Model Performance Between Linear Vector Autoregressive and Markov Switching Vector Autoregressive Models on Modelling Structural Change in Time Series Data

Phoong Seuk Wai  
Department of Economics  
Faculty of Business and Finance UTAR  
Kampar Perak Malaysia  
e-mail: wai_phoong@yahoo.com

Mohd Tahir Ismail  
School of Mathematical Sciences  
Universiti Sains Malaysia  
Penang, Malaysia.  
e-mail: m.tahir@usm.my

Sek Siok Kun  
School of Mathematical Sciences  
Universiti Sains Malaysia  
Penang, Malaysia.  
e-mail: sksek@usm.my

Samsul Ariffin Abdul Karim  
Department of Fundamental and Applied Sciences  
Universiti Teknologi PETRONAS (UTP)  
Tronoh Perak Malaysia  
e-mail: samsul_ariffin@petronas.com.my

Abstract—Real financial time series data always exhibit structural change, jumps or breaks. Thus, in this paper, the performance of the linear vector autoregressive model (VAR), mean adjusted Markov switching vector autoregressive model (MSM-VAR) and mean adjusted heteroskedasticity Markov switching vector autoregressive model (MSMH-VAR) are applied in order to examine the oil price return and the gold price return effect on stock market returns. The two break point tests indicate the existence of break dates in the data. In addition, a comparison among the three model’s performance show that the two Markov switching vector autoregressive models with first autoregressive order able to provide the most significance, reliable and valid results as compared to linear vector autoregressive.

Keywords- linear VAR, Markov switching VAR, model performance

I. INTRODUCTION

Vector autoregressive (VAR) model is a statistical method that is able to capture the linear interdependencies among multiple series, predict the impulse response functions and forecast the variance decompositions for the empirical model. VAR is easier to justify and does not need expert knowledge due to the characteristics of VAR that requires less restrictive assumptions in estimating as compare to other multivariate methods ([1]; [2]).

The main usages of VAR are to analysis the economic shocks and predict the economic variables such as inflation and depression. However, there are argument on how the lag length should be determined in VAR model. Thus, information criterion is used to decide the number of lag length. Most commonly use criterion are the Akaike information criterion (AIC), Schwarz information criterion (SC) and Hannan-Quinn information criterion (HQ).

Although there are a number of critics on VAR, but the most important weakness of VAR is a lack of theoretical foundation of VAR analysis. Since all effects of the omitted variables will be in the residuals and the empirical model is estimated by using the impulse response function. So choosing the wrong variables or selecting the wrong restrictions or an inappropriate data selection might cause the misspecification on the assumption during the analytic process and therefore the impulse response function may not reflect to the implication on the theory ([3]). Moreover, if a structural change or switching regime occurs, the parameter will shift which can be represented by significant changes in time series data and lead to misspecification of VAR models. Thus, regime switching models are used to estimate this structural change feature.

Regime switching models are time series models where different values of the parameter can be taken in each of some fixed number of regimes. A stochastic process is assumed to have generated the regime shifts and allows predicting the probability of future regime shifts. In general, the regime is unobserved. Moreover, inference about which regime of the process belonged to in the past are needed to conduct before analysis the model.

Reference [4] was first proposed the unstable regression switching model. Reference [5] described that the Quandt’s model which just used square root for each equation to represent the subsample and then solve it by calculating is not efficient since this might contribute to missing data happens in the empirical model. Reference [6] then reviewed the model by introducing \( \lambda \) method to avoid wasting information or missing observations in estimate the regression switching model. The regression switching model that proposed by [6] using \( \lambda \) method are then reviewed by [7] to obey a Markov chain behavior. Reference [7] found that regime change in the model cannot be observed directly. Besides, [7] also stated that the current state in empirical model is independent with the other states; and the probability of certain states is not changing over time. Likelihood function in a Markov switching regression model that introduced by [7] in econometric field was first correctly by [8]. Then a recursive algorithm on computation for evaluation of likelihood under the assumption of no
lagged latent dependent variables in regression equations was suggested. Although recursive algorithm that suggested by [8] are efficient in estimation and inference with likelihood function, but is not relevant for a complicated lagged latent dependent variables model. Reference [9] used this algorithm as a basic idea and then suggested a Hamilton’s filter.

Markov switching autoregressive (MS-AR) model is an extension of the basic Markov switching model. This model is developed based on the ideas of (i) [7] in extending the Markov switching model to Markov switching autoregressive model; (ii) [10] in proposed a filter algorithm based on the assumption of Nefci that the more recent turning point influences the density function for current observations and (iii) [8] algorithm.

In general, Markov process allows any order visited the regimes and visited more than once. Moreover, the transition probabilities of the Markov chain are allowed to be practically determined by lagged values of the series. The Markov switching vector autoregressive (MS-VAR) model is proposed by Krolzig as an alternative to the constant-parameter and linear time series models of the earlier [11] model to allow the regime shift of the process of generating the time series.

Reference [12] has studied the statistical analysis of MS-VAR model and the application of the model to a dynamic multivariate system ([13], [14]). Based on the studies, the MS-VAR model is more general to characterize a non-linear data generating process as piecewise linear by restricting the process to be linear in each regime. In a MS-VAR model, there are many parameters which can switch across regimes by condition. Means, intercepts, autoregressive parameters, variance or covariance are examples of parameters that are may all be regime-dependent and only a discrete number of regimes are feasible.

The objective of this paper is to determine the best statistical model to estimate the economic model. Linear VAR model, two regime means adjusted Markov switching vector autoregressive model with first autoregressive order, MSM(2)-VAR(1) and two regime means adjusted heteroskedasticity Markov switching vector autoregressive model with first autoregressive order, MSMH(2)-VAR(1) are selected to examine the economic time series data. Oil price, gold price and four ASEAN countries’ stock market indices are used in this study. The purpose of selecting the linear VAR model is a linear VAR model is a simplest statistical model in measuring the economic and financial data. While mean adjusted Markov switching model is selected because in economic time series models, there might have one time immediate jump of the observed time series vector to its new level after a permanent regime shift in the process of mean. Besides, the MSMH-VAR model also included in the study since variation might exist in real economic data.

This paper is organized as follows, the section 2 reviews the previous studies that related to this study. Section 3 discusses the research design and methodology used. Section 4 presents the results of the study and the paper end with a conclusion and recommendation.
applied in the analysis and detect the turning points of the business cycle phases. Examples like [15] used the two state Markov switching range-based volatility and correlation models to estimate the Nasdaq index and S&P 500 index. Results shown that Markov switching models are suitable to apply to study of the stock market. While [16] also used a three regimes mean and heteroskedasticity vector autoregressive (MSMH-VAR) on examining the dynamic relationship between stock returns and inflation in Greece. The result of the finding was MSMH-VAR model outperforms than Ordinary Least Squares estimation and performs better than the linear VAR model in modeling the real stock returns. Furthermore, [17] compare the simple VAR model and MS-VAR model outputs on the studied and found that only a small portion of all of the VAR estimated is able to study in long-run relationship when compete to the Markov switching model in a simple and recursive portfolio experiment.

III. METHODOLOGY

In VAR model, all variables are commonly treated as priori endogenous and has a symmetrical structure ([18]). Each variable in the model is explained based on its own lags and the lags of other variables that inside the model. While explanatory variables are commonly assumed as exogenous variables in a single equation approach and the endogenous variables are able to describe by other variables in the system.

The basic VAR with one lagged values and k-th order autoregressive series for every variable in a set of p time series variables is as the following form,

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_k y_{t-k} + \varepsilon_t$$  \hspace{1cm} (1)

where $A_i$ is a $p \times p$ coefficient matrix for every $i = 1, \ldots, p$ and $\varepsilon_t$ is a $p \times 1$ unobservable error term that generally assumed to be a zero-mean independent white noise process with time–invariant, $E(\varepsilon_t) = 0$. For non-zero $p$, $E(\varepsilon_t, \varepsilon_{t-p})$ in the series are equal to zero which means that no serial correlation across time in individual error terms happened and the error term also known as independent stochastic vectors which will distribute with the zero mean and covariance structure, $\Sigma_w$.

The general idea of the MS-VAR model is $n$-dimensional time series vectors, $y_t = (y_{t,1}, \ldots, y_{t,n})$ depend upon an unobserved regime variable, $s_t \in \{1, \ldots, n\}$ represents the probability of being in a particular state to the switching mechanism in various states. While the conditional probability density of the observed $y_t$ is given by

$$P(y_t | Y_{t-1}, s_t) = h(y_t | Y_{t-1}, \beta_1) + \cdots + h(y_t | Y_{t-1}, \beta_n)$$  \hspace{1cm} (2)

where $\beta_n$ is the VAR parameter vector with the regime $n$ and $Y_{t-1}$ is the past information. Thus, the parameter vector, $\beta$ depend on the regime at time $t$.

This study was undertaken to design the linear VAR, MSM-VAR and MSMH-VAR models to estimate the oil price return and the gold price return effect on Malaysia, Singapore, Thailand and Indonesia stock market returns; and evaluate the most suitable model in analysis the economic time series data. Akaike Information Criterion (AIC) Test, Schwarz Information Criterion (SC) Test, Hannan-Quinn Information Criterion (HQ) Test are used in this study to compare the models’ performance. The formulas of these three criterion tests are

$$AIC = 2k - 2\log(L)$$
$$SC = p\ln(n) - 2\log(L)$$
$$HQ = 2p\ln\ln(n) - 2\ln(L)$$  \hspace{1cm} (3)

Prior to running the models, two breakpoint tests will be executed to examine the existence of structural breaks. The two tests are Bai-Perron ([19]) test and Chow ([20]) test. Bai-Perron test is used to disclose the unknown break date, while the Chow test is used to examine the break date that found using Bai-Perron test since the Chow test is able to examine the structural change and potential breaks in the series. Monthly data of oil price (OP), gold price (GP), Malaysia stock market index (KLCI), Singapore stock market index (STI), Thailand stock market index (SETI), and Indonesia stock market index (JCI) are collected from Thomson Reuters DataStream. The data are collected from December 1989 until May 2012. Prior to analysis all data are transformed to return series.

Prior to running the models, two breakpoint tests will be executed to examine the existence of structural breaks. The two tests are Bai-Perron ([19]) test and Chow ([20]) test. Bai-Perron test is used to disclose the unknown break date, while the Chow test is used to examine the break date that found using Bai-Perron test since the Chow test is able to examine the structural change and potential breaks in the series. Monthly data of oil price (OP), gold price (GP), Malaysia stock market index (KLCI), Singapore stock market index (STI), Thailand stock market index (SETI), and Indonesia stock market index (JCI) are collected from Thomson Reuters DataStream. The data are collected from December 1989 until May 2012. Prior to analysis all data are transformed to return series.

Figure 1 shows the plot of OP, GP, KLCI, STI, SETI and JCI original data. The plots give a rough idea for the properties of the series being used. All the series that are sketched in Figure 1 are fluctuating suggesting that the series is not linear. These series are seen to be subjected to structural change and regime shift.
IV. RESULTS AND DISCUSSION

In a financial time series model, there may have many break dates in the series. So, the breakpoint test is applied in the study to shed some light on the co-movement of the commodity and stock market returns.

<table>
<thead>
<tr>
<th>Table 1: Bai-Perron Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bai-Perron test</td>
</tr>
<tr>
<td>Reject $H_0$.</td>
</tr>
<tr>
<td>There are 5 breaks exist in the data when estimate the time series data at 5% significance level.</td>
</tr>
</tbody>
</table>

Table 1 show that there are breaks exists in the data and there are 5 break dates detected by using Bai-Perron test. Therefore, the Chow breakpoint test is used to derive the validity of these dates. According to the output in Table 2, it is concluded that there is a structural change in September 1994, February 1998, August 2001, September 2005 and January 2009 at 5% significance level. Thus, we conclude that all the series shows structural breaks behavior.

Next all the series are models using linear VAR and Markov switching VAR. The number of lag use for both models are 1 in order to utilize as many data as possible. Higher lags will reduce the number of data. While the next step is to compare the models accordingly to the log-likelihood ratio test and information criterion tests.

<table>
<thead>
<tr>
<th>Table 2: Chow Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
</tr>
<tr>
<td>September 1994</td>
</tr>
<tr>
<td>February 1998</td>
</tr>
<tr>
<td>August 2001</td>
</tr>
<tr>
<td>September 2005</td>
</tr>
<tr>
<td>January 2009</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3: Criterion Tests Results on the Linear VAR and MS-VAR Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion</td>
</tr>
<tr>
<td>Linear VAR</td>
</tr>
<tr>
<td>Log-likelihood</td>
</tr>
<tr>
<td>AIC</td>
</tr>
<tr>
<td>SC</td>
</tr>
</tbody>
</table>

Smaller information criterion statistics represent better model performance. AIC statistics represent better goodness of fit of the estimated model, while small SC statistics mean better model selection of different number of parameters in a class of parametric models and HQ is an alternative information criterion test to AIC and SC statistics in estimate the consistently order of the statistical model. Therefore, the smallest AIC, SC and HQ statistics represents the better statistical model in capturing the changes of the time series data.

According to the outputs in Table 3, smaller AIC, SC and HQ information criterion statistics are reported by MS-VAR model. Moreover, the log-likelihood function shows a largest log-likelihood value for the MS-VAR model. This can be concluded that the MS-VAR model able to perform better than the linear VAR model in examining the oil price return and the gold price return effect on Malaysia, Singapore, Thailand and Indonesia stock market returns when structural breaks happen.

V. CONCLUSION

In this paper, linear VAR and Markov switching VAR models are used in modelling the economic time series data to determine the best model performance since both models are different type of statistical model but share the same basic idea. A linear VAR model designed to examine the linear interdependencies among multivariate series while MS-VAR model designed to capture the structural or regime change of the multivariate series through time. However, these two types of statistical model have the similarity statistical properties and idea, that is the MS-VAR model used the VAR model as the basis. The reason for comparing the linear VAR and MS-VAR model are to determine which model able to provide the most significance and reliable results when estimate the economic data. MSM-VAR model and MSMH-VAR model are selected to represent the MS-VAR model since the average growth or the mean adjusted after a regime shift either with regime dependent heterogeneous or regime independent of the series transition are examined and compare with the linear VAR model. Results have shown that both models have the similar values but MSMH-VAR model has the smallest information criterion statistics and the largest log-likelihood statistics. Thus, it can be concluded that MSMH-VAR model able to provide the more significance, reliable and valid findings than linear VAR model and MSM-VAR model when estimate the oil price and gold price effect on stock market index.

ACKNOWLEDGMENT

We would like to acknowledge support for this project from Universiti Sains Malaysia (USM Short-term grant 304/PMATHS/6313045).

REFERENCES


