PREDICTING THE EVOLUTION OF BET INDEX, USING AN ARIMA MODEL

Florin Dan Pieleanu [1*]

ABSTRACT

Trying to predict the future price of certain stocks, securities or indexes is quite a common goal, being motivated by different reasons and being based on various techniques. The present article has the same purpose, employing an ARIMA model, due to its proven effectiveness and success. Used data is comprised of monthly values for the mentioned index, on a four-year period, from 2010 to 2014, which lead to 60 recordings. The main steps for the analysis are identifying the model, estimating the parameters and the prediction itself.

After each one of them is carefully conducted, a comparison is made: the predicted values for BET versus the real values for BET, in order to see if any resemblances exists, or if the differences are significant. Those resemblances or differences are explained, while the conclusion will highlight ARIMA's capacity or incapacity of forecasting in an accurate way, in the presented context.

KEYWORDS: *ARIMA, BET, prediction, moving average, autoregressive*

I. INTRODUCTION

Trying to predict the price of securities or financial indexes was always a tough mission. The reasons for that are multiple, and they show the complex nature of the capital market. But the continuous wish of the investors to get easy profits led to the development of new forecasting models. Among those, several proved to be more efficient, like the neural networks, who succeeded in the process of "learning" data structures. Also, ARMA and ARIMA models were remarked, especially for shorter periods of time.

The present article shows the building of an ARIMA model, then its utility for short term predictions, which may help the investors in their decisions in the capital market. The rest of the article is organized as such: section II reminds the main characteristics of ARIMA models, section III highlights the used methodology, and section IV discusses the results and also draws conclusions.

II. THE ARIMA MODELS AND LITERATURE REVIEW

ARIMA (autoregressive integrated moving average model) was introduced by Box and Jenkins in the '70, and represents a generalization of ARMA (autoregressive moving average model). Its main purpose is either to better comprehend the used data, either to develop predictions of the variables for future periods of time. ARIMA knows success even when the series are not stationary, because applying a difference of some degree leads to reducing or eliminating non-stationarity [1-4, 6, 11].

[1*] corresponding author, Adjunct Assistant, PhD, School of Cybernetics, Statistics and Economic Informatics, Academy of Economic Studies Bucharest, Romania, florin.dan.pieleanu@gmail.com

The model can be written ARIMA (p,d,q), where p,d and q parameters mean: p = the order of the AR autoregressive model, d = the difference degree and q = the order of the MA moving average model. ARIMA constantly proved its capacity to generate short term predictions, topping many other models.

The future value of a variable is, according to ARIMA, a linear combination of the past values and residual terms, as such [13]:

$$Y_t = \Phi_0+\Phi_1*Y_{t-1}+ \Phi_2*Y_{t-2}+ \Phi_3*Y_{t-3}+\ldots\ldots+ \Phi_p*Y_{t-p}+\xi_t-\theta_1*\xi_{t-1}- \theta_2*\xi_{t-2}- \theta_3*\xi_{t-3}-\ldots\ldots- \theta_q*\xi_{t-q}$$

where:

Y_t is the time series for the studied variable
Φ_i, θ_j are the coefficients' series
ξ_t is the residual terms' series

The steps for building the ARIMA model are: identifying the model, estimating the parameters and the prediction itself.

The model was highly used in studies along time, and below can be found only a brief list of them:

- [5] used ARIMA on the Korean market
- [7] considered ARIMA for the Indian market
- [8] used ARIMA for water consumption forecast, finding it fit for such a purpose
- [10] analyzed the market in Malaysia using ARIMA
- [13] compared ARIMA with other models in the Indonesian market
- [14] successfully used ARIMA for ozone consumption forecast

III. METHODOLOGY

Identifying the model

In this study, the used data represents monthly values for BET index (one of the indexes used on Bucharest Stock Exchange), from January 2010 to December 2014. This leads to a number of 60 recordings. Figure 1 shows the graph of the series, in order to check for stationarity:

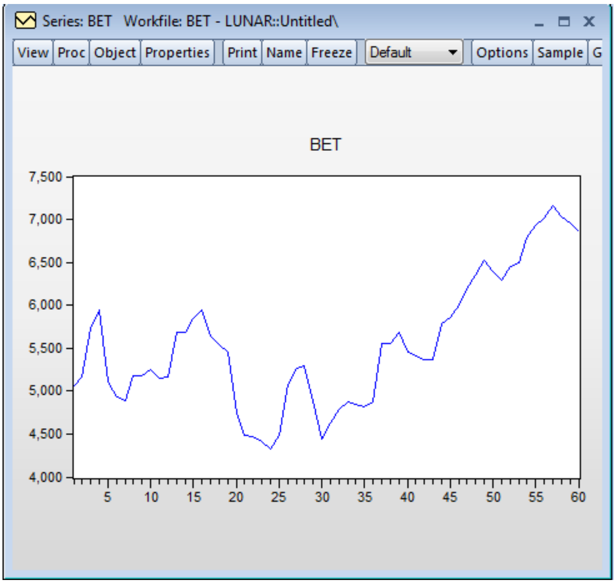
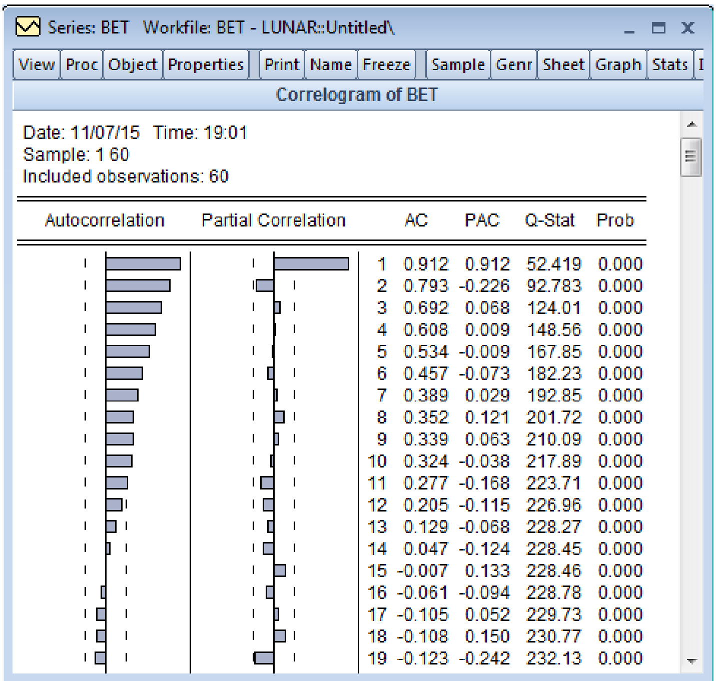


Figure 1. Graph of the original time series

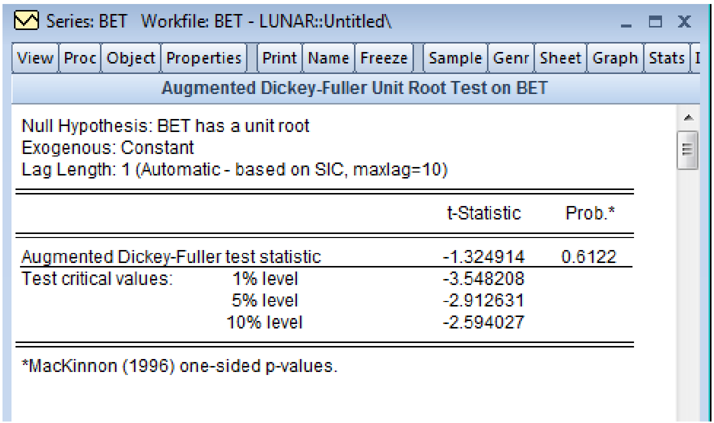
It can be observed that the series follows a random walk pattern. The correlogram confirms this:



	AC	PAC	Q-Stat	Prob
1	0.912	0.912	52.419	0.000
2	0.793	-0.226	92.783	0.000
3	0.692	0.068	124.01	0.000
4	0.608	0.009	148.56	0.000
5	0.534	-0.009	167.85	0.000
6	0.457	-0.073	182.23	0.000
7	0.389	0.029	192.85	0.000
8	0.352	0.121	201.72	0.000
9	0.339	0.063	210.09	0.000
10	0.324	-0.038	217.89	0.000
11	0.277	-0.168	223.71	0.000
12	0.205	-0.115	226.96	0.000
13	0.129	-0.068	228.27	0.000
14	0.047	-0.124	228.45	0.000
15	-0.007	0.133	228.46	0.000
16	-0.061	-0.094	228.78	0.000
17	-0.105	0.052	229.73	0.000
18	-0.108	0.150	230.77	0.000
19	-0.123	-0.242	232.13	0.000

Figure 2. Correlogram of the original time series

Also, the Dickey-Fuller tests shows:



Series: BET Workfile: BET - LUNAR::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | I

Augmented Dickey-Fuller Unit Root Test on BET

Null Hypothesis: BET has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.324914	0.6122
Test critical values:	1% level	-3.548208	
	5% level	-2.912631	
	10% level	-2.594027	

*MacKinnon (1996) one-sided p-values.

Figure 3. Dickey-Fuller test for the original time series

All these prove that the original series is not stationary. In order to make it so, using first difference may help. The graph becomes now:

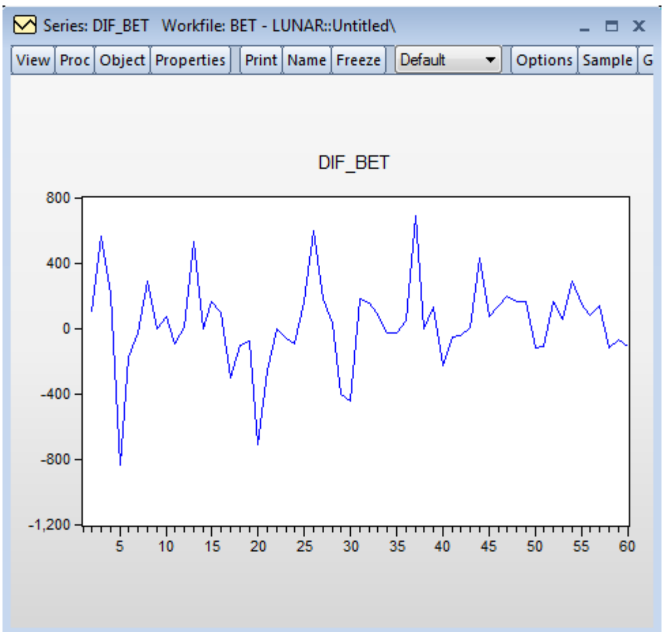
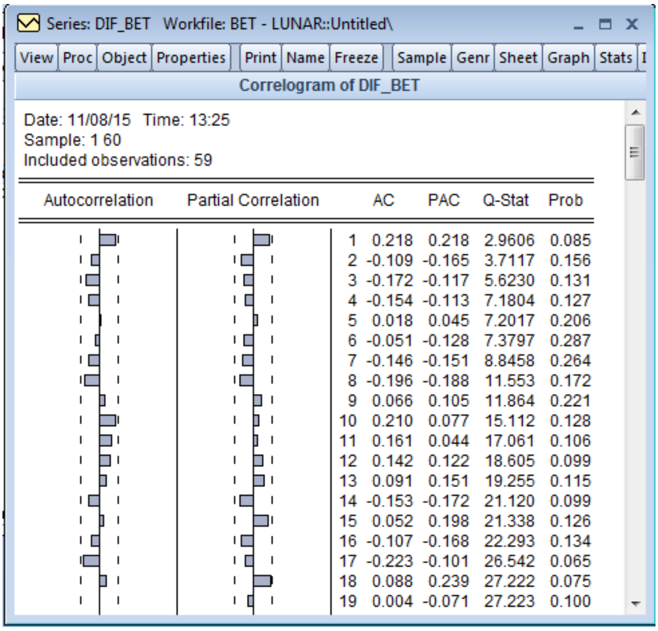


Figure 4. Graph of the time series after applying first difference

,and the correlogram:



Series: DIF_BET Workfile: BET - LUNAR::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats

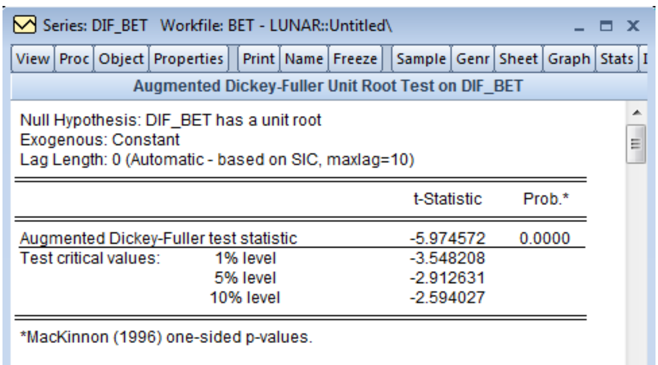
Correlogram of DIF_BET

Date: 11/08/15 Time: 13:25
Sample: 1 60
Included observations: 59

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.218	0.218	2.9606	0.085
		2	-0.109	-0.165	3.7117	0.156
		3	-0.172	-0.117	5.6230	0.131
		4	-0.154	-0.113	7.1804	0.127
		5	0.018	0.045	7.2017	0.206
		6	-0.051	-0.128	7.3797	0.287
		7	-0.146	-0.151	8.8458	0.264
		8	-0.196	-0.188	11.553	0.172
		9	0.066	0.105	11.864	0.221
		10	0.210	0.077	15.112	0.128
		11	0.161	0.044	17.061	0.106
		12	0.142	0.122	18.605	0.099
		13	0.091	0.151	19.255	0.115
		14	-0.153	-0.172	21.120	0.099
		15	0.052	0.198	21.338	0.126
		16	-0.107	-0.168	22.293	0.134
		17	-0.223	-0.101	26.542	0.065
		18	0.088	0.239	27.222	0.075
		19	0.004	-0.071	27.223	0.100

Figure 5. Correlogram of the time series after applying first difference

,with the Dickey-Fuller test:



Series: DIF_BET Workfile: BET - LUNAR::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats

Augmented Dickey-Fuller Unit Root Test on DIF_BET

Null Hypothesis: DIF_BET has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.974572	0.0000
Test critical values:	1% level	-3.548208	
	5% level	-2.912631	
	10% level	-2.594027	

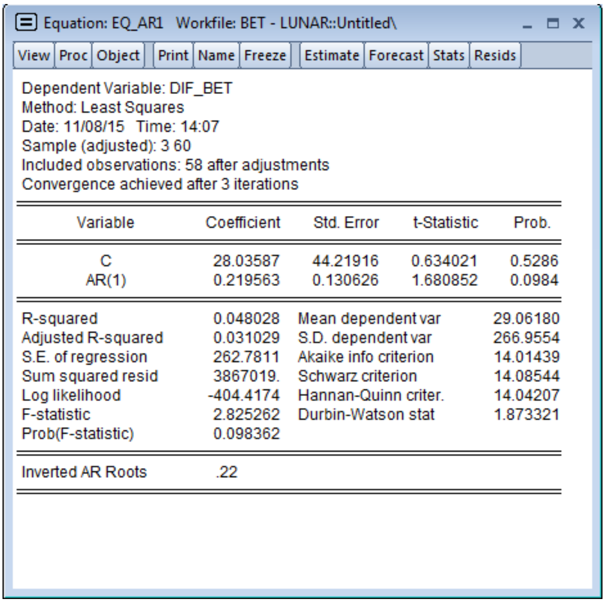
*MacKinnon (1996) one-sided p-values.

Figure 6. Dickey-Fuller test for the time series after applying first difference

The series is now first-order integrated, because the first difference made it stationary. So the model will be ARIMA (p,1,q), with the next step being to determine the p and q orders.

Estimating parameters

Considering that the autocorrelation (AC) and partial autocorrelation (PAC) from the correlogram are slowly, not sharply, decaying towards zero, probably an ARMA model will describe best the evolution of the studied variable. If AC would sharply fall, and PAC would slowly fall, then a moving average (MA) model would fit best. On the other hand, if AC would slowly fall, and PAC would sharply fall, then an autoregressive (AR) model would be the best. But, for certainty reasons, several AR and MA models will be tested, in order to establish which one serves entirely the goal. Firstly, an AR(1) model will be approached, with the following results:



Equation: EQ_AR1 Workfile: BET - LUNAR::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: DIF_BET
Method: Least Squares
Date: 11/08/15 Time: 14:07
Sample (adjusted): 3 60
Included observations: 58 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	28.03587	44.21916	0.634021	0.5286
AR(1)	0.219563	0.130626	1.680852	0.0984

R-squared	0.048028	Mean dependent var	29.06180
Adjusted R-squared	0.031029	S.D. dependent var	266.9554
S.E. of regression	262.7811	Akaike info criterion	14.01439
Sum squared resid	3867019.	Schwarz criterion	14.08544
Log likelihood	-404.4174	Hannan-Quinn criter.	14.04207
F-statistic	2.825262	Durbin-Watson stat	1.873321
Prob(F-statistic)	0.098362		

Inverted AR Roots .22

Figure 7. Results for AR(1) model

,not being satisfactory, due to lack of significance. An AR(2) is being obvious that this one isn’t optimal either, because its parameters are not statistically significant.

The next array of models: AR(3), MA(1), MA(2), MA(3), ARMA(1,1), ARMA (2,1), ARMA (1,2), ARMA (2,2), ARMA (2,3), ARMA (3,2), ARMA (3,3) will be also employed. The criteria selection are: relatively small Akaike, Schwarz and S.E (standard error) values, relatively high adjusted R2 value. The table below summarizes the obtained results:

Table 1. Different ARMA models and their values for chose criteria

Model	**Akaike**	**Schwarz**	**Statistic significance**	**Adjusted R^2**	**S.E**
AR(1)	14,014	14,085	no	0,031	262,781
AR(2)	13,964	14,072	no	0,040	254,138
AR(3)	14,002	14,147	no	0,028	256,752
MA(1)	13,984	14,054	yes	0,044	258,892
MA(2)	14,012	14,117	no	0,032	260,447
MA(3)	14,027	14,168	no	0,032	260,433
ARMA(1,1)	14,019	14,126	yes	0,041	261,302

ARMA(2,1)	13,996	14,139	no	0,025	256,088
ARMA(1,2)	13,994	14,136	yes	0,080	255,919
ARMA(2,2)	13,846	14,025	partially	0,174	235,686
ARMA (2,3)	13,877	14,092	partially	0,161	237,535
ARMA (3,2)	13,885	14,102	partially	0,163	238,212
ARMA (3,3)	13,849	14,102	partially	0,205	232,167

Only MA(1), ARMA (1,1) and ARMA (1,2) have statistically significant parameters, so the choice will be made between them. For this purpose, analyzing each of them is helpful:

ARMA (1,1)

To determine the adequacy of this model, a comparison will be made: theoretical versus empirical values for AC and PAC. The situation looks like this:

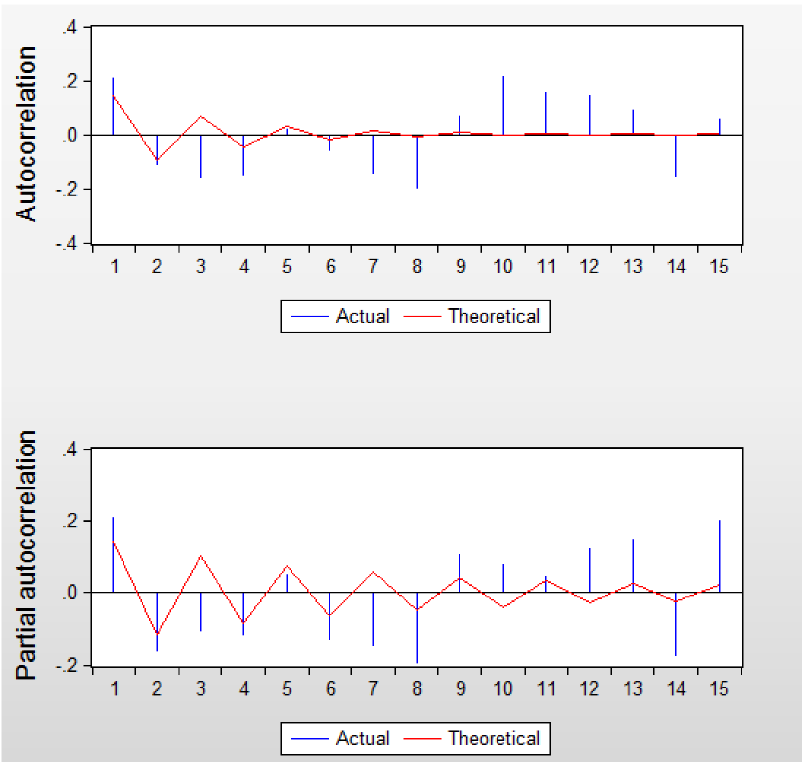
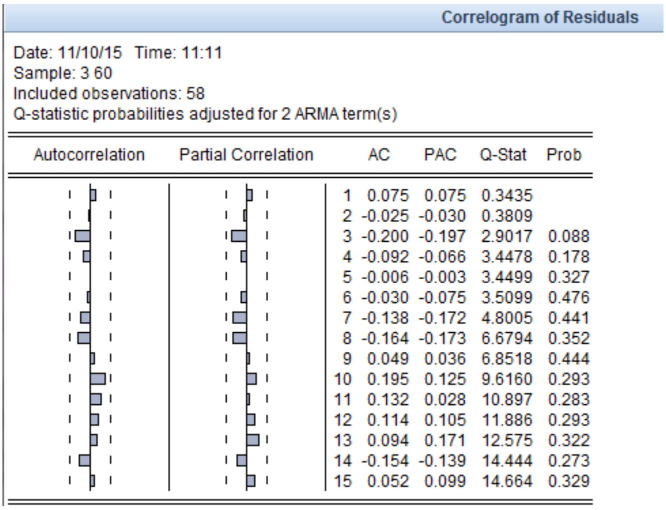


Figure 9. Actual vs. theoretical ACF and PACF for ARMA(1,1)

Also, the correlogram of the residuals, which care establish if they are independent, shows:



Correlogram of Residuals

Date: 11/10/15 Time: 11:11
Sample: 3 60
Included observations: 58
Q-statistic probabilities adjusted for 2 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.075	0.075	0.3435	
		2	-0.025	-0.030	0.3809	
		3	-0.200	-0.197	2.9017	0.088
		4	-0.092	-0.066	3.4478	0.178
		5	-0.006	-0.003	3.4499	0.327
		6	-0.030	-0.075	3.5099	0.476
		7	-0.138	-0.172	4.8005	0.441
		8	-0.164	-0.173	6.6794	0.352
		9	0.049	0.036	6.8518	0.444
		10	0.195	0.125	9.6160	0.293
		11	0.132	0.028	10.897	0.283
		12	0.114	0.105	11.886	0.293
		13	0.094	0.171	12.575	0.322
		14	-0.154	-0.139	14.444	0.273
		15	0.052	0.099	14.664	0.329

Figure 10. Correlogram of residuals for ARMA(1,1)

We can surely say that the residuals represent white noise. So, the ARMA(1,1) model can be suitable for our goal.

ARMA (1,2)

The same steps will be followed. The AC and PAC graph shows:

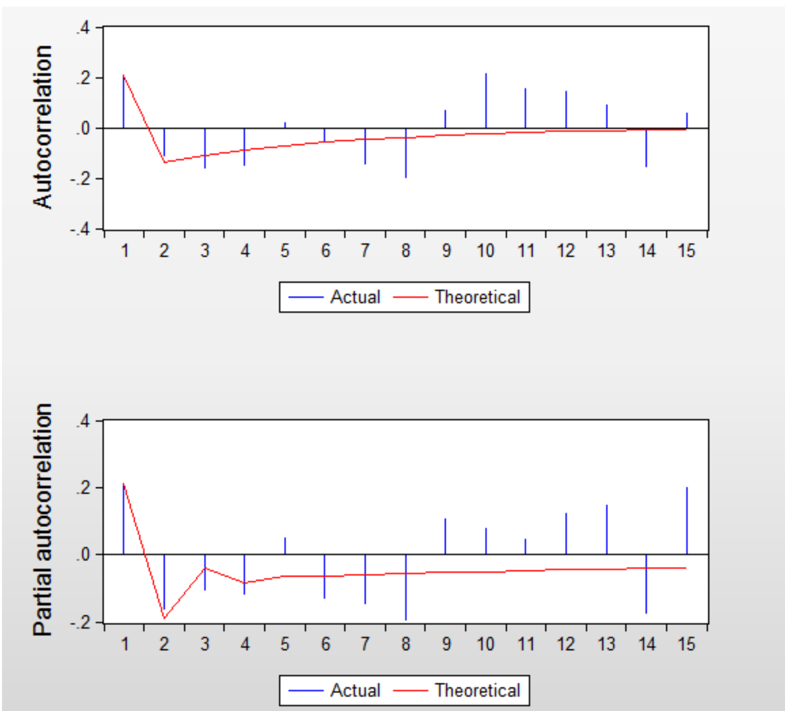
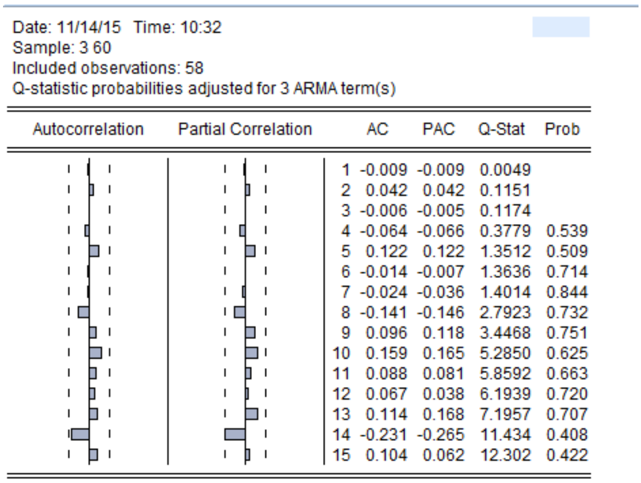


Figure 11. Actual vs. theoretical ACF and PACF for ARMA(1,2)

,not being as exact as ARMA(1,1). Correlogram of the residuals:



Date: 11/14/15 Time: 10:32
Sample: 3 60
Included observations: 58
Q-statistic probabilities adjusted for 3 ARMA term(s)

	AC	PAC	Q-Stat	Prob
1	-0.009	-0.009	0.0049	
2	0.042	0.042	0.1151	
3	-0.006	-0.005	0.1174	
4	-0.064	-0.066	0.3779	0.539
5	0.122	0.122	1.3512	0.509
6	-0.014	-0.007	1.3636	0.714
7	-0.024	-0.036	1.4014	0.844
8	-0.141	-0.146	2.7923	0.732
9	0.096	0.118	3.4468	0.751
10	0.159	0.165	5.2850	0.625
11	0.088	0.081	5.8592	0.663
12	0.067	0.038	6.1939	0.720
13	0.114	0.168	7.1957	0.707
14	-0.231	-0.265	11.434	0.408
15	0.104	0.062	12.302	0.422

Figure 12. Correlogram of residuals for ARMA(1,2)

Also the residual terms are white noise.

MA (1)

For this third model, we obtain:

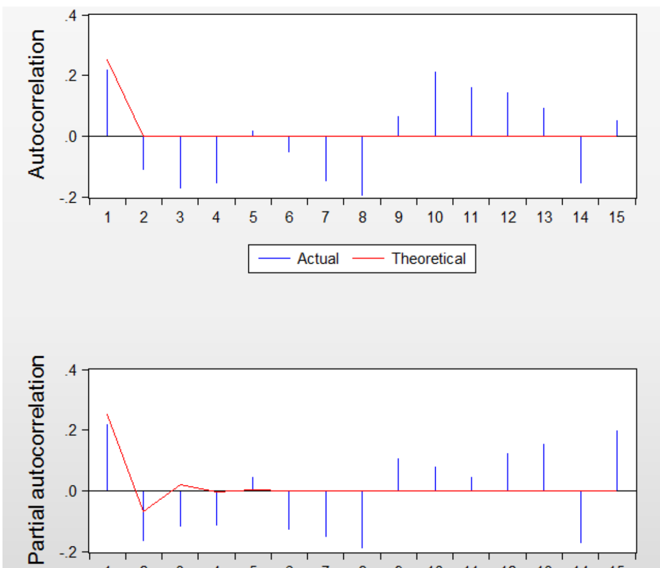
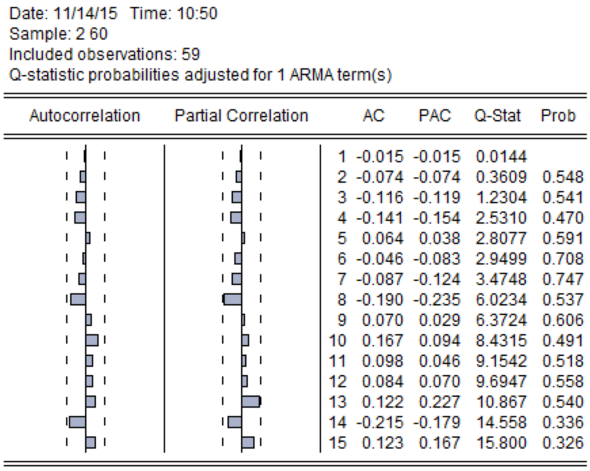


Figure 13. Actual vs. theoretical ACF and PACF for MA(1)

,and:


Date: 11/14/15 Time: 10:50
Sample: 2 60
Included observations: 59
Q-statistic probabilities adjusted for 1 ARMA term(s)

	AC	PAC	Q-Stat	Prob
1	-0.015	-0.015	0.0144	
2	-0.074	-0.074	0.3609	0.548
3	-0.116	-0.119	1.2304	0.541
4	-0.141	-0.154	2.5310	0.470
5	0.064	0.038	2.8077	0.591
6	-0.046	-0.083	2.9499	0.708
7	-0.087	-0.124	3.4748	0.747
8	-0.190	-0.235	6.0234	0.537
9	0.070	0.029	6.3724	0.606
10	0.167	0.094	8.4315	0.491
11	0.098	0.046	9.1542	0.518
12	0.084	0.070	9.6947	0.558
13	0.122	0.227	10.867	0.540
14	-0.215	-0.179	14.558	0.336
15	0.123	0.167	15.800	0.326

Figure 14. Correlogram of residuals for MA(1)

The moving average proves itself to be fit, because it is adequate, and the residuals are shown to represent white noise. The selection is to be made between thee MA(1) and ARMA(1,1) models, both being more appropriate than ARMA(1,2). Additional, MA(1) showed smaller values for Akaike, Schwartz and standard error (S.E.) criteria, and a higher value for adjusted R2 criteria. As a consequence, it is chosen as the best model in the presented conditions.

Predicting the variable

Not forgetting that our model is also first degree differentiated, its form to be used for predicting the future values of BET index is an ARIMA(0,1,1) one, with the following forecasting equation:

$$\Delta BET_t = \xi_t - \theta_1 * \xi_{t-1}$$

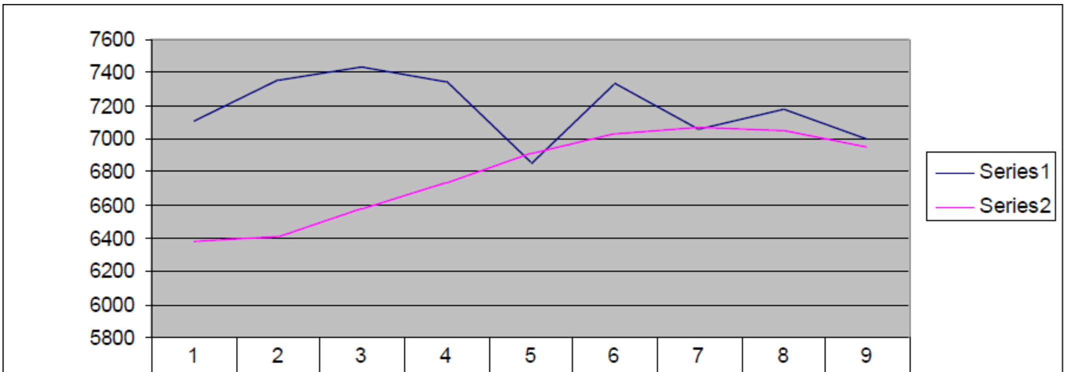
IV. RESULTS AND CONCLUSIONS

The below table presents both the estimated and the real BET values, for the first 9 months of year 2015:

Table 2. Real and predicted values for BET index, first 9 months of 2015

Period	RealBETvalues	Predicted BET values
January 2015	7003, 421	6.376,55
February 2015	7176, 002	6.409,89
March 2015	7060,128	6.576,49
April 2015	7338,491	6.738,57
May2015	6851,280	6.910,46
June 2015	7343,580	7.032,96
July2015	7437,072	7.066,68
August 2015	7357,447	7.050,14
September 2015	7104,513	6.949,77

,and the common plot of the two series (real and estimated) looks like this:



	1	2	3	4	5	6	7	8	9
Series1	7104,51	7357,45	7437,07	7343,58	6.851,28	7338,49	7060,13	7176	7003,42
Series2	6.376,55	6.409,89	6.576,49	6.738,57	6.910,46	7.032,96	7.066,68	7.050,14	6.949,77

Figure 15. Plot of real and predicted values for BET index, first 9 months of 2015

It can be seen that the prediction's performance is satisfactory, so we can say that the used model behaves acceptable. This result proves one more time the good potential of ARMA/ARIMA models to forecast the evolution of prices, on short and medium term. So, this category of models can successfully compete against other estimation techniques [9, 12, 13].

REFERENCES

[1] A. Adebiyi, A. Adewumi, C. K. Ayo – ”Stock Price Prediction Using the ARIMA Model”, UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 2014

[2] M. Al Wadia, I. Tahir - “Selecting Wavelet Transforms Model in Forecasting Financial Time Series Data Based on ARIMA Model”, Applied Mathematical Sciences, Vol. 5, 2011, no. 7, 315 – 326

[3] C. Javier, E. Rosario, J.N. Francisco, J.C. Antonio - “ARIMA Models to Predict Next Electricity Price”, IEEE Transactions on Power Systems vol. 18 no.3, pp.1014-1020, 2003

[4] J. Jeffrey,E. Kyper - "ARIMA modeling with intervention to forecast and analyze Chinese stock prices", International Journal of Engineering Business Management 3.3 (2011), 53-58

[5] K.Mitra - “Optimal combination of trading rules using neural networks”, International Business Research, vol. 2, no. 1, pag. 86–99, 2009

[6] P. Mondal, L. Shit, S. Goswami – “Study of Effectiveness of Time Series Modeling (ARIMA) in Forecasting Stock Prices”, International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.4, No.2, April 2014

[7] M. Mostafa – „Forecasting stock exchange movements using neural networks: empirical evidence from Kuwait”, Expert Systems with Applications, vol. 37, no. 9, pag. 6302–6309, 2010

[8] A. Napitupulu, B. Wijaya, S. Kom - “Stock price prediction: Comparison of Arima and artificial neural network methods—an Indonesia stock’s case”, Proceedings of the 2nd International Conference on Advances in Computing, Control and Telecommunication Technologies, pag. 176–179, December 2010

[9] F. Pieleanu - „Comparative study in estimating Volkswagen’s price: ARIMA versus ANN”, sent for publishing, 2016

[10] V. Prybutok, J. Yi, D. Mitchell - “Comparison of neural network models with ARIMA and regression models for prediction of Houston’s daily maximum ozone concentrations”, European Journal of Operational Research, vol. 122, no. 1, pp. 31–40, 2000

[11] N. Rangan and N. Titida, “ARIMA Model for Forecasting Oil Palm Price”, Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications, University Sains Malaysia, 2006

[12] P. Saxena, N. Merh, R. Pardasani - “A comparison between hybrid approaches of ANN and ARIMA for Indian stock trend forecasting”, Journal of Business Intelligence, vol. 3, no. 2, pag. 23–43, 2010

[13] B. Wijaya, S. Kom, A. Napitupulu - “Stock price prediction: Comparison of Arima and artificial neural network methods—an Indonesia stock’s case”, Proceedings of the 2nd International Conference on Advances in Computing, Control and Telecommunication Technologies, pag. 176–179, Jakarta, December 2010

[14] G. Zhang, B. Patuwo, Y. Hu - “Forecasting with artificial neural networks: the state of the art”, International Journal of Forecasting, vol. 14, no. 1, pag. 35–62, 1998

Links:

http://www.ktd.ro

http://statistici.insse.ro

http://www.bvb.ro/FinancialInstruments/Indices/IndicesProfiles

http://people.duke.edu/~rnau/411arim.htm