Proceeding of The 5th Annual Basic Science International Conference
Received: Feb 13th, 2015; Accepted: Mar 23rd, 2015; Published online: April 11th, 2015.

Identification of Self-Exciting Threshold Autoregressive Model by Using Genetic Algorithm

Maulida Nurhidayati[1]*, Irhamah[1]

[1] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya
* Corresponding authors: Maulida_nurhidayati@yahoo.co.id

Abstract – Self-Exciting Threshold Autoregressive (SETAR) models are models that can be applied to nonlinear time series models. SETAR models are models that partition data into multiple regimes in each regime follow autoregressive model. SETAR model has problem in terms of the identificaton of model. The best model usually obtained by trial and error which allows the best model is not a model of the global optimum. According this problem, in this study used Genetic Algorithm (GA) is a search technique is applied to the target-oriented optimization process to find a global optimum solution. Search techniques performed simultaneously on a number of solutions using three operations: selection, crossover, and mutation. Excellence GA in the optimization process, among others, working on the set of solutions, search based on population of solutions is not only one solution, and use of fitness information to obtain a global optimum solution. Identification model by using GA applied to stock return BMRI. Result from SETAR model using GA are compared with ARIMA model based on AIC. This result shows that SETAR model using GA provides better estimation for daily stock return than ARIMA model, because SETAR-GA produces less AIC then ARIMA.

Keywords : SETAR, nonlinear time series, GA, daily stock return

1. INTRODUCTION

Self-Exciting Threshold Autoregressive (SETAR) models are piecewise linear models where the linear relationship changes according to the value of the process. The model contains many interesting features, such as limit cycles, amplitude dependent frequencies, and jump phenomena, which cannot be described by a linear time series process [1]. A SETAR models consists of two or more AR models which are referred to as regimes between which the system switches. The switching occurs when the threshold variable passed one of the thresholds [2]. SETAR models have some problems in terms of model construction. The problem occurs when the determination of the value of the threshold, the delay, and the AR order corresponding to each regime. In addition, the number of parameters to be searched is large so the process will take a long time. Based on the problems, this study proposes a procedure for SETAR model identification used genetic algorithm (GA). GA is stochastic search techniques based on the mechanism of nature selection and natural genetics. GA uses a search procedure based on the value of the objective function, there is no used of gradient or calculus techniques. GA works with a coding of the solution set, not be solution themselves and search from a population of solution, not a single solution [4]. So, GA can find the global or nearly global solution for SETAR model.

2. METHODS

2.1 Self-Exciting Threshold Autoregressive

A time series Z_t is a self-exciting threshold autoregressive (SETAR) process if it follows the model

$$Z_t = \phi_0^j + \sum_{i=1}^{p_j} \phi_i^j Z_{t-i} + a_t{}^j; \text{ if } Z_{t-d} \in R_j \tag{1}$$

Where j=1,2,...,k, d is a positive integer and is known as the delay parameter, thresholds are $-\infty < r_1 < \ldots < r_{k-1} < \infty$, and $a_t{}^j$ is a sequence of independently identically distributed random noise with mean zero and variance $\sigma_j{}^2$ [1]. Since the SETAR model is a locally linear model, conditional least square techniques are useful in studying the process [3]. Parameter for SETAR model can be selected by using minimum Akaike's Information Criteria (AIC)[1], defined by

$$AIC(M) = -2\ln\left[maksimum\, likelihood\right] + 2M \tag{2}$$

2.2 Genetic Algorithm

Genetic Algorithms initiated by Holland in the 1970s as stochastic search and optimization techniques based on the mechanism of natural selection and natural genetics. Genetic algorithms start with an initial population of individuals generated at random. Each individual in the population represents a potential solution to the problem under consideration. The individuals evolve through successive iterations, called generations. During each generation, each individual in the population is evaluated using some of fitness. Then the population of the next generation is created through genetic operators. The procedure continues until the termination condition is satisfied [4].

Main genetic operator in genetic algorithms are selection, crossover, and mutation usually used to create the next generation. For many problems, it is nearly imposible to represent their solutions with binary encoding. During the last 10 years, various encoding methods have been created for particular problem to provide effective implementation of genetic algorithms. According to what kind of symbol is used as the alleles of a gene, the encoding methods can be classified as follows: binary encoding, real-number encoding, integer or literal permutation encoding, and general data structure encoding [5]. Genetic algorithms differ from conventional optimization because (1) genetic algorithms work with a coding of solution set, not the solutions themselves, (2) genetic algorithms search from a population of solutions, not a single solution, (3) genetic algorithms use fitness information, not derivatives or other auxiliary knowledge, and (4) genetic algorithms use probabilistic transformation rules, not deterministic ones [4].

2.3 Procedures

The step of the analysis consist of : (1) test for nonlinearity, (2) determine maximum order of autoregressive used PACF, (3) GA process to find the best model for SETAR using conditional least square as estimation method, (4) calculate AIC from model SETAR-GA. (5) Result from SETAR model using GA are compared with ARIMA model based on AIC. The data used in the study are simulation study SETAR(2,1,1) model, simulation study SETAR(2,2,2) model, and daily stock return BMRI from Jan 5, 2009 until Oct 31, 2014.

3. RESULTS AND DISCUSSION

3.1 Simulation Study SETAR (2,1,1)

In this simulation was set up to compare the performance of Information criteria (AIC) in case of a SETAR model and linear AR model. The simulation with sample size 500 from the following SETAR(2,1,1) model can be expressed as

$$Z_t = \begin{cases} 2 + 0.6Z_{t-1} + a_t & ; Z_{t-1} > 5 \\ 4 + 0.3Z_{t-1} + a_t & ; Z_{t-1} \le 5 \end{cases} \tag{3}$$

The results are estimation parameter for SETAR-GA and AR presented in Table 1 and Table 2.

Table 1 Estimation parameter simulation study SETAR(2,1,1) using SETAR-GA

Threshold	Regime 1 (lower)	Regime 2 (Upper)	AIC
4.836976	$\phi_0^1 = 1.6791929$	$\phi_0^2 = 3.39850951$	1445.21
	$\phi_1^1 = 0.7028865$	$\phi_1^2 = 0.0194467$	
		$\phi_2^2 = 0.39393435$	

Table 2 Estimation parameter simulation using AR

AR parameter	AIC
$\phi_0 = 5.4983$	1459.11
$\phi_1 = 0.5918$	

When comparing the results from Table 1 and Table 2, it can be concluded that SETAR model using GA provides better estimation for this simulation study than AR model, because SETAR-GA produces less AIC then AR.

3.2 Simulation study SETAR (2,2,2)

In simulation study was set up to compare the performance of AIC in case SETAR and linear A model with the model is following SETAR(2,2,2) as

$$Z_t = \begin{cases} 2 + 0.3Z_{t-1} + 0.6Z_{t-2} + a_t & ; Z_{t-1} > 2.5 \\ 2 - 0.3Z_{t-1} + 0.6Z_{t-2} + a_t & ; Z_{t-1} \le 2.5 \end{cases} \tag{4}$$

The result for this model when sample size 300 show in Table 3 and Table 4. From this Table can be concluded that that SETAR model using GA provides better estimation than AR model, because SETAR-GA produces less AIC then AR.

Table 3 Estimation parameter simulation using SETAR-GA

Threshold	Regime 1 (Lower)	Regime 2(Upper)	AIC
2.985543	$\phi_0^{\ 1} = 1.7398784$	$\phi_0^{\ 2} = 3.3993710$	896.5073
	$\phi_1^{\ 1} = 0.4072105$	$\phi_1^{\ 2} = 0.0989279$	

Table 4 Estimation parameter simulation using AR

AR paremater	AIC
$\phi_0 = 3.3229$	917.68
$\phi_1 = 0.4201$	
$\phi_2 = 0.0710$	

3.3 Empirical Study

In This empirical study would use the data is daily stock return BMRI from Jan 5, 2009 until Oct 31, 2014. Figure 1 show time series plot of stock return BMRI.

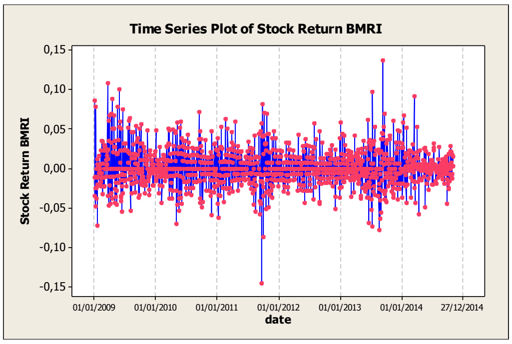

Figure 1 Time Series Plot of Stock Return BMRI

Table 5 Parameter of Genetic Algorithm

The population size	30
The fitness function AIC(d,p_1,p_2,r)	AIC(p_1)+AIC(p_2)
The crossover probability Pc	0.8
The mutation probability Pm	0.1
Number of iteration	50
Number of Bits	16

From this data, to detect nonlinearities in the data used terasvirta test. The null is the hypotheses of linearity in mean. The result of terasvirta test is X^2= 6.8524 with p-value=0.03251. From this result, terasvirta test successfully rejects the null hypotheses of linearity and conclude that data follow nonlinear process. The maximum order of autoregressive determined using PACF plots. From PACF plot got maximum order of autoregressive is 4. This order will used in selection SETAR model using GA with assume the highest order of each regime is 4. In this study limited regime is 2. Parameters of GA process show in Table 5.

The best chromosome from SETAR-GA obtained in 50th generation with fitness value is 7090.103. The best chromosome is 10100000000101 stands for d=3, p_1=3, p_2=3, and $r = -0.03448343$. Construct the parameter of SETAR model according to d,p_1,p_2 and r value using conditional least square as estimation method. Equation (5) show result parameter of SETAR model using GA with $AIC = -7090.103$.

$$Z_t = \begin{cases} 0.001572 - 0.000563Z_{t-1} - 0.02105Z_{t-2} - 0.058172Z_{t-3} + a_{1t} & ; Z_{t-3} > -0.03448343 \\ -0.03269 - 0.01114Z_{t-1} - 0.06804Z_{t-2} - 0.7659Z_{t-3} + a_{2t} & ; Z_{t-3} \le -0.03448343 \end{cases} \quad (5)$$

For comparison, we used ARIMA family to fit this data and find the best fitted ARIMA model is ARIMA(3,0,2) with $AIC = -7040.39$.

$$Z_t = 0.0014 + 1.1434Z_{t-1} - 0.5543Z_{t-2} - 0.0398Z_{t-3} - 1.156a_{t-1} + 0.542a_{t-2} + a_t \quad (6)$$

4. CONCLUSIONS

This study shows that SETAR model using GA provides better estimation for simulation study and daily stock return than ARIMA model, because SETAR-GA produces less AIC then ARIMA.

5. REFERENCES

1. W. W. Wei, 2006, **Time Series Analysis: Univariate and Multivariate Methods**, Pearson, New York.
2. E. Hamaker, 2009, *Journal of Mathematical Psychology*, **53**, 518-529.
3. R. S. Tsay, 1989, *Journal of American Statistical Association*, **84** (405), 231-241.
4. M. Sawaka, 2002, **Genetic Algorithms and Fuzzy Multiobjective Optimization**, Kluwer Academic Publishers, Boston.
5. M. Gen and R. Cheng, 2000, **Genetic Algorithms and Engineering Optimization**, John Wiley & Sons, Canada.