

Subset ARMA Model Identification for Monthly Electricity Consumption Data

Amaal El Sayed Abd El Ghany Mubarak

Department of Applied Statistics and Insurance, Faculty of Commerce, Damietta University, Egypt

Article history

Received: 03-12-2018

Revised: 31-01-2019

Accepted: 06-03-2019

Email: prof_amaal2010@yahoo.com

Abstract: Subset models can always be highly influential in series analysis, particularly when the data demonstrate a sort of form in periodic behavior with miscellaneous natural period's ranges, specifically; days, weeks, months and years. Subset models can also be effective as they let the number of parameters lower allowing only the really needed ones to be present in the model. Though subset autoregressive moving-average (ARMA) models always receive much attention, their identification is computationally cumbersome. This paper aims at the identification of Subset ARMA model through utilizing two methods of identification; innovation regression method and genetic algorithm method. The innovation regression method is a traditional one whilst the genetic algorithm methodologies represent a relatively modern approach for identifying Subset ARMA models in recent decades. After encoding every ARMA model as a binary string in the latter method, the iterative algorithm tries tracing the natural evolution of the population in those strings through letting strings to reproduce, producing newer models for competing for survival within upcoming populations. The aim of this research is to show the procedures for identifying the most appropriate order of subset ARMA models for the monthly electricity consumption data in Damietta governorate.

Keywords: ARMA Model, BIC, Genetic Algorithm, Identification, Innovation, Subset Models

Introduction

The family of Autoregressive model (AR) of order P for a zero-mean stationary stochastic process $\{X_t\}$ may be written in operator notation as $\phi(B)X_t = W_t$, where $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, $W_t \sim NID(0, \sigma_w^2)$.

It is assumed that the parameters are in the admissible region, i.e., all roots of the equation $\phi(B) = 0$ lie outside the unit circle.

As Haggan and Oyetunji (1984), a process $\{X_t\}$ is said to be a subset autoregressive (SAR) model of order $= (k_1, \dots, k_p)$, where $k_1 < \dots < k_p$, if it satisfies the stochastic difference equation:

$$X_t - \phi_{k_1} X_{t-k_1} - \dots - \phi_{k_p} X_{t-k_p} = W_t \quad (1)$$

where, $\{W_t\}$ is the white noise parallel to zero mean and variance $\sigma_w^2 > 0$. That model represents a special

case of the autoregressive model of order P with zero coefficients in specific intermediate lags. This means that the usual family of SAR(P) models is obtained by taking a subset of the parameters ϕ_1, \dots, ϕ_p , where $\phi(B) = 1 - \phi_{k_1} B - \dots - \phi_{k_m} B^{k_m}$. The family of SAR models may be denoted by $AR_p(k_1, \dots, k_m)$, where $0 \leq k_1 \leq k_2 \leq \dots \leq k_m \leq p$.

For a linear time series model, the selection of the subset model can be approximated by the function:

$$\log \hat{\sigma}_m^2 + m \frac{C(T)}{T} \quad (2)$$

As m is the number of nonzero parameters in the linear stochastic difference Equation (1), $\hat{\sigma}_m^2$ is the conditional maximum possibility for the estimation of the residual variance. Also $C(T)$ is a function of the number T of observations. Akaike Information Criterion

(AIC) as well as the Bayesian information criterion (BIC) correspond to $C(T) = 2$ and $C(T) = \log T$ can be used in model selection.

Initially, with no consideration of the computational hardships effort, a better SAR model of size m can be reached through the comparison of $\hat{\sigma}_m^2$ in all SAR models with available lag combinations and choosing the one that has the lowest $\hat{\sigma}_m^2$. The function (2) can subsequently be employed to decide the optimal SAR through allowing m to vary. This identification gets harder in situations like those of subset ARMA models.

Many recent searches such as Goldberg (1990), Chatterjee *et al.* (1996), Gaetan, (1998) and Baragona *et al.* (2004) employed the genetic algorithm as a new algorithm for selecting the optimal subset ARMA models to overcome its computationally cumbersome.

In the present work, two selection procedures for subset ARMA models are introduced based on innovation regression techniques together with stochastic binary search algorithms methodology. Many causes are present for selecting genetic algorithms. Firstly, the space of the solutions; any probable models, is finite, discrete however, it is still large. Secondly, any model is easily coded through binary strings. Additionally, the final population of models produced through that algorithm permits considering several other models.

The present paper has been planned according to the following. Part 2 contains a description of innovation regression technique for selection procedure. Part 3 contains an introduction of genetic algorithms and its application for identifying problem while in part 4 we use the monthly electricity consumption in Damietta governorate as a real data and apply the proposed model.

Innovation Regression Method for Selection Procedure

Keep in mind Durbin-Levinson recursion for the $AR(P)$; see Durbin (1960):

$$\phi_{j,k+1} = \phi_{j,k} - \phi_{k+1,k+1}\phi_{k+1-j,k}, j = 1, \dots, k. \quad (3)$$

where, $k = 1, \dots, p-1$ and $\zeta_{k+1} = \phi_{k+1,k+1}$ denotes the partial autocorrelation at Lag $k+1$. Such recursion is used for defining one-to-one transformation:

$$(\phi_1, \dots, \phi_p) \leftrightarrow (\zeta_1, \dots, \zeta_p). \quad (4)$$

A new family of models denoted by $AR_z(k_1, \dots, k_m)$, is obtained by selecting $\zeta_{k_1}, \dots, \zeta_{k_m}$ as parameters and constraining the other partial autocorrelations as zero. This model forms a subset of the $AR(P)$ model that

may be written as $\phi(B)X_t = W_t$, where the parameters ϕ_1, \dots, ϕ_p are parameterized by Equation (4). In the $AR_z(1,3)$ model, for example, $\phi_1 = \zeta_1, \phi_2 = -\zeta_1\zeta_3$ and $\phi_3 = \zeta_3$, since $\zeta_2 = 0$. Although $AR_p(k_1, \dots, k_m)$ and $AR_z(k_1, \dots, k_m)$ are similar, however, those models are distinct. In the $AR_p(1,3)$, $\zeta_1 = \phi_1 / (1 - \phi_2 - \phi_3^2)$, $\zeta_2 = \phi_2 / (1 - \phi_3^2)$ and $\zeta_3 = \phi_3$.

The plot of the usual partial autocorrelations together with the confidence limits is far more effective to identify AR_z subset models (see, for example, Yu and Lin, 1991). For the maximum sequence of p of the model, estimated partial autocorrelations $\hat{\zeta}_1, \dots, \hat{\zeta}_p$ are computed with their estimated standard errors. According to McLeod and Zhang (2006), the estimated standard errors can be calculated from the sample covariance matrix of $\hat{\zeta}_1, \dots, \hat{\zeta}_p$. The nonzero subset of partial autocorrelations $\zeta_k, k = 1, \dots, p$ for the AR_z model is chosen.

As Zhang and Terrell (1997) inform a new statistic and a new algorithm for selecting the optimal SAR, there may always be preference for using automatic model choice techniques (see Brockwell and Davis, 1996). The model space in the subset model selection of high-order AR may be quiet big hence, the usual AIC/BIC criteria might choose over parameterized models. The UBIC criterion was prepared by Chen and Chen (2008) to address that large model space problem. The approximate concentrated log-likelihood function for an $AR_z(k_1, \dots, k_m)$ evaluated at $\hat{\zeta}_1, \dots, \hat{\zeta}_m$ is as follows:

$$\mathcal{L} = -n \log \left[\prod_{k \in \{k_1, \dots, k_m\}} (1 - \hat{\zeta}_k^2) \right]. \quad (5)$$

where, n is the series length. The UBIC criterion may be written as:

$$UBIC = -2\mathcal{L} + m \log n + 2 \log \binom{p}{m}. \quad (6)$$

where, p is the maximum possible lag.

Denote $k = (k_1, \dots, k_p), i = (i_1, \dots, i_q)$, also, mind a zero-mean stationary stochastic process $\{X_t\}$ produced through $ARMA(k, i)$ model:

$$X_t - \phi_{k_1} X_{t-k_1} - \dots - \phi_{k_p} X_{t-k_p} = W_t - \theta_{i_1} W_{t-i_1} - \dots - \theta_{i_q} W_{t-i_q}. \quad (7)$$

which is assumed to be invertible.

As Hannan and Rissanen (1982), innovation regression method for ARMA modeling is composed of a couple of steps for determining ranks together with evaluation parameters. Step one aims at fitting $AR(n)$ model with high order employing Yule-Walker equations:

$$X_t = \sum_{i=1}^n \alpha_i X_{t-i} + W_t. \quad (8)$$

After that, estimated innovations are computed:

$$\hat{Z}_t = X_t - \sum_{l=1}^n \hat{\alpha}_l X_{t-l}, t = n+1, \dots, T \quad (9)$$

The second step is applying lowest squares technique on observations in addition to the estimated innovations obtained in Equation (9) of the $AR(n)$ model for estimating the ARMA parameters.

Let P and Q be sufficiently large, i.e., these may be higher than the real maximum order for k and i , then calculate the least squares estimates, $\hat{\phi}$ and $\hat{\theta}$, of

$$\phi = (\phi_{k_1}, \dots, \phi_{k_p})', \theta = (\theta_{i_1}, \dots, \theta_{i_q})', \text{ minimizing:}$$

$$\sigma^2(\phi, \theta) = \frac{1}{T} \sum_{t=t_0}^T \left(X_t - \sum_{j=1}^p \phi_{k_j} X_{t-k_j} + \sum_{l=1}^q \theta_{i_l} \hat{Z}_{t-i_l} \right)^2 \quad (10)$$

As $t_0 = \max(n + k_p, n + i_q)$, k^* is generally selected and i^* minimizing:

$$BIC(k, i) = \ln \sigma^2(\hat{\phi}, \hat{\theta}) + (p + q) \frac{\ln T}{T} \quad (11)$$

Genetic Algorithms and its Application to Identification Problem

Genetic algorithms can be defined as haphazard search techniques pivoted on principles of natural selection for obtaining either maximum or higher values for objective functions. Genetic Algorithms (GAs) employ a form of chromosomal representation for problem space, in addition to evaluating data of different points in the search space simultaneously, additionally, they employ recombination operators for preserving obtained data. GAs begin at a population of size N of binary encoded strings, chromosomes, of length L . Every one of the L positions illustrates a gene; and every gene's value is selected haphazardly to become a zero or 1. Every string demonstrates a solution for search space. A fitness function f can be computed as a monotone increasing function of g for every chromosome in the generation. Later, N parents for the next generation can be chosen with replacement utilizing probability p_j of selecting the j chromosome in the present population that is proportional to its fitness f_j :

$$p_j = \frac{f_j}{\sum_{i=1}^N f_i} \quad (12)$$

The parents may be counted in pairs; for each pair a crossover operator chooses a point that extends between 1 and L with probability p_c haphazardly. A couple of kids can be got through exchanging the parent's genetic

structure after the crossover point. Upon applying a mutation operator to every kid's genes, the operator flips the bits with probability p_m . Kids, then, replace the parents in the population and the breeding cycle can be iterated for fixed numbers of generations or until a convergence criterion can be performed. The string in the last population with the best value of g may be restored as the solution of the optimization problem.

Implementation of the GAs, as described above, first needs a string representation. Every chromosome $z = (z_1, \dots, z_{p+q})$ illustrates an ARMA model and its length is equal to $P + Q$ gene. The first P genes correspond to the Autoregressive parameters, the last Q to the Moving-Average ones and a gene is 1 provided the corresponding parameters are unequal to zero. For instance, when $P = Q = 5$, the chromosome $z = (0, 1, 0, 0, 0, 1, 0, 0, 1, 0)$ is in correspondence with the ARMA model:

$$X_t - \phi_2 X_{t-2} = W_t - \theta_1 W_{t-1} - \theta_4 W_{t-4} \quad (13)$$

There is another decision to take in utilizing a GAs, after deciding on encoding, this represents the way of performing choice. The function f is to offer enough options for ensuring that algorithm's preference of superior solutions finally reaching either optimal or near-optimal answer. Because the BIC could have some positive values, a simple selection $f = -BIC$ cannot be suitable due to providing negative fitness. Goldberg (1989) offers employing adaptive shapes of the fitness function, in situations including negative values of f , as:

$$f_j = M + 1 - BIC_j \quad (14)$$

BIC_j is the BIC value for the j th chromosome for the present population, M is the maximum value of BIC for the present population and the constant 1 can be added for ensuring a non-zero probability for the choice. Experiments demonstrated that the fitness function (14) may sustain the diversity in the population for above one hundred generations however, it is unable to avail higher levels for selection during final phases for moving in the direction of best solutions.

Another selection scheme, where a progressively varying temperature $\delta > 0$ controls choice rate, in this scheme the fitness function is given by:

$$f_j^\delta = \exp\left(-\frac{BIC_j}{\delta}\right) \quad (15)$$

High beginning temperatures involve that all chromosomes have accepted probabilities of reproduction. As GAs runs, the temperature is reduced in a gradual manner to let the GAs located nearer to the

most suitable part of the search space while defending more fit degrees for diversity.

The Application to Real Data

We consider the monthly electricity consumption in Damietta governorate from January 2002 through December 2016. A plot of the original data, Fig. 1, shows an obvious trend and nonstationary. We transformed the data by getting natural logarithms together with differencing at lag 12 as shown in Fig. 2. Figure 3 shows both sample autocorrelations and partial autocorrelations for transformed data. Innovation regression method was used first for Subset ARMA model selection procedure. According

to the UBIC criterion, the best model was the $AR_2(1,2,3,12,13)$. This identification agrees well with the model suggested by the partial autocorrelations plot (Pacf Plot) since the partial autocorrelation coefficients are out of the 95% confidence limits at lags 1, 2, 3, 12 and 13. The parameter estimates of the suggested model and their standard errors are reported in Table 1. Inspecting the standard errors of the parameter estimates ensures that all parameters are significant. The sample autocorrelations and partial autocorrelations of the residuals from $AR_2(1,2,3,12,13)$ model (Fig. 4) indicate that the residuals are random and an adequate, parsimonious model was obtained.



Fig. 1: Monthly electricity consumption plot

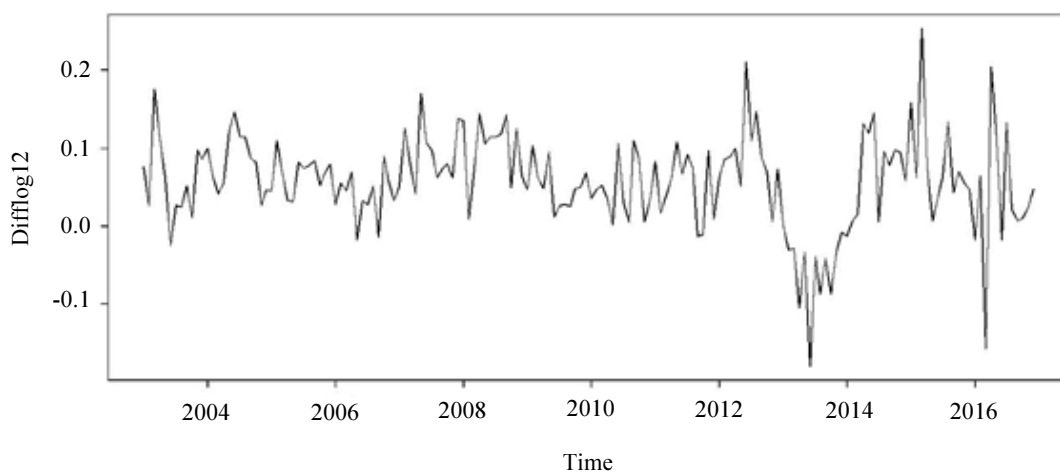


Fig. 2: Time series plot for transformed data

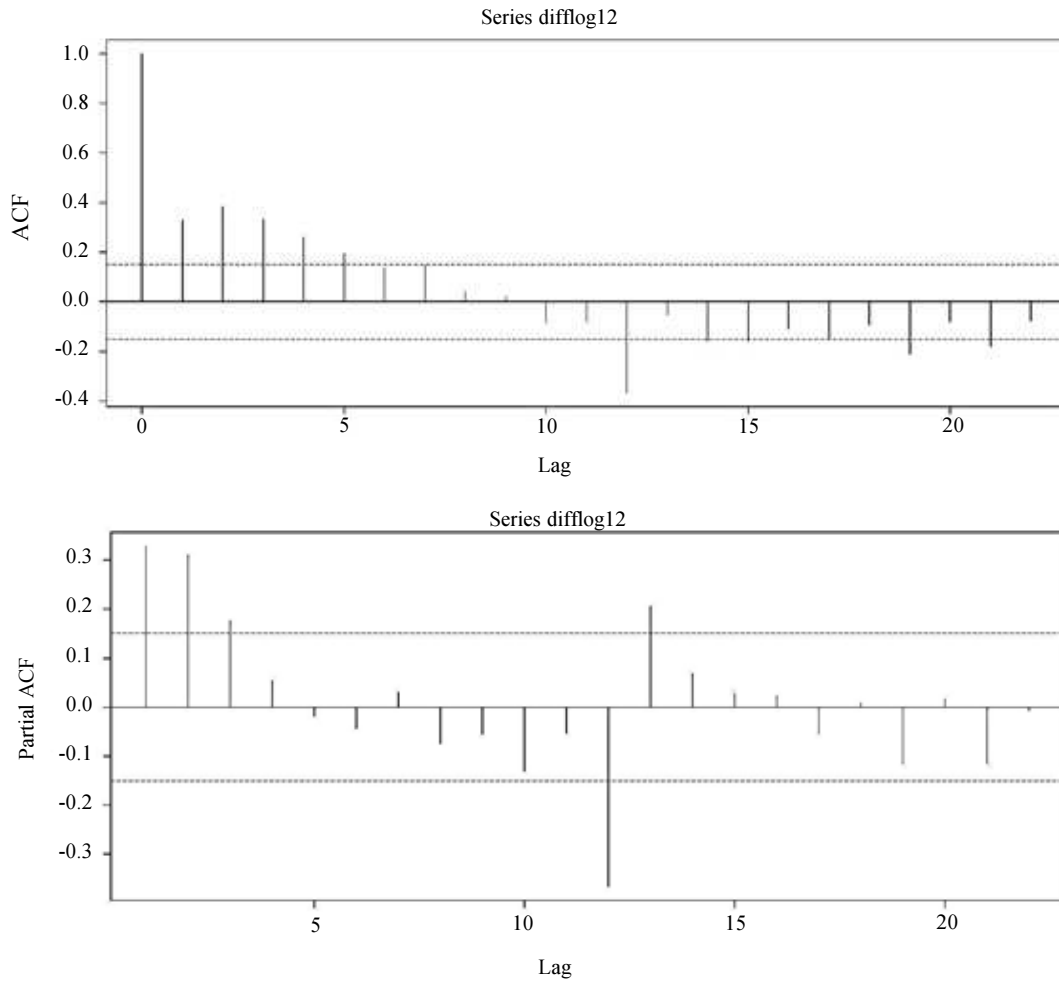


Fig. 3: Sample autocorrelation and partial autocorrelation functions of transformed data

Table 1: The estimated parameters for $AR_{\alpha}(1,2,3,12, 13)$ model

	MLE	SD	Z-ratio
Zeta(1)	0.22158780	0.093641499	2.366342
Zeta(2)	0.23138493	0.058528569	3.953367
Zeta(3)	0.20323856	0.053485514	3.799880
Zeta(12)	-0.49939073	0.070195714	-7.114262
Zeta(13)	0.30879179	0.073381246	4.208048

log likelihood = 516.314, AIC = -1020.6, BIC = -1001.9, UBIC = -987

An illustration of applying GAs to identify the appropriate Subset ARMA model, is presented hereby. A steady temperature decrement of the form $\delta_k = \delta_0 \alpha^k$, with $\delta_0 > 0$ is employed and $0.85 < \alpha < 0.99$. Upon experimenting with specific values, the researcher set $\alpha = 0.95$ and $\delta_0 = 1$. Selecting the crossover probability, p_c , is around 0.5- 0.9. Little values of p_c led to a decrease in the capability of searching the whole space of solutions and early convergence. Lower values for the probability of mutation p_m have always been favored for avoiding better schemata distraction. $p_m = 0.05$ was selected.

Inspecting the results in Table 2 obtained when we ran the Genetic Algorithm (GA) for the transformed data, we find the proposed subset model is $AR(1,4,7,12,15)$. This model has the smallest conditional maximum likelihood estimate of the residual variance $\hat{\sigma}^2 = 1663.9$ compared to other alternative models suggested by GAs. We can also refer to the proposed model as $GA:AR(15)$ with zero coefficients for some intermediate lags. The standard errors of the parameter estimates show that not all parameters are significant. Some of the sample autocorrelations and partial autocorrelations of residuals (Fig. 5) outside the significance limits indicate that the residuals are not completely random.

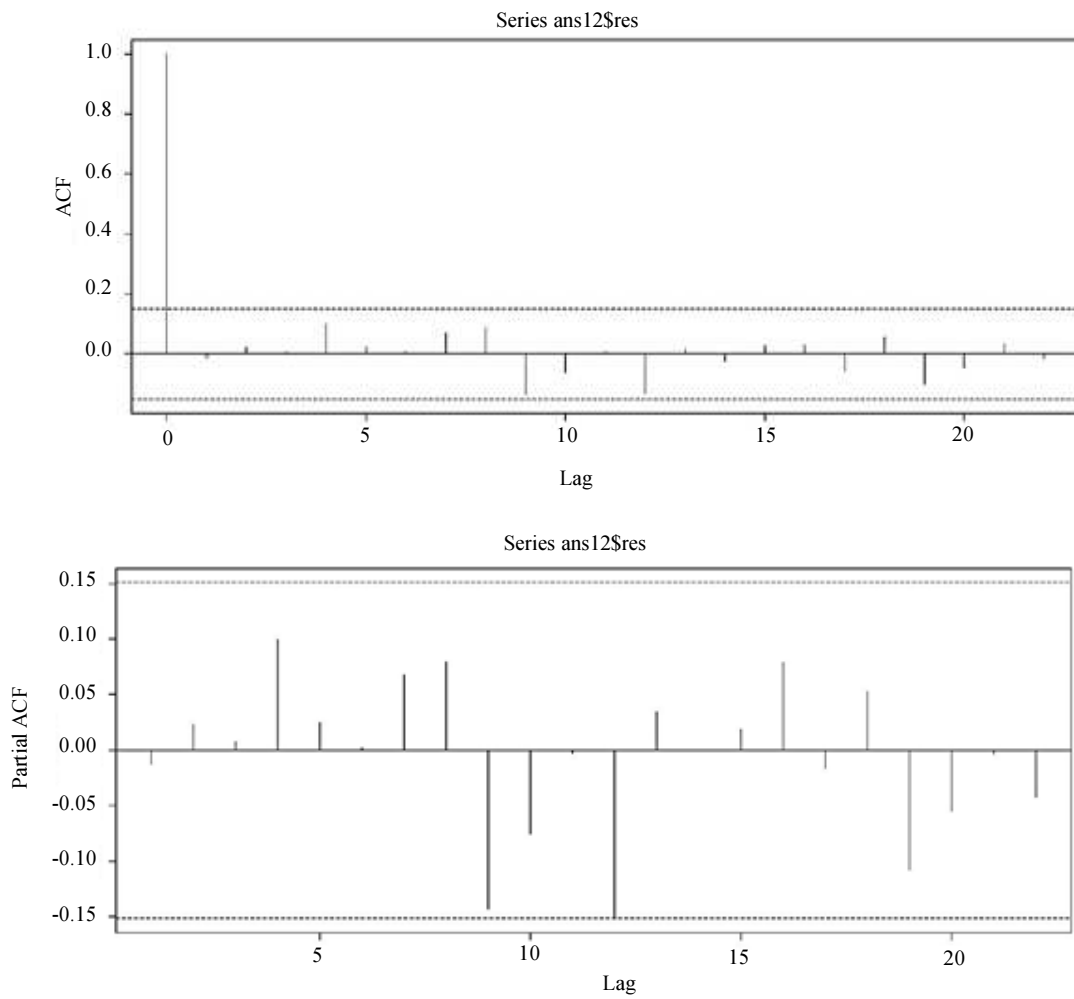
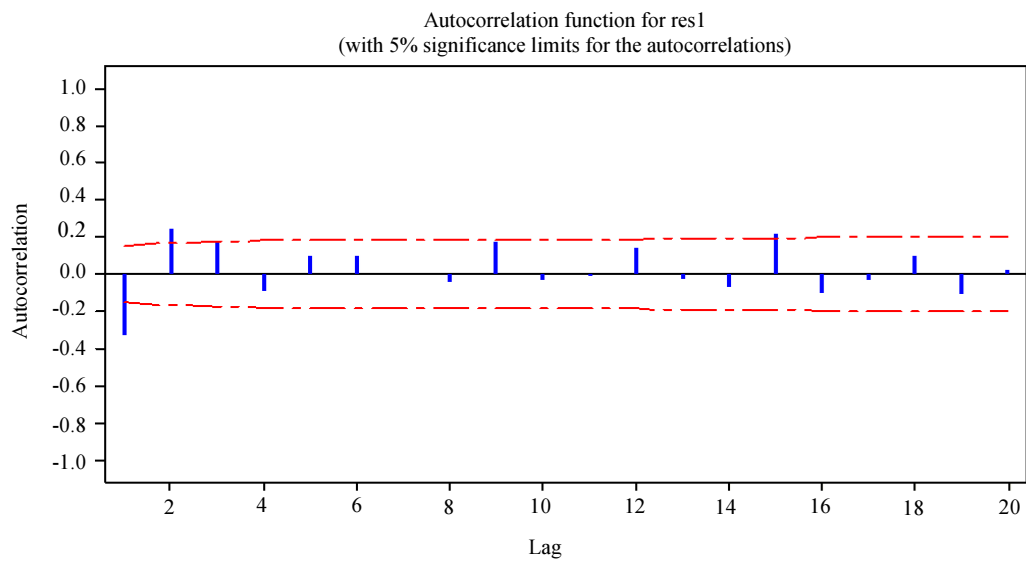


Fig. 4: Autocorrelation and partial autocorrelation functions of residuals for $AR_2(1,2,3,12,13)$ model



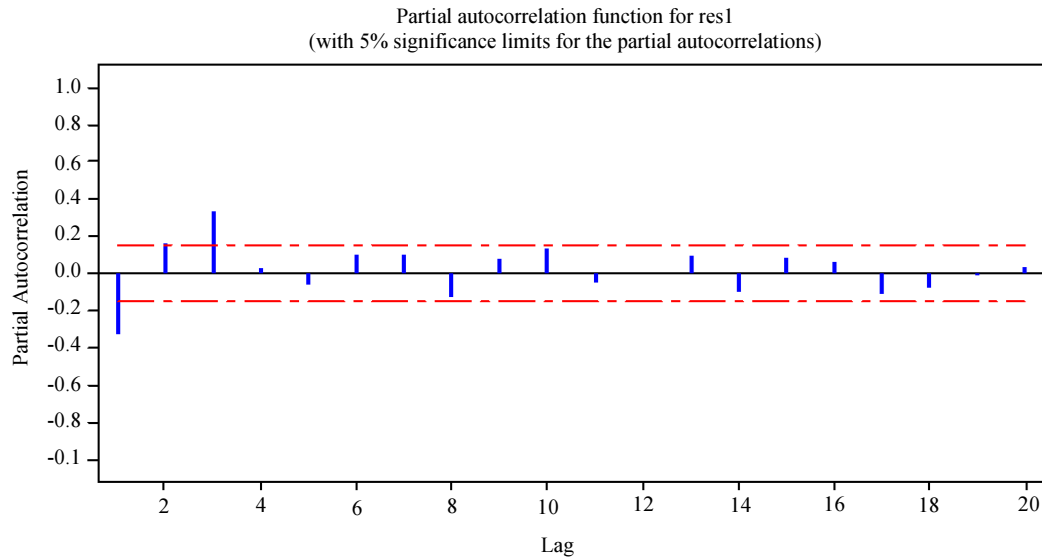


Fig. 5: autocorrelation and partial autocorrelation functions of residuals for $GA:AR(15)$ model

Table 2: The estimated parameters for $GA:AR(15)$ model (standard deviation in parentheses)

$\hat{\phi}_i, i = 1, \dots, 15$				
0.2666 (0.1961)	0.0000 (0.0000)	0.0000 (0.0000)	0.4883 (0.0767)	0.0000 (0.0000)
0.0000 (0.0000)	0.8491 (0.0576)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
0.0000 (0.0000)	0.0123 (0.0691)	0.0000 (0.0000)	0.0000 (0.0000)	0.0864 (0.1912)

The results of performance evaluation show that $AR_2(1,2,3,12,13)$ model obtained by the innovation regression method yields better forecast at in-sample data set than $GA:AR(15)$ model.

Conclusion

Subset models are usually helpful in the analysis of stationary time series. Lately, a class of subset ARMA models has been suggested by many researchers as parsimonious and special cases of ARMA models. This paper is interested in the selection of the optimal subset model so we represent two methods of subset ARMA models selection. The first which is usual is the Innovation regression method, the second is the genetic algorithms (GA_s). We use the monthly electricity consumption in Damietta governorate as a real data in our application. According to the Innovation regression method, the best model was the $AR_2(1,2,3,12,13)$ while the best one was the $AR(1,4,7,12,15)$. According to the genetic algorithms method. The results of performance evaluation show that $AR_2(1,2,3,12,13)$ model was the superior.

We hope to report a successful application of these two methods to the subset selection of ARMA models. Finally, we expect that the extensions in the use of genetic algorithms to the subset multivariate ARMA models would be very useful.

Acknowledgment

The author gratefully acknowledge with thanks the very thoughtful and constructive suggestions and comments of the Editor-in-Chief and the reviewers which resulted in much improved paper.

Ethics

The author declares that there is no conflict of interests regarding the publication of this article.

References

- Baragona, R., F. Battaglia and D. Cucina, 2004. Estimating threshold subset autoregressive moving-average models by genetic algorithms. *Int. J. Stat.*, 62: 39-61.
- Brockwell, P.J. and R.A. Davis, 1996. *Introduction to Time Series and Forecasting*. 3rd Edn., Springer, New York, ISBN-10: 0387947191.
- Chatterjee, S., M. Laudato and L.A. Lynch, 1996. Genetic algorithms and their statistical applications: An introduction. *Comput. Stat. Data Anal.*, 22: 633-51. DOI: 10.1016/0167-9473(96)00011-4
- Chen, J. and Z. Chen, 2008. Extended Bayesian information criteria for model selection with large model space. *Biometrika*, 95: 759-771. DOI: 10.1093/biomet/asn034

- Durbin, J., 1960. The fitting of time series models. *Rev. Int. Stat. Inst.*, 28: 233-243.
- Gaetan, C., 1998. Subset ARMA model identification using genetic algorithms. *J. Time Series Anal.*, 21: 559-570. DOI: 10.1111/1467-9892.00198
- Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. 1st Edn., Addison-Wesley, Reading, MA., ISBN-10: 0201157675, pp: 432.
- Goldberg, D.E., 1990. A note on Boltzmann tournament selection for Genetic algorithms and population-oriented simulated annealing. *Complex Syst.*, 4: 445-60.
- Haggan, V. and O.B. Oyetunji, 1984. On the selection of subset autoregressive time series models. *J. Time Series Anal.*, 5: 103-13.
DOI: 10.1111/j.1467-9892.1984.tb00380.x
- Hannan, E.J. and J. Rissanen, 1982. Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69: 81-94.
DOI: 10.2307/2335856
- Mcleod, A.L. and Y. Zhang, 2006. Partial autocorrelation parameterization for subset autoregression. *J. Time Series Anal.*, 27: 599-612.
DOI: 10.1111/j.1467-9892.2006.00481.x
- Yu, G. and Y. Lin, 1991. A methodology for selecting subset autoregressive time series models. *J. Time Series Anal.*, 12: 363-73.
DOI: 10.1111/j.1467-9892.1991.tb00090.x
- Zhang, X. and R.D. Terrell, 1997. Projection modulus: A new direction for selecting subset autoregressive models. *J. Time Series Anal.*, 18: 195-212.
DOI: 10.1111/1467-9892.00046