Decision Science Letters 1 (2012) 59–68

Contents lists available at GrowingScience

Decision Science Letters

homepage: www.GrowingScience.com/dsl

A New Hybrid Model for Improvement of ARIMA by DEA

Reza Narimani* and Ahmad Narimani

[a]*Department of Financial Engineering, University of Economic Sciences, Tehran, Iran*
[b]*Department of Economics, University of Allameh Tabatabae'i, Tehran, Iran*

A R T I C L E I N F O

Article history:
Received September 9, 2010
Received in revised format June 5, 2012
Accepted 31 July 2012
Available online
August 17 2012

Keywords:
DEA
ARIMA
Information criteria
Model selection
Prediction

A B S T R A C T

The classic ARIMA models use the information criteria for lag selection since 1990s. The information criteria are based on the summation of two expressions: a function of Residual Sum of Squares (RSS) and a penalty for decrease of degrees of freedom. However, the information criteria have some disadvantages since these two expressions do not have the same scale, so the information criteria are mainly based on the first expression (because of its bigger scale). In this paper, we propose a hybrid ARIMA model, which uses the Data Envelopment Analysis (DEA) model to select the best lags of AR and MA process called DEA-ARIMA. DEA is a linear programming technique, which computes a comparative ratio of multiple outputs to multiple inputs for each Decision Making Unit (DMU), which is reported as the relative efficiency score. We identify inputs as the number of AR and MA terms and outputs of the model are inverse of Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). In fact, in our proposed model, inputs consider as resources, so we are looking for some models with fewer resources and high efficiency. The DEA unlike the information criteria may have more than one solution and all of them are efficient so to compare this two models selection the mean of best DMUs is calculated. Experimental results demonstrate DEA-ARIMA will not trap in over fitting problem in contrast to classic ARIMA models because of considering a set of efficient ARIMA models.

1. Introduction

In univariate time series models attempt to predict economic and financial variables is based on its past values including random error terms. The ARIMA model is one of the most popular models for time series forecasting. This model has been originated from the autoregressive model (AR) proposed by Yule, 1927, the moving average model (MA) invented by Walker, 1931 and the combinations of the AR and MA. The ARMA model can be used when the time series is stationary, but the ARIMA model does not have that limitation.

* Corresponding author. Tel : +98 09196275190
E-mail addresses: reza_narimani@yahoo.com(R. Narimani)

doi: 10.5267/j.dsl.2012.08.002

In recent years, hybrid forecasting models have been proposed applying an ARIMA model as one element of their model such as ARIMA and artificial neural networks (ANN) (Asadi et al., 2012), ARIMA and probabilistic neural network (PNN), ARIMA and Particle swarm optimization (PSO)(Wang & Zhao (2009), ARIMA and Genetic Algorithm (GA) (Salcedo-Sanz et al., 2002). We believe one of the most major weaknesses of ARIMA model is that it uses information criteria to select the best lags of AR and MA process. Information criteria are based on the summation of two expressions: The first expression is a function of Residual Sum of Squares (RSS) and the other expression considers the loss of decrease of degree of freedom for entering additional lags of AR and MA. However, it seems, by adding these two expressions, which have different scales, it does not make an appropriate criterion to select the best ARIMA model.

In this paper, we propose a hybrid ARIMA model, which uses the Data Environment Analysis (DEA) to select the best lags of AR and MA process. DEA is a linear programming technique, which computes a comparative ratio of multiple outputs to multiple inputs for each Decision Making Unit (DMU), which is reported as the relative efficiency score. The efficiency score is usually expressed as a number between 0% and 100%. A DMU with a score less than 100% is deemed inefficient relative to others (Avkiran, 2006). In this case, DMUs are different ARIMA models with the number of AR and MA lags as inputs and inverse of Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) as outputs.

2. Related Work

ARIMA, if well-defined, can eliminate the auto correlation and partial correlation of residual. Thus, many researchers use this model as a part of their prediction proposed model. Khashei et al. (2012) use a hybrid model of ARIMA and PNN in order to yield more accurate results than traditional ARIMA models. Many researchers have implemented to identify the best order of ARIMA model. Valenzuela et al. (2008) propose the use of fuzzy rules to elicit the order of the ARIMA model, without the intervention of a human expert. In this area, the main tool that used for selection of order the ARIMA model is GA, which can be used for feature selection in other models. Salcedo-Sanz et al. (2002) present a GA for feature selection. They propose a genetic operator, which fixes the number of selected features. Minerva and Poli (2001) propose an automatic approach to the model selection problem, based upon evolutionary computation. They use genetic algorithm to choose both the orders and the predictors of the model. Many researchers such as Wang and Zhao (2009) and Asadi et al. (2011) present an ARIMA model, which uses PSO algorithm for model estimation.

In selection of the best order of ARIMA model, the key performance measure is AIC or its various modifications/extensions, including BIC procedures. Spanos (2010) argues that these model selection procedures invariably give rise to unreliable inferences, primarily because their choice within a pre-specified family of models. This paper replaces trading goodness of fit against parsimony with statistical adequacy as the sole criterion for when a fitted model accounts for the regularities in the data. AIC not only applies on ARIMA model but also applies to the other forecasting models. Zhao (2008) proposes an ensemble neural network algorithm based on the AIC. At first, the AIC-based ensemble neural network searches the optimum weight configuration of each component network first, then it uses the AIC as an automating tool to find the best combination weights of the ensemble neural network.

3. Methodology

3.1 ARIMA model

In the traditional time series models, the future values of a variable are a linear function of its past values including random error terms. This model can be expressed as follows

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + u_t - \sum_{j=1}^{q} b_j u_{t-j} \tag{1}$$

where y_t represents the measurement at time *t* in the time series; $u_{t-j};(j = 1,2,...,q)$ represents the effects of random factors that are independently and identically distributed (i.i.d.) with a mean of zero and a constant variance of σ_t^2. Model (1) called ARMA(p,q) that if y_t did not stationary, with determined to make stationary series, we achieve the ARIMA(p,d,q) model. The model (1) can be rewritten as follows:

$$\varphi(L)\nabla^d (y_t - \mu) = \theta(L)u_t \ ; \ u_t \sim N(0,\sigma_t^2) \tag{2}$$

where L is lag operator; $\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$ and $\theta(L) = 1 - \sum_{i=1}^{q} \theta_j L^j$ are the polynomials in L of degrees p and q respectively. $\nabla = (1 - L)$ is also the first differential of time series, therefore ∇^d means the d times difference to make stationary series. Parameters estimation of ARIMA model can be implemented by Box and Jenkins (1976) approach that has three stages: identification, estimation and diagnostic checking. Identification can be held by minimizing the information criteria. Information criteria are composed by two terms: a function of residual sum of square and a penalty for the loss of degree of freedom. The most information criteria that is recommended (Shibata, 1976), is Akaike's information criterion (AIC) that is given by following

$$AIC = -2\ln(\theta) + 2k \ , \tag{3}$$

where θ and *k* are the maximum likelihood estimator (MLE) and the number of unknown parameters of model 1, respectively. In the special case of least squares (LS) estimation with normally distributed errors, AIC can be expressed as

$$AIC = T\log(\hat{\sigma}^2) + 2k \tag{4}$$

where

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{T} \hat{\varepsilon}_t}{T} \ , \tag{5}$$

where $\hat{\varepsilon}_t;(t = 1,2,...,T)$ are the estimated residuals from the fitted model (Burnham & Anderson, 2004). However, for small samples, AIC often leads to over fitting. Therefore, attempts to deal with this problem gave rise to some modifications of the AIC, including Schwartz Bayesian Information Criterion (SBIC) and Hannan-Quinn Information Criterion (HQIC) as follows.

$$SBIC = T\log(\hat{\sigma}^2) + k\ln(T) \ , \tag{6}$$

$$HQIC = -2\ln(\theta) + 2k\ln(\ln(T)). \tag{7}$$

3.2.Data envelopment analysis

DEA is a methodology for evaluating the relative efficiencies of decision-making units (DMUs) within a relatively homogenous set (e.g. Sun & Lu, 2005). In fact, DEA provides a comprehensive analysis of relative efficiencies for inputs ($i = 1,2,...,m$) and outputs ($r = 1,2,...,t$) situations by evaluating each DMU and measuring its performance relative to an envelopment surface composed of other DMUs ($j = 1,2,...,n$).

In the basic DEA model that introduced by Charnes, Cooper and Rhodes (CCR) (1978), the objective is to maximize the efficiency value of a test DMU (j_0) subject to maximum efficiencies of all the DMUs are constrained to 1. In this model, decision variables are associated weights of input and output measures. CCR mode is shown as follows.

$$\text{Maximaize}\, Z_{j_0} = \frac{\sum_{r=1}^{t} u_{rj_0} Y_{rj_0}}{\sum_{i=1}^{m} v_{ij_0} X_{ij_0}}$$

$$\text{subject to: } Z_j = \frac{\sum_{r=1}^{t} u_{rj_0} Y_{rj}}{\sum_{i=1}^{m} v_{ij_0} X_{ij}} \le 1 \; ; j = 1,2,\dots,n \tag{8}$$

$$u_{rj_0}, v_{ij_0} \ge 0$$

where Z_{j_0} is the efficiency score of DMU_{j_0}; Y_{rj_0} $(r = 1,2,\dots,t)$ and X_{ij_0} $(i = 1,2,\dots,m)$ are the values of outputs and inputs for DMU_{j_0} respectively; u_{rj_0} and v_{ij_0} are the weight assigned to DMU_{j_0} for outputs and inputs respectively. This non-linear programming is the equivalent to the linear programming problem represented by following optimization model

$$\text{Maximaize}\, Z_{j_0} = \sum_{r=1}^{t} u_{rj_0} Y_{rj_0}$$

$$\text{Subject to: } \sum_{r=1}^{t} u_{rj_0} Y_{rj} - \sum_{i=1}^{m} v_{ij_0} X_{ij} \le 1 \; ; j = 1,2,\dots,n \tag{9}$$

$$\sum_{i=1}^{m} v_{ij_0} X_{ij_0} = 1$$

$$u_{rj_0}, v_{ij_0} \ge 0$$

The transformation is completed by constraining the efficiency ratio denominator from Eq. (8) to a value of 1, represented by the constraint $\sum_{i=1}^{m} v_{ij_0} X_{ij_0} = 1$. The result of the model Eq. (9) is an optimal simple efficiency $Z^*_{j_0}$ that is at most equal to 1. If $Z^*_{j_0} = 1$, it means DMU_{j_0} lie on the optimal frontier and then no other DMU is more efficient than DMU_{j_0} for its selected weights determined by Eq. (9). If $Z^*_{j_0} < 1$, there is at least one other DMU that is more efficient than DMU_{j_0} for the optimal set of weights. This process must perform n times for each DMU. However, this model has $n+1$ constrained, which is hard to solve. Therefore, we use the dual of the model (9) that represented by model (10) as follows,

$$\text{Minimize}\, E$$

$$\text{subject to: } \left(\sum_{i=1}^{n} \lambda_j X_{ij}\right) - X_{ij_0} E \le 0; i = 1,2,\dots,m$$

$$\left(\sum_{j=1}^{n} \lambda_j Y_{rj}\right) - Y_{rj_0} \le 0; r = 1,2,\dots,t \tag{10}$$

$$\lambda_j \ge 0; j = 1,2,\dots,n$$

$$E: \text{free}$$

where E and λ_j $(j = 1,2,\dots,n)$ are secondary variables of equality constraint and inequalities constraints of model (9), respectively. The CCR model has an assumption of constant returns to scale (CRS) for the inputs and outputs. In CRS conception, the output changes proportionally to input. In CRS model, if a DMU determines as an efficient unit, it is a scale efficient DMU. However, with variable returns to scale (VRS), a change in the input leads to a disproportional change in the output. To take into consideration variable returns to scale, a model introduced by Banker, Charnes and Cooper (BCC)

(1984) is utilized. The BCC model aids in determining the technical efficiency of a set of DMUs. This model has an additional convexity constraint is given by:

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{11}$$

The use of the CCR and BCC models together helps determine the overall technical and scale efficiencies of the DMU and whether the data exhibits varying returns to scale (Sarkis, 2000).

3.3 Proposed models

The classic ARIMA model selects the best lags of AR and MA process by minimizing the information criteria. As we said, when the data is small, AIC can be used. In Eq. 4, we mention AIC is based on the summation of two expressions: The first expression is a function of Residual Sum of Squares ($T\log(\hat{\sigma}^2)$) and the second expression (2k) considers the loss of decrease of degree of freedom for entering additional lags of AR and MA. But we think the summation of these two expressions, which have different scales, cannot be an appropriate criterion to select the best ARIMA model.

Therefore, we propose a hybrid ARIMA model, which uses the DEA model to select the best lags of AR and MA process. Our model consists of two stages:

In the first stage, we use AR and MA term from 1 to 5. Thus, we can constitute the reference set $W = \{AR(1), AR(2), \ldots, AR(5), MA(1), MA(2), \ldots, MA(5)\}$ which any of them can (be a potential ARIMA model) participate in ARIMA model. So we have 2^{10} different models. For each of these models, we calculate the Akaike's information criteria, Root Mean Square Error (RMSE) and Mean Absolute Percent Error (MAPE).

In the second stage, we use DEA approach to identify the best model. The inputs of the DEA are the number of AR and MA terms. For example, if ARIMA model is constructed by AR(3), AR(5) and MA(4) then the input vector is [2 1]. The first element indicates that the number of AR terms is equal to 2 and the second element indicates that the number of MA term is equal to 1. To determine the output of DEA, we use two performance measure of each model. These two different criteria are: RMSE and MAPE that is given by Eq. 12 and Eq. 13 respectively:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}}, \tag{12}$$

$$MAPE = 100\sum_{i=1}^{N}\frac{\left|\frac{\hat{y}_i - y_i}{y_i}\right|}{N}, \tag{13}$$

where y_i is i^{th} observation of response data and $\hat{y}_i$ is the forecasted value of y_i .*N* indicates the number of observations. In order to use RMSE and MAPE in DEA output, we inverse these performance indicators to obtain the efficiency of the model. Thus, the output of DEA is given by:

$$output_{1,j} = \frac{1}{RMSE_j} \quad j = 1,2,\ldots,1024 \tag{14}$$

$$output_{2,j} = \frac{1}{MAPE_j} \quad j = 1,2,\ldots,1024 \tag{15}$$

where *j* indicates the *j*th DMU in iteration.In each iterations the outputs are scaled between 0 and 1 by Eq. 16 to Eq. 18

$$\mathrm{Min}_{\mathrm{Outputs}} = \min(\mathrm{output}_{i,1}, \mathrm{output}_{i,2}, \ldots, \mathrm{output}_{i,1024}) \quad i = 1,2 \tag{16}$$

$$\mathrm{Max}_{\mathrm{Outputs}} = \max(\mathrm{output}_{i,1}, \mathrm{output}_{i,2}, \ldots, \mathrm{output}_{i,1024}) \quad i = 1,2 \tag{17}$$

$$\mathrm{output}_{i,j} = \frac{\mathrm{output}_{i,j} - \mathrm{Min}_{\mathrm{Outputs}}}{\mathrm{Max}_{\mathrm{Outputs}} - \mathrm{Min}_{\mathrm{Outputs}}} \quad i = 1,2, \text{and } j = 1,2,\ldots,1024 \tag{18}$$

where j indicates the j^{th} DMU in iteration and i indicates the i^{th} outputs. In the first iteration, DEA determines some of the potential ARIMA models as efficient (efficiency is equal to 1) and some of them as inefficient. Therefore, after each iteration we remove DMUs that their efficiency is less than one. When the number of efficient DMUs is equal in two successive iterations, the process will be stopped. At the final iteration, we maybe have more than one DMU that their efficiencies are equal to one. We show that the average of these solutions has the higher performance than conventional method. This method is shown on Fig. 1.

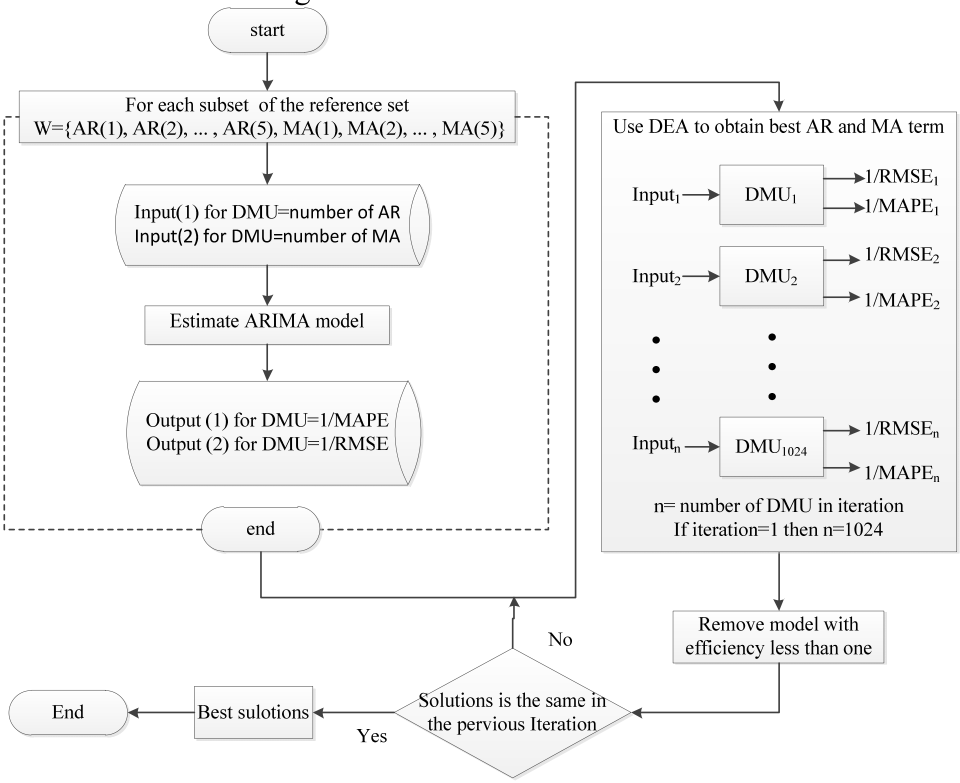


Fig. 1. The proposed model

The conventional method is based on the model that has minimum Akaike's information criteria. This method is shown on Fig. 2.

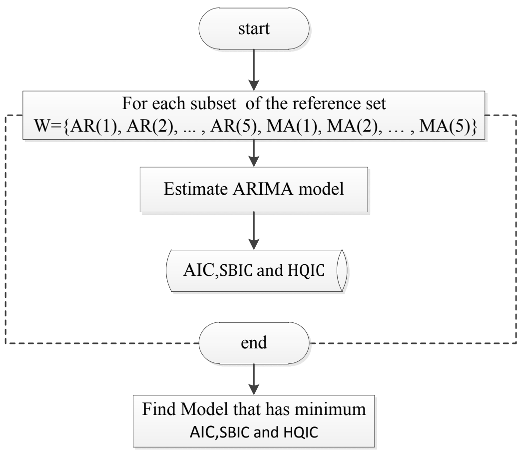


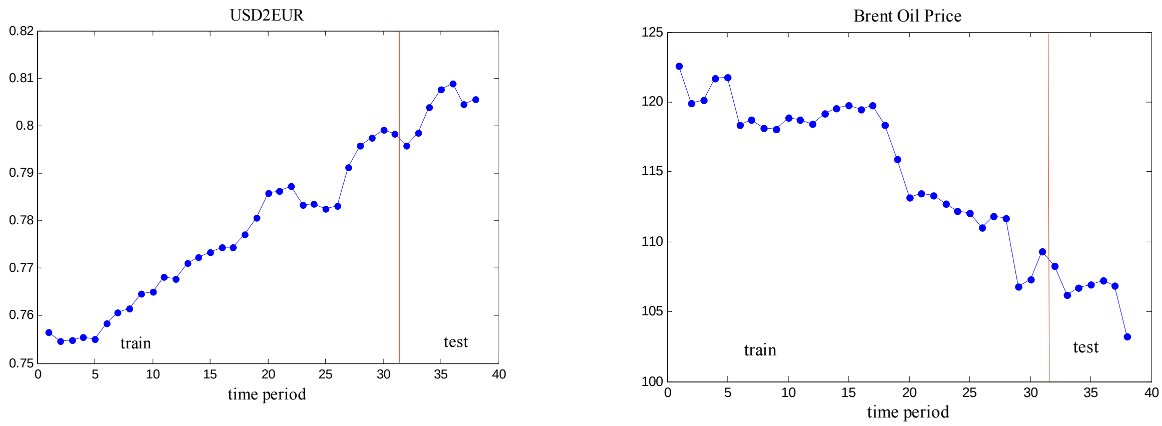
Fig. 2. The conventional method based on AIC, SBIC and HQIC

4. Data Sources

For evaluation of our proposed model, we use two dataset: the exchange rate of US Dollar to Euro (USD2EUR) obtained from http://www.oanda.com and Brent Oil Price (BP) taken from http://www1.investis.com. We divide the data to two sections for training and testing. Each dataset covers 31 daily observations for training (81%) and 7 daily observations for testing (19%) that are given by table 1 and Fig.3:

Table 1
The date for training and testing of two dataset

dataset	Train	Test
USD2EUR	28-Apr-12– 28-May-12	29-May-12– 04-Jun-12
BP	09-Apr-12– 21-May-12	22-May-12– 30-May-12



(a) (b)

Fig. 3. (a) The pattern of Brent oil price and (b) USD2EUR exchange rate

5. Empirical Results

5.1 Unit Root Test

In general, since many economic time series have non-stationary characteristics, the variables must be tested for stationary process. Therefore, in order to avoid the incorrect conclusions, the Augmented Dickey-Fuller (ADF) test proposed by Dickey and Fuller (1981), whose null hypothesis is that there is a unit root, is adopted. Table 1 shows results of unit root tests for four variables. The results indicate that the series are non-stationary when the variables are defined in levels. By first-differencing the series, in all cases, the null hypothesis of the non-stationary process is rejected at the 1% significance level.

Table 2
Augmented Dickey-Fuller test results

Variables	Intercept	Intercept and Trend
USD2EUR(log)	-1.265645	-2.747497
∇USD2EUR(log)	-16.49908***	-16.53787***
BP(log)	-2.474680	-2.152321
∇BP(log)	-21.84100***	-22.02063***

Notes: (1) ∇ means 1stdifference. (2) *, ** and *** refer to the rejection of the null hypothesis of the presence of a unit root at 10%, 5% and 1% levels, respectively.

In ARIMA model we use first difference of log(p) as the dependent variable, where p is the main series and after the estimation, we transform forecasted data to main series (p).

5.2 The performance of proposed model

In this section, we present the computational results of two dataset preformed to assess the behavior of the proposed model for lag selection problems. The maximum of AR and MA Term is 5 in this paper. With this configuration, we have 2^{10} models to compare. The two performance measures are inverse of RMSE and MAPE.

At the first, we perform our proposed model and classic ARIMA to USD2EUR dataset. In this case, the number of best DMUs in each iteration is presented in table 3. As you see, we have 8 efficient models in the 3rd iteration.

Table 3
The number of best DMUs in each iteration (USD2EUR)

Iteration	1	2	3
Number of Best DMUs	11	8	8

Now we estimate 8 ARIMA models (best DMUs) which are obtained after 4 iterations and 3 ARIMA models by minimizing AIC, SBIC and HQIC. As estimating 3 ARIMA models by minimizing AIC, SBIC and HQIC lead to the same model (AR(2) AR(4) AR(5)MA(3)MA(4)), we report this models as classic ARIMA.AR and MA lags of the best DMUs and classic ARIMA in train and test data are represented in table 4. As you see, the classic ARIMA has the minimum RMSE and MAPE in compare with each of the best DMUs in the train data, but it has the worth result in the test data which it is as result of over fitting problem in classic ARIMA.

Table 4
The performance of the DEA-ARIMA and classic ARIMA (USD2EUR)

Model	DMU	AR and MA term	Train		Test (average)	
			RMSE	MAPE	RMSE	MAPE
ARIMA (DEA)	1	MA(4)	0.0018	0.1845	0.0032	0.3249
	2	MA(1) MA(4)	0.0018	0.1778	0.0029	0.2717
	3	AR(5) MA(1) MA(2) MA(3) MA(4) MA(5)	0.0009	0.0933	0.0140	1.3177
	4	AR(4)	0.0020	0.1896	0.0028	0.2959
	5	AR(3) AR(4) AR(5)	0.0019	0.1800	0.0029	0.3001
	6	AR(2) AR(4) AR(5) MA(3) MA(4)	0.0005	0.0493	0.0103	0.8528
	7	AR(2) AR(3) AR(4) AR(5)	0.0019	0.1862	0.0031	0.3343
	8	AR(1) AR(2) AR(3) AR(4) AR(5) MA(4) MA(5	0.0005	0.0500	0.0107	0.8736
Classic ARIMA		AR(2) AR(4) AR(5) MA(3) MA(4)	0.0005	0.0493	0.0103	0.8528

According to previous section, we present the result of DEA-ARIMA and classic ARIMA for Berent oil price.

Table 5
The number of best DMUs in each iteration (Brent oil price)

Iteration	1	2	3
Number of Best DMUs	10	8	8

Table 6
The best DMUs obtained by the proposed model (Brent oil price)

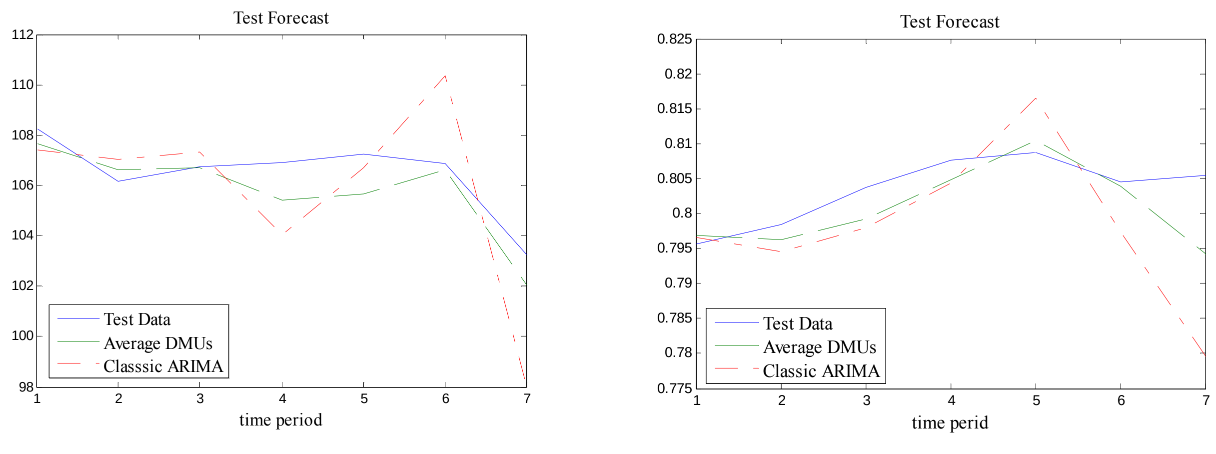
Model	DMU	AR and MA term	Train		Test (average)	
			RMSE	MAPE	RMSE	MAPE
ARIMA (DEA)	1	MA(1) MA(2) MA(3) MA(5)	0.9362	0.5863	1.0729	0.7678
	2	AR(2)	1.3231	0.9122	1.1527	0.7637
	3	AR(2) AR(5)	1.3156	0.8794	1.3384	0.7911
	4	AR(2) AR(3)	1.3114	0.8919	1.2169	0.7941
	5	AR(2) AR(3) AR(5)	1.3089	0.8712	1.3668	0.8526
	6	AR(2) AR(3) AR(4) AR(5) MA(3) MA(4) MA(5)	0.4105	0.2766	3.5200	2.2316
	7	AR(1) AR(2) AR(3) AR(5) MA(3) MA(5)	0.4693	0.3313	2.9156	1.7902
	8	AR(1) AR(2) AR(3) AR(4) AR(5) MA(3) MA(5)	0.4687	0.3269	2.7816	1.7419
Classic ARIMA		AR(2) AR(3) AR(5) MA(3) MA(4) MA(5)	0.4138	0.2868	2.4961	1.7104

Table 7 summarizes the improvement results of our proposed model. As you see, in the both of datasets, the classic ARIMA has a good performance in the train data, but DEA-ARIMA has an excellent performance in the test data.

Table 7
The performance of the proposed model and classic ARIMA

Data	Model	Train		Test	
		RMSE	MAPE	RMSE	MAPE
USD2EUR	Classic ARIMA	0.0005	0.0493	0.0103	0.8528
	DEA- ARIMA (average of the best DMUs)	0.0012	0.1233	0.0045	0.3828
BP	Classic ARIMA	0.4138	0.2868	2.4961	1.7104
	DEA- ARIMA (average of the best DMUs)	0.8181	0.5231	0.9281	0.6937

Fig. 4 compares the results of real values of both variables with their estimated values when applying the Classic-ARIMA and DEA-ARIMA methods.



(a) (b)

Fig. 4. Forecasted and Test data for (a) Brent oil and (b) USD2EUR

6. Conclusion and future works

In this paper, we propose a new effective approach for lags selection in ARIMA models using by DEA approach. DEA obtains DMUs with high efficiency and lower resources. The number of AR and MA terms is as input set and inverse of RMSE and MAPE are the output set. Increasing of inputs can role as loss of decrease of degree of freedom. Our results demonstrate DEA-ARIMA, because of considering a set of efficient ARIMA models, will not trap in overfitting problem in contrast to classic ARIMA. In addition, this research also provides a new direction in the model selection area and the new application of DEA technique.

Our future works include: 1) Apply DAE approach to other forecasting methods like Neural Network. 2) Use other criteria like the importance variable in input. 3) Use other criteria like the hit rate (identify the direction of movement) as output. 4) Use efficient ensemble method to obtain a weighted predictor from the best DMUs.

References

Asadi, S., Tavakoli, A., &Hejazi, S. R. (2012). A new hybrid for improvement of auto-regressive integrated moving average models applying particle swarm optimization. *Expert Systems with Applications*, 39, 5332-5337.

Avkiran, (2006). Developing foreign bank efficiency models for DEA grounded in finance theory. *Socio-Economic Planning Sciences*, 40, 275–296

Box, P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-day Inc.

Burnham, K., & Anderson, D. (2004). Multimodel inference: understanding {AIC} and {BIC} in model selection.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429-444.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica* 49, 1057-1072.

Khashei, M., Bijari, M., & RaissiArdali, G. A. (2012). Hybridization of autoregressive integrated moving average (ARIMA) with probabilistic neural networks (PNNs). *Computers & Industrial Engineering*, 63, 37-45.

Minerva, T., & Poli, I. (2001). Building ARMA models with genetic algorithms. Applications of Evolutionary Computing. In E. Boers (Ed.), 2037, 335-342, Springer Berlin / Heidelberg.

Salcedo-Sanz, S., Prado-Cumplido, M., Pérez-Cruz, F., &Bousoño-Calzón, C. (2002). *Feature Selection via Genetic Optimization*. Artificial Neural Networks — ICANN 2002. In J. Dorronsoro (Ed.), 2415, 44-144, Springer Berlin / Heidelberg.

Sarkis, J. (2000). An analysis of the operational efficiency of major airports in the United States. *Journal of Operations Management*, 18, 335-351.

Spanos, A. (2010). Akaike-type criteria and the reliability of inference: Model selection versus statistical model specification. *Journal of Econometrics*, 158, 204-220.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, 63, 117-126.

Sun, S., & Lu, W. (2005). Evaluating the performance of the Taiwanese hotel industry using a weight slacks-based measure. *Asia–Pacific Journal of Operational Research*, 22, 487–512.

Walker, G. (1931). On periodicity in series of related terms. Monthly Weather Review, 59, 277-278.

Valenzuela, O., Rojas, I., Rojas, F., Pomares, H., Herrera, L. J., Guillen, A., Marquez, L., &Pasadas, M. (2008). Hybridization of intelligent techniques and ARIMA models for time series prediction. *Fuzzy Sets and Systems*, 159, 821-845.

Wang, H., & Zhao, W. (2009). ARIMA model estimated by particle swarm optimization algorithm for consumer price index forecasting. *Artificial Intelligence and Computational Intelligence*. In H. Deng, L. Wang, F. Wang & J. Lei (Eds.), 5855, 48-58, Springer Berlin / Heidelberg.

Yule, U. (1927). On a method for investigating periodicities in disturbed series, with special reference to Wolfer’s sunspot numbers. *Philosophical Transactions of the Royal Society of London*, 226, 267 - 298.

Zhao, Z., Zhang, Y., & Liao, H. (2008). Design of ensemble neural network using the Akaike information criterion. *Engineering Applications of Artificial Intelligence*, 21, 1182-1188.