



Wroclaw University  
of Economics and Business

**Faculty of Economics and Finance**

**Grzegorz Tratkowski**

ABSTRACT OF DOCTORAL DISSERTATION

**DYNAMIC INVESTMENT PORTFOLIO OPTIMIZATION WITH  
REINFORCEMENT LEARNING**

Supervisor: **dr hab. Krzysztof Piontek, prof. UE**

Wrocław 2022

One of the basic human activities in the field of economics is investing [Jajuga and Jajuga, 2005]. From the perspective of managing the entire set of investments, it is defined as an ongoing capital commitment to obtain future payments that compensate the investor for time and his uncertainty [Reilly and Brown, 2018]. The entire collection of investments in the form of financial or real assets is called an investment portfolio [Ostrowska, 2011], the composition of which changes over time as the investor's expectations of future payouts change. This requires a continuous process of investment portfolio management [Reilly and Brown, 2018], which is divided into: 1) drawing up an investment policy statement, 2) selecting an investment strategy, 3) constructing the portfolio, and 4) evaluating its performance. Using new technologies, especially artificial intelligence tools, in the investment portfolio management process, some elements of the process can be improved and automated, as will be shown in this paper.

Artificial intelligence is defined as a system's ability to correctly interpret data from external sources, learn from it, and use that knowledge to perform specific tasks and achieve goals through flexible adaptation [Kaplan and Haenlein, 2019]. Definitions of artificial intelligence are usually divided by the perception of humans as targets for machines to emulate or by rationality in decision-making [Russell and Norvig, 1995]. Humans by nature make mistakes, are driven by emotions and their perception is limited, resulting in the fact that they are not always able to correctly interpret data and make the best decision. Thus, the very process of teaching a machine that improves its efficiency in performing an assigned task with experience on the pages of this paper is called machine learning.

There are various machine learning techniques, but the most common division is made on the basis of the type of learning process itself, i.e. the availability of feedback [Flach, 2012]. On this basis, machine learning is divided into three types: supervised learning, unsupervised learning and reinforcement learning [e.g. Murphy, 2012]. Each type is responsible for teaching the machine to solve different types of problems. Supervised learning is responsible for classification and regression problems, unsupervised learning is responsible for clustering, while reinforcement learning is responsible for finding the optimal sequential decision-making policy given the environment the decision-maker is in.

The application of machine learning, as an area of artificial intelligence, has recently been changing every aspect of our lives. With the increase in computational capabilities and the

growing size of available data, machine learning is increasingly taking over tasks that were previously performed by experts in the field [Lopez de Prado, 2018]. A similar phenomenon is also occurring in finance, where examples include algorithmic trading, credit scoring or bank fraud detection.

When using machine learning issues in finance, one needs to combine knowledge from financial econometrics, statistics, algorithmics and programming [Dixon et al., 2020]. This knowledge is included in this paper, which justifies its interdisciplinary nature combining economics and finance along with computer science.

Interest in the use of machine learning methods in the investment portfolio management process is enormous. This is evidenced by the number of available publications on forecasting asset prices, creating investment strategies, generating trading signals or optimizing investment portfolios. A systematic review of the literature conducted as part of the study showed that the vast majority of academic papers focus on asset price forecasting and the creation of trading signals using supervised learning. The popularity of this type of learning is due to the simplicity of the problem definition itself and the application of even a very complex algorithm. Unsupervised learning, on the other hand, is no longer as intuitive and simple to apply as supervised learning. This translates into far fewer works studying such issues in the field of investment portfolio management. On the other hand, the most difficult to apply is reinforcement learning, which requires not only the selection of an appropriate algorithm, but also the conceptual design of a solution to a given problem.

The use of reinforcement learning in the investment portfolio management process is far less researched compared to the use of the other two types of machine learning. Reilly and Brown [2018] emphasize that the entire investment portfolio management process is a continuous process over time. It requires constant monitoring and decision-making about portfolio composition given the current situation and market forecasts. This translates directly into sequential decision-making under uncertainty related to the characteristics of financial instruments. For this reason, the author believes that reinforcement learning is an unexplored yet ideal tool for application in the investment portfolio management process. This has also been confirmed in the conducted literature research.

The first main objective of the paper is to conduct a literature review on the application of machine learning in the investment portfolio management process. The systematic review of

the literature is aimed at identifying the research gap and determining the current state of art on the issues under consideration and verifying the methodology for testing the suitability of using machine learning tools.

Conducting a literature review made it possible to clarify the second main objective of the paper, that is, to verify the effectiveness of constructing a dynamic (multi-period with transaction costs) investment portfolio using a reinforcement learning algorithm in comparison with the classical Markowitz approach. A second literature review already focused only on the use of reinforcement learning methods in investment portfolio optimization allowed the selection of a specific G-learning algorithm, which, according to the author, is the most adapted to the characteristics of financial time series from those currently available. The effectiveness of this algorithm was verified in an empirical study.

Due to the complexity of the second main objective of this work, the following specific objectives were defined:

- Verification of the effectiveness of the selected algorithm on different asset classes (stocks, bonds, currencies, commodities).
- Verification of the impact of investor risk aversion on portfolio efficiency for different asset classes.
- Verification of the effectiveness of different methods of investment portfolio construction depending on the state of the market (ups, downs).
- Polemics with the results obtained on too few instruments or a short period of time, which lead to overly optimistic interpretations on the use of the given tools.

In addition, the indicated main and specific objectives are summarized in the thesis set forth by the author: the use of reinforcement learning methods and consideration of multi-stage decision-making leads to better financial results. The author decided to forgo statistical significance verification due to the high bias of test data selection and the theory of false strategy, which is the reason why it is often not possible to draw generalized conclusions about the effectiveness of a tool in finance [Lopez de Prado and Bailey, 2021]. Nevertheless, the author tried to construct a test to verify the effectiveness of a given tool, which tries as best as possible to reduce the bias of data selection and the receipt of falsely optimistic results.

The empirical test, prepared to meet the second main objective of the paper and the specific objectives, included synthetic data generated by the multivariate jump-diffusion model and empirical data. The actual data included four asset classes that a hypothetical investor might include in his investment portfolio: stocks, bonds, currencies and commodities. The research period for which daily returns were prepared was from 2005 to 2020. For each asset class, 100 time series reflecting the characteristics of the financial instruments in question were selected. The author of this paper significantly extended the designed test for verifying the efficiency of an investment portfolio proposed by Lopez de Prado [2016] by adding elements of randomization of the test period, selection of time series and inclusion of 400 empirical time series representing different asset classes. Three measures of investment portfolio evaluation were chosen as performance criteria: the rate of return, standard deviation as portfolio risk, and the adopted main criterion in the form of return to risk ratio. Transaction costs were also taken into account in the evaluation of portfolios, emphasizing the dynamic nature of the problem. All adopted performance criteria were tested out-of-sample, that is, on data that were not used to estimate the models.

Due to the interdisciplinary nature of this work, it was divided into chapters on both investment portfolio management and machine learning. The work has been structured in such a way that the successive issues presented constitute a coherent sequence leading to the fulfilment of all the stated goals of the work.

The first chapter contains the theoretical foundations of investment portfolio management. In the process of investment portfolio management, the portfolio manager plays an important role, as he decides the composition of the portfolio taking into account the investment policy statement, market situation and potential financial instruments. Due to the manager's approach, the investment portfolio management process can be divided into passive and active management [Grinold and Kahn, 2000; Swensen, 2009; Elton et al, 2014].

The above division is based on the investor's or manager's belief in the potential possibility of outperforming the market - active management, or the lack of such a possibility - passive management. The possibility of beating the market is unresolved from a scientific perspective, and the academic community itself is sharply divided on the issue [Grinold and Khan, 2000; Pedersen, 2015]. This problem directly relates to the efficient market hypothesis defined by

Fama in 1970, according to which markets immediately incorporate all new information into prices.

If the market is assumed to be efficient, a rational approach will be passive management of the investment portfolio. In passive management, the composition of the portfolio does not change due to changing expectations of the market, hence the origin of the very name "passive," meaning unresponsive [Maginn et al., 2007]. In turn, active management of the investment portfolio aims to achieve above-average performance compared to the benchmark, that is, to generate as much alpha as possible [Niedziółka and Czapiewski, 2016; Reilly and Brown, 2018]. At the same time, the portfolio manager will actively react to price changes and the market situation.

The composition of an investment portfolio and an investor's expectations of future payouts change over time, requiring a continuous process of investment portfolio management. Reilly and Brown [2018] identify four steps in this process:

- setting an investment policy statement,
- selecting an investment strategy,
- construction of an investment portfolio,
- continuous monitoring of portfolio performance and investor needs.

The first step in the investment portfolio management process is to establish an investment policy statement (IPS) based on investor preferences, which is still a planning step. An investor's preferences are investment objectives, which indicate expected returns and risk acceptance, and constraints, which reduce the investor's ability to reap full or partial benefits from particular investments [Maginn et al., 2007]. The investment policy itself imposes a framework for the entire investment process and reduces the potential for inadequate decisions by the manager. The two main reasons for constructing an IPS are to help set realistic investment goals for the investor after becoming familiar with the financial markets and their characteristics; and to create standards for evaluating the portfolio manager's performance [Reilly and Brown, 2018]. Without the information contained in the investment policy, investors would not be able to adequately communicate their needs to the manager, who constructs the investment portfolio based on this very data.

Once the investment policy is drawn up, the portfolio manager defines the investment strategy, which is the manager's approach to investment analysis and the rules for selecting the securities included in the investment portfolio [Maginn et al., 2007]. In this step, the manager makes predictions about the market and combines them with the investment policy to form the investment strategy. Economies are dynamic and are buoyed by numerous business cycles, politics, demographic changes and social attitudes. Therefore, an investment portfolio needs to be constantly monitored and investment strategies updated if they are no longer performing as expected [Reilly and Brown, 2018].

The next step in the investment portfolio management process, after drawing up the policy and investment strategy, is the construction of the investment portfolio. The portfolio manager, on the basis of the policy, strategy and forecasts, decides how to allocate available capital in individual assets. This boils down to constructing a portfolio that minimizes risk while meeting return requirements and needs [Reilly and Brown, 2018].

The criteria in the form of expected rate of return and acceptable level of risk and the problem of finding the best portfolio that meets these criteria boils down to an optimization problem. From the point of view of an investor or manager building a portfolio, this problem boils down precisely to choosing such financial instruments that the entire portfolio meets the specific requirements while sticking to the set constraints. In order to determine the investor's propensity to risk, with a given rate of return and constraints, tools based on utility theory are used [Jajuga and Jajuga, 2005].

An extension of utility theory from the perspective of an investor building an investment portfolio is modern portfolio theory (MPT), also known as mean-variance analysis, which was proposed by Harry Markowitz in 1952. From the perspective of portfolio theory, the important aspect is not only the individual risk of particular investments, but the risk of the entire investment portfolio. Markowitz showed that the variance of the returns of the entire portfolio is an important measure of portfolio risk under the following assumptions [Reilly and Brown, 2018]: the investor views each investment as a probability distribution of potential returns over a specified time horizon. The investor maximizes expected utility, and his utility curve is characterized by decreasing marginal utility. The investor estimates portfolio risk based on the variety of potential rates of return. The investor makes decisions solely on the basis of expected returns and their risk, which means that his utility curves are a function of

expected returns and their variance (or standard deviation). For a given level of risk, the investor prefers a higher expected rate of return, and similarly, for a given return the investor will choose an investment with less risk.

Markowitz's portfolio theory and his approach to the investment portfolio management process usually serves as a starting point in determining the composition of a portfolio [Fabozzi et al., 2007]. The original approach is often developed due to various practical constraints [Jajuga and Jajuga, 2005] that take into account the investor's specific investment preferences [Fabozzi et al., 2007]. These extensions may involve different optimality criteria [Jajuga and Jajuga, 2005], modeling the behavior of market returns [Ostrowska, 2011], estimating expected returns [Schmidt, 2021] or taking into account additional constraints [Fabozzi et al., 2007]. An additional factor when solving an optimization problem is the choice of an appropriate algorithm tailored to the nature of the objective function and constraints [Kochenderfer and Wheeler, 2019].

The final step in the investment portfolio management process is the continuous monitoring of the portfolio by assessing its performance against the expectations and requirements set out in the investment policy [Reilly and Brown, 2018]. The performance measurement itself is a quality control of the investment portfolio management process and provides the necessary information for the manager and the investor to evaluate how the capital is invested [Bacon, 2008].

Each measure of investment portfolio performance has its own advantages and disadvantages. They are usually simple to calculate, but not always easy to interpret. The choice of an appropriate indicator should be tailored to the investor's goals and preferences, so that it fulfils its purpose as a measure to assess the effectiveness of investment portfolio management [Bacon, 2008]. Studies on the impact of the choice of a given measure for assessing portfolio performance indicate that in a sufficiently large sample of portfolios, the choice has no significant effect on the created ranking of portfolios [Eling, 2008], and the rank correlation coefficient between measures is at least 90% [Eling and Schuhmacher, 2007]. In addition, Razafitombo [2010] concludes that multiple measures should be used simultaneously when assessing performance in order to achieve unencumbered results.

The second chapter describes theory from the field of machine learning as an area of artificial intelligence dedicated to algorithms that learn through data analysis. The role of machine

learning is discussed from the perspective of the concept of artificial intelligence, as the two concepts are often equated. Understanding the important terms and mechanisms behind machine learning algorithms is key to their correct application.

The most common division is made by the type of learning process itself, i.e. the availability of feedback [Flach, 2012]. On this basis, machine learning is divided into three types: supervised learning, unsupervised learning and reinforcement learning [e.g., Murphy, 2012; Raschka and Mirjalili, 2017]. Each of the above-mentioned types of machine learning has its own characteristics and applications. However, this division does not exclude the use of different learning methods to solve an identical problem, but the results obtained may be different. It is also possible to combine methods to form hybrids, as exemplified by the use of semi-supervised learning with reinforcement by Finn et al [2016].

Supervised learning is the most common method of machine learning used in practice [Murphy, 2012]. This learning involves using both input data and output data to serve as a template for training the algorithm. During this process, the supervised series, or output of the function, is accessed and recognized as the correct (or approximately correct) value of the function of the input variables. This series plays the role of supervisor, so as to best represent the relationships in the data. Learning involves changing and adjusting the function using feedback from the learning process [Russell and Norvig, 2002].

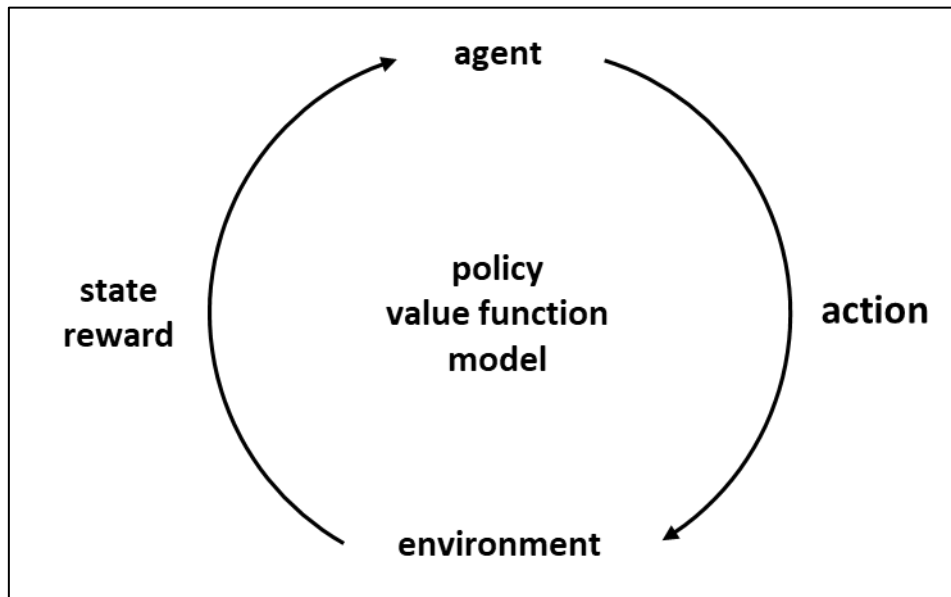
The goal of supervised learning is to find a function  $h$  that, for a new pair  $(x, y)$  belonging to a similar distribution as the training pairs, will reproduce  $h(x) = y$  with high probability. Depending on the characteristics of the label space, supervised learning can be categorized [Marsland, 2015]. If label space  $C$  is contained in a two-element set, e.g.  $\{0, 1\}$ , then it is a binary classification. If  $C$  is contained in a  $K$ -element set, where  $K$  is greater than two, then it is a multiclass classification. The last type of supervised learning is regression, when  $C$  belongs to real numbers.

Unsupervised learning, unlike supervised learning, has no existing labels, or series of data, to which the feature would be matched. This type of learning offers an opportunity to explore the structure of the data and can often reveal features that were not previously expected [Michie et al, 1994]. Jain et al [2000] call unsupervised learning clustering, a name that refers to many methods of grouping multidimensional data sets. Examples of the use of this type of learning include partitioning space by linearly separating characteristic clusters of variables,

or simply determining the region with the highest density compared to the background [Jain et al., 2000]. Functional definitions of clusters are expressed by the characteristic that similarities within a cluster are more pronounced than those between clusters, or by stating that a cluster consists of a relatively high density of points separated from other clusters with a relatively low density of points [Sammut and Webb, 2011]. Jain et al [2000] believe that clustering algorithms are mainly based on two techniques: hierarchical methods and iterative clustering methods based on partitioning with quadratic error. Hierarchical methods involve sequential clustering of clusters, which can be represented as a tree. The second method involves finding a split that would minimize intra-group distances or maximize inter-group distances.

The third type of machine learning is reinforcement learning (RL). The very concept of "learning" from a human perspective is defined by the PWN dictionary of the Polish language as the process of acquiring knowledge and gaining skills by learning from experience. Similarly, reinforcement learning is defined as the learning process of choosing the appropriate action depending on the situation in order to maximize the numerically expressed reward [Sutton and Barto, 2018]. The learner, who is called an agent in RL nomenclature, is not told what action it should perform. The agent must discover, based on repeated trials, which action will provide the greatest current reward and subsequent future rewards depending on the state it is in. Jansen [2020] calls reinforcement learning most similar to people taking actions in the real world and observing the consequences of their actions. The premise of learning with reinforcement is that the agent can read the current state of the environment to some degree. Based on this observation, the agent has the ability to take actions that affect this state, and for taking specific actions, the agent can earn a reward which is its overarching goal [Sutton and Barto, 2018]. This goal is completed by finding the optimal policy which maps the state of the environment with agent's actions. Figure 1 presents a diagram of reinforcement learning along with its most important elements.

Figure 1: Reinforcement learning scheme



Source: own elaboration.

The next part of chapter two describes methods for evaluating the effectiveness of machine learning algorithms according to the division into supervised learning, unsupervised learning and reinforcement learning.

Already having knowledge of the theory of investment portfolio management and the theory of machine learning, in the last part of the second chapter a systematic literature review on the use of machine learning in investment portfolio management was conducted. This review was aimed at identifying the research gap, determining the current state of knowledge regarding the issues under consideration and verifying the methodology for testing the suitability of using machine learning tools.

The first main goal of the paper was achieved, and the main conclusions of the systematic literature review are:

1. the most common type of machine learning used in the field of finance is supervised learning. It is conceptually the simplest, as well as the easiest to implement.
2. Most of the authors of the publications focused on forecasting the given values and create investment strategies based on the forecasts.

3. A wide variety of algorithms were used, dominated by artificial neural networks and models based on decision trees.
4. Although the evaluation of algorithms by a single measure is insufficient, this option was the most frequently chosen.
5. Stock price forecasting is the most popular problem considered among the publications reviewed, and easy-to-obtain technical analysis indicators and lagged price series were the most common input variables.
6. Daily frequency data are usually used.
7. Investment strategies are most often compared with strategies generated by other models or other configurations of the selected model, and with the "buy and hold" strategy.
8. Cumulative return is the most popular measure evaluating a strategy, but it often appears as the only measure of performance, which may indicate a lack of knowledge of the authors of the publications.

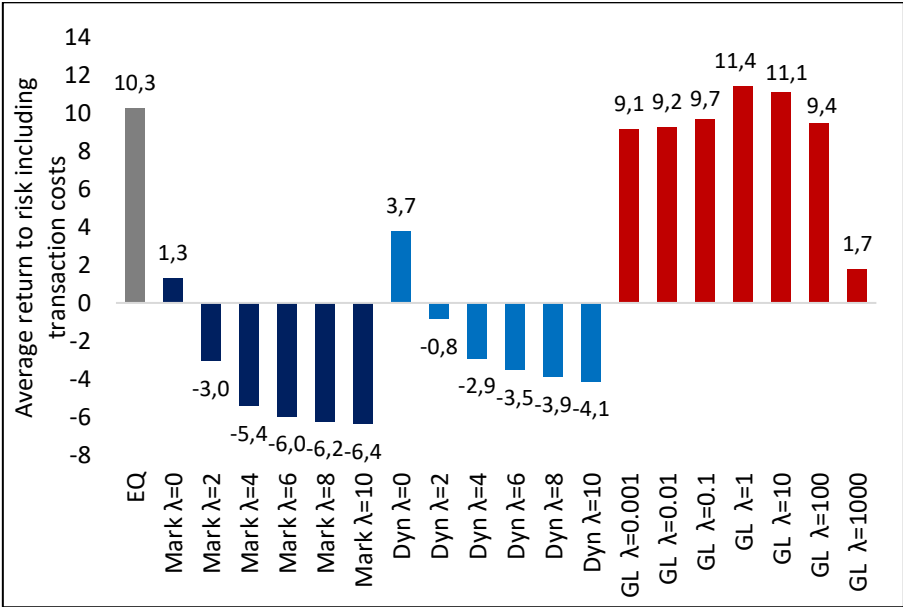
The systematic literature review also identified a research gap in the absence of a survey of applications of reinforcement learning algorithms in the construction of investment portfolios. In addition, it identified the current state of knowledge of machine learning methods in solving problems related to the construction of investment portfolios and how to evaluate the effectiveness of both algorithms and strategies and investment portfolios.

Chapter three described reinforcement learning in detail, along with a technical introduction to selected methods. This made it possible to conduct a second critical review of the literature on the use of reinforcement learning for investment portfolio optimization. According to the author of the paper, almost all of the reviewed publications contained a heavily truncated way of testing the algorithms. Selecting individual financial instruments from a specific period is insufficient to draw general conclusions about the effectiveness of the constructed portfolio and the application of a given algorithm. On the basis of the review, the G-learning algorithm was also selected, which, in the author's opinion, is characterized by suitable features for application to the investment portfolio optimization problem. In the last parts of the third chapter, the operation of the G-learning algorithm is described in detail, followed by a proposal for solving the portfolio construction problem using this algorithm.

Chapter four presents an empirical study. The author independently designed a test for testing the efficiency of investment portfolios considering: synthetic data, empirical data on different asset classes, randomly selected test periods and randomly selected financial instruments. The performance evaluation of the G-learning algorithm also took into account the classical Markowitz model, its dynamic extension that takes into account transaction costs, and an equal-weighted portfolio that acted as a benchmark. Synthetic data was generated by a multivariate jump-diffusion model, which allowed the creation of a set of randomly correlated time series with random price shocks. Based on the performance of the portfolios on the synthetic data, the values of the hypothetical investor's risk aversion parameter were determined for both Markowitz models and the reinforcement learning algorithm.

The empirical part of the study included an assessment of the efficiency of portfolios in terms of the rate of return, the amount of transaction costs, risk and a combined criterion taking into account all these parameters, namely the rate of return taking into account transaction costs relative to risk. The study was carried out on five datasets of different asset classes: bonds, currencies, stocks and commodities, as well as on a set combining all classes. An example of the results of tests conducted on market data aggregating all four asset classes is shown in Figure 2.

Figure 2: Average return to risk including transaction costs for bonds, currencies, stocks and commodities.



Source: own elaboration. EQ – equally weighted portfolio, Mark – classical Markowitz model, Dyn – dynamic extension of Markowitz model, GL – G-learning,  $\lambda$  – risk aversion parameter.

The results obtained allowed to achieve the second main objective of this work and the specific objectives. It was verified that, compared with the classical approach, portfolios constructed using the G learning algorithm obtained, on average, better results for each considered asset class separately and all together. It was also verified that the risk aversion parameter influences the efficiency of portfolios in terms of various evaluation criteria, and at the same time its influence depends on the characteristics of the asset class. It was also verified that the efficiency of actively managed investment portfolios decreases when the market is in an uptrend, while portfolios show better efficiency during a downtrend. The other conclusions of the empirical study are:

- compared to the classical approach, portfolios constructed using the G learning algorithm achieved better results on average for each asset class considered separately and all of them combined.
- The results obtained for different values of the risk aversion parameter reach the highest values differently for each asset class. Typically, the riskier the asset class, the higher the value of risk aversion is required to achieve the highest return-to-risk ratio.
- An increase in risk aversion lowers the standard deviation regardless of the method and asset class.
- In three out of five cases considered for different data sets, actively managed portfolios perform better when the overall market is in a downtrend, and better during an uptrend. In the other two cases, the relationship was insignificant or weakly positive.
- Taking into account the dynamics of the problem allows built portfolios to produce better results.
- Strategies following the leader, that is, not taking into account the risk, have the highest volatility.
- The best results are obtained by presenting the investment portfolio optimization problem as a sequential decision-making problem, which takes into account potential decisions in subsequent periods.
- The specifics of reinforcement learning are matched to the characteristics of the problem of constructing optimal investment portfolios.
- It is worth using advanced analytical tools in active investment portfolio management.

- The equally weighted portfolio, treated as a passive investment approach, achieved comparable results to the best performance of actively managed portfolios, which is consistent with the literature.
- It is necessary to test the effectiveness of the investment portfolio on different periods and different financial instruments due to different time series characteristics.

Additionally, based on two literature reviews, the author concludes that there is no established and controlled methodology for testing the effectiveness of investment strategies and portfolios. The bias of data selection, the false-strategy theory and the drawer effect of published studies make it impossible to properly evaluate these efficiencies and draw general conclusions about the use of a given tool. Currently available publications lead to overly optimistic assumptions from the data use of machine learning algorithms in finance. According to the author of the paper, this justifies the need to create generalized rules for testing the effectiveness of given tools in finance, similar to what is done in testing the effectiveness of drugs in medicine.

On the basis of this work, the author concludes that the process of investment portfolio management is a continuous process over time and requires sequential decisions on the composition and optimal distribution of capital in the portfolio. The sequential nature of decision-making justifies the choice of a reinforcement learning method for solving the dynamic optimization problem of an investment portfolio. The results of the empirical study support the thesis adopted in the paper that the use of reinforcement learning methods and consideration of multi-stage decision-making leads to better financial results.

The results and conclusions of the above work can be used by individual and institutional investors to increase the efficiency of decision-making in the context of asset allocation in investment portfolios. The use of reinforcement learning methods can support or, in some cases, replace the method of investment portfolio optimization. This can translate into a reduction in transaction costs and an increase in the rate of return relative to the accepted level of risk.

The author notes that the topic of applying reinforcement learning methods to investment portfolio optimization is not exhausted and requires further research exploration. A potential extension of the above study could be to include external variables, such as macroeconomic data, market consensus, sentiment or technical indicators, and then examine their impact on

the efficiency of the considered portfolios. In addition, the set of optimized objective functions can be expanded for both classical and reinforcement learning models. The author also points out the need to study the effectiveness of other algorithms in a similarly expanded efficiency test, since no conclusions can be drawn about their suitability based on the current literature.

In this paper, the author presented an original solution to the investment portfolio optimization problem using the G-learning reinforcement learning algorithm. In addition, he included in the paper an extensive theoretical introduction in the aspects of investment portfolio management, machine learning and reinforcement learning, which allowed him to conduct two literature reviews of the subject: a systematic one and a critical one. The author independently designed the empirical study, programmed the solution, conducted an extensive simulation and drew conclusions based on the results. The above made it possible to achieve the goals of the dissertation and confirm the thesis stated in the introduction.