Artigo Original

textos de economia

# DOES MANAGING A RETIREMENT PORTFOLIO VIA RATE OF RETURN, SHARPE RATIO AND SOCIAL INTERACTION GENERATE GOOD RETURNS? AN ANALYSIS FOR THE YEARS 2017 TO 2020

**Gerenciar uma carteira de aposentadoria via taxa de retorno, índice de Sharpe e interação social geram bons retornos? Uma análise para os anos de 2017 a 2020**

Nicollas Mateus Gomes
nicollasgomes10@gmail.com
Graduação em Ciências Econômicas
Universidade Federal de Santa Catarina

Helberte João França Almeida
helberte.almeida@ufsc.br
Universidade Federal de Santa Catarina

Rafael Jasper Feltrin
rafafeltrin5@gmail.com
Mestre em Ciências Econômicas
Universidade Federal de Santa Catarina

Guilherme de Oliveira
oliveira.guilhermede1@gmail.com
Professor do Programa de Pós Graduação em Economia, PPGECO
Universidade Federal de Santa Catarina

## RESUMO

Motivado pelo contexto da reforma previdenciária no Brasil e pelo comportamento dos investidores individuais brasileiros no mercado de fundos de aposentadoria, este estudo elabora um algoritmo para gerenciar uma carteira real de um fundo de aposentadoria. O gestor teórico pode alocar seus recursos entre quatro títulos pertencentes a uma mesma instituição financeira: dois fundos de renda fixa; e dois fundos de crédito privado. O algoritmo de aprendizado de máquina otimiza a alocação de portfólio utilizando taxa de retornos, que recompensa boas decisões e pune decisões ruins e também índice de sharpe com parâmetros de sensibilidade que depende da perfomance de outros fundos (interação social). Em seis cenários de simulação, o modelo superou o portfólio real por retornos médios consideráveis e significantes taxas de retornos.
**PALAVRAS-CHAVE:** Aprendizagem por reforço. Reforma da Previdência. Gerenciamento de Portfólios.

## ABSTRACT

Motivated by context social security reform in Brazil, and the behavior of Brazilian individual investors in the retirement-fund market, this study elaborates an algorithm to manage a real retirement-fund portfolio. The theoretical manager can allocate its resources among four securities belonging to the same financial institution: two fixed income funds; and two private credit funds. The machine learning algorithm optimizes the portfolio allocation using rate of return, which rewards good decisions and punishes bad decisions and too shape ratio with parametric of sensibility that depends on the performance of others funds (social interaction). In six scenarios of simulation, the model outperforming the actual portfolio by considerable and significant average returns.
**KEYWORDS:** Reinforcement learning. Social security. Portfolio management.

**Classificação JEL: C63. G11. G17**

Recebido em: 19-11-2021. Aceito em: 12-12-2023.

## 1 INTRODUCTION

The slow pace of the Brazilian economy since 2011 and the recession of 2015-6, inter alia, have made low inflation persistent. Consequently, interest rates have been reduced to historically low levels by monetary policy, thus dropping the average return of relatively conservative investment positions. This scenario, along with the social security reform of 2018, has raised interest in private retirement funds in Brazil. In the last ten years, however, individuals that invested in at least one of the ten biggest Brazilian pension funds – managed by big financial market institutions – yielded smaller returns than risk-free assets. Motivated by this context and the behavior of Brazilian individual investors, this study designs an algorithm to manage an allocation resource problem in a retirement-fund portfolio.

Literature on computational finance has made progress in recent years, but specific studies on the problem of managing retirement portfolios via algorithms are scarce (Forsyth and Vetzal, 2019). This study intends to contribute in this regard. One of our main contributions is the modeling of social interaction within an algorithm that looks at the competitors' performances in order to choose the aggressiveness level in resource allocation schedules. The rules and constraints of our theoretical fund are inspired in a real retirement fund.

In said model, the manager of the fund can allocate its resources among four securities belonging to the same financial institution: two fixed income funds; and two private credit funds. Innovatively, we built an algorithm that optimizes the portfolio allocation via reinforcement learning, which rewards good decisions and punishes bad decisions. The algorithm takes into account feedback effects from private indicators (profitability and the Sharpe ratio) and a social interaction measure (the performance of the biggest funds in the same market). The model is compared using data on the Brazilian financial market.

This theoretical mechanism is in line with the empirical evidence. Choi *et al*. (2009) remark that individual investors base themselves on their previous results, which means investors with better returns tend to raise their investment ratio by a bigger rate compared to investors with worse experiences. This is analogous to reinforcement learning's behavior and the authors remark about the possibility of using reinforcement learning to create trading algorithms. Machine learning is a fruitful tool to model reinforcement learning once it purports to better an agent's performance by doing specific tasks and gaining experience.

In this context, our contribute in research about the results of reinforcement learning in portfolio management, taking into account the complexity of social interaction in financial markets. To do so, we employ data from 2017 to 2020 and evaluate the chosen hypothesis. All funds used have their composition publicly available in CVM's website[1]. In other words, the contribution of this study is to verify whether, when introducing social interaction in portfolio management methodologies, precisely, reinforcement learning (return rate), and Sharpe index (portfolio balancing) generates better returns in the evaluated funds.

Besides this introduction, the present study is structured in four more sections. Section two being the literature review, section three presents the data, the computational model used and section four shows its implementation and results obtained. Last of all, section five brings our concluding remarks.

## 2. REINFORCEMENT LEARNING AND SOCIAL INTERACTION

Mullainathan and Spiess (2017) remark that with the emergence of bigger datasets, machine learning is poised to become a cornerstone of economic modelling. Murphy (1998) defines reinforcement learning is a problem of maximizing results of an agent in a certain scenario. Sutton (1992) reveals that the idea of reinforcement learning is quite old: the first researches were published by Minsky (1961), followed by Waltz and Fu (1965). However, only in the near end of the 1980s relevant studies in the area were published again, including Werbos (1987) and Watkins (1989), which connected the subject to dynamic programming and showed the link between artificial intelligence and machine learning.

According to Dias Júnior (2012), in reinforcement learning, the agent is inserted into an environment and reacts to it with some possible actions. In a first moment, the agent finds him/herself in an initial state, before the action is taken. After the action, this state is altered and this new state generates a feedback determined by a certain value for the algorithm, in order for it to discern, according to predefined criteria, if that output was desirable or not, called reinforcement. The reinforcement normally is given by {0,1} or by real numbers. Depending on the result given by the reinforcement, the algorithm will define a bigger or smaller probability of taking that decision again when it is again in that initial state. Afterwards, the cycle repeats, so it keeps learning.

Nevmyvaka, Feng and Kearns (2006), for instance, present a large-scale empirical application of reinforcement learning to optimize financial market operations, utilizing high-

[1] Funds can be looked up in: **conteudo.cvm.gov.br/menu/regulados/fundos/consultas/fundos**. In addition, 2020 one of the fixed income funds was disbanded, so the analysis was carried out until this year, in order to be faithful to the model.

frequency data microstructures of securities traded on NASDAQ. Matsui, Goto and Izumi (2009) show that reinforcement learning was able to make profit in the Japanese government bond market, which is impressive as it is bond trading conducted in a stable economy of low interest rates.

Jiang and Liang (2017), in turn, built a neural network model, training an algorithm with little more than six months of data from a cryptocurrency broker, utilizing reinforcement learning methods. The authors tested their algorithm, putting it randomly in three different moments of the last six months' market during 30 minutes. Both algorithms yielded better returns than the simpler buying and selling policies that were actually carried out. Both contributions illustrate the potential of this research program. Wang and Zhou (2020) approach a continuous-time mean-variance portfolio selection with reinforcement learning and are also able to outperform conventional mean-variance portfolio selection and even an alternative deep learning neural network algorithm.

A set of relevant contributions also looks at the role of social interaction in individual behavior in order to approximate the economic theory of empirical evidence. In this context, Flieth and Foster (2002) highlight that interactive expectations involve discussion between agents and observation of a given group's behavior. Kaufman (1999) remarks that emotional reactions to events, which emerges from social interaction, are causes of biases and irrationality in decisions. Hommes (2006), in turn, observes that choices of agents are based also on influences from their peers, not only indirectly through the market behavior, but also directly through imitation, learning, peer pressure, information sharing or any other externality.

A specific example of social interaction in financial markets is Lux (1998). Using an agent-based model to separate the foreign exchange traders between fundamental and technical analysts, the study finds chaotic paths with appreciation peaks and huge drops in exchange rate, showing the importance of social interaction to the behavior of these trajectories. The social interaction model of Shanta *et al*. (2018) predicts that the desire for learning enable investors to inquire of the reasons and underlying strategies of others' behaviors, resulting unconscious or incidental revisions to their biased perspectives relating to stock trading.

## 3. MATERIALS AND METHODS

This section presents the building blocks of the model. The simulation environment illustrates a real decision problem involving the managing of a retirement portfolio fund. The portfolio is composed by four fixed assets of private credit and fixed income. The manager of the fund starts with an exogenous given amount of resources. The algorithm uses information on private indicators and social measures as feedbacks in selecting the best individual funds options. The strategy is reevaluated, and the resources are reinvested in time frequencies which vary according to six different scenarios detailed below. The theoretical results are compared to the actual performance of the biggest fixed income and private credit funds in Brazil.

### 3.1 Investment Funds in Brazil

Investment funds have an intermediary role between businesses and investors, thus being important products in financial markets. The funds are heterogeneous, each one having certain perks and guidelines for many kinds of investors. The degree of risk and variety of assets in its portfolio are the main features in rating them. Another important characteristic is related to the fund manager: a human, a robot, or a mixture of both. A quantitative fund is an investment fund that selects securities by using the capabilities of advanced quantitative analysis. Managers make use of technology to automate their work in order to exploit the market's behavior patterns and its inefficiencies.

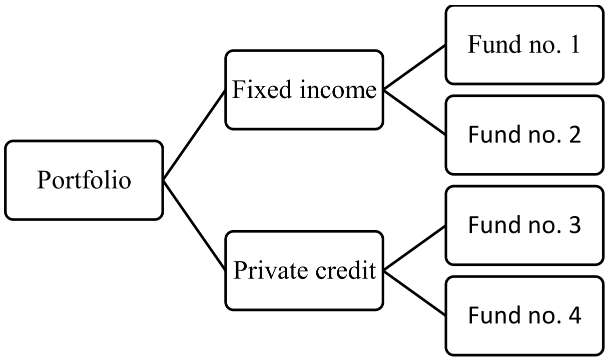
In this study, the simulation environment illustrates a real decision problem involving the management of a retirement portfolio fund. The main feature and role of this portfolio fund is to allocate its resources among alternative individual funds. It represents an investment fund in quotas of investment funds. By the Brazilian Securities Exchange Commission assortment, this class of investment has the obligation to possess at least 95% of its net worth allocated in quotas of funds from the same class. The only exception to this rule is the multimarket fund, which can invest in quotas of funds with different asset types.

In turn, our theoretical fund allocates its resources in two fixed and different kinds of funds of fixed income and private credit. As defined by the Brazilian Securities Exchange Commission assortment, fixed income funds must have the variation of interest rate or price index as the main risk factor in their portfolio, and also have at least 80% of their net worth allocated in treasury bonds and/or low risk fixed income securities. On the other hand, private credit funds must have at least 50% of their net worth in any kind of private-issued securities, such as debentures.

### 3.2 Investment Structure

The previous subsection described the fund's eligibility criteria. It is important to emphasize that the asset choice is not within the scope of the model. Based on this exogenous investment policy, the fund can allocate its resources between the two fixed income and the two private credit funds, each one of them under a different third-party manager. The investment structure embodied in the model is shown in Figure 1.

Figure 1 – Theoretical portfolio structure.




Source: Own Elaboration.

The management of a portfolio requires a complex structure of management and custody, which demands a series of third-party services with non-negligible costs that have a direct impact on profitability. Our work already takes these costs in consideration, discounting the administration and audit taxes in the simulated net worth.

Another important issue is on the existence of an allocative binding constraint: an asset cannot make up more than 97% of the net worth, neither can they be less than 1%. This constraint illustrates the policy allocation of the real portfolio fund and guarantees that all the four funds will take part in all of the simulation's periods. Furthermore, this fund cannot be leveraged nor does it leave any idle resources.

### 3.3 Feedbacks and Sensitivity

The algorithm solves the portfolio allocation problem by selecting the share of each individual funds through an endogenous mechanism of negative and positive feedbacks (reinforcement learning process). If a certain fund sends a positive feedback to the theoretical manager, a greater share of the net worth will be allocated in that fund in the next period. Conversely, in case of a negative feedback, the asset will receive a smaller share of resources in the next period. The algorithm uses two private criteria to attribute positive and negative feedbacks: the profitability criterion and the Sharpe ratio.

The profitability criterion runs as follows. At each period, the algorithm will rank each one of the four funds according to their respective rate of return and assigns positive feedback on the fund with the highest returns and negative feedback for all other funds. In the case of Sharpe ratio-based criterion, the algorithm will check whether the portfolio's Sharpe ratio decreased or increased. In case of decreasing, the algorithm will attribute a positive feedback towards the fund with the best Sharpe ratio and a negative feedback towards the others. The Sharpe ratio used in this criterion is the standard of the literature and considers the CDI rate (Brazil's interbank deposit certificate) as the risk-free asset.

In the reinforcement learning process, as stressed out by Roth and Erev (1995, p. 165), a decision strategy that presented a positive reward in the t period has its probability of being chosen in the t + 1 period reinforced. Thus, the fund that presented the highest profitability in the present period, will receive greater resources in the next period. All other funds will have less resource. This relationship is formally given by:

$$if\ \pi_{i,t} > \pi_{j,t} \rightarrow (1+\alpha)\ a_{i,t+1}, \tag{1}$$

$$if\ \pi_{i,t} < \pi_{j,t} \rightarrow (1-\alpha)\ a_{i,t+1}, \tag{2}$$

in which $\pi_{i,t} \in \boldsymbol{R}$ stands fund *i* profitability, measured by the fund's return; $\pi_{j,t} \in \boldsymbol{R}$ stands profitability of other funds; $\alpha \in \boldsymbol{R}$ represents sensitivity of adjustments in asset allocation. For greater return, allocation in period *t + 1* increases. Otherwise, allocation in the next period decreases; $a_{it} \in \boldsymbol{R} \subset (0,1)$ is the share of the total resources allocated in asset *i* at time.

In order to revise the Sharpe ratio (SR), the algorithm performs a check of whether the portfolio's Sharpe ratio on period t was greater or equal to the Sharpe ratio on period t – 1. Thus, the fund that results in the lowest Sharpe index in the t period, the allocation will be increased in the next the period. For all other funds the allocation is reduced in the next period. The relationship between the Sharpe ratio and the portfolio's allocation is formalized as follows:

$$if\ SR_{i,t} > SR_{j,t} \rightarrow (1 - \alpha)\, a_{i,t+1}, \tag{3}$$

$$if\ SR_{i,t} < SR_{j,t} \rightarrow (1 + \alpha)\, a_{i,t+1}, \tag{4}$$

In summary, profitability and Sharpe index are the private attributes that influence the agent in allocating the portfolio for the next period. However, the decision problem of the algorithm also selects the share of individual funds, considering what we call sensitivities - a proxy for the investor's mood. In the present study, sensitivities are related to the social interaction of the algorithm and the performance of different funds.

The role of social interaction is essential to the mechanism, as without it the algorithm cannot evaluate its performance in comparison to the environment. In fact, real-life managers look at market benchmarks and that is the motivation behind the third criterion. Hence, the algorithm includes a profitability comparison, in which the portfolio's profitability in the period is compared to the average market profitability. The market return is comprised of the ten biggest fixed income funds and the ten biggest private credit funds in Brazil, as listed by Ferreira (2019).

Sensitivity can vary between bold, moderate and conservative and is related to the performance comparison of the portfolio and the ten largest investment funds. Thus, if the return of the studied portfolio is greater than the performance of the ten largest funds, the algorithm will be conservative. On the other hand, if the portfolio's profitability is below the performance of the ten largest funds, the algorithm presents a bold behavior, seeking profits even if the risk exposure is greater.

In the first simulation period the fund starts as conservative, with the classification varying according to the fund's performance in each period. Our model evaluates two feedback magnitudes: one with 0.6% as the standard positive feedback and other one with 1.2% instead. We have to work with such diminutive rebalancing percentages because that is the way the pension fund market works, with small portfolio rebalancings.

To illustrate the process, with the smaller feedbacks: if the algorithm had a better profit than the market funds, for the profitability feedback the algorithm allocates 0.6% to the asset that had the best return and all other assets have a reduction of 0.2%. If the profitability is lower than the market performance, the algorithm will accept more volatility to make up for its underperformance and change its sensitivity by 1, with values inside the range [1, 3], which would mean a positive feedback of 1.8% and negative feedbacks of 0.6% at most. The exact same reasoning is valid for the Sharpe ratio feedback, except for the sensitivities inverted, as shown in Table 1.

Table 1. Corresponding sensitivities for both private attribute feedbacks

| Sensitivity | Profitability | Sharpe ratio |
|---|---|---|
| $\alpha$ | Conservative | Bold |
| $2\alpha$ | Moderate | Moderate |
| $3\alpha$ | Bold | Conservative |

Source: Own Elaboration.

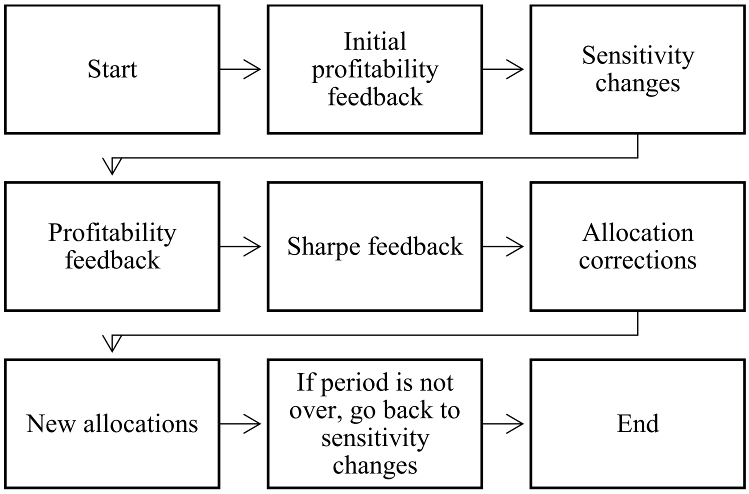
A common challenge in finance is to outperform a random strategy. Therefore, we compare our machine learning performance with the performance of an alternative algorithm that chooses each fund randomly, with equal probability ($p = 0.25$). In addition, we conducted paired $t$ tests between the actual results and each of our simulations, in order to check whether the mean difference between their returns is statistically different from zero.

## 4. COMPUTATIONAL IMPLEMENTATION AND EMERGENT PROPERTIES

### 4.1 Computational Implementation

Initial conditions in period $t = 0$ are the same as the allocation share of the real fund on 29 December of 2016. The only difference in period $t = 0$ is our profitability feedback. From $t > 0$ onwards, the algorithm always executes the same five-step series: Checks for sensitivity changes; feedbacks; allocative corrections; allocation decisions and a time check, as described previously and now shown in Figure 2.

Figure 2. Algorithm structure




Source: Own Elaboration.

### 4.2 Emergent Properties

In this section, we present the portfolio's results in different reallocation frequencies (1, 10 or 25 workdays) and feedback magnitudes (smaller or bigger). The simulation starts at 29 December 2016 and ends at 31 December 2020. The real retirement-portfolio fund yielded a profit of 32.29% over the period, or a CAGR of roughly 7.25%. This result is used as a benchmark in all the rebalancing frequencies and feedback magnitude simulation scenarios. There are six possible scenarios involving each frequency and feedback magnitudes.

Tables 2 through 6 presents the results of the six tests organized by the combinations of frequency and feedback magnitudes, comparing performances from the reinforcement learning algorithm against the real retirement-portfolio fund. Also, we calculate paired t-tests to check the statistical significance of the algorithm's mean difference to the real results. Table 2 shows the cumulative results from the entire periods, while the next four show the yearly comparisons.

Table 2. Performance of the algorithm against the real retirement-portfolio fund, considering six rebalancing frequencies, 2017-2020

| **Reallocation frequency** | **Simulated results** | **Difference to the actual results** | **t-test** |
|---|---|---|---|
| Daily, 0.6% | 32.77% | +0.48 | 0.3626 |
| Daily, 1.2% | 32.28% | -0.01 | -0.0071 |
| 10-day, 0.6% | 35.24% | +2.95 | 5.1347 *** |
| 10-day, 1.2% | 35.77% | +3.48 | 4.6457 *** |
| 25-day, 0.6% | 34.12% | +1.83 | 3.5418 *** |
| 25-day, 1.2% | 33.76% | +1.47 | 2.2139 ** |

Source: Own Elaboration.

The critical values are sampled from the t table. *, ** and *** respectively correspond to 90, 95 and 99% confidence levels.

Table 3. Performance of the algorithm against the real retirement-portfolio fund, considering six rebalancing frequencies, 2017

| **Reallocation frequency** | **Simulated results** | **Difference to the actual results** | **t-test** |
|---|---|---|---|
| Daily, 0.6% | 9.45% | -0.04 | -0.6845 |
| Daily, 1.2% | 9.42% | -0.04 | -0.8597 |

| | | | |
|---|---|---|---|
| 10-day, 0.6% | 9.93% | +0.01 | 5.0123 *** |
| 10-day, 1.2% | 9.87% | +0.00 | 4,8266 *** |
| 25-day, 0.6% | 9.99% | +0.01 | 3,6919 *** |
| 25-day, 1.2% | 10% | +0.02 | 2,0112 ** |

Source: Own Elaboration.

Table 4. Performance of the algorithm against the real retirement-portfolio fund, considering six rebalancing frequencies, 2018

| Reallocation frequency | Simulated results | Difference to the actual results | t-test |
|---|---|---|---|
| Daily, 0.6% | 8.66% | +0.08 | 0.0365 |
| Daily, 1.2% | 8.70% | +0.09 | 0.0075 |
| 10-day, 0.6% | 8.63% | +0.08 | 5.9124 *** |
| 10-day, 1.2% | 8.66% | +0.08 | 4.2189 *** |
| 25-day, 0.6% | 8.72% | +0.09 | 2.0934 ** |
| 25-day, 1.2% | 8.79% | +0.10 | 2.0120 ** |

Source: Own Elaboration.

Table 5. Performance of the algorithm against the real retirement-portfolio fund, considering six rebalancing frequencies, 2019

| Reallocation frequency | Simulated results | Difference to the actual results | t-test |
|---|---|---|---|
| Daily, 0.6% | 6.96% | 0.00 | 0.2348 |
| Daily, 1.2% | 6.68% | -0.03 | -0.0123 |
| 10-day, 0.6% | 6.55% | -0.05 | -5.5412 *** |
| 10-day, 1.2% | 7.71% | 0.07 | 4.8954 *** |
| 25-day, 0.6% | 7.70% | 0.07 | 3.9862 *** |
| 25-day, 1.2% | 7.56% | 0.05 | 2.4432 ** |

Source: Own Elaboration.

Table 6. Performance of the algorithm against the real retirement-portfolio fund, considering six rebalancing frequencies, 2020

| Reallocation frequency | Simulated results | Difference to the actual results | t-test |
|---|---|---|---|
| Daily, 0.6% | 4.54% | 0.02 | 0.0234 |
| Daily, 1.2% | 4.25% | -0.01 | -0.0953 |
| 10-day, 0.6% | 5.48% | 0.11 | 5.8231 *** |
| 10-day, 1.2% | 5.88% | 0.15 | 2.7638 ** |

| | | | |
|---|---|---|---|
| 25-day, 0.6% | 4.68% | 0.03 | 2.0983 ** |
| 25-day, 1.2% | 4.39% | 0.00 | 2.1043 ** |

Source: Own Elaboration.

In all the simulations except for one, the algorithm performed better than the actual results. The middle scenarios (10-day reallocation strategy) performed the best, followed by the 25-day scenarios, which math common practices of real-life pension funds. The worst scenario is the daily, which is the least feasible due to a lack of liquidity in the retirement fund market. Besides, except for the daily frequency algorithms, the others had at least 95% statistical significance, suggesting the algorithm can be useful as a support to allocation decisions involving a retirement fund.

In all of the simulations, the algorithm's results started distancing themselves from the actual results after May 2018, a very sensitive month for the Brazilian economy due to the national truckers' strike. Uncertainty ramped up in the following months and the algorithm gained more net worth than the actual portfolio. In this period, the asset manager might have been too much cautious due to a fear of big losses. But the daily frequency algorithms, which performed the worst, had their downward trend during the COVID-19 pandemic, showing that not necessarily a robot will outperform human managers in moments of crisis.

Still, these results are very stimulating as they show a conclusion similar to that of Matsui, Goto and Izumi (2009): a reinforcement learning framework for portfolio selection needs not to be very complicated and pick between hundreds of different securities, as a real-life fund could have its performance improved by a simple algorithm juggling through only a few other funds it could choose from.

It is also important to remind that working with computational mathematics and complexity science does not necessarily mean a radical departure from mainstream finance, as also shown by Wang and Zhou (2020), which show the potential of portfolio management framework (mean-variance selection, in that case) with reinforcement learning. This study, by employing a Sharpe ratio variant to compare our results to market benchmarks and a risk management that tries to find an optimal allocation by navigating through different volatility levels through trial and error, still keeps a bridge towards traditional financial economics.

## 5. CONCLUDING REMARKS

Motivated by the economic context, the social security reform in Brazil, and the behavior of Brazilian individual investors in the retirement-fund market, this study designs an algorithm to manage a pension fund. The study contributes to an emerging literature on computational finance by remarking the role of reinforcement learning and social interaction mechanism in allocation resource decisions.

The profitability and Sharpe ratio criteria were used by the machine to evaluate its own performance compared to the previous period. Social interaction, which is given by the observation of returns from other funds in the same market on the same period, aimed to improve the algorithm's realism, trying to emulate an asset manager's real-life behavior. Without this mechanism, the algorithm might have been satisfied by improving its performance compared to its own past even if it performed worse than the market.

The algorithm presented a consistent and stable performance in all the six scenarios of simulation, outperforming the actual portfolio by considerable and significant average return values. The simulated portfolio started to distance themselves from the actual one after May 2018, a moment of uncertainty in Brazil due to the national truckers' strike with significant macroeconomic consequences. A possible cause for this difference is the algorithm's impersonal and objective behavior, free of pressures that might hinder the work of a human manager in moments of great stress and doubt.

Even though this machine learning-based algorithm is relatively simple, it supplies a starting point for more complex decision-making structures, involving both more decision criteria (individual and social learning) and large sample contexts. Not only computational finance seems to be a fruitful research agenda, the results from the slower-paced algorithms (which mimic real-life pension funds with a certain lack of liquidity) are encouraging towards an eventual spread in real-life usage of these algorithms in a financial market sector which still shies away from employing them. Replicating this model for larger funds requires more attention, as these funds tend to have a big amount of public bonds in their portfolios.

## REFERENCES

Choi, J. J.; Laibson, D.; Madrian, B. C.; Metrick, A. Reinforcement learning and savings behavior. **The Journal of finance**, 64(6), 2515-2534, 2009.

Comissão de Valores Mobiliários. (2023). https://conteudo.cvm.gov.br/menu/regulados/fundos/consultas/fundos.html. Acesso em Março de 2023.

Fama, E. F.; French, K. R. A five-factor asset pricing model. **Journal of financial economics**, 116(1), 1-22, 2015.

Ferreira, G. The ten biggest pension funds in Brazil perform below CDI. Valor Econômico. Available in: https://www.valor.com.br/financas/6244401/maiores-fundos-de-previdencia-rendem-menosque-o-cdi, 2019.

Flieth, B.; Foster, J. Interactive expectations. **Journal of Evolutionary Economics**, 12(4), 375-395, 2002.

Forsyth, P. A.; Vetzal, K. R. Optimal asset allocation for retirement saving: Deterministic vs. time consistent adaptive strategies. **Applied Mathematical Finance**, 26(1), 1-37, 2019.

Gale, D.; Kariv, S. Bayesian learning in social networks. **Games and economic behavior,** 45(2), 329-346, 2003.

Hommes, C. H. Heterogeneous agent models in economics and finance. **Handbook of computational economics**, 2, 1109-1186, 2006.

Jiang, Z.; Liang, J. Cryptocurrency portfolio management with deep reinforcement learning. In 2017 **Intelligent Systems Conference (IntelliSys)** (pp. 905-913). IEEE, 2017.

Kaufman, B. E. Emotional arousal as a source of bounded rationality. **Journal of Economic Behavior & Organization**, 38(2), 135-144, 1999.

Lux, T. The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. **Journal of Economic Behavior & Organization**, 33(2), 143-165, 1998

Matsui, T.; Goto, T.; Izumi, K. Acquiring a government bond trading strategy using reinforcement learning. **Journal of Advanced Computational Intelligence and Intelligent Informatics**, 13(6), 691-696, 2009.

Mullainathan, S.; Spiess, J. Machine learning: an applied econometric approach. **Journal of Economic Perspectives**, 31(2), 87-106, 2017.

Nevmyvaka, Y.; Feng, Y.; Kearns, M. Reinforcement learning for optimized trade execution. In **Proceedings of the 23rd international conference on Machine learning** (pp. 673-680), 2006.

Roth, A. E.; Erev, I. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. **Games and economic behavior**, 8(1), 164-212, 1995.

Shantha, K. V. A.; Xiaofang, C.; Gamini, L. P. S. A conceptual framework on individual investors' learning behavior in the context of stock trading: An integrated perspective. **Cogent Economics & Finance**, 6(1), 1544062, 2018.

Topa, G. Social interactions, local spillovers and unemployment. **The Review of Economic Studies**, 68(2), 261-295, 2021.

Wang, H.; Zhou, X. Y. Continuous-time mean–variance portfolio selection: A reinforcement learning framework. **Mathematical Finance**, 30(4), 1273-1308, 2020.