# RANDOM WALK HYPOTHESIS ON BUCHAREST STOCK EXCHANGE

**Sorina GRAMATOVICI, Corina-Mihaela MORTICI**

Bucharest University of Economic Studies, Romania (sorina.gramatovici@csie.ase.ro, corinamortici@gmail.com)

DOI: 10.19062/1842-9238.2018.16.2.7

***Abstract:*** *The aim of this paper is to examine the random walk in two of the stock indexes of , the Bucharest Stock Exchange (BET and BET Plus). Random walk hypothesis states that stock prices move randomly; as a result, the expected profit for the speculator is zero. Many economists believe that random walk can be applied to test the efficient market hypothesis in the weak level. Early literature used stochastic processes to test whether prices precluded everyone from easy profit and whether prices were following those processes or not. When stock prices do not fluctuate randomly, some investors can use past stock prices to gain an abnormal return. Assuming rationality and risk neutrality, a version "of the efficient market hypothesis states that information observable to the market prior to week t should not help to predict the return during week t". In other words, stock returns are not correlated to one another; consequently, the statistical model of the efficient market hypothesis holds and changes in returns are independent We employ several tests, such as econometric tests, Monte Carlo simulation using AI methods: Naive Bayes' Classifier, K Nearest Neighbors, Support Vector Machines. Daily data on returns covered the period February 2016 – November 2017. These tests support the common results that the random walk theory is valid for the two indexes therefore the Bucharest Stock Market is weak-form efficient.*

***Keywords:*** *Random Walk, Stock Market, Efficient Market Hypothesis, Weak form efficiency, K Nearest Neighbors, Support Vector Machines*

## 1. INTRODUCTION AND MOTIVATION

The Efficient Market Hypothesis (EMH), also known as Random Walk Theory, refers to the efficiency of information on stock markets. In economic literature, the term efficient market is used to explain the dependence between available information and stock price. These concepts were introduced and defined by Eugene Fama [3] in 1970, whose perspective has been that financial market efficiency is being driven by the speed of response and the adjustment of prices to new information in the market.

More specifically, in the context of an efficient market, the prices of the current period's shares should fully reflect the relevant information in order to be able to forecast future prices so that there is no possibility of generating further profits using this information. Therefore, the main criteria in terms of efficiency with respect to what was described are: the extent to which the information is absorbed, the time it takes for it to accumulate, and the type of information so incorporated.

The price of an asset reflects the current value of the revenue it is speculated to generate in the forthcoming period. The expected revenue is influenced by determinants such as risks of volatility, liquidity, or bankruptcy.

While prices are determined and set reasonably, price changes are expected to be random and unpredictable, because new information is unpredictable by its nature. Thus, capital market prices are considered to follow a random walk process.

As examples of random walks, we mention throwing a coin or selecting a sequence of numbers from a random number table. Looking back to the financial markets, the current price is independent and uncorrelated with other evolution patterns of the past price.

Let X be a stochastic variable which follows a random process defined by the following equation

$$X(t+1) = \delta + X(t) + \varepsilon_{t+1} \tag{1}$$

where $\delta$ is the drift parameter, $\varepsilon_{t+1}$ identically and independently distributed prediction error.

The test of efficiency in its weak form has been widely studied in financial literature. Ayadi and Pyun investigated in [1] the prices of stock traded on Korean stock market between January 1984 and December 1988 and proved that The Korean stock market is a random walk. Kim and Shamsuddin [4] report the existence of a random walk for Hong Kong, Japon, Korea and Taiwan and rejected the random walk hypothesis for Indonesia, Malaysia and Phillipines. Lim et al. tested in [5] the efficiency of Shangai and Shenzhen stock markets and concluded that China's stock market has a week form efficiency. More recently, Chaibi [2] tested the weak form efficiency according to two indices of the Hong Kong stock exchange between July 1997 and December 2012 and rejected the random walk hypothesis pursuant to both of these indices. Mishra et al. [7] tested the random walk hypothesis for the Indian stock market using 19 years data on six indexes from National Stock Exchange and Bombay Stock Exchange. They used a unit root test that simultaneously accounts for heteroskedasticity and structural breaks and proved that Indian stock indexes are mean reverting.

The aim of the present paper is to analyze the validity and correspondence with the real markets of the theoretical concept of random walk process. In this respect, the practical illustration of the Romanian capital market case was considered relevant, namely by studying the evolution of two of the most important indices evaluated by the Bucharest Stock Exchange, i.e. BET and BET Plus using daily data throughout a period of two years. The random nature of these time series strictly corresponds to the real case if and only if the market is efficient in the weak form. Through the current paper, we intend to study the efficiency of the Romanian capital market, contributing to the results obtained in the existing literature.

Evaluating shares and stock indices is a crucial function of the financial markets, as it leads to the possibility of players' making investment strategies. Evaluating the value of shares is of particular importance to determine the behavior of markets, its behaviour being possible if and only if the type of efficiency is known. Remark that the available information is crucial because it can lead to arbitrage. Acquisition and sale of similar share simultaneously on two different markets as a result of its differences in price conceptualizes the idea of arbitrage. The effect of arbitration plays an essential role in the efficiency of a market because this phenomenon brings prices back to their intrinsic value.

If markets are efficient in a weak sense, it is not possible for players to buy a share whose price is underestimated and to sell them on other markets, where they are fairly valued or overestimated. The very event described makes it impossible for the players involved to „beat the market”.

In this respect, the obvious question is still: why do investors analyze the market with sophisticated and time-consuming tools if their efforts are futile?

This seems to be the main concern of the current paper, since totally rational investors would not play or invest if they did not have the chance to beat the market. As Lo, Mamaysky and Wang [6] asserted, "With the help of sophisticated non-parametric techniques ... [analysts] would only enjoy a modest prediction power", often insufficient to play based on the fundamentals of these overwhelming strategies.

This paper is structured as follows. Section 2, which is preceded by the present introduction, covers the theoretical and methodological aspects that are the solid foundation of the case study undertaken in Section 3. Both sections deal in a multidisciplinary manner with the characteristics of the time series studied in order to test the existence of the "random walk" phenomenon, then to make predictions using econometric methods, but also using artificial intelligence. The final section summarizes the conclusions of the analysis – its significance and validity, and the author's contribution to the literature.

The period of February 15, 2016 - November 2, 2017 has been analyzed, namely the data on BET and BET Plus stock exchange index values from the Bucharest Stock Exchange. The data was collected from the official website of the Bucharest Stock Exchange (www.bvb.ro) and processed by the authors to perform the relevant tests in order to achieve the above mentioned objective.

This paper is outlined by the analysis and tests performed on the BET and BET Plus time series assuming rationality and risk neutrality, and not taking into account transaction costs and other costs that may be charged for players on the capital market. The analyzed period (435 observations) was the subjective decision of the author, and no attempt was made in order to identify an optimal dimension of time to observe indexes' patterns.

## 2. METHODOLOGY

**2.1 The phenomenon of "ARCH Effect".** Financial time series are frequently characterized by volatility, a phenomenon that is modeled by processes such as ARCH. The ARCH effect defines the hypothesis of financial market's speculators estimating the variance over a certain period of time with the information that has appeared in the previous period, and included in the model by the term ARCH. This context describes the well-known concept of "volatility clustering", which means that periods of high magnitude change are followed by periods of small fluctuations. In other words, significant changes in financial time series tend to cluster together, and low magnitude changes are of the same behavior. If the data series is affected by the ARCH, term, they might be predictable to a certain extent and respond to market speculation (a well-known example would be the weekend effect).

By testing the presence of this phenomenon with the ARCH LM test, we introduce the null test hypothesis

$H_0$: *There is no arch effect*

and the alternative

$H_1$: *There is an arch effect*

The regression to be estimated is given by the following equation.

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2 + v_t \tag{2}$$

where $u_t$ represents the residue of the initial regression estimated by the ordinary least squares method.

The statistic test is $T * R^2$, where T is the number of observations included in the analysis, and $R^2$ is the coefficient of determination of the initial regression. It follows a Chi-square distribution with p degrees of freedom.

**2.2 Naïve Bayes Classifier.** The naive Bayesian algorithm is a classification technique based on Bayes' theorem, based on the hypothesis of independence of predictors. In other words, the naive Bayes Classifier assumes that the presence of a certain characteristic in a class is uncorrelated with the presence of any other characteristic. For example, a fruit is identified as an apple if it is red, round, and has a diameter of about 5 cm. Although these characteristics depend on one another or, more clearly, the existence of each characteristic is dependent on the existence of the other, all these properties contribute, in an independent manner, to the likelihood that this fruit being called apple - and that is why this classifier is called "naive".

The naive Bayesian classifier is useful for large data sets, and, despite the simplicity of the assumptions on which it is based, it is known that it provides better performance than other complex classification techniques, and therefore has been included in this study.

Bayes' theorem represents a method to calculate the posterior probability P (c/x) based on the posterior probability of the class P(c), the posterior probability of the predictor P(x) and the probability of a predictor in a given class P(x/c), as follows.

$$P(c|x) = \frac{P(x|c) * P(c)}{P(x)} \tag{3}$$

Where

$$P(c|X) = P(x_1|c) * P(x_2|c) * \ldots * P(x_n|c) * P(c) \tag{4}$$

**2.3 K Nearest Neighbors.** The KNN method is a classification algorithm that includes all the observations and classifies the new observations generally based on a measure of similarity, and in most cases, this is a function of distance. The KNN has been used since the 1970s in statistical estimations and in the pattern recognition as a non-parametric technique.

The algorithm assumes that a new observation is classified based on the vote of its nearest K neighbors. A new instance is assigned to the most common class among its neighbors based on the distance function. The most widely used distance functions are: Euclidean distance, Manhattan distance and Minkowski distance, for continuous variables, and Hamming distance for discrete variables.

The decision to determine the optimum value for K is to be taken by inspecting the data series. Generally, a high K value leads to more accurate results, as it reduces noise. The cross validation procedure is another method by which the value of K is computed.

Based on the empirical studies already performed, for most data sets, a value of K between 3 and 10 produces satisfactory results and, moreover, more efficient results than the 1NN method.

**2.4 Monte-Carlo Simulation.** The Monte-Carlo simulation method is used in many branches of science in order to quantify the expectations of the evolution of a variable of an index whose behavior is similar to a random walk process.

In this case, however, no analytical function can describe its evolution, and the optimal decision is to generate random samples describing these variables course of action. The accuracy of the estimates obtained by the Monte-Carlo simulation method is inversely proportional to the number of extractions.

In the current study, the type of simulation used is called "time-driven", in the sense that, for this given period, we have built different scenarios that can lead to the fluctuations in the analysis.

If a stock market player uses the Monte-Carlo simulation for a past period, for which the evolution of the index in the analysis is already known, they will be aware of the various trajectories that this index could have followed, and thus the magnitude of the risk assumed it by choosing a discreet strategy among all possible. Computing a strategy, they certainly take need to take into account both the risk prize and the magnitude of a potential loss. Loss has to be taken into account, and it plays a crucial role in the computation of the expected profit, because by visualizing the many trajectories that an index may follow, ~~it~~ the high probability of the monetary loss involved can obviously be deduced. However, the more players are involved in capital market's game, the greater the risk, and so, the importance of using the Monte-Carlo method is obvious. The Monte-Carlo simulation is even superior to "What If" analysis, because in many cases, it is very difficult to identify or test the determinants of fluctuations in an index on the stock market. Moreover, when making a decision, it is especially important to include a graphical view of the various scenarios, so that the decision maker may become~~s~~ aware of the probability associated with the occurrence of each state of nature.

**2.5 Support Vector Machine.** Support vector machine is a supervised learning algorithm that can be used in regressions, but is especially suitable for classifications. In this method, each observation is represented in an n-dimensional space (where n is the number of states of the variable), the value of each state being represented by a coordinate. Classification is carried out by determining the hyperplane that segregates these classes, and in our investigation, the separation of the two classes.

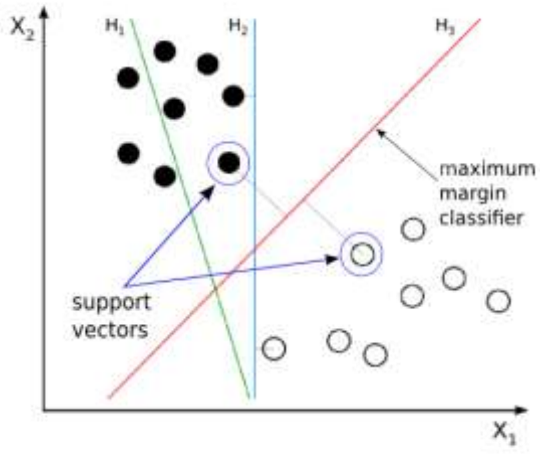


**FIG. 1.** Support Vector Machine representation [10]

Support vectors are simply the coordinates of an individual observation, and the support vector machine is the hyperplan, the border that optimally segregates the two classes.

In order to correctly determine this hyperplane, the shape of the nucleus must be computed. For most financial data sets, it is linear. Also, the gamma coefficient is associated to the kernel; the higher the value of the kernel, the more the algorithm will try to classify a new observation only if it has an almost perfect identification with a particular class.

However, the disadvantage of a too high gamma coefficient is that some of the observations will not be classified.

## 3. EMPIRICAL RESULTS

**3.1 ARIMA-Modelling and Seasonality.** ARIMA-modelling the evolution of stock indices is appropriate, as, mainly, in the context of financial time series there are trends, seasonality, errors, shocks, all these factors being taken into account in this type of model. To perform the procedure, we used the stationarized data series. For the analyzed period, we included monthly BET and BET Plus indices in order to capture the phenomenon of seasonality along with the influence on the evolution of time series. Moreover, we considered it important to add dummy variables corresponding to each month of the year in the analysis - to test whether they are statistically significant or not. Although, generally, we have to include 12 seasonal factors, we chose to exclude December, because, including the intercept, the dummy trap would have occurred.

The ultimate goal of the analysis is to predict the average values of the indices over the next two months based on the chosen model.

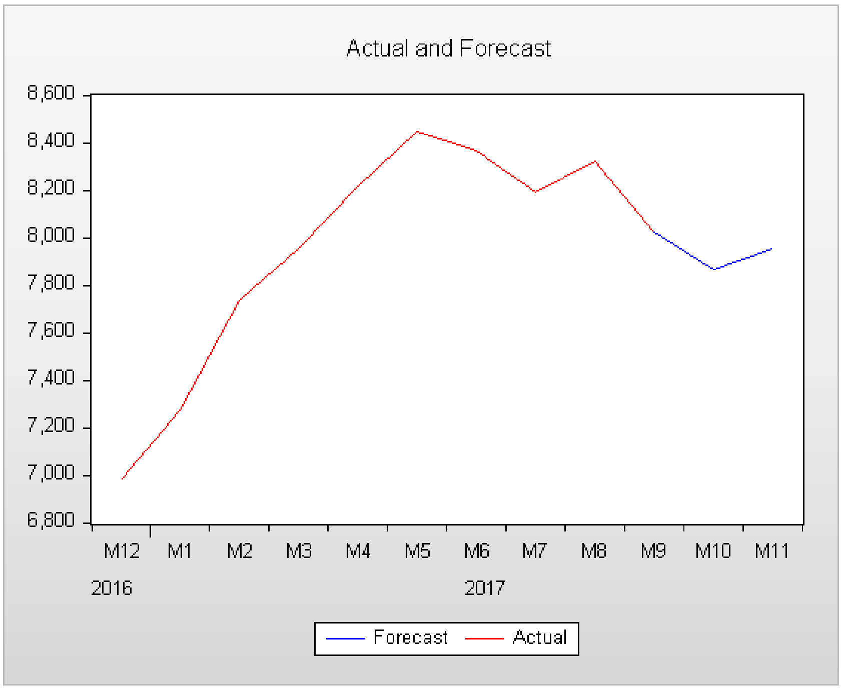


**FIG.2.** ARIMA Modelling

The most appropriate model was chosen using the Akaike informational criterion, out of 100 estimated candidate models. The minimum value (13.6714) corresponds to the most effective model. Proceeding with this analysis, the current values and the predicted values are presented in the table below.

Table 1 - Real values vs. predicted values

| Month | Average real value | Average predicted value |
|---|---|---|
| October | 7946.01 | 7868.38 |
| November | 7783.36 | 7955.372 |

As these values differ significantly from the real ones, and moreover, if the BET index actually shows, on average, a decrease from October to November, and using ARIMA modelling, it was forecasted to increase by almost 100 we can safely conclude that, even if the presence of the seasonal factors in the analysis is taken into account, the evolution of the Bucharest Stock Exchange still cannot be predicted, and therefore, there is a random walk.

**3.2 Correlation Between BSE and international stock markets**. Furthermore, we consider crucial to analyze the correlation between the evolution of the main indices of the Bucharest Stock Exchange and the evolution of the international stock markets. In this respect, it is relevant to observe the fluctuations of the US S&P500 index, for the same period.

By standardizing the data, we achieved a very strong correlation of 91.67% with the BET index and 92.29% with BET Plus.

Initially, not considering the seasonal factors, we attempted to predict the BET and BET Plus indices based on their own previous day value (taking into account the short-term dependency, the concept called volatility clustering) and the value of the previous day of the S&P500 index. The hypothesis is that, including a small lag over time, the indices studied follow, however, faithfully, the evolution of the main stock exchanges on the international level.

We present below the estimated models along with their most important properties.

$$DBET_t = 3.08 + 0.26 * DSP_{t-1} + 0.99 * DBET_{t-1} + \varepsilon_t$$

$$DBETPLUS_t = 1.27 + 0.26 * DSP_{t-1} + 0.99 * DBET_{t-1} + \varepsilon_t$$

Table 2 - Statistical properties of models

| Index | Root Mean Square Error | F-statistic | Probability of Fischer-Snedecor test |
|---|---|---|---|
| BET | 0.107 | 29463.05 | 0.000 |
| BET Plus | 0.108 | 32068.32 | 0.000 |

Although the estimated models are statistically significant, the forecast of the two indices based on them was not performing well. It is obvious that, using this type of technical analysis of the market, one cannot beat the market. It is to be noticed that the predicted values correspond to the actual values in very few days of the period considered. Moreover, for both indices, the upward trend of their evolution from the beginning of the period studied until the beginning of the third quarter of 2017 was validated. Otherwise, neither the magnitude nor the sign of the fluctuations were correctly predicted.

However, in order to improve this analysis, we explicitly admitted the existence of trends and included in the two models the effect of the seasonal factors in the attempt to explain precisely the fluctuations of the data series. We excluded the December analysis as a seasonality factor to avoid the dummy trap. The models obtained are as follows.

$$DBET_t = 7.47 + 0.06 * DSP_{t-1} + 0.01 * TR_t + 0.02 * mth1 + 0.06 * mth2 + 0.09 * mth3 + 0.09 * mth4 + 0.07 * mth5 + 0.05 * mth6 + 0.03 * mth7 + 0.04 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11 + \varepsilon_t$$

$$DBETPLUS_t = 3.49 + 0.04 * DSP_{t-1} + 0.08 * TR_t + 0.02 * mth1 + 0.01 * mth2 + 0.05 * mth3 + 0.09 * mth4 + 0.06 * mth5 + 0.01 * mth6 + 0.03 * mth7 + 0.02 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11 + \varepsilon_t$$

Both models are statistically significant, but what is relevant in the analysis is that the estimation confirmed the significant effect of the seasonal factors. However, based on the probability associated with the t-Student test, it is to be noted that the seasonal effects of October and November are not significant. Based on these models, the results of the forecast are graphically represented below.


**FIG. 3.** Prediction of BET index based on short-past evolution of S&P500


**FIG. 4.** Prediction of BET Plus index based on short past evolution of S&P500

The graphs presented certainly point out to a more qualitative forecast, certainly due to the inclusion of seasonal factors, but also to the fact that the chosen method is static, which means that the predicted value at a certain iteration is based only on past real values of the index, and not the previously predicted values. Thus, in quantitative terms, we conclude that the predicted values of the BET index in the analyzed period correspond to 96.64% with the real values, and the predicted values of the BET Plus index are correlated with the real values in a higher proportion of 96.91%.

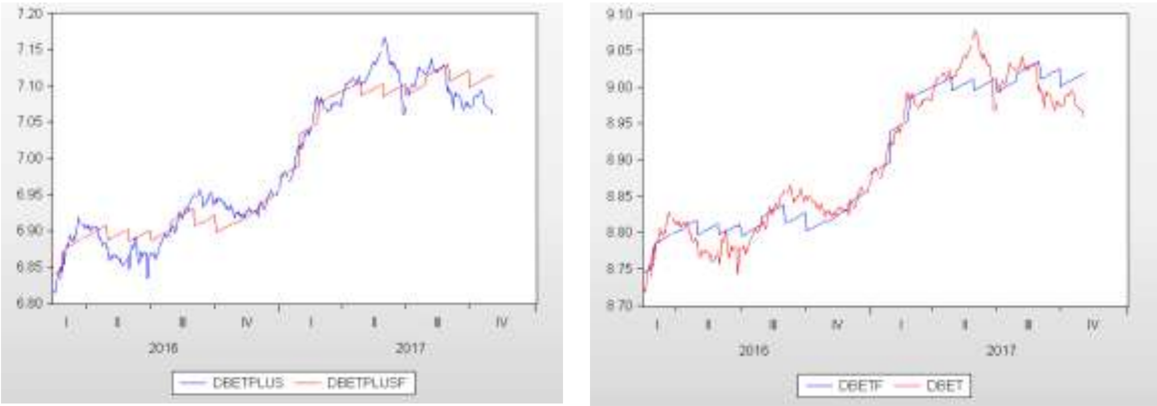


**FIG. 5.** BET Plus and BET index prediction based on seasonality and short term evolution of NYSE index

Based on these results, the international players' evolution is certainly an important determinant of the evolution of the Bucharest Stock Exchange and we naturally want in a more detailed analysis to test whether the BET and BET Plus indices are influenced by the recent-short or long-term evolution of the S&P500 US market index. In this attempt, we have improved the model by taking into account the values for the last 5 days of the S&P500 index, thus obtaining the following estimation.

$$DBET_t = 8.92 + 0.04 * DSP_{t-1} + 0.05 * DSP_{t-2} + 0.04 * DSP_{t-3} - 0.02 * DSP_{t-4} + 0.12 * DSP_{t-5} + 0.01 * TR_t + 0.02 * mth1 + 0.06 * mth2 + 0.09 * mth3 + 0.09 * mth4 + 0.07 * mth5 + 0.05 * mth6 + 0.03 * mth7 + 0.04 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11 + \varepsilon_t$$

$$DBETPLUS_t = 7.18 + 0.03 * DSP_{t-1} + 0.04 * DSP_{t-2} + 0.04 * DSP_{t-3} - 0.02 * DSP_{t-4} + 0.11 * DSP_{t-5} + 0.01 * TR_t + 0.02 * mth1 + 0.06 * mth2 + 0.09 * mth3 + 0.09 * mth4 + 0.07 * mth5 + 0.05 * mth6 + 0.03 * mth7 + 0.04 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11 + \varepsilon_t$$

In this approach, we note that, based on the Wald test, these newly introduced coefficients are not significantly different from 0 for both of the time series. Actual values and predicted values are shown below.

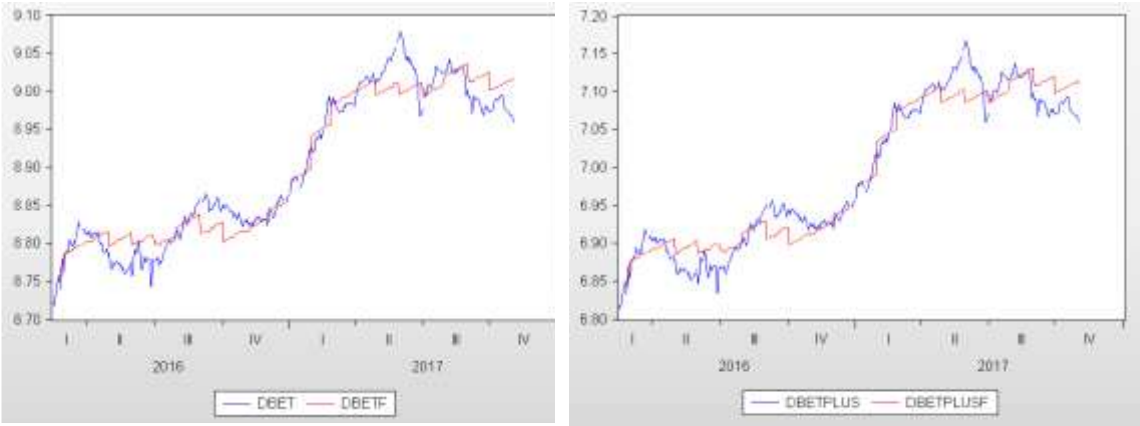


**FIG. 6.** Predictions on BET and BET Plus index based on past 5-days evolution of NYSE index

For the BET index, the correlation coefficient between the actual values and the predicted values is 96.57% and for the BET Plus index it is 96.85%. In both cases, predictions based on the recent past value are better than considering a longer period. This behavior is probably due to the dynamic evolution of stock market indices; so, even if some shocks emerge, they quickly disappear and it would not be effective to include more past days in predicting the indexes in the analysis.

Finally, the most important determinant in the forecast being clearly the inclusion of the seasonality phenomenon, we added to the model that includes the seasonal factors a moving average term, an autoregressive term, and finally an AR and a MA term simultaneously. Based on the initial model and the three additional models, we have computed the average predicted values of the BET index evolution.

The models obtained are statistically significant, considering the validity of the Fischer-Snedecor test, and they take the following form.

$$
\begin{aligned}
DBET_t = 3.31 &+ 0.98 * DBET_{t-1} + 0.25 * DSP_{t-1} + 0.01 * TR_t + 0.01 * mth1 + 0.01 \\
&* mth2 + 0.01 * mth3 + 0.01 * mth4 + 0.01 * mth5 + 0.01 * mth6 \\
&+ 0.01 * mth7 + 0.01 * mth8 + 0.01 * mth9 + 0.01 * mth10 + 0.01 \\
&* mth11 + \varepsilon_t
\end{aligned}
$$

$$
\begin{aligned}
DBET_t = 6.43 &+ 0.85 * \varepsilon_{t-1} + 0.11 * DSP_{t-1} + 0.01 * TR_t + 0.01 * mth1 + 0.01 \\
&* mth2 + 0.01 * mth3 + 0.01 * mth4 + 0.01 * mth5 + 0.01 * mth6 \\
&+ 0.01 * mth7 + 0.01 * mth8 + 0.01 * mth9 - 0.01 * mth10 - 0.01 \\
&* mth11 + \varepsilon_t
\end{aligned}
$$

$$
\begin{aligned}
DBET_t = 3.45 &+ 0.98 * DBET_{t-1} + 0.01 * \varepsilon_{t-1} + 0.24 * DSP_{t-1} + 0.01 * TR_t + 0.01 \\
&* mth1 + 0.01 * mth2 + 0.01 * mth3 + 0.01 * mth4 + 0.01 * mth5 \\
&+ 0.01 * mth6 + 0.01 * mth7 + 0.01 * mth8 + 0.01 * mth9 + 0.01 \\
&* mth10 + 0.01 * mth11 + \varepsilon_t
\end{aligned}
$$

The overall representation of these time series together with the predictions obtained are presented below.

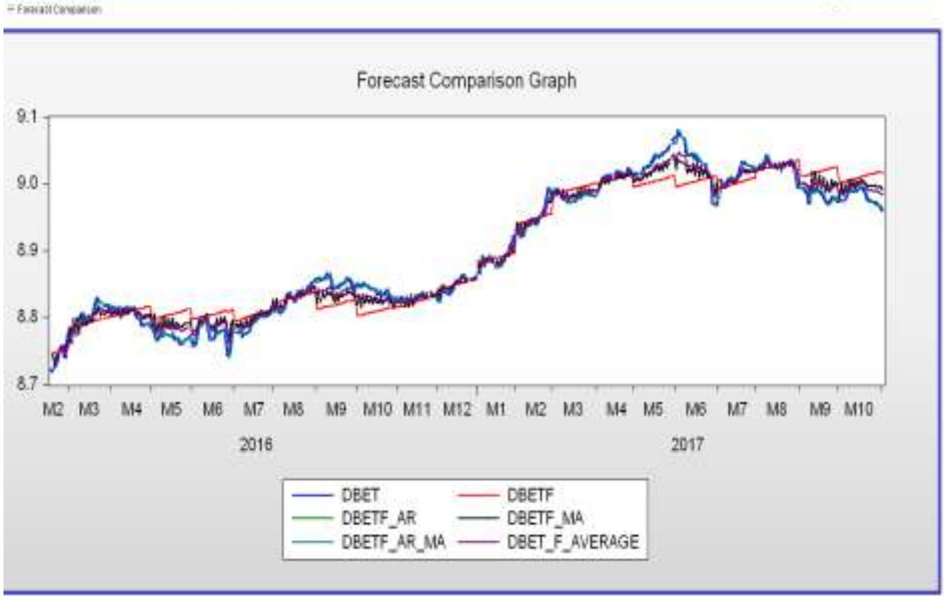


**FIG. 7.** Comparison between methods

In an attempt to achieve higher performance, predictions were computed by means of the static method, so only real values were taken into account. The average of the predictions is thus the most elaborate methodology among those tested in this study to predict the BET and BET Plus indices. The average of predictions along with real values are shown in FIG. 8.

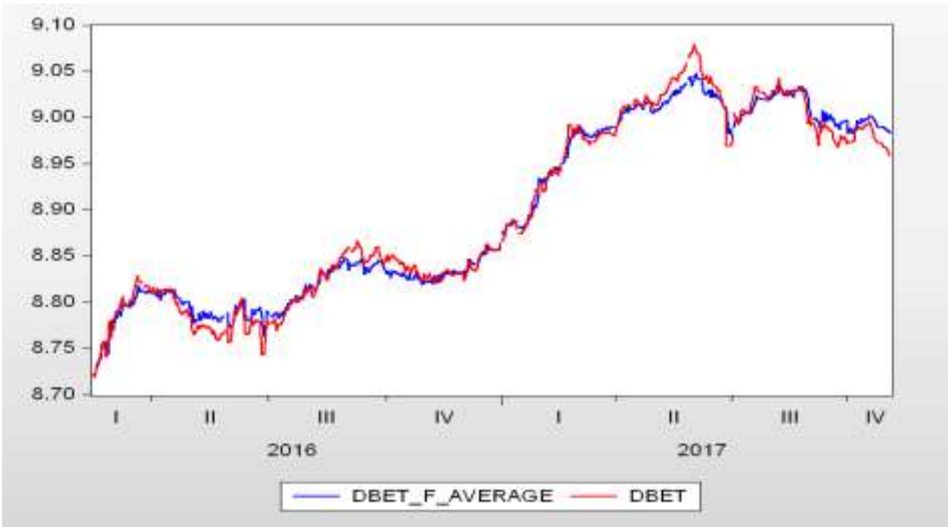


**FIG. 8.** Actual vs predicted values based on Forecasting Average best method

Although the Forecasting Average model based on the four sub-models is the most qualitative, the differences between the actual and the predicted values are still significant, with no significant profitability to be obtained if one uses analytical forecasting techniques. The predicted values increase when the real ones increase as well, and they decrease when the real ones show a negative evolution, but the magnitude of the fluctuation differs, sometimes significantly.

Moreover, as we have previously proved, it is not more significant to consider the last 5 days of the S&P500 than the previous day exclusively; so, any forecast obtained using analytical methods is anyway valid in the short term, but the evolution will instantly change significantly as compared to the one expected by speculators.

Once again, we have demonstrated that there are no profitable opportunities for analysts. Not even including correlation to international stock markets, we cannot predict the evolution of stock indices in order to gain surprofit.

As the quality of the forecast has increased considerably when considering the seasonal effect, we conclude that, it is important to thoroughly study and take into account all the determinants that exist in capital market. Even in this ideal case, shocks will remain unpredictable. However, the costs of these analyzes, as well as the transaction costs and other costs that appear on the real financial market, make these analyzes meaningless when players want ~~a~~ significant profit.

**3.3 Predictions based on learning techniques – Naïve Bayes' Classifier and KNN.** The naive Bayesian classifier is a prediction technique whereby the data that compose the training set (70% of the data set) is converted into a frequency table. The probability associated with each event (selling or keeping the portfolio in the same structure – hold – as considered optimal strategy) is calculated and then computed the probability table based.

Using the Bayes' posterior probability formula, these probabilities are computed for all classes – the posterior probability for the sell strategy and the posterior probability for the hold strategy. Being the first index of Bucharest Stock Exchange, we considered relevant to include the BET index in the analysis.

A simple strategy is as follows: if there are positive fluctuations, the player will choose to choose the Buy event; otherwise, one will choose the "Hold" event. In fact, considering the Naive Bayes classifier, at each iteration, the class with the highest posterior probability will be chosen. Based on these assumptions, the observations in the test set (30% of the dataset) predicted the correct strategy in 62% of the cases.

```
Confusion Matrix and Statistics


preds  Buy Hold
  Buy   80   50
  Hold   0    1

               Accuracy : 0.6183
                 95% CI : (0.5294, 0.7018)
    No Information Rate : 0.6107
    P-Value [Acc > NIR] : 0.4669

                  Kappa : 0.0238
 Mcnemar's Test P-Value : 4.219e-12

            Sensitivity : 1.00000
            Specificity : 0.01961
         Pos Pred Value : 0.61538
         Neg Pred Value : 1.00000
             Prevalence : 0.61069
         Detection Rate : 0.61069
   Detection Prevalence : 0.99237
      Balanced Accuracy : 0.50980

       'Positive' Class : Buy
```

**FIG. 9.** The performance of Naïve Bayes classification algorithm

Furthermore, based on the confusion matrix in FIG. 9., 80 observations were predicted in the correct "Buy" class, and only one observation was predicted correctly in the "Hold" class. However, Cohen's coefficient shows a very low value of 2%. This shortcoming can be corrected by trying, for instance, to redefine the category variable (i.e., the strategy), adding more variables in the analysis, or even selectively choosing the determinants that certainly the strategy decision.

As for the k nearest neighbors method, we used the same test set and training set, and for both methodologies, we set the same seed to ensure the possibility of comparing their performance. The KNN method is one of the pattern recognition procedures, therefore, based on observations from the training set, for which we consider the real class known, we calculated the Euclidean distance between the observations in the test set and those in the training set. By initially setting k = 3, a new observation is classified based on the simple majority of the classes of the nearest three neighbors. For k = 10, we proceeded analogously and the results are provided below.

```
Confusion Matrix and Statistics

preds  Buy Hold
  Buy   42   19
  Hold  38   32

               Accuracy : 0.5649
                 95% CI : (0.4755, 0.6512)
    No Information Rate : 0.6107
    P-Value [Acc > NIR] : 0.87753

                  Kappa : 0.1428
 Mcnemar's Test P-Value : 0.01712

            Sensitivity : 0.5250
            Specificity : 0.6275
         Pos Pred Value : 0.6885
         Neg Pred Value : 0.4571
             Prevalence : 0.6107
         Detection Rate : 0.3206
   Detection Prevalence : 0.4656
      Balanced Accuracy : 0.5762

       'Positive' Class : Buy
```

```
Confusion Matrix and Statistics

preds  Buy Hold
  Buy   44   24
  Hold  36   27

               Accuracy : 0.542
                 95% CI : (0.4527, 0.6293)
    No Information Rate : 0.6107
    P-Value [Acc > NIR] : 0.9547

                  Kappa : 0.0762
 Mcnemar's Test P-Value : 0.1556

            Sensitivity : 0.5500
            Specificity : 0.5294
         Pos Pred Value : 0.6471
         Neg Pred Value : 0.4286
             Prevalence : 0.6107
         Detection Rate : 0.3359
   Detection Prevalence : 0.5191
      Balanced Accuracy : 0.5397

       'Positive' Class : Buy
```

**FIG. 10.** Performance of the KNN method (K=3 and K=10)

Thus, we conclude that the KNN algorithm is more efficient than the naive Bayes classifier. Although accuracy is lower, it is important that the value of Cohen's coefficient increases considerably, so that for every class the ratio between the correctly predicted values and all values belonging to that class is balanced.

However, it is remarkable that considering only three neighbors, the quality of the algorithm was higher than choosing ten neighbors. This may be, for instance, due to the fact that the choice of the class for a non-classified observation by the simple majority determines that choosing an odd number of closest neighbors is certainly a rational decision.

**3.4 Monte-Carlo Simulation of the behavior of Bucharest Stock Exchange.** In an attempt to establish a stock market optimal strategy, a player must consider all the trajectories that a stock index may follow in a given period of time. In this analysis, we chose to track the potential fluctuations of the BET index, being the first index of the Romanian stock exchange, which has highlighted the performance of the most liquid and active ten traded companies since the beginning of the BSE.

We considered it important not only to observe its past values but, above all, the vast trajectories that it could have followed in the context of different states of nature.

By illustrating them as shown in FIG. 11, a decision maker whose purpose is to obtain profit from stock market games, will make rational decisions, because by conducting this analysis, one can compute the probability of assuming that a particular event, of all possible events, will take place and will generate the expected profit.

In the context of such a dynamic system, this assumption is crucial, because if another event occurs, it will not bring the expected revenue or, in a pessimistic way of thinking, it can even produce significant losses.

In fact, the player should propose a strategy, quantifying the effects of a possible error in which the expected scenario will not occur.

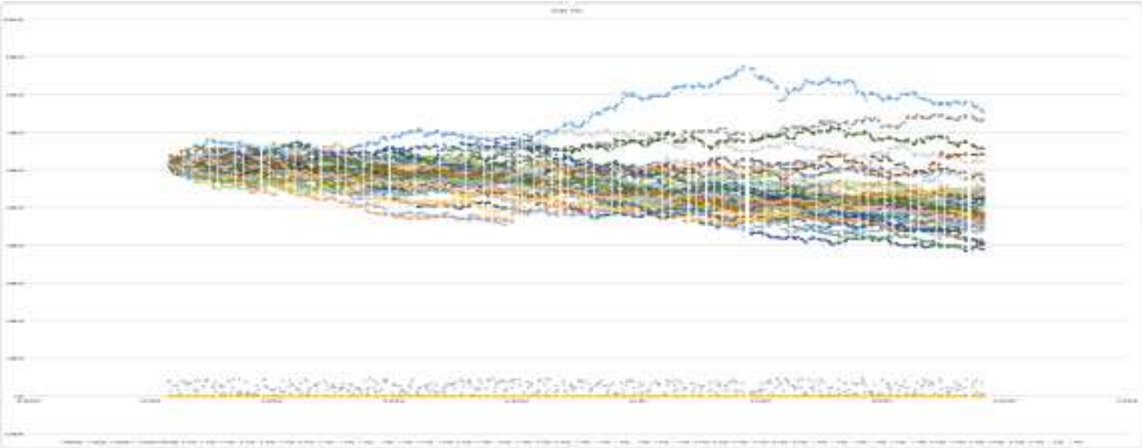

**FIG. 11.** Monte Carlo simulation of possible trajectories of BSE

**3.5 Predictions based on Support Vector Machine learning technique.** Predictions based on vector-based machines imply the division of the dataset into two subsets, as follows: one-third represents the test set, and two-thirds constitute the training set. To train the VSM, we considered the kernel to be linear, as in almost all cases, this type of kernel is suitable for the financial data series.

For the entire dataset, an UPDOWN categorial variable was included, so that when the BET index shows positive fluctuations, the variable takes the value Up, and when decreases occur, the variable takes the value Down.

Based on the training set, the support vector machine was trained to be able to predict the daily BET index fluctuations for the period included in the test set.

Subsequently, the results obtained based on the supervised learning algorithm were compared with the actual fluctuations.

```
[1] 145
> na.omit(settestare[,2])
  [1] Up   Up   Up   Up   Down Up   Down Down Up   Down Down Down Up   Up   Down
 [16] Up   Down Up   Down Up   Up   Up   Up   Down Up   Up   Down Down Up   Up
 [31] Down Up   Up   Up   Up   Down Down Up   Down Up   Down Down Up   Down Down
 [46] Up   Up   Down Up   Up   Up   Down Up   Down Down Down Up   Down Down Up
 [61] Up   Up   Down Down Up   Up   Up   Down Down Down Up   Down Down Down Down
 [76] Down Down Down Up   Up   Up   Down Up   Up   Down Up   Down Down Down Up
 [91] Up   Up   Down Up   Down Up   Down Down Up   Up   Up   Up   Up   Up   Up
[106] Down Down Up   Down Up   Up   Up   Up   Up   Up   Down Down Down Up   Down
[121] Down Up   Down Up   Up   Down Down Up   Down Up   Up   Down Down Up   Down
[136] Up   Up   Down Down Down Up   Up   Up   Up   Down
Levels: Down Up
> classAgreement(confuzie)
$diag
[1] 0.537931
```

**FIG. 12.** Performance of Support Vector Machine learning algorithm

The prediction in FIG. 12. coincide 53.79% with the actual results for the 145 observations included in the test set. Although the algorithm has made accurate predictions in more than half of all cases, it does not guarantee that there will be profitable opportunities.

However, the purpose of this analysis was to prove the possibility of predicting the sign of the Bet indexes, and not the magnitude of these fluctuations.

This distinction should be highlighted as a decision-maker cannot compute the expected monetary value of making a decision (choosing a particular strategy), even if assuming that they ~~he~~ know for sure whether or not the index will show a positive or negative trend the next day. Stock market players generally have a varied portfolio of shares that can include both shares whose evolution follows the market, and other that do the contrary, so the magnitude of a change in the evolution of the BET index is of considerable importance. We can therefore conclude that the accurate forecast of 53.79% of the BET index trajectory does not add value to the market players' strategies.

## CONCLUSIONS

As proven in this paper, the challenge of the random walk theory on the Romanian stock market, to its players is outlined by the following context: if markets are efficient, then the prices of the shares at any time will represent the consistent estimation of their intrinsic value. In this respect, the fundamental analysis is useful if and only if the analyst has new information, which is not yet available on the market, so it was not considered in the formation of current prices. If the analyst does not have new information or contexts not yet exploited in the market, then the optimal decision should be choosing shares in a portfolio or transaction through a purely random procedure.

In essence, the tests performed were not able to reject the hypothesis of describing the evolution of prices on the Romanian stock market as a random walk process. Further work may be developed. The present study could be extended in any of the following directions.

- A random walk study can be carried out in the context of efficient market using data from the Sibiu Stock Exchange, in order to supplement the framework for the Romanian stock market.
- The same analysis can be carried out using other analysis tools to improve the results and conclusions obtained in this paper.
- The period under analysis can be extended in order to achieve more consistent results.

## REFERENCES

[1] Ayadi, O.F., Pyun, C.S., *„An application of variance ratio test to the Korean securities market"*, Journal of Banking and Finance, 18(4),pp.643-658

[2] Chaibi L.F., *„ Testing the random walk: The case of Hong Kong stock exchange"*, Journal of Empirical Studies,2014, 1 (2), pp.54-61

[3] Fama E., *„ Efficient capital market: A reviw of theory and empirical work"*, Journal of Finance, 25(2), 1970, pp.387-417

[4] Kim, J.H., Shamsuddin A., *„Are Asian stock market efficient? Evidence from new multiple variance ratio tests"*, Journal of Empirical Finance,15(3), 2008, pp.518-532

[5] Lim, P.K., Habibullah, M.S., Hinich, M.J., *„ The weak form efficiency of Chinese stock markets: Thin trading, nonlinearity and episodic serial dependencies"*, Journal of Emerging Market Finance, 8(1), 2008, pp.133-163

[6] Lo, A.W., Mamaysky,H., Wang, J., *„Foundations of Technical Analysis: Computational Algorithms, Statistical Inference and Empirical Implementation"*, The Journal of Finance, 55(4), Papers and Proceedings of the Sixtieth Annual Meeting of the American Finance Association, Boston, Massachusetts, January 7-9, 2000 (Aug., 2000), pp. 1705-1765

[7] Mishra, A., Mishra,V., Smyth R., *„ The random walk hypothesis on the Indian stock market"*, Journal of Emerging Markets Finance and Trade, 51 (5), 2015, pp. 879-892

[8] Wooldridge, J., *„Introductory Econometrics. A Modern Approach",* Pearson, 2009

[9] www.bvb.ro.

[10] http://scikit-learn.org/stable/modules/svm.html.