

Pemodelan Data Return Saham PT. Bank Republik Indonesia dengan *Self-Exciting Threshold Autoregressive* dan Algoritma Genetika

Maulida Nurhidayati¹

IAIN Ponorogo¹, NurhidayatiMaulida@gmail.com¹

DOI:<https://doi.org/10.15642/mantik.2018.4.1.16-21>

Abstrak

Model deret waktu nonlinier adalah model deret waktu yang diterapkan pada data-data yang memiliki pola nonlinier. Salah satu model deret waktu nonlinier adalah model *Self-Exciting Threshold Autoregressive* (SETAR). Model SETAR adalah model deret waktu yang pemodelan dari datanya dilakukan dengan membagi data menjadi beberapa *regime* dengan masing-masing *regime* mengikuti suatu model *autoregressive* (AR). Pembagian *regime* tersebut didasarkan pada nilai *delay* dan *threshold* dari data itu sendiri. Banyaknya parameter pada model SETAR mengakibatkan proses pencarian menghasilkan model SETAR yang belum optimum. Berdasarkan temuan tersebut, pada penelitian ini digunakan Algoritma Genetika untuk menghasilkan model SETAR terbaik yang merupakan model SETAR yang optimum. Pada penelitian ini dilakukan pemodelan data simulasi SETAR dan data return saham Bank Rakyat Indonesia (BRI). Metode yang digunakan untuk memodelkan data tersebut adalah *Grid Search* (GS) dan Algoritma Genetika (GA). Hasil analisis pada data simulasi SETAR menunjukkan bahwa metode GA memberikan hasil pemodelan yang lebih baik dibandingkan metode GS. Nilai AIC metode GA untuk jumlah data 200 sebesar -3,976178 yang lebih kecil dari AIC metode GS sebesar 1,361723. Untuk jumlah data 500 nilai AIC metode GA juga lebih kecil dari AIC metode GS. Pada data return saham BRI, metode GA juga memberikan hasil pemodelan lebih baik dibandingkan dengan GS. Hal ini ditandai dengan nilai AIC metode GA sebesar -11147,66 kurang dari -11146,26 yang merupakan AIC metode GS. Jadi, hasil analisis pada data simulasi model SETAR dan return saham BRI menunjukkan bahwa metode GA memberikan hasil pemodelan yang lebih baik dibandingkan dengan metode GS berdasarkan nilai AIC yang dihasilkan.

Kata kunci: Nonlinier, SETAR, Grid Search, Algoritma Genetika, Return saham

Abstract

Nonlinear time series model is a time series model applied to data that has the nonlinear pattern. One of the nonlinear time series models is Self-Exciting Threshold Autoregressive (SETAR). The SETAR model is a time series model that data modeling is done by dividing data into multiple regimes, whereas each regime following an autoregressive (AR) model. The division of the regime based on the score of the delay and threshold of the data itself. The number of SETAR model parameters not only resulted from the best model search process but also resulted in a SETAR model that is not yet optimum. Based on these findings, this study used Genetic Algorithm (GA) to produce the best and optimum SETAR model. In this research, using SETAR simulation data modeling and return data of Bank Rakyat Indonesia (BRI) were performed. The method used to model the data is Grid Search (GS) and Genetic Algorithm (GA). The result of analysis of SETAR simulation data shows that GA method gives better modeling result than GS method. The GA motive AIC value for the amount of 200 data is -3.976178 which is smaller than the AIC GS method of 1.361723. For the amount of data of 500 AIC values, GA method is also smaller than AIC GS method. In BRI stock return data, GA method also gives better modeling result compared to GS. It is marked by the GA AIC method value of -11147.66 less than -11146.26 which is the AIC method of GS. Thus, the result of analysis of SETAR model simulation data and BRI stock return shows that GA method gives better modeling result compared to GS method based on generated AIC value.

Key words: Nonlinear, SETAR, Grid Search, Genetic Algorithm, Stock Return

1. Pendahuluan

Model dari data deret waktu secara umum terdiri dari model deret waktu linier dan model deret waktu nonlinier. Model deret waktu linier adalah model deret waktu yang menggambarkan suatu data deret waktu yang membentuk suatu pola linier. Sedangkan model deret waktu nonlinier adalah model deret waktu yang menggambarkan suatu data deret waktu yang membentuk pola tidak linier. Pada data-data keuangan seperti data return saham dan inflasi biasanya menunjukkan fenomena kluster volatilitas, yaitu periode dimana data menunjukkan perubahan yang bergantian untuk periode yang panjang dan diikuti periode yang menunjukkan keadaan yang stabil [1]. Keadaan ini membuat data menjadi tidak stabil dan pemodelan data tersebut mulai didekati dengan menggunakan model deret waktu nonlinier. Kelebihan dari model ini dapat menangkap fenomena yang tidak dapat ditangkap oleh model deret waktu linier. Salah satu model deret waktu nonlinier yang dapat digunakan adalah model *Self-Exciting Threshold Autoregressive* (SETAR). Model SETAR adalah kasus khusus dari model *Threshold Autoregressive* (TAR) ketika *threshold* yang digunakan diambil dari nilai lag pada data deret waktu itu sendiri[2]. Prinsip dasar dari model SETAR adalah membagi data menjadi beberapa *regime* berdasarkan suatu nilai *delay* dan *threshold*. Masing-masing *regime* yang terbentuk mengikuti suatu model *Autoregressive*(AR) lokal.

Model SETAR terbaik yang diperoleh memiliki masalah dalam penentuan parameter-parameternya. Yaitu nilai *threshold*, *delay*, serta order AR untuk masing-masing *regime*. Banyaknya parameter yang dicari mengakibatkan proses pencarian model terbaik terkadang menghasilkan model SETAR yang belum optimum. Berdasarkan masalah tersebut pada tahun 2002, [3] menggunakan algoritma genetika (GA) untuk mengestimasi model *threshold autoregressive* dengan kesimpulan bahwa algoritma genetika secara simultan dapat memilih subset yang sesuai untuk parameter *delay* dan *threshold* pada model TAR. Hal ini dapat mengurangi waktu pencarian yang sangat besar dibandingkan dengan penggunaan metode estimasi tradisional. Keuntungan lain dari penggunaan algoritma genetika ini terletak pada fleksibilitas untuk mendapatkan hasil yang paling dekat dengan cara evolusi [3].

Algoritma genetika (*genetic algorithm* atau biasa disingkat GA) adalah suatu teknik pencarian berorientasi target yang biasa diterapkan pada proses optimasi mencari nilai-nilai ekstrim universal. Pada algoritma genetika, teknik pencarian dilakukan sekaligus atas sejumlah solusi yang dikenal dengan populasi. Operasi-operasi yang digunakan pada algoritma genetika antara lain: seleksi, *crossover* dan mutasi. Algoritma genetika menggunakan prosedur pencarian yang didasarkan pada nilai fungsi tujuan, tidak ada pemakaian gradien atau teknik kalkulus. Algoritma genetika memiliki beberapa keunggulan dalam proses optimasi antara lain: (1) Algoritma bekerja pada kumpulan solusi; (2) Algoritma genetika mencari berdasarkan populasi dari solusi, bukan hanya satu solusi; dan (3) Algoritma genetika menggunakan informasi *fitness* yang ingin dicari, bukan turunan atau pengetahuan khusus lainnya [4].

Berdasarkan temuan yang telah dipaparkan diatas, pada penelitian ini disajikan pemodelan data simulasi model SETAR dan data return saham Bank Rakyat Indonesia (BRI) dengan menggunakan model *self-exciting threshold autoregressive* dengan metode untuk pemodelannya menggunakan algoritma genetika (GA) dan *grid search* (GS).

2. Tinjauan Pustaka

2.1 Proses Deret Waktu

Proses deret waktu dapat dituliskan sebagai berikut

$$Z_t = \sum_{i=0}^{\infty} \psi_i a_{t-i} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_{i,j} a_{t-i} a_{t-j} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_{i,j,k} a_{t-i} a_{t-j} a_{t-k} + \dots \quad (1)$$

Jika proses Z_t pada persamaan (1) hanya terdiri dari suku pertama maka proses yang terjadi adalah proses linier. Jika proses Z_t pada persamaan (1) terdiri dari suku pertama dan suku lain maka proses yang terjadi merupakan suatu proses nonlinier [5].

2.2 Model *Threshold Autoregressive*

Model *Threshold Autoregressive* (TAR) pertama kali diusulkan oleh Tong dan Lim pada tahun 1980 [6]. TAR adalah alternatif model untuk mendeskripsikan deret waktu periodik. Model ini dimotivasi oleh beberapa karakteristik nonlinier yang biasa ditemukan dalam kehidupan sehari-hari seperti adanya asimetri

dalam pola turun dan naik suatu proses, fenomena lompatan, serta frekuensi ketergantungan amplitudo yang tidak dapat ditangkap oleh model deret waktu linier[6].

Model TAR menggunakan *threshold* untuk meningkatkan pendekatan linier pada modelnya. Model TAR dapat dituliskan sebagai berikut

$$Z_t = \phi_0 + \sum_{i=1}^p \phi_i Z_{t-i} + \left(\alpha_0 + \sum_{i=1}^p \alpha_i Z_{t-i} \right) I \left(\frac{Z_{t-d} - \tau}{\delta} \right) + a_t \quad (2)$$

Dengan Z_t adalah data deret waktu, d adalah parameter *delay*, τ adalah parameter lokasi (*threshold*), δ adalah parameter skala, dan $I(\cdot)$ adalah fungsi penghalus yang dapat berupa suatu fungsi logistik, eksponensial, maupun suatu fungsi indikator [5].

2.3 Model Self-Exciting Threshold

Autoregressive

Model *Self-Exciting Threshold Autoregressive* (SETAR) adalah kasus khusus dari model *Threshold Autoregressive* (TAR) ketika *threshold* yang digunakan diambil dari nilai lag pada data deret waktu itu sendiri [2]. Suatu proses Z_t mengikuti suatu proses SETAR jika mengikuti model deret waktu berikut

$$Z_t = \phi_{0,j} + \sum_{i=1}^{p_j} \phi_{i,j} Z_{t-i} + a_{t,j}; \text{ jika } Z_{t-d} \in R_j \quad (3)$$

Dengan $j = 1, 2, \dots, k$, d adalah bilangan bulat positif dan merupakan parameter *delay*, $a_{t,j}$ adalah barisan peubah acak yang identik, independen, dan mengikuti distribusi tertentu dengan mean nol dan varian σ_j^2 [7].

Berdasarkan persamaan (3) dapat dibentuk model $SETAR(2, p_1, p_2)$ yang merupakan model SETAR dengan 2 *regime* dimana p_1 menunjukkan order AR pada *regime* bawah sedangkan p_2 menunjukkan order AR pada *regime* atas. Bentuk persamaan model $SETAR(2, p_1, p_2)$ dapat dilihat pada persamaan (4) dengan d adalah *delay* dan r_1 adalah *threshold*.

$$Z_t = \begin{cases} \phi_{0,1} + \sum_{i=1}^{p_1} \phi_{i,1} Z_{t-i} + a_{t,1} & , \text{ jika } Z_{t-d} \leq r_1 \\ \phi_{0,2} + \sum_{i=1}^{p_2} \phi_{i,2} Z_{t-i} + a_{t,2} & , \text{ jika } Z_{t-d} > r_1 \end{cases} \quad (4)$$

2.4 Kriteria Pemilihan Model Terbaik

Akaike's Information Criteria (AIC) adalah suatu kriteria pemilihan model terbaik yang diperkenalkan oleh Akaike dengan mempertimbangkan banyaknya parameter dalam model. Semakin kecil nilai AIC yang diperoleh semakin baik model yang digunakan. Kriteria AIC dapat dirumuskan sebagai berikut

$$\begin{aligned} AIC(M) &= -2 \ln [\text{maksimum likelihood}] + 2M \\ &= N \ln \left(\frac{SSE}{N} \right) + 2M \end{aligned} \quad (5)$$

Dengan M adalah banyak parameter dalam model [5].

2.5 Algoritma Genetika (GA)

Algoritma genetika pertama kali dikembangkan oleh John Holland dari Universitas Michigan pada tahun 1975. John Holland mengatakan bahwa setiap masalah yang berbentuk adaptasi (alami maupun buatan) dapat diformulasikan dalam terminologi genetika. Algoritma genetika adalah simulasi dari proses evolusi Darwin dan operasi genetika atas kromosom [8].

Algoritma genetika masuk dalam kelompok *evolutionary Algorithm*. Algoritma genetika didasarkan pada genetika dan seleksi alam. Elemen-elemen ini yang dipakai dalam prosedur algoritma genetika. Algoritma genetika banyak dipakai dalam menyelesaikan masalah kombinatorial seperti TSP (*traveling salesman problem*), *crew scheduling* untuk *airline* hingga masalah kontrol. Algoritma genetika merupakan temuan penting dalam bidang optimasi, dimana suatu algoritma diciptakan dengan meniru mekanisme evolusi dalam perkembangan makhluk hidup. Dalam algoritma genetika prosedur pencarian hanya didasarkan pada nilai fungsi tujuan, tidak ada pemakaian gradient atau teknik kalkulus [9].

3. Metode Penelitian

Langkah-langkah analisis dalam penelitian ini dijabarkan dalam 2 tahap. Tahap pertama adalah studi simulasi model SETAR. Pada tahap ini, pencarian model SETAR terbaik dilakukan menggunakan metode *Grid Search* (GS) dan algoritma genetika (GA). Untuk tahap kedua adalah mengaplikasikan metode GS dan algoritma genetika untuk pemodelan data return saham Bank Rakyat Indonesia. Data return saham yang digunakan dari 3 Januari 2011 sampai dengan 30 Desember 2016:

4. Hasil dan Pembahasan

4.1 Simulasi Model SETAR

Studi simulasi model SETAR dilakukan untuk membangkitkan model SETAR. Nilai parameter model SETAR serta *threshold* yang digunakan dapat dilihat pada persamaan (6). Model SETAR yang digunakan merupakan model SETAR dengan *delay*=1 dan merupakan model yang simulasikan oleh Barogona, Battaglia dan Cucina [10]. Banyaknya data yang dibangkitkan 200 dan 500.

$$Z_t = \begin{cases} -1,2Z_{t-1} - 0,7Z_{t-2} + a_{t,1} & \text{jika } Z_{t-1} \leq 0 \\ 0,8Z_{t-3} + a_{t,2} & \text{jika } Z_{t-1} > 0 \end{cases} \quad (6)$$

Sebelum melakukan pemodelan dengan algoritma genetika perlu didefinisikan *fitness* yang digunakan. Hal ini dikarenakan nilai *fitness* yang sangat mempengaruhi hasil yang akan diperoleh. Selain *fitness*, parameter-parameter seperti peluang crossover, mutasi, banyak generasi, dan panjang kromosom yang digunakan juga harus ditentukan. Nilai dari masing-masing parameter yang digunakan pada pemodelan dengan algoritma genetika dapat dilihat pada Tabel 1.

Tabel 1. Parameter Algoritma Genetika (GA)

untuk Model SETAR

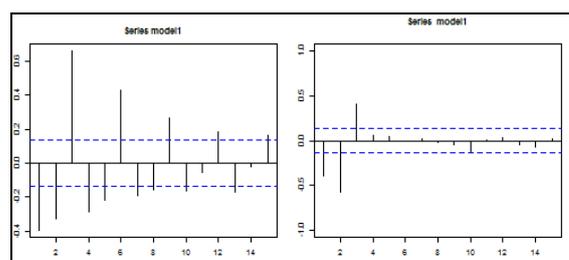
Parameter	Nilai
PopSize	200
Pc	0,8
Pm	0,1
Maksimum iterasi	200
<i>Fitness</i> (AIC)	$N * \log\left(\frac{SSE}{N}\right) + 2(p_1 + p_2 + 2)$

Pemodelan data dengan SETAR dilakukan dengan beberapa tahapan. Tahapan pertama adalah melakukan uji nonlinieritas Terasvirta. Uji ini dilakukan sebagai syarat suatu data dapat dianalisis dengan menggunakan model SETAR.

Tabel 2. Uji Nonlinieritas Terasvirta dan White Simulasi Model Subset SETAR

N	Terasvirta	
	F	p-value
200	13,6304	2,856e-06
500	40,3998	2,2e-16

Berdasarkan hasil yang ditunjukkan pada Tabel 2 diketahui bahwa hasil simulasi model SETAR memenuhi sifat nonlinieritas karena nilai F_{hitung} pada model tersebut lebih dari 3,04175.



Gambar 1. Plot ACF dan PACF Data Simulasi

Setelah diketahui bahwa data memiliki sifat nonlinier, langkah selanjutnya adalah menentukan orde maksimum dari model. Orde maksimum dapat dilihat dari hasil plot PACF yang dibuat. Plot PACF pada Gambar 1 menunjukkan bahwa data *cuts off* pada lag 3 sehingga pendefinisian orde awal untuk identifikasi model dilakukan dengan mengambil nilai d adalah 1, 2, atau 3 dengan $p_1 = 3$ dan $p_2 = 3$

Tabel 3. Identifikasi Model SETAR dengan Metode GS dan GA

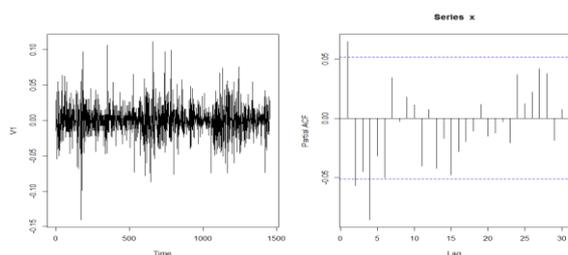
N	Metode	d	p1	p2	Threshold	AIC
200	GS	1	2	3	0,062794647	1,361723
	GA	1	2	[3]	0,06279465	-3,976178
500	GS	1	2	3	0,01980077	40,34689
	GA	1	2	[3]	0,01980077	33,7398

Hasil identifikasi model SETAR dengan metode GS dan GA ditunjukkan pada Tabel 3. Berdasarkan hasil yang ditunjukkan pada Tabel 3, Metode GS dan GA memiliki hasil yang sama

untuk orde d, p_1 , dan *threshold*. Tetapi terdapat perbedaan untuk orde p_2 yang dihasilkan. Metode *GS* memberikan hasil $p_2 = 3$ sedangkan metode *GA* memberikan hasil $p_2 = [3]$. Berdasarkan hasil tersebut, metode *GA* memberikan hasil identifikasi yang lebih sesuai karena model yang terbentuk sesuai dengan model yang disimulasikan serta nilai *AIC* yang dihasilkan dengan *GA* lebih kecil dibandingkan dengan metode *GS*.

4.2 Aplikasi Pada Data Return Saham

Sebelum melakukan pemodelan pada data return saham BRI perlu dilakukan pengujian nonlinieritas dari. Pengujian ini dilakukan untuk membuktikan secara statistik bahwa data yang dianalisis benar merupakan suatu data yang memiliki sifat nonlinier. Pengujian dilakukan dengan menggunakan metode Terasvirta dan diperoleh hasil sebesar 9,5459 dengan $p\text{-value} = 0,008456$. Karena $p\text{-value} = 0,008456 < 0,05$ dapat dibuat suatu kesimpulan bahwa data return saham BRI mengikuti sifat nonlinier sehingga model SETAR dapat diterapkan pada data return saham tersebut.



Gambar 2. Plot deret waktu dan PACF Data Return Saham BRI

Gambar 2 menunjukkan plot deret waktu dan plot PACF. Plot PACF yang terbentuk menunjukkan bahwa data *cuts off* pada lag 4 sehingga pemodelan data return saham selanjutnya didekati dengan menggunakan orde tertinggi adalah 4. Berdasarkan hal tersebut, pendefinisian orde untuk identifikasi model dilakukan dengan mengambil nilai d adalah 1, 2, 3, atau 4 dengan $p_1 = 4$ dan $p_2 = 4$.

Tabel 4. Identifikasi data return saham dengan Metode *GS* dan *GA*

Metode	d	p1	p2	Threshold	AIC
GA	2	2	[4]	0,006741573	-11147,66

Metode	d	p1	p2	Threshold	AIC
GS	2	2	4	0,006667	-11146,26

Tabel 4 memberikan hasil identifikasi model SETAR pada data return saham BRI. Hasil yang diperoleh menunjukkan nilai *AIC* dari metode *GA* lebih kecil dibandingkan dengan metode *GS*. Berdasarkan hasil tersebut, selanjutnya dilakukan estimasi pada model dan diperoleh model SETAR dengan identifikasi model dengan menggunakan algoritma genetika sebagai persamaan (7) berikut

$$Z_t = \begin{cases} -0,00291 - 0,1883Z_{t-4} + a_{t,1} & \text{jika } Z_{t-2} > 0,006741573 \\ -0,0008396 + 0,07369Z_{t-1} - 0,08637Z_{t-3} + a_{t,2} & \text{jika } Z_{t-2} \leq 0,006741573 \end{cases}$$

Model terbaik pada persamaan (7) menunjukkan bahwa suatu data return saham saat ini akan masuk *regime* 1 ketika return saham pada 2 hari sebelumnya kurang dari 0,006741573 atau dapat dituliskan sebagai berikut $Z_{t-2} \leq 0,006741573$. Apabila data return saham pada 2 hari sebelumnya lebih dari 0,006741573 maka data return saham saat ini masuk pada *regime* 2.

Nilai forecast untuk data pada *regime* 2 dipengaruhi oleh data return saham 1 hari sebelumnya dan 2 hari sebelumnya masing-masing sebesar 0,07369 dan -0,08637. Artinya return saham pada *regime* 1 akan mengalami kenaikan sebesar 0,07369 rupiah ketika data return saham pada 1 hari sebelumnya naik sebesar 1 rupiah dengan asumsi bahwa data return saham pada 2 hari sebelumnya adalah tetap. Nilai forecast untuk data pada *regime* 1 dipengaruhi oleh data return saham 4 hari sebelumnya sebesar -0,1883. Artinya return saham pada *regime* 2 akan mengalami penurunan sebesar 0,1883 rupiah ketika data return saham pada 1 hari sebelumnya naik sebesar 1 rupiah dengan asumsi bahwa faktor yang lain tetap.

5. Simpulan dan Saran

Pemodelan data simulasi SETAR yang dilakukan dengan algoritma genetika (*GA*) memberikan hasil pemodelan yang lebih baik dibandingkan dengan metode *Grid Search* (*GS*). Metode *GA* memberikan nilai *AIC* yang lebih kecil dibandingkan metode *GS* dengan hasil identifikasi yang lebih sesuai dengan simulasi

model yang dibangkitkan. Pemodelan data return saham BRI dengan metode GA juga memberikan hasil AIC lebih kecil dibandingkan dengan menggunakan metode GS.

Pada penelitian ini dibatasi regime yang digunakan adalah 2 sehingga pada penelitian selanjutnya dapat digunakan regime lebih dari 2. Selain itu, dapat juga dicoba menggunakan model time series nonlinier lain seperti *logistic smooth transition Autoregressive* (LSTAR) maupun *Exponential smooth transition Autoregressive* (ESTAR).

Referensi

- [1] D. N. Gujarati dan D. C. Porter, *Dasar-dasar Ekonometrika*, 5 ed. Jakarta: Salemba Empat, 2013.
- [2] P. H. Franses dan D. van Dijk, *Nonlinear Time Series Models in Empirical Finance*. New York: Cambridge University Press, 2003.
- [3] B. Wu dan C.-L. Chang, “Using Genetic Algorithms to Parameters (d;r) Estimation for Threshold Autoregressive Models,” *Comput. Stat. Data Anal.*, vol. 38, hlm. 315–330, 2002.
- [4] M. Sawaka, *Genetic Algorithms and Fuzzy Multiobjective Optimization*. Boston: Kluwer Academic Publishers, 2002.
- [5] W. W. S. Wei, *Time Series Analysis Univariate and Multivariate Methods*. New York: Pearson, 2006.
- [6] H. Tong dan K. S. Lim, “Threshold Autoregression, Limit Cycles and Cyclical Data,” *J. R. Stat. Soc. Ser. B Methodol.*, vol. 42, no. 3, hlm. 245–292, 1980.
- [7] R. S. Tsay, “Testing and Modeling Threshold Autoregressive Process,” *J. Am. Stat. Assoc.*, vol. 84, no. 405, hlm. 231–240.
- [8] D. Sri Kusuma dan H. Purnomo, *Penyelesaian Masalah Optimasi dengan Teknik-teknik Heuristik*. Yogyakarta: Graha Ilmu, 2005.
- [9] B. Santosa dan P. Willy, *Metode Metaheuristik, Konsep dan Implementasi*. Surabaya: Guna Widya, 2011.
- [10] R. Barogana, F. Battaglia, dan D. Cucina, “Estimating Threshold Subset Autoregressive Moving-Average Models by Genetic Algorithms,” *METRON - Int. J. Stat.*, vol. LXII, hlm. 39–61, 2004.