DIGITALES ARCHIV

ZBW – Leibniz-Informationszentrum Wirtschaft
ZBW – Leibniz Information Centre for Economics

Radu, Stefan-Constantin; Anghel, Lucian Claudiu; Ermiş, Ioana Simona

Article

Improving Wealth Management Strategies Through the Use of Reinforcement Learning Based Algorithms. A Study on the Romanian Stock Market

Provided in Cooperation with:
National University of Political Studies and Public Administration, Bucharest

Reference: Radu, Stefan-Constantin/Anghel, Lucian Claudiu et. al. (2021). Improving Wealth Management Strategies Through the Use of Reinforcement Learning Based Algorithms. A Study on the Romanian Stock Market. In: Management dynamics in the knowledge economy 9 (3/33), S. 405 - 416. http://www.managementdynamics.ro/index.php/journal/article/download/427/379. doi:10.2478/mdke-2021-0027.

This Version is available at:
http://hdl.handle.net/11159/6431

Kontakt/Contact
ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics
Düsternbrooker Weg 120
24105 Kiel (Germany)
E-Mail: *rights[at]zbw.eu*
https://www.zbw.eu/econis-archiv/

Standard-Nutzungsbedingungen:
Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.

https://zbw.eu/econis-archiv/termsofuse

Terms of use:
This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Mitglied der Leibniz-Gemeinschaft

sciendo
National University of Political Studies and Public Administration
Faculty of Management

Improving Wealth Management Strategies Through the Use of Reinforcement Learning Based Algorithms. A Study on the Romanian Stock Market

Ștefan-Constantin RADU[1], Lucian Claudiu ANGHEL[2], Ioana Simona ERMIȘ[3]
[1] Romanian Economic Studies Academy, Faculty of Finance and Banking, Bucharest, RO; stefan.radu98@yahoo.com
[2] National University of Political Studies and Public Administration, 30a Expoziției Blvd., Sector 1, Bucharest, 012104, RO; lucian.anghel@facultateademanagement.ro (corresponding author)
[3] Romanian Economic Studies Academy Bucharest, RO; simonaermis@yahoo.com

Abstract: In the context of the growing pace of technological development and that of the transition to the knowledge-based economy, wealth management strategies have become subject to the application of new ideas. One of the fields of research that are increasing in influence in the scientific community is that of reinforcement learning-based algorithms. This trend is also manifesting in the domain of economics, where the algorithms have found a use in the field of stock trading. The use of algorithms has been tested by researchers in the last decade due to the fact that by applying these new concepts, fund managers could obtain an advantage when compared to using classic management techniques. The present paper will test the effects of applying these algorithms on the Romanian market, taking into account that it is a relatively new market, and compare it to the results obtained by applying classic optimization techniques based on passive wealth management concepts. We chose the Romanian stock market due to its recent evolution regarding the FTSE Russell ratings and the fact that the country is becoming an Eastern European hub of development in the IT sector, these facts could indicate that the Romanian stock market will become even more significant in the future at a local and maybe even at a regional level
.
Keywords: reinforcement learning; wealth management strategies; stock market; East European economies. .

Introduction

The recent developments in the field of information technology have made the application of iterative-based mathematical methods a feasible possibility, by taking advantage of the increase in processing power and the data available, researchers have developed reinforcement learning-based algorithms. These algorithms can be used for the management of a trading account actively, or the least as the motivation behind an investment decision. The present article tries to apply these techniques to the Romanian market, this decision was based on the idea that it is interesting to see the results obtained by the algorithms on a relatively new market, that has a small number of traded stocks and a lower volume compared to the bigger markets such as the London Stock Exchange from the United Kingdom and the NYSE or NASDAQ markets from the United States of America. Today's investment opportunities should be represented by the Eastern European markets due to their novelty and the fact that in the past two decades they have become more stable. Another important fact is that the Bucharest Stock Exchange is classified, starting with September 2020, as a Frontier Market in the category of Secondary Emerging Market by the rating agency FTSE Russell makes investing in this market more attractive for possible investors. These changes have made the Romanian market an attractive one for international investors and big investment funds.

In this article, we will test the way in which a series of active investment strategies represented by: the "turtle" algorithm, the simple moving average algorithm, and the Dual

How to cite
Radu, S. C., Anghel, L. C. & Eremis, I. S. (2021). Improving Wealth Management Strategies Through the Use of Reinforcement Learning Based Algorithms. A Study on the Romanian Stock Market. *Management Dynamics in the Knowledge Economy.* 9(3), 405-416, DOI 10.2478/mdke-2021-0027
ISSN: 2392-8042 (online)
www.managementdynamics.ro
https://content.sciendo.com/view/journals/mdke/mdke-overview.xml

Q-Learning strategy, compare to the returns obtained by two portfolios (composed out of five shares traded on the Romanian Stock Exchange) obtained by minimizing the Standard Deviation and the Mean Absolute Deviation of the returns of the stock prices. At the end of the study, we will be able to conclude if when trading on the Romanian Stock Exchange we should use an active investment strategy or a passive investment strategy.

The real-time algorithms used will trade every day for an analyzed period of ten years between the 7th of September 2011 and the 7th of September 2021, the algorithm will have three possible options to buy, to sell, or do not take any action, also the algorithm is limited to buying or selling only one share per day. This limitation was put in place, to keep the time intervals equal for all the analyzed shares (in the case in which an algorithm runs out of money, it will not be able to buy anymore). Another limitation of the program is that the algorithm is not allowed to short sell a stock, meaning that it has to own a stock before it can sell it, this was done also due to the idea that the time intervals have to be equal, to allow the comparison of the returns, in any case, the short-selling operation is not allowed on the Romanian capital market. The selected stocks are from 3 different business segments. Two of the selected shares are issued by companies that are in the energy segment: OMV Petrom S.A. (SNP), S.N.T.G.N. and Transgaz S.A. (TGN), other two are issued by financial institutions: Fondul Proprietatea S.A. (FP) and Banca Transilvania S.A. (TLV) and one company in the industrial sector: Teraplast S.A. (TRP). The stocks were chosen based on the fact that on the 9th of September 2021 they had the largest traded volume, this being of interest due to the way the algorithm works, by buying or selling on a daily basis.

An idea that has to be taken into account when dealing with the application of trading algorithms in particular and when working with historical data related to stock markets, in general, is the level of efficiency of a certain market. Market efficiency as it was defined by Fama (1970) consists in the fact that all the information available in the market is reflected in the evolution of the price of the share. This idea has many implications regarding the use of models and the most important one is that if the market is efficient, the previous prices of the share should have no effect on the current evolution of the share's price. This hypothesis is unlikely due to many factors, such as the existence of certain information that is known only by a few individuals that activate on the stock market and not by all the traders. A similar effect could also be obtained if the available information is used in a manner that is deficient by all the market participants. These hypotheses are not likely to occur due to the many studies that have the subject of efficient markets, one of which being the one by Dreman and Berry(1995) that have proved that an efficient market in a strict sense is unlikely to occur.

Literature review

The reinforcement learning-based algorithms have become in the last few years a more common method of analysis which has led to their use for solving several non-trivial problems. In economics, this type of algorithm has been taken into consideration since the end of the 20th century for use in the field of portfolio management. One of the first applications of the reinforcement learning algorithms in the domain of stock optimization were the two articles written by Neuneier in 1995 and in 1997, in which the author managed an asset portfolio by implementing the use of a Q-learning-based algorithm. At the same time, several notable studies in terms of their approach and the used methodology have been released, the first one being written by Moody, Wu, Liao, and Saffell (1998) and the other one being made by Moody and Saffell (2001), they used reinforcement learning for buying and selling a share, these studies were instrumental in introducing the ratios developed by Sharpe and Sterling when comparing the strategies

Vol.9 (2021) no.3, pp.405-416; DOI 10.2478/mdke-2021-0027

and the relative performance of the algorithms for the analyzed period, thus developing a framework for tracking the way the programs perform in certain situations. Another article that was considered instrumental at the time of its launch, was the one written by Jangmin, Lee, Lee, and Zhang (2006) which developed a transaction system that was based on a Q-learning algorithm, which classified a series of prices of the shares in four categories, and applied different rules for each category. The article used this strategy to buy and sell the South Korean index KOSPI, between the years 2002 and 2003 and has registered for the analyzed period a return of 258%.

An important article in the last decade was the one written by Eilers, Dunis, Mettenheim, and Breitner (2014), the authors develop an algorithm using the Q-Learning methodology adding conditions for a seasonal adjustment in the behavior of the algorithm when buying and selling a single asset from the market. The research is applied to the S&P 500 and the DAX index for a period between 2000 and 2012. In the same time period and using similar techniques Zhang et al. (2015) developed a program that generates transaction rules by using a classification system combined with a reinforcement learning type algorithm to take part in the activities of the ProShares Short S&P500 (SH) index and NASDAQ market for the period between 2001 and 2013. For the mentioned period the algorithms generated an average return amounting to 205.65% for the SH index and 77.85 % for the NASDAQ index.

An article that implements advanced reinforcement learning techniques for the stock market (i.e. Deep Q-Learning) is „Adaptive stock trading strategies with deep reinforcement learning methods" which was written by Xing Wu et al. (2020) which tests several trading strategies on the markets of the United Kingdom, the United States of America and China comparing the results of the algorithms to other iterative measures that are considered simpler (i.e. "turtle" algorithm). The study concludes that the best strategies for the analyzed markets are Gated Deep Q Learning and Gated Deterministic Policy Gradient had the best performance. Several studies were also of interest in developing the algorithms and interpreting their results such as: Brock et al (1992), which establishes the basis of the application of reinforcement based learning theory in finance, and Dempster et al (2006), Kouwenberg (2001) and Mansini et al (2003) where we can see the implementation of reinforcement based algorithms on different markets and the obtained results.

Other articles that were important in the development of this paper are the ones written by Fama (1970) in which the background regarding the theory of the efficient market hypothesis is set out, and the article of Dreman and Berry (1995) which criticizes the ideas of Fama's article and proposes the ideas that the efficient market hypothesis is not consistent with the empirical evidence regarding the market. Derman et al. (1995) article suggests that shares are mispriced due to overreaction to events that are presented in the news. The overview on the Romanian market and its evolution has been obtained by reading the following studies: Fleanță et al (2018) and Mihalcea et al (2018), which describe the evolution of the Romanian stock exchange over the years.

Wealth management strategies

When considering the use of reinforcement learning-based algorithms in the economy, most authors suggest that their best use is in the field of wealth management strategies. In this way, we can state that a possible implementation can be the automated management of a trading account; this would be done by leaving the algorithm to trade in the limits of certain indicators, such as the number of daily transactions and a maximum transaction value. This application can be of interest to fund managers, in the case of

developing a strategy involving an automated trading algorithm. A question remains regarding the way the algorithm can implement the buying and selling of certain stocks, it should be taken into consideration that the program could buy or sell only stocks that are considered to be investment grade by the investor.

All the strategies that are tested in this article will be compared to the minimum variance portfolio calculated using the method developed by H. Markowitz (1952), and in the case of active portfolio management, the "turtle" algorithm and the simple moving average strategy will be used as benchmarks for the tested strategy. The "turtle" algorithm was developed in the second half of the 20th century, by Richard Dennis and William Eckhart in the project called the „Turtle Trading Experiment", explained in a brief way this algorithm follows the perceived trend of the market, as described in Carr (2021). The algorithm functions in a way that when the current price of the share is smaller than the price average of the last 126 days (this number was obtained by calculating 10% out of the number of total observations) the algorithm buys the share and when it is smaller the program sells, the algorithm is limited to buying or selling only a share per day, this limit is imposed to facilitate the comparison by providing equal time intervals for each share.

The trading algorithm based on the simple moving average works by taking two intervals and calculating the moving average for each of them. Then the obtained results are compared to decide if a certain stock is to be bought or sold. In this article, the short window was calculated for a period of 32 days (0.025 of the chosen time interval) and the long window for a period consisting of 63 days (0.05 of the analyzed interval). The algorithm decides that when the short window average is bigger than the long window average, the algorithm will buy a unit of the share and when the short window is smaller the program will sell a unit of the share.

The obtained results have been calculated with the use of the Jupyter Notebook package for Python and the riskfolio-lib library, developed by Cajas (n.d) and were based on the calculations made in the Stock Prediction Models repository made by Zolkepli (n.d).

The 5 companies whose shares were chosen to be analyzed by the use of the algorithm are the most liquid as of the 9th of September 2021, considering the daily transactions value, on the Bucharest Stock Exchange market: Teraplast S.A. (TRP), Banca Transilvania S.A. (TLV), OMV Petrom S.A. (SNP), S.N.T.G.N. Transgaz S.A. (TGN) and Fondul Proprietatea S.A. (FP), the data used is the daily closing values of the respective shares for the last ten years. At first, we calculated the composition of the portfolio with the idea to minimize two risk measures first the Standard Deviation of the time series, and secondly the Mean Absolute Deviation.

The mentioned risk measures were calculated by using the following mathematical formulas: in the case of the Standard Deviation the formula that was used is:

$$DS = \sqrt{\frac{1}{N-1} * \sum_{i-1}^{N} (x_i - \bar{x})^2} \quad (1)$$

And for the Mean Absolute Deviation the formula used for calculation is:

$$MAD = \frac{\sum_{i=1}^{N} |x_i - \bar{x}|}{N} \quad (2)$$

For the portfolios that minimize the Standard Deviation (SD) and the Mean Absolute Deviation (MAD) the compositions were the following, this was obtained by applying the methodology stated by Markowitz (1952) and minimizing for each of them the risk

Vol.9 (2021) no.3, pp.405-416; DOI 10.2478/mdke-2021-0027

measure. The results represent the best possible outcome and are calculated with all the values from the analyzed period, from this we can conclude that in the best possible case the returns for a portfolio composed at the start of the period will be the one obtained. In the following table, we present the percentage of the sum that needs to be invested in each share, to obtain the portfolio that minimizes that certain risk measure.

Table 1. The composition of the calculated portfolios

Share	TGN	FP	TLV	TRP	SNP	Portofolio return
SD	6,93%	39,72%	30,38%	17,94%	5,04%	19,45%
MAD	5,89%	42,97%	26,82%	18,84%	5,47%	20,25%

Source: own calculation

In Figure 1 we can see the composition of the portfolio that minimizes the standard deviation of the shares returns, the graph presents the ratios in which to invest the available resources, we can conclude that the investment should be made in the following proportions: 39.7% in the FP shares, 30.4 in the TLV shares, 17.9% in TRP, 6.9% in TGN and 5.04% in SNP shares

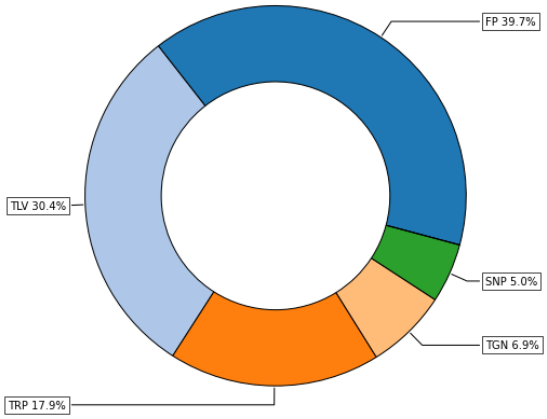

Figure 1. The composition of the portfolio that minimizes the Standard Deviation
(Own calculation)

In the next segment, we will present the results obtained by applying active portfolio management strategies, for the analyzed period on each share that was taken into consideration. The algorithm was limited at buying and selling a maximum of one share per day and it is not able to do short selling, selling of an asset that it's not owned having in view also that this operation is not possible on the Romanian capital market. The limitation regarding short selling is put in place due to the fact that the algorithm could decide to short sell and may end the money available for trading in a shorter period of time, this will lead to time intervals of different lengths that would be difficult to compare. By applying the "turtle" algorithm for the analyzed shares we obtained the following returns for the analyzed period.

Table 2. Results obtained by applying the "turtle" algorithm

Share	*TGN*	*FP*	*TLV*	*TRP*	*SNP*	*Average return*
Return	*-64,88%*	*0,02%*	*0,07%*	*0,01%*	*0,02%*	*-12,95%*

Source: own calculation

The results of the "turtle" algorithm in the case of the TLV share are presented in Figure 2, where the red triangles represent the selling signals and the green triangles the buying signals.

Figure 2.The results of the "turtle" algorithm for the TLV share
(Own calculations)

The results obtained by applying the "turtle" algorithm for the TGN issued shares represented a loss of 64.88% of the invested sum, this was the only loss registered by the algorithm in the analyzed period. For the rest of the shares, the algorithm registered returns under 1% for the ten-year period which was taken into account. These results seem to suggest that the application of real-time investment strategies when trading on the Romanian Stock Market is difficult due to the nature of the evolution of the share prices.

In the next part of the article, we decided to apply the algorithm based on the simple moving average calculation, which was described in detail at the start of this chapter. By applying the SMA-based algorithms for the shares in the selected interval we obtained the following results.

Table 3. Results obtained by applying the SMA algorithm

Share	*TGN*	*FP*	*TLV*	*TRP*	*SNP*	*Average return*
Return	*0,61%*	*-0,01%*	*-0,02%*	*-0,0005%*	*-0,05%*	*0,11%*

Source: own calculation

The results obtained when applying the simple moving average have been similar to the ones obtained by the "turtle" strategy, the only difference being the fact that the algorithm has not registered significant losses the biggest loss being registered when trading the SNP share and amounting to 0.05% of the invested sum. The largest return was registered by the TGN share and represented 0.61% of the invested sum, we can state by looking at the results that the SMA strategy is not suitable on the Romanian Stock Market, at least for the shares that were analyzed in this article. The used algorithm is limited at buying or selling only one share per day, this being decided due to the fact that the algorithm could have bought a considerable number of shares at the start of the interval and run out of the available money before the time interval ended, this would lead to the only the possibility of selling, letting the program run with these settings could have led to the spending of all the available funds in a shorter period than the analyzed period of ten years. This would make the intervals of uneven lengths for different shares, making the results difficult to be compared.

The use of the Simple Moving Average algorithm has been noted to be limited due to the fact that it's difficult to say if the first days that are part of the long window are really relevant to the current trend of the market, taking into account the fact that we are talking

Vol.9 (2021) no.3, pp.405-416; DOI 10.2478/mdke-2021-0027

about daily closing prices, at a long window consisting of 126 days, the first day in the interval is as important in the average as the last, this can lead to possible biases. Also, we can state that all the Simple Moving Average based algorithms shouldn't be useful if the market is efficient, as stated by Fama (1970), because if markets are efficient, using previously recorded data regarding the financial performance of the asset should not indicate the future direction of the asset prices.

In Figure 3 we can see how the algorithm performed when applied to data represented by the prices of the TGN share.

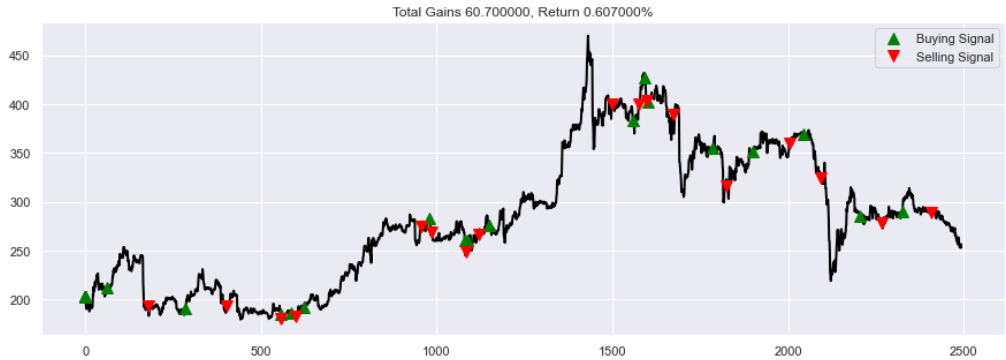

Figure 3.The results of the SMA algorithm for the TGN share
(Own calculations)

In the figure, the green arrows represent the buying signal as perceived by the program and the red arrows represent the selling signal. We can see that the algorithm was influenced by the price of the share which has been on a decline, a fact that can be observed by the trajectory of the black line that is drawn on the graph.

Q-Learning is a process defined by its capacity to learn in an iterative manner. This learning is done by maximizing the value of the reward function that is programmed within the algorithm. The algorithm has a period in which it explores the function, this is called the training period. After that the algorithm selects a random action that has an assigned probability of ε, the probability of the opposing action is 1-ε, this is why the process is defined as being "epsilon-greedy". In the algorithm that was applied, we chose the value of ε to be 0.5, by setting this value we get equal probabilities for both options (i.e. exploring or using a solution that worked before). By using these functions, the program estimates the Q value of certain actions, in the training phase the agent updates the Q value for each state-action combination, in this way: the Q values matrix is updated at each iteration. These values are updated by using the following function:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha * [R_{t+1} + \delta * Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (3)$$

The Q value of a state-action pair is generally considered a good estimation of the usefulness of a certain action at the given state. With the increase of the Q value, the value of the reward function (which is notated with r) increases. The algorithm learns to solve the problem in the most efficient manner by taking into account the value of the reward function at each given state. In the formula notated with 3, we can observe the way in which the function's present value is calculated by taking into account the value of the next period, this is possible because when the algorithm is initialized the Q values are defined as random values for all the states that are taken into consideration. In the first iteration of the training phase, the program is updating the Q value in the present state by considering the value of the randomly generated variable that was initialized in the next state. Because the algorithm works by maximizing the reward and not the Q value, the program will obtain the best result after running several iterations.

As described in the preceding paragraph the algorithm based Q learning algorithm has the following structure: at the beginning of the program it randomly generates the Q values in the matrix according to formula 4:

$$Q(s,a) = n, \forall\, s \in S, a \in A \tag{4}$$

The final state is generated by the use of formula 5:

$$Q(s - final, a) = 0 \tag{5}$$

In the presented formulas S represents the set of states and A represents the set of actions, the algorithm then chooses the action a from set A (which represents the set of all possible actions), and then it takes into consideration the r value for the next state s_(t+1) . From all the actions that are considered possible the algorithm then selects the one that has the greatest Q value, and then the states are updated according to the 3rd formula. After that, the steps are repeated until the reaching of final state. All these steps represent one iteration of the algorithm, but to obtain results that can be considered significant it is necessary to run multiple iterations.

In time more advanced methods of Q-Learning have been developed: Double Q-Learning, Dueling Q-Learning, and Recurrent Q-Learning. The Double Q Learning trading algorithm uses two separate and identical neural network models. One of the models learns while running in the same way as the Q-Learning-based algorithm and the other is a copy of the last iteration of the first model, this one being used by the program to calculate the Q value. In this case, the concept behind this algorithm is to reduce the overestimation bias that is done by decomposing the maximization operation into two separate stages: selection of the right action to perform and the evaluation of the results of the action.

In the case of the applied Double Q-Learning algorithm, there were 500 iterations used, out of which 100 were used for the test. In Table 4 we present the results of the implementation of the Double Q-Learning-based algorithm for the chosen time interval, the returns presented are those obtained at the end of the time interval, after ten years.

Table 4. The returns used by applying Double Q Learning

Share	***TGN***	***FP***	***TLV***	***TRP***	***SNP***	***Average return***
Return	*-5,25%*	*-0,01%*	*-19,20%*	*-0,69%*	*-0,05%*	*-5,04%*

Source: own calculation

By analyzing the results, we can state that the Double Q Learning algorithm is not suitable to be applied when operating on the Romanian market, as we can see all tested real-time algorithms have obtained lower returns than the passive investment-based strategies. In Figure 4 we present the results obtained by the algorithm in the case of the FP share, which obtained the greatest return for the analyzed period when using Double Q Learning.

Vol.9 (2021) no.3, pp.405-416; DOI 10.2478/mdke-2021-0027

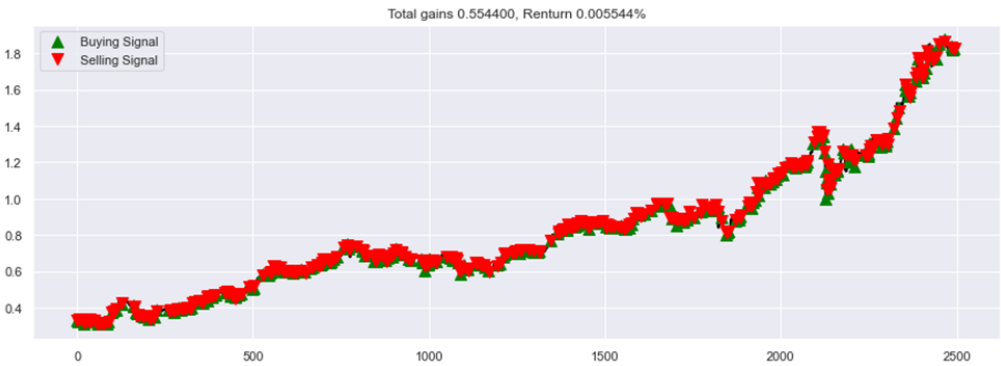

Figure 4. The returns of the algorithm of the Double Q Learning for the FP share
(Own calculations)

In Figure 4 the green arrows represent the buying signals and the red arrows represent the selling signals when they are perceived by the algorithm. We can see that for the analyzed period the algorithm bought or sold the FP share almost daily, these results also suggest the fact that the use of the algorithm by an individual investor could lead to a higher transaction cost because most brokers charge a commission for every transaction, this being added to the fact that the algorithm registered only a 0.005% return, discourages the use of Double Q-Learning-based algorithm for the Romanian stock market.

By analyzing the performance of the algorithm for all the shares that are taken into account, we can state that the worst performance has been obtained by the TLV share which has registered a loss of 19.2% of the invested sum, all the shares except FP have registered losses when the Double Q-Learning was applied.

Conclusions

For the analyzed period the results obtained by the real-time transaction algorithms have been modest, this is due to the fact that the market is not a very liquid one, even though we selected the shares with the largest traded volume. In the following table, we summarized the results obtained for the given period by investing in the selected stocks with both active and passive investment strategies. The passive investment strategies were calculated by investing in the portfolios that minimize the Standard Deviation and the Mean Absolute Deviation of the evolution of the shares prices, at the start of the period and calculating the return at the end of it.

Table 5. The returns obtained by using the investment strategies

Passive investment		***Active investment***		
SD	***MAD***	***"Turtle"***	***SMA***	***DQL***
19.45%	20.25%	-13.95%	0.11%	-5.04%

Source: own calculation

In conclusion, we can state that when taking into account the Romanian stock market, at least in the last ten years, a passive investment strategy is the most suitable if long-term gains are to be taken into consideration. The results obtained by applying a passive investment strategy yield a return of almost 20% which is better than the returns registered by the active investment strategies tested in this article, which have all been less than 1%. We can propose that the results of the active and passive investment strategies could have been influenced by the COVID-19 health crisis that generated a wave of financial instability, although the global markets have been in a relative recovery period

since the end of 2020. By taking into account the results obtained we can recommend the use of a passive investment strategy when trading on the Romanian stock market.

Even though the returns of the algorithms were not that significant, the conclusion does not lead to the idea that reinforcement learning-based algorithms are not suitable for wealth management but to the conclusion that the Romanian market has not reached yet the level of liquidity at which the use of reinforcement learning-based algorithms becomes a viable alternative to traditional wealth management strategies. Adding in the consideration the fact that in most cases for every transaction made a commission fee is paid, the idea of applying active investment strategies on the market for a long term period seems to become more unlikely, due to the possible hidden costs that come with applying this strategy. When comparing our results to the return obtained by Jangmin, Lee, Lee, and Zhang (2006) for the South Korean index KOSPI, for the time period between 2002 and 2003, which was 258%, we can state that the application of reinforcement learning-based algorithms on the Romanian stock market is at present not a viable alternative to traditional wealth management strategies. The use of the reinforcement-based learning algorithms in the context of the Romanian stock market can be related to the analysis of an investment decision as a preliminary study that could be used to justify investing in a certain stock and not as a way to manage active investment portfolios.

An idea that could explain the lack of financial performance registered by the algorithms is that the financial market on which the assets are present, the Bucharest Stock Exchange, is an efficient market according to the financial market hypothesis developed by Fama (1970). According to this theory when a market is efficient the prices of the asset are reflective of all the information that is available in the market, a possible idea that is derived from the hypothesis is that the prices of the assets react only to new information that is present on the market regarding the companies that issued the shares. This implies that the historical values registered by the price of a share are not a measure of the future developments of its price. If this is indeed the case for the Bucharest Stock Exchange the hypothesis should be further tested.

Limitations and future research

The study is limited by the fact that the algorithms have been applied to only five shares (that were considered to be the most liquid on the market) and are on the Romanian Stock Exchange, those being the ones issued by: Teraplast S.A. (TRP), Banca Transilvania S.A. (TLV), OMV Petrom S.A. (SNP), S.N.T.G.N. Transgaz S.A. (TGN) and Fondul Proprietatea S.A. (FP), this leads to a lack of diversity when it comes to analyzed business segments due to the fact that two of them are in the energy sector and two in the financial sector. Another possible limitation of the study is the fact that the analyzed period contains the COVID-19 health crisis which generated a degree of financial instability on the financial markets, this can be the cause of the low returns that we have been seeing for the analyzed companies.

Future research should concentrate on trying to apply the reinforcement learning-based trading algorithms on the transaction of an Exchange Traded Fund (ETF) that follows the BET index of the Bucharest Stock Exchange. This approach could lead to an increased return due to the fact that the index contains a greater number of stocks which represent companies from different business segments, and in this way, the buyer is not exposed to risks that are specific only to one business sector. In future studies an interesting development should be done by investigating if the Bucharest Stock Exchange is indeed similar to an efficient market as stated in the "Efficient Capital Markets: A Review of Theory and Empirical Work" by Fama (1970), although we could conclude that this possibility is

not likely, the ideas behind the study could at least imply a lack of correlation between the past prices of the asset and the future registered prices.

In the end, we can state that the study accomplished its initial objectives by studying the use of reinforcement learning-based algorithms in the management of a trading account by buying and selling a share issued on the Bucharest Stock Exchange for a period of ten years.

References

Brock, W., Lakonishok, J., & Lebaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance, 47*, 1731–1764. https://doi.org/10.2307/2328994

Cajas D. (n.d.). Riskfolio library. Retrieved from https://github.com/dcajasn/Riskfolio-Lib

Carr M. (2021). Turtle Trading: A Market Legend. Retrieved from https://www.investopedia.com/articles/trading/08/turtle-trading.asp

Dempster, M. A. H., & Leemans, V. (2006). An automated FX trading system using adaptive reinforcement learning. *Expert Systems with Applications, 30*, 543–552. https://doi.org/10.1016/j.eswa.2005.10.012

Dreman, D., & Berry, M. (1995). Overreaction, Underreaction, and the Low-P/E Effect. *Financial Analysts Journal*, *51*(4), 21–30. https://doi.org/10.2469/faj.v51.n4.1917

Eilers, D., Dunis, C. L., Mettenheim, H. J., & Breitner, M. H. (2014). Intelligent trading of seasonal effects: A decision support algorithm based on reinforcement learning. *Decision Support Systems, 64*, 100–108. https://doi.org/10.1016/j.dss.2014.04.011

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 25(2), 383–417. https://doi.org/10.2307/2325486

Fleanță, S., & Anghel, L. C. (2018). The Romania's Capital Market Chances of Becoming an Emerging Market. In C. Bratianu et al. (Eds.), P*roceedings of Strategica. Challenging the Status Quo in Management and Economics* (pp. 180-189), Tritonic.

Jangmin, O., Lee, J., Lee, J. W., & Zhang, B. T. (2006). Adaptive stock trading with dynamic asset allocation using reinforcement learning, *Information Sciences, 176*(15), 2121-2147. https://doi.org/10.1016/j.ins.2005.10.009

Kouwenberg, R. (2001). Scenario generation and stochastic programming models for asset liability management. *European Journal of Operational Research, 134*, 279–292. https://doi.org/10.1016/S0377-2217(00)00261-7

Mansini, R., Ogryczak W., & Speranza, M. G. (2003). On lp solvable models for portfolio selection. *Informatica, 14*, 37–62. https://doi.org/10.15388/Informatica.2003.003

Markowitz, H. M. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91. https://doi.org/10.2307/2975974

Mihalcea, A., & Anghel, L. (2018). Romanian Capital Market: On the Road toward an Emergent Market Status. In *Proceedings of Strategica. Challenging the Status Quo in Management and Economics* (pp. 168-179), Tritonic.

Moody, J., & Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks 12*(4), 875-89. https://doi.org/10.1109/72.935097

Moody, J., Saffell, M., Liao, Y., & Wu, L. (1998). Reinforcement Learning for Trading Systems and Portfolios: Immediate vs Future Rewards. In A. P. N. Refenes, A. N. Burgess, & J. E. Moody (Eds.), *Decision Technologies for Computational Finance. Advances in Computational Management Science* (vol. 2), Springer. https://doi.org/10.1007/978-1-4615-5625-1_10

Neuneier, R. (1995). Optimal Asset Allocation using Adaptive Dynamic Programming. *NIPS*. https://doi.org/10.5555/2998828.2998962

Neuneier, R. (1997). Enhancing Q-Learning for Optimal Asset Allocation. *NIPS*. https://doi.org/10.5555/3008904.3009035

Zhang, X., Hu, Y., Xie, K., Zhang, W., Su, L., & Liu, M. (2015). An evolutionary trend reversion model for stock trading rule discovery. *Knowledge-Based Systems, 79,* 27–35. https://doi.org/10.1016/j.knosys.2014.08.010

Zolkepli H. (n.d), Stock Prediction Models. https://github.com/huseinzol05/Stock-Prediction-Models.

Received: June 08, 2021
Accepted: August 25, 2021

© 2021 Faculty of Management (SNSPA), Author(s). This is an open-access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (http://creativecommons.org/licenses/by-nc-nd/4.0/).