PREDICTING THE EVOLUTION OF BET INDEX, USING AN ARIMA MODEL

Florin Dan Pieleanu

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].
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]:

\[
Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p} + \xi_t - \theta_1 \xi_{t-1} - \theta_2 \xi_{t-2} - \ldots - \theta_q \xi_{t-q}
\]

where:
- \(Y_t\) is the time series for the studied variable
- \(\Phi_i\), \(\theta_j\) are the coefficients’ series
- \(\xi_t\) 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:
It can be observed that the series follows a random walk pattern. The correlogram confirms this:
Also, the Dickey-Fuller tests shows:

![Dickey-Fuller test for the original time series](image1)

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:

![Graph of the time series after applying first difference](image2)

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

![Figure 5. Correlogram of the time series after applying first difference](image1)

, with the Dickey-Fuller test:

![Figure 6. Dickey-Fuller test for the time series after applying first difference](image2)

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:

![Figure 7. Results for AR(1) model](image)

...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>
<thead>
<tr>
<th>Model</th>
<th>Akaike</th>
<th>Schwarz</th>
<th>Statistical significance</th>
<th>Adjusted $R^2$</th>
<th>S.E</th>
</tr>
</thead>
<tbody>
<tr>
<td>$AR(1)$</td>
<td>14,014</td>
<td></td>
<td>no</td>
<td>0.031</td>
<td>262,781</td>
</tr>
<tr>
<td>$AR(2)$</td>
<td>13,964</td>
<td></td>
<td>no</td>
<td>0.040</td>
<td>254,138</td>
</tr>
<tr>
<td>$AR(3)$</td>
<td>14,002</td>
<td></td>
<td>no</td>
<td>0.028</td>
<td>256,752</td>
</tr>
<tr>
<td>$MA(1)$</td>
<td>13,984</td>
<td>14,054</td>
<td>yes</td>
<td>0.044</td>
<td>258,892</td>
</tr>
<tr>
<td>$MA(2)$</td>
<td>14,012</td>
<td>14,117</td>
<td>no</td>
<td>0.032</td>
<td>260,447</td>
</tr>
<tr>
<td>$MA(3)$</td>
<td>14,027</td>
<td>14,168</td>
<td>no</td>
<td>0.032</td>
<td>260,433</td>
</tr>
<tr>
<td>$ARMA(1,1)$</td>
<td>14,019</td>
<td>14,126</td>
<td>yes</td>
<td>0.041</td>
<td>261,302</td>
</tr>
<tr>
<td>Model</td>
<td>AIC</td>
<td>BIC</td>
<td>Significance</td>
<td>Parameter Type</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>------</td>
<td>--------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>ARMA(2,1)</td>
<td>13,996</td>
<td>14,139</td>
<td>no</td>
<td>0,025</td>
<td>256,088</td>
</tr>
<tr>
<td>ARMA(1,2)</td>
<td>13,994</td>
<td>14,136</td>
<td>yes</td>
<td>0,080</td>
<td>255,919</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>13,846</td>
<td>14,025</td>
<td>partially</td>
<td>0,174</td>
<td>235,686</td>
</tr>
<tr>
<td>ARMA(2,3)</td>
<td>13,877</td>
<td>14,092</td>
<td>partially</td>
<td>0,161</td>
<td>237,535</td>
</tr>
<tr>
<td>ARMA(3,2)</td>
<td>13,885</td>
<td>14,102</td>
<td>partially</td>
<td>0,163</td>
<td>238,212</td>
</tr>
<tr>
<td>ARMA(3,3)</td>
<td>13,849</td>
<td>14,102</td>
<td>partially</td>
<td>0,205</td>
<td>232,167</td>
</tr>
</tbody>
</table>

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:

![ACF and PACF for ARMA(1,1)](image)

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

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

![Graph showing actual vs. theoretical ACF and PACF for ARMA(1,2)](image)

Figure 11. Actual vs. theoretical ACF and PACF for ARMA(1,2)
, not being as exact as ARMA(1,1). Correlogram of the residuals:

![Figure 12. Correlogram of residuals for ARMA(1,2)](image)

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)](image)

, and:
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 the 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:

\[ \Delta \text{BET}_t = \xi_t - \theta_1 \cdot \xi_{t-1} \]

**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>
<thead>
<tr>
<th>Period</th>
<th>Real BET values</th>
<th>Predicted BET values</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 2015</td>
<td>7003.421</td>
<td>6376.55</td>
</tr>
<tr>
<td>February 2015</td>
<td>7176.002</td>
<td>6409.89</td>
</tr>
<tr>
<td>March 2015</td>
<td>7060.128</td>
<td>6576.49</td>
</tr>
<tr>
<td>April 2015</td>
<td>7338.491</td>
<td>6738.57</td>
</tr>
<tr>
<td>May 2015</td>
<td>6851.280</td>
<td>6910.46</td>
</tr>
<tr>
<td>June 2015</td>
<td>7343.580</td>
<td>7032.96</td>
</tr>
<tr>
<td>July 2015</td>
<td>7437.072</td>
<td>7066.68</td>
</tr>
<tr>
<td>August 2015</td>
<td>7357.447</td>
<td>7050.14</td>
</tr>
<tr>
<td>September 2015</td>
<td>7104.513</td>
<td>6949.77</td>
</tr>
</tbody>
</table>
and the common plot of the two series (real and estimated) looks like this:

![Plot of real and predicted values for BET index, first 9 months of 2015](image)

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


Links:

http://www.ktd.ro

http://statistici.insse.ro

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

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