic and conventional indices using indicator saturation approach has not been discovered yet (at the time of writing).

This research does not aim to find whether the Islamic indices outperform the conventional index, or vice versa. However, by following (Rejeb and Arfaoui, 2019), this research studies the relationship between Islamic indices and conventional indices during financial crisis period. This study also includes the relationship between Islamic index in developed and developing economies during crisis period.

2.8 Summary

In this chapter, the definition of outliers and structural breaks and currently available methods to detect these disturbance in time series has been discussed. In addition, this chapter reviewed on the indicator saturation approach in detecting outliers and structural breaks in various fields with different properties used. Besides, this chapter also reviewed the method of detecting of outliers and structural breaks used in the Islamic and its conventional stock indices. The next chapter will discuss in detail on the indicator saturation approach in detecting outliers and structural breaks.

CHAPTER 3:

METHODOLOGY

This chapter discuss the properties of both impulse and step indicator saturation, automatic model selection based on the general-to-specific modelling, and the performance analysis of the indicator saturation approach. From here-on, impulse indicator saturation will be abbreviated as IIS and step indicator saturation will be abbreviated as SIS.

3.1 Indicator Saturation in General-to-Specific Modelling

General-to-specific modelling consist of four main aspects of modelling a time series process, which are: regressors significance testing, backward elimination, diagnostic testing and information criteria (Pretis, Reade and Sucarrat, 2018). This type of modelling process commences with a general model that includes all the regressors, without a prior knowledge or assumption either the regressors are significance or insignificance to the response variable that are being investigated. Ordinary least square (OLS) estimation method is used for parameters estimation during the regressors significance testing. Subsequently, backward elimination plays role in regressors significance testing, by removing the least significant regressors one at a time, based on the considered significance level, α. From this strategy, a general model is simplified into a more adequate model that fit to the research's objective framework (Campos, Ericsson and Hendry, 2005). This methodology is used when applying saturation indicator (IIS and