Indicator Saturation in Autoregressive Model using gets in R: A Computational Simulation and Empirical Evidence in Shariah Compliant Index

M Azim M Nasir\(^1\) and Mohd Tahir Ismail\(^1\)

\(^1\)School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia.

azimnasir11@gmail.com

Abstract. Structural changes is very important in analysing an economic situation such as an economic boom or recession. There has been a great demand to introduce an efficient method of detecting structural changes in a data series. One of the most recent method called step indicator saturation, uses the extension of general-to-specific modelling framework to detect the presence of structural changes. We believed from the Monte Carlo simulation results that step indicator saturation approach can also be applied to dynamic model: Autoregressive of Order One, AR(1). gets package in R provides an alternative software to Autometrics and it is outperform Autometrics in term of time processing. We apply this method to detect structural changes presence in Malaysia Shariah compliant index market and its conventional counterpart index. The results shows that, there are interdependence between Shariah compliant index and its conventional counterpart. The retained step indicators collide with global financial crisis, so this implied that Shariah compliant index does not immunized against the recession crisis.

1. Introduction

Structural changes are example of abrupt alterations and shocks in a time series data. In finance, their presence may represent an economic situation such as financial crash, sudden changes in policies or there are a data processing error has occurred. Therefore, there is a demand to define an impressive method to detect the occurrence of structural changes in order to investigate the causes, effects, and plan for the recovery actions including policies.

In early 21\(^{st}\) century, Hendry [1] and Hendry, Johansen and Santos [2], introduced impulse indicator saturation to detect outliers in a time series data. Their studies involving a full and complete set of impulse indicator: \(1_{(t=j)} = 1\), to a model. This approach is then expanded by Johansen and Nielsen [3], Doornik, Hendry and Pretis [4] and Castle, Doornik and Hendry [5], introducing step indicator saturation to detect structural changes in a time series data. Step indicator saturation approach takes each observation in a data series to be the zero-one dummy variables collectively, and their significance are being tested in regressors significance testing process. Specifically, they add step indicators of the form: \(1_{(t \leq j)} = 1\), to a model. The main part of this approach is that, it utilizes model selection based on the general-to-specific modelling technique in which, a general model is used in the beginning before it is simplified to the most satisfactory model that compromised the research framework.

Autometrics, a module from Ox programming [6] has been used by previous researchers [1-5,7-13]
in their research on investigating general-to-specific modelling and the indicator saturation approach. However, performance of step indicator in R programming has not been investigated yet. In 2018, as the indicator saturation literature grows, Sucarrat [14] introduced getts package in R programming, focussed on automatic model selection based on general-to-specific modelling and indicator saturation method. While previous researchers in this literature use stationary static regression model in their studies, Marczark and Proietti [15] use structural time series model, and Santos and Hendry [16] study on impulse indicator saturation (only) in autoregressive model. At the time of writing, adding step indicator saturation to the dynamic model: autoregressive model has not been studied yet.

The most important goals of the this paper is to fill some gaps in the structural changes detection literature and find some added values to this approach, by providing perceptive analysis and impressive data interpretation. To achieve these, we will apply step indicator saturation in dynamic model: stationary autoregressive model of order one, AR(1) and the investigation utilizes gets package in R. What makes this paper differs to the existence studies is that, we apply step indicator of the form: \( \sum_{t=j}^{T} \frac{t}{T} = 1 \) to the AR(1) model. Monte Carlo simulations will be used to analyze the performance of the step indicator saturation in this model. The investigation proceed with real data of the considered model which is Malaysia Shariah Compliant Index and its conventional index namely; FTSE Bursa Malaysia EMAS Shariah Index and FTSE Bursa Malaysia EMAS Index respectively.

The direction of this paper are as follows: Section 2 is about structural changes detection by step indicator saturation in details, Section 3 is about Monte Carlo simulations to investigate the performance of step indicator saturation, Section 4 is applying step indicator saturation to the real data and Section 5 conclude.

2. Structural Changes Detection by Step Indicator Saturation

Step indicator saturation (SIS) works as additional variables to the model, which is for one particular independent variable, a SIS is defined to each observation. Thus, for univariate AR(1) data series with T observations, T number of SIS will be added to the model. Specifically, value 1 is define at a period of time, from \( t = j \) until \( t = T \), and zeroes otherwise. These SIS will act as regressors and will go through a significance testing in general-to-specific modelling process. However, this process rises a problem of more regressors than number of observations in which lead to degree of freedom deficiency.

To overcome this problem, Hendry, Johansen and Santos [2] proposed block-splitting estimation method based on general-to-specific modelling in indicator saturation literature. Through this process, SIS are divided into \( m \) partitions of \( w_a \) regressors such that \( \sum_{a=1}^{m} w_a = N \). The key point of the partition splitting process is, it makes the number of regressors in each partition will always less than the number of observations, \( w_a \ll T \). To illustrate the process, assume \( m = 2 \). The process commence with the SIS in the first partition are added to the model, selects and records time location of the significance SIS, and then repeat the process with SIS in the second partition; finally combine all the significance SIS from both partitions and re-selects and records time location of the retained SIS. The selection of a retained significant indicator is made based on absolute value of t-statistics estimator for that indicator is greater than a critical value, \( |t_a| > c_a \), where \( c_a \) is the t-distribution critical value of the chosen significance level \( \alpha \). For the case of no structural changes presence in a time series data, on average, there are \( \alpha T \) saturation indicator will be retained by chance after the combination and re-estimation stage. These \( \alpha T \) SIS are irrelevant regressors that inadvertently retained. By setting small value of \( \alpha \), for example, \( \alpha \leq 1/N \), the number of SIS that are mistakenly retained by chance can be controlled.

Selecting the significant indicators in one-cut split that has been discussed so far is following non-sequential selection process. An alternative approach of selection process called sequential selection where an insignificance indicator is eliminated one at a time starts from the least significant indicator in each partition, and the process continues until only the significance ones are retained. According to Prentis, Schneider, Smerdon and Hendry [17], to lower the estimators’ variance of the indicators due to multicollinearity and non-orthogonality, sequential selection from multiple splits that allowing multi-path searches of model estimation are suggested to be implemented.
3. Monte Carlo Simulations

Monte Carlo simulation will be used in order to investigate the performance of SIS in detecting structural changes in univariate AR(1) model. The data generating process (DGP) of univariate AR(1) with SIS is:

\[ y_t = \rho y_{t-1} + \sum_{j=1}^{m} \delta_j y_{t-j} + \varepsilon_t \quad j=1, \ldots, T, \quad d=1, \ldots, m \]  

where, \( \varepsilon_t \sim \text{N}[0, \sigma^2] \), \( m \) is number of partitions and \( \delta \) is SIS coefficient. Following Markzak and Castle [15], the average retention frequency of SIS, \( \tilde{f}_j \) in Monte Carlo simulations is:

\[ \tilde{f}_j = \frac{1}{M} \sum_{r=1}^{M} \left[ \delta_{jr}^* \neq 0 \right], \quad j=1, \ldots, T, \]  

where \( M \) is the number of Monte Carlo simulations. If a SIS is statistically significant, the element in the square brackets is true, then, \( 1[\cdot] \) takes the value 1, and 0 otherwise. If the location of the level shift are deterministic, the number of relevant indicator, \( n \) is known and hence the total number of retained irrelevant indicators is: \( Tn \). Based on the retention frequency in (2), the performance of SIS will be analysed using potency and gauge. Potency is the average retention frequency of significant indicator(s) that is(are) relevant to the model:

\[ \text{Potency} = \frac{1}{n} \sum_{j} \tilde{f}_j, \quad j \in R_n \]  

where \( R_n \) is the set of time indices correspond to the relevant indicators. Meanwhile, gauge is the average retention frequency of significant indicator(s) but irrelevant to the model:

\[ \text{Gauge} = \frac{1}{T-n} \sum_{j} \tilde{f}_j, \quad j \in R_{T,n} \]  

where \( R_{T,n} \) is the set of time indices correspond to irrelevant indicators.

The first part of the Monte Carlo simulation investigated the case of no structural changes occurs in AR(1) data series. By using R language and gets package, the experiment apply \( M = 1000 \) simulations, number of observations, \( T = 500 \), significance level, \( \alpha = \{0.001, 0.01, 0.05\} \), and coefficient of AR(1), \( \rho = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\} \). From the results in Table 1, it seems that sequential selection or multi-path approach is essential in reducing the retention frequency of irrelevant indicators close to the significant level \( \alpha \). From \( \rho = 0.1 \) to \( \rho = 0.9 \), the gauge values using multi-path approach clustered around the significant level, \( \alpha \), except for significance level \( \alpha = 0.05 \). Therefore, multi-path approach with \( \alpha \leq 0.01 \) is the best practice.

**Table 1.** Retention Rate of SIS for No Structural Breaks.
Second part of the Monte Carlo experiment investigated the performance of SIS in detecting the true structural change, using both potency and gauge analysis. Consider a level shift occurs in a data series with magnitude \( \lambda \) is equal to 4 times the standard deviation of the error term \( \varepsilon \) and apply 

\[
\begin{align*}
    y_t &= \begin{cases} 
    \rho y_{t-1} + \varepsilon_t & t < j \\
    \rho y_{t-1} + \lambda + \varepsilon_t & t \geq j
    \end{cases}
\end{align*}
\]

(5)

The level shift occurs in a period of time series, such as in the last \( r = 20\% \) of the series where \( r = (T - j)/T = 0.2 \) [16]. Table 2 shows the results of the experiment when a level shift starts at \( j = 400 \) using \( T = 500 \) and significance level \( \alpha = 0.01 \).

### Table 2. Retention Frequency of SIS in the presence of a level shift.

| \( \rho \) | \( \alpha = 0.001 \) | \( \alpha = 0.01 \) | \( \alpha = 0.05 \) |
| --- | --- | --- | --- |
| 1-cut Multi Path | 1-cut Multi Path | 1-cut Multi Path | 1-cut Multi Path |
| 0.1 | 0.00139 | 0.00090 | 0.02290 | 0.00824 | 0.12854 | 0.05804 |
| 0.2 | 0.00145 | 0.00091 | 0.02369 | 0.00829 | 0.12950 | 0.05827 |
| 0.3 | 0.00156 | 0.00094 | 0.02477 | 0.00833 | 0.13065 | 0.05835 |
| 0.4 | 0.00163 | 0.00105 | 0.02582 | 0.00841 | 0.13256 | 0.06009 |
| 0.5 | 0.00179 | 0.00109 | 0.02661 | 0.00844 | 0.13499 | 0.06027 |
| 0.6 | 0.00181 | 0.00110 | 0.02775 | 0.00936 | 0.13508 | 0.06093 |
| 0.7 | 0.00195 | 0.00112 | 0.02871 | 0.01017 | 0.13265 | 0.06136 |
| 0.8 | 0.00216 | 0.00114 | 0.02986 | 0.01043 | 0.13282 | 0.06209 |
| 0.9 | 0.00223 | 0.00143 | 0.03115 | 0.01063 | 0.13225 | 0.06253 |

From Table 2 there are not much difference in average potency between one-cut and multi-path approaches; around 75% to 93%. However, the performance of multi-path selection is outperform one-cut selection can be seen in the average gauge values: average gauge values in multi-path approach much smaller than one-cut approach. In fact, most of the gauge values in one-cut approach are greater than the significance level \( \alpha \). Thus, multi-path approach is crucial in detecting structural change by SIS in order to eliminate any irrelevant indicators that are inadvertently retained. On the other hand, from both selection methods, as AR(1) coefficient increases, average gauge of SIS increases, which means, more irrelevant indicators has been retained by chance as the serial correlation increases.

### 4. Applications to Real Data

The performance of SIS in detecting structural changes will be investigated in real data: FTSE Bursa Malaysia EMAS Shariah Index and FTSE Bursa Malaysia EMAS Index. These data has been approximately modelled by AR(1) model based upon the Bayesian Information Criterions, using function auto.arima in R. The data are collected from Datastream spans daily from January 03, 2007 to June 28, 2019 and using the return values that are calculated from the log difference of daily market price.

As the leading research shows Autometrics has been performing well in detecting structural changes using indicator saturation, this experiments are repeated using Autometrics in OxMetrics. The aims are
to confirm the structural changes found by gets package in R collides with the retained indicators found by Autometrics.

Table 3 recorded the number of structural changes and occurrence dates from both data at significance level $\alpha = 1/T$, where $T = 3068$. The experiment uses multi-path selection as simulation proved that multi-path is crucial to reduce the irrelevant indicators.

Table 3. Retention Frequency of SIS in the presence of a levelshift Index

| Index       | EMAS Shariah | EMAS |
|-------------|--------------|------|
| Number of Breaks | 46           | 46   |
| Dates of Break   |              |      |
| 27-Feb-2007, 05-Mar-2007, 06-Mar-2007, 09-Mar-2007, 06-Aug-2007, 07-Aug-2007, 15-Aug-2007, 20-Aug-2007, 21-Aug-2007, 22-Aug-2007, 24-Aug-2007, 30-Nov-2007, 04-Dec-2007, 16-Jan-2008, 17-Jan-2008, 18-Jan-2008, 24-Jan-2008, 28-Jan-2008, 10-Mar-2008, 11-Mar-2008, 13-Mar-2008, 18-Mar-2008, 27-Feb-2007, 05-Mar-2007, 06-Mar-2007, 09-Mar-2007, 06-Aug-2007, 07-Aug-2007, 14-Mar-2007, 15-Mar-2007, 06-Aug-2007, 07-Aug-2007, 15-Aug-2007, 20-Aug-2007, 21-Aug-2007, 22-Aug-2007, 24-Aug-2007, 16-Jan-2008, 17-Jan-2008, 18-Jan-2008, 24-Jan-2008, 28-Jan-2008, 10-Mar-2008, 11-Mar-2008, 13-Mar-2008, 18-Mar-2008, 06-Jun-2008, 09-Jun-2008, 19-Sep-2008, 22-Sep-2008, 19-Sep-2008, 22-Sep-2008, 10-Oct-2008, 13-Oct-2008, 13-Oct-2008, 15-Oct-2008, 20-Oct-2008, 20-Oct-2008, 22-Oct-2008, 30-Oct-2008, 30-Oct-2008, 4-Nov-2008, 04-Nov-2008, 02-Jan-2009, 02-Jan-2009, 06-Jan-2009, 06-Jan-2009, 06-Jan-2009, 31-Mar-2009, 05-Aug-2011, 10-Aug-2011, 22-Sep-2011, 27-Sep-2011, 27-Sep-2011, 10-Aug-2011, 22-Sep-2011, 06-May-2013, 07-May-2013, 01-Dec-2014, 02-Dec-2014, 06-May-2013, 07-May-2013, 04-Aug-2015, 25-Aug-2015, 28-May-2018, 31-May-2018 |

The results show that the number of breaks found in shariah compliant is equal to its conventional counterpart. Most of the dates of breaks found in both data are similar with majority were detected between the period 2007-2009. This is the period of global financial crisis. Figure 1 and Figure 2 show coefficient paths of the breaks found from EMAS Shariah index and EMAS index respectively; which indicates the direction of the structural break. The figures also show similar pattern between Shariah compliant index and its conventional counterparts.

Table A1 recorded the retained SIS from gets and compare them to the ones obtained from Autometrics, using the same significance level $\alpha = 1/T$. The main outcomes from this comparison is that, almost 98% of the retained SIS found in gets also found in Autometrics. This means, gets in R can be an alternative programming to Autometrics in detecting structural changes using SIS. Indeed, for $T = 3068$, the computation process by gets is much faster than Autometrics.
The results from the real data claims that there are interdependence between Shariah compliant index and its convention counterpart in Malaysia. As most of the retained SIS collide with the global financial crisis period, we claim that the Shariah compliant index in Malaysia does not immunized against the economic recession.

5. Conclusion
This paper has studied the structural changes detection by step indicator saturation in stationary Autoregressive order one, AR(1). Simulation and application results proved that this method can be applied to the AR(1) model. Multi-path selection is essential in reducing the irrelevant indicators and significance level $\alpha_{0.01}$ is the best practice to get gauge values close the significance level $\alpha$. Even though the magnitude of autocorrelation parameter, $\rho$ might cause some deviations, however, the deviations are always very small. Besides, structural changes detection by SIS using gets in R is outmatch the Autometrics in term of time processing when large $T$ is used.

As structural changes detection is important in financial time series data, the experiment apply FTSE Bursa Malaysia EMAS Shariah Index and FTSE Bursa Malaysia EMAS Index to investigate the performance of SIS in real data. Both data has been modelled by AR(1). Results from the real data shows that there are interdependent between Shariah compliant index and convention index. The retained SIS during the global financial crisis between 2007 to 2009 shows that the Shariah compliant index does not immunized against the economic recession.

As this research applies only a single level shift in the simulation, in the future, we would like to investigate the performance of SIS using multiple shift in the simulation. Plus, we would like to investigate the performance of SIS in higher order of ARIMA models in the future research.

6. Appendices
Table A1. Retention Frequency of SIS in Islamic and Conventional Index from gets and Autometrics
Acknowledgements

The authors would like to extend their sincere gratitude to the Ministry of Higher Education Malaysia (MOHE) for the financial support received for this work under FRGS grant (203/PMATHS/6711604).

Reference

[1] Hendry D F 1999 An Econometric Analysis of US Food Expenditure, 1931-1989. In J. R. Magnus and M. S. Morgan (eds), Methodology and Tacit Knowledge: Two Experiments in Econometrics Chichester: John Wiley and Sons pp. 341-361.

[2] Hendry D F, Johansen S, Santos C 2008 Automatic selection of indicators in a fully saturated regression. Computational Statistics. 23: 317.

[3] Johansen S, Nielsen B 2009 An analysis of the indicator saturation estimator as a robust regression estimator. In J L Castle, and N Shephard (eds) The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry. Oxford: Oxford University Press. pp.1-36.

[4] Doornik J A, Hendry D F, Pretis F 2013 Step-indicator saturation (Discussion Paper No. 658). University of Oxford, Department of Economics.

[5] Castle J L, Doornik J A, Hendry D F 2012. Model selection when there are multiple breaks. Journal of Econometrics, 69(2),pp 239-46
[6] Doornik J A 2009 *Object-Oriented Matrix Programming Language Using Ox 7*. London: Timberlake Consultants Press.

[7] Johansen S, Nielsen B 2013. Outlier detection in regression using an iterated one-step approximation to the Huber-skip estimator. *Econometrics* 1(1). pp 53-70.

[8] Castle J L, Doornik J A, Hendry D F, Pretis F 2015 Detecting location shifts during model selection by step-indicator saturation. *Econometrics* 3(2). pp 240-64.

[9] Hendry D F, Mizon G E 2011 Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics*. 3(1).

[10] Ericsson N R 2017 How biased are US government forecasts of the federal debt?. *International Journal of Forecasting*. 33(2). pp 543-59.

[11] Reade J J, Volz U 2011. From the General to the Specific-Modelling Inflation in China. *Applied Economics Quarterly*. 57.

[12] Castle J L, Hendry D F 2019 The Polymath: Combining Theory and Data. In: *Modelling our Changing World*. Palgrave Texts in Econometrics. Palgrave Pivot, Cham. pp 85-99.

[13] Frydman R, Stillwagon J 2016 *Stock-Market Expectations: Econometric Evidence that both REH and Behavioral Insights Matter*. Institute for New Economic Thinking Working Paper Series(44).

[14] Sucarrat G 2018 *gets: An R Package for General-to-Specific (GETS) Modelling and Indicator Saturation Methods*. R package version 0.16.

[15] Marczak M, Proietti T 2016. Outlier detection in structural time series models: The indicator saturation approach. *International Journal of Forecasting*. 32(1). pp 180-202.

[16] Santos C, Hendry D F 2006. *Saturation in Autoregressive Models*. Notas EconÃşmicas. Faculdade de Economia da Universidade de Coimbra(24). pp 8-19

[17] Pretis F, Schneider L, Smerdon J E, Hendry D F 2016 Detecting Volcanic Eruptions in Temperature Reconstructions by Designed Break-Indicator Saturation. *Journal of Economic Surveys* 30(3). pp 403-29.