

Reinforcement learning applied to a cryptocurrency portfolio

Abstract

In recent years, cryptocurrencies have been utilized as financial assets and have presented positive returns, albeit their volatility is high. This paper aims to elaborate a hypothetical cryptocurrency portfolio and to do so, uses machine learning and an optimization algorithm to define the ideal amount to be allocated in each asset. The results show the hypothetical portfolio presents superior returns and lesser volatility compared to other allocation strategies.

Keywords: Cryptocurrencies, reinforcement learning, return on assets.

JEL: C61; G11.

1. INTRODUCTION

Recently, cryptocurrencies have become more popular as a portfolio choice. As high volatility is intrinsic to this market, there is a need to develop a good strategy to build a portfolio with positive returns. However, the classic approaches to portfolio management in finance studies might face hardships in this market because of its instability.

Because of such instability, it might be said that an agent purporting to maximize his returns investing in cryptocurrencies finds himself inside a complex system. According to Arthur et al. (1997), in complex systems, interactions occur in a diffuse way between adaptative and heterogeneous economic agents. New features are introduced all the time, which maintains agents in a perpetual process of learning, adaptation and evolution, therefore, out of equilibrium.

Given this scenario, the methodology chosen is machine learning – particularly, reinforcement learning – to find the weights of each asset in the portfolio and so, the best allocation. In order for the results to be more faithful, the portfolio will go through daily rebalancing. It will be composed by the five cryptocurrencies with the biggest capitalization: Bitcoin (BTC), Ether (ETH), Ripple (XRP), Bitcoin Cash (BCH) and Litecoin (LTC).

In addition to this introduction, the paper has four more sections. Section two presents reference about cryptocurrencies and reinforcement learning. In turn, section three presents the structure and the main features of the adopted model. Section

four shows the results found and their discussion. To cap it off, section five presents the concluding remarks.

2. FRAMEWORK

2.1 Cryptocurrencies

According to Nakamoto (2009), a virtual coin is a chain of digital signatures, where two agents transfer assets by validating a public and a private key, both cryptographed. The entire ideology behind the creation of cryptocurrencies revolves around having a financial system which doesn't depend on trust in institutions and that safeguards the privacy of the agents involved.

According to Chiu and Koepl (2017), in a scenario with virtual coins, the means of payment are nothing but a set of bits. Specifically, big volumes of cryptographed data compose a transaction, and they compose a block. When a block is validated, it becomes part of a chain (hence the name blockchain).

Even though bitcoin is the most famous cryptocurrency, there are many more, but always revolving around the same blockchain concept. According to CoinMarketCap, in March 2018 there were 1592 listed cryptocurrencies, moving around 25 billion dollars every day, with a total market value of more than 400 billion dollars.

A specific feature of this kind of asset is high volatility. Yermack (2013) shows that correlation between Bitcoin and other coins is near zero, which mostly writes off its use as medium of exchange. So, the currency ends up more utilized as a speculative asset rather than a medium of exchange.

2.2 Reinforcement learning

According to Murphy (1998), 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.

By and large, the usage of machine learning in economics is related to the idea of complexity. To Arthur (2010), complexity in economics is linked to a tendency towards disequilibrium. In turn, Stodder (1995) relates

complexity to a web of complicated inter-relationships in which the interaction between involved parts in this complex system affects the individual behavior of each one of them. Simon (1962) understands that in complex systems, the whole is more than the sum of all parts, which means that it is hard to infer the properties of the whole.

To model complex systems, a popular approach is machine learning. Arthur (2006) argues that while neoclassical economics is used to studying actions, strategies and expectations of agents in equilibrium, reinforcement learning models allow the researchers to create environments and observe how agents react out of steady state. Moreover, Pyka and Fagiolo (2005) point out that computational models allow the insertion of heterogeneity between agents, emphasizing bounded rationality, bottom-up approaches and microfoundations of economics. Therefore, learning, adapting and evolving is a most important feature of these models.

According to Honda, Facure and Yaohao (2017), machine learning aims to enhance an agent's performance through conducting a task and acquiring experience. One method of machine learning is reinforcement learning, which employs trial and error, based in an entry function, and evaluates the results of each action taken. The process happens by way of attributing rewards and penalties to each decision, without need to necessarily determine how to undertake the task. In other words, this is a method that allows the algorithm to work through trial and error, being punished when it is wrong and rewarded when it is right. This kind of algorithm was used to develop the robots that play chess against professional payers.

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 himself 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.

3.CHOOSING THE CRYPTOCURRENCY PORTFOLIO BASED ON REINFORCEMENT LEARNING

3.1 Cryptocurrency selection

Consider an agent that aims to allocate his net worth, initially in \$1.000.000,00, among five different cryptocurrencies. The eligibility criterion was their capitalization. According to CoinMarketCap (2018), the five cryptocurrencies with the biggest capitalization were: *Bitcoin* \$142.478.043.110; *Ethereum* \$53.085.811.446; *Ripple* \$26.970.313.426; *Bitcoin Cash* \$16.502.623.412; *Litecoin*. \$8.735.794.413. These cryptocurrencies represent more than 85% of the capitalization in circulation.

By each time period t , the agent allocates his net worth in such a way that the sum of the modulus of the weight of each coin is equal to 100%. Precisely, the agent might, by using future contracts, realize buy (if he believes the price will rise) or sell (if he believes the price will fall) operations. To do so, he will have to leave the money counterpart to the total operated as insurance, so the net worth cannot be leveraged¹.

Besides choosing the position (buy/sell), the agent needs to define the intensity of each negotiation, which means choosing the percentage of the net worth that should be allocated in each asset. Assuming that the agent has the intent to diversify his portfolio, minimum and maximum ranges (limits) of the net worth percentage that can be allocated in each cryptocurrency are established. As seen next, the implemented reinforcement learning increases (reduces) the ranges in case the previous strategy has obtained success (failure). Lastly, it is worth mentioning that transaction costs associated to each portfolio reallocation are incremented daily. Therefore, by the end of each day, it is possible to calculate the portfolio's total net worth.

3.2 Data, portfolio composition and the reinforcement learning algorithm

This paper will observe the empirical series of real cryptocurrency prices from January 2018 to March 2018 with daily periodicity². After the price of the assets is obtained, it is possible to calculate the returns. Formally:

$$R_t = \sum_{i=1}^5 (P_{i,t} - P_{i,(t-1)}) * Q_{i,t}, \quad (1)$$

¹ Not allowing the agent to leverage his net worth makes the cryptocurrency portfolio returns more trustworthy for comparison.

² Data was obtained from CoinMarketCap (2018).

in which R is the asset return; P is the price; Q is the quantity; and t is the period analyzed.

Aiming to make the model the most realistic possible, the exchange costs of the BitMex exchange, one of the biggest cryptocurrency brokers in the world, will be considered. The transaction cost (TC) of an order is 0,05% of the order's value. Therefore, TC is given by:

$$TC_t = \sum_{i=1}^5 (|Q_{i,t} - Q_{i,(t-1)}| * P_{i,t} * 0.0005). \quad (2)$$

In the initial period, the portfolio's net worth is expressed by:

$$PL_0 = \sum_{i=1}^5 Q_i * P_{i,t}. \quad (3)$$

In turn, the portfolio's net worth in the ($t \geq 1$) periods given by:

$$PL_t = PL_{(t-1)} + R_t - TC_t. \quad (4)$$

The strategy review process occurs between the end of the t period and the t+1 period. While it happens, the agents evaluate the past performances of their strategies (allocations) and search for the strategies that offer the best rewards inside the possible ranges. In this adaptative process, the agents utilize the reinforcement learning mechanism to reevaluate and find the best allocation ranges.

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 the t+1 period reinforced. However, in the case of the payoff result being undesirable in the t period, the probability of this strategy being picked in the t+1 period is weakened. Formally, this relationship is expressed by:

$$(1 + 0.10) \text{ if } P_t > P_{(t-1)}, \quad (5)$$

$$(1 - 0.10) \text{ if } P_t < P_{(t-1)}. \quad (6)$$

By the allocation range updating function, it can be inferred that given a strategy that results in a positive return in the t period, the allocation range will be increased for the t+1 period. But if the chosen asset's price is reduced in the t period, its allocation range will be reduced in t+1 period, this way the decisionmaker tends to allocate less resources in the asset with the lesser reward.

Once the new allocation ranges are defined for the following period, the SLSQP optimization algorithm is utilized to find the best weights distribution for each cryptocurrency in the portfolio in order to maximize the present period's return, and these weights will be allocated in the asset selection process in the t+1 period.

Precisely, the optimization mechanism adopted aims to obtain the largest net worth from the previous day and return the optimal weights of each asset. This result is found by many tries of weight distributions and analysis of their results, returning the best one afterwards. These tries follow two constraints, namely: the first is related to the sum of the weights. The sum of the absolute value of the assets (bought/sold) must be equal to 1 (or 100%).

The second is related to the maximum and minimum ranges (limits) of each asset's allocation. All the currencies start the study with limits of (-40%, 40%). However, these limits are altered due to the implemented reinforcement learning mechanism. Once the optimal allocation for each asset in the next period is defined, a new round is initiated.

3.2.1 Computational implementation

To implement the agent's choice model, the software Jupyter Notebook – a text editor for the Python programming language – will be utilized. It must be reminded that for each t period of time, the agent will be faced with the same set of cryptocurrencies and has to find a combination (buy/sell) in such a way that all his net worth is utilized.

In order to define the model's initial condition and avoid possible selection biases, for the $t=0$ period, it is assumed that the five cryptocurrencies are adopted in a uniform way (20% of the net worth in each asset) and only the buying position is able to be carried out in this initial period.

After the designation of initial conditions, it is possible to obtain the return of assets (equation 1) and subsequently calculate the operation costs (equation 2). Afterwards, the SLSQP optimization algorithm is utilized to find the best weight distribution for each cryptocurrency in the portfolio in order to maximize the return in the present period, and these weights will be allocated in the process of asset choice in the next period. Therefore, it is possible to calculate the portfolio's net worth (equation 4).

All the procedures described here for the optimal weight choice of assets in the $t=1$ period can be applied for any other $t \geq 2$ period.

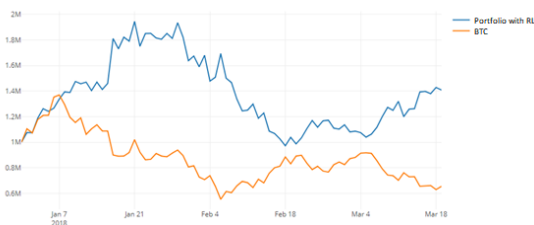
4. RESULTS

4.1 Portfolio results

This section presents the portfolio results with an initial net worth of \$1.000.000,00 and the analysis period from January 2018 to March 2018. Both for the theoretical portfolio and for Bitcoin (BTC), the cumulative financial returns (eq.4), the cumulative percentage returns, the total costs (eq.3), the variance, the largest drawdown (largest fall in a day) and the average daily returns were calculated (eq.1).

Figure 1 shows the net worth of the cryptocurrency portfolio proposed and all the net worth invested in Bitcoins.

Figure 1 – Comparing with Bitcoin

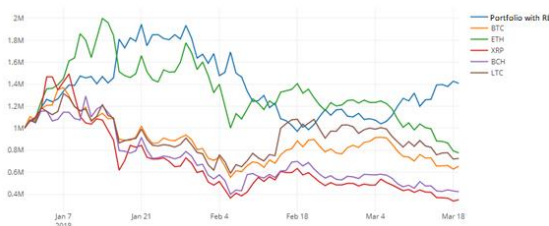


Own elaboration.

It is possible to see that the theoretical portfolio keeps constantly above Bitcoin, except for a few days in the sample's beginning. At the period's end, the portfolio's cumulative financial return is a\$ 455.508,57 while Bitcoin's is a loss of \$ 344.421,05. That means a return of 45,55% for the portfolio, against -34,44 for Bitcoin. It is worth mentioning that this return already accounts for transaction costs, which totaled \$ 86.000,12. As there is no brokerage when acquiring only Bitcoin with the starting net worth, the transaction cost in that case is zero.

Not only against Bitcoin is the theoretical portfolio superior. Figure 2 shows that, even though Ethereum (ETH) was superior during some moments, at the period's end all of the cryptocurrencies lost out to the proposed portfolio.

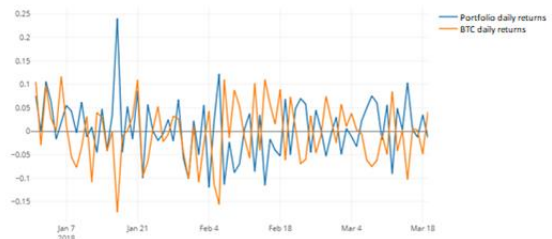
Figure 2 – Comparing with the five biggest cryptocurrencies



Own elaboration.

With regard to the risk-associated variables, the chosen measures were the portfolio's variance and maximum drawdown against Bitcoin's. In the study's period, the proposed portfolio's variance was 0,59% and Bitcoin's was 0,41%, showing the portfolio had a bigger risk. However, the maximum daily drawdown was bigger in Bitcoin's case. While the portfolio's biggest loss in a day was 15,74%, Bitcoin's was 17,23%. To measure the daily returns of both assets, the average daily returns were utilized. Figure 3 shows this data in a visual way.

Figure 3 – Daily returns with reinforcement learning



Own elaboration.

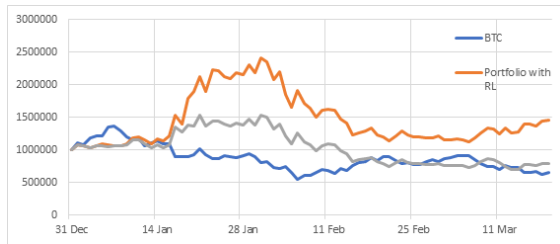
The portfolio's average results were much superior to Bitcoin, as the portfolio had a 0,76% average daily return while Bitcoin had -0,33%. It is possible to see in the Figure that the returns of both assets were, to a certain extent, inversely correlated in most days.

The results obtained show that the value invested in the theoretical portfolio would have been superior to any cryptocurrency. The portfolio's biggest drawdown was inferior to Bitcoin and the portfolio's returns were much bigger than Bitcoin's returns – even considering transaction costs. The portfolio had a bigger variance (hence, it was riskier) but the difference was very small.

4.2 Comparing to other strategies

For comparison purposes, two more strategies were tested, in order to understand how the resources directed to allocation decisions impact results. The first one was the same model as the portfolio but keeping the allocation limits fixed from -40% to 40%, therefore, without the reinforcement learning mechanism. This strategy's result compared to Bitcoin can be observed in Figure 4.

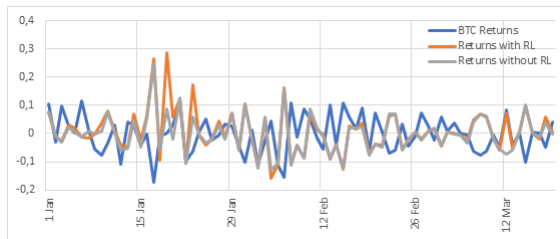
Figure 4 – Comparing to Bitcoin without reinforcement learning



Own elaboration.

This strategy had much worse returns, demonstrating reinforcement learning’s value to improve allocation decisions as time goes on. This strategy’s cumulative return was a loss of \$ 206.582,75 – equivalent to -20,66%, which is closer to Bitcoin’s result that was shown in the previous sub-section. The variance was also closer to Bitcoin, with a value of -0,43%. There was, too, a rather insignificant reduction in the maximum daily drawdown, which was 12,84%. The Figure 5 shows the daily return for each day of the study for the model without reinforcement learning compared to Bitcoin. The average daily return was negative, of -0,08% against -0,33% from Bitcoin.

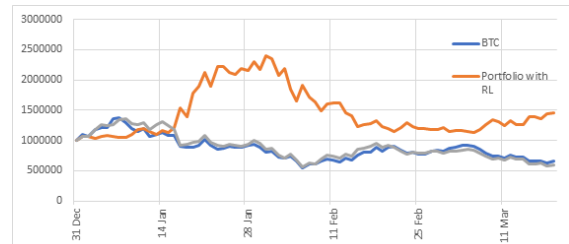
Figure 5 – Daily returns without reinforcement learning



Own elaboration.

The second model for comparison is a portfolio with fixed weights (20% for each currency). This is the most ingenu model among the tests, as it not only does not learn with the right and wrong decisions from before but does not even take allocation decisions. The Figure 6 shows the comparison to Bitcoin.

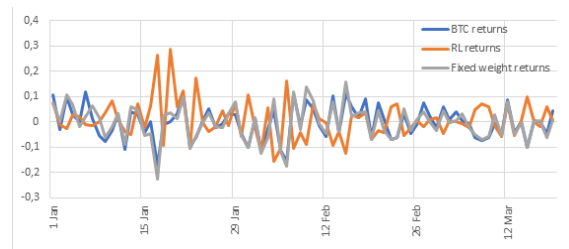
Figure 6 – Comparing fixed weights to Bitcoin, no reinforcement learning



Own elaboration.

The period’s cumulative financial return is a loss of \$ 410.371,29, superior even to the loss from putting everything in Bitcoin for the same period. Even so, the variance was superior to Bitcoin, with a value of 0,46%. The portfolio’s biggest daily loss was 22,59%, much superior to Bitcoin and the biggest among the analyzed strategies. The Figure 7 shows the returns for each day of the series for the fixed weights strategy and Bitcoin.

Figure 7 – Daily returns with fixed weights, no reinforcement learning



Own elaboration.

This strategy’s average return was -0,43%, the worst among all strategies, and the maximum drawdown was also the biggest. An interesting observation from these results is that with fixed weights of 20% for each currency, the daily returns are very close to Bitcoin’s, which shows high correlation of the other currencies with it.

For a more comprehensive analysis of the more ingenu strategies and the proposed portfolio, the approach used will be inspired on the Sharpe ratio, very famous in traditional financial markets. Formally, it is given by:

$$RR = \frac{PR - BR}{PV} \tag{7}$$

In which: RR is Risk/return ratio; PR is Portfolio’s return; BR is Bitcoin’s return; PV is Portfolio’s variance.

For all estimates, Bitcoin’s return is - 34,44%. For the theoretical portfolio proposed, in which the period’s return is 45,55%, and variance is 0,59%, the ratio is:

$$RR = \frac{0,4555 - (-0,3444)}{0,0059}$$

$$RR = 135,58$$

For the comparative portfolio without reinforcement learning, in which the period's return is -20,66%, and variance is 0,43%, the ratio is:

$$RR = \frac{(-0,2066) - (-0,3444)}{0,0043}$$

$$RR = 32,05$$

For the comparative portfolio with fixed weights of 20%, in which the period's return is -41,03%, and variance is 0,46%, the ratio is:

$$RR = \frac{(-0,4103) - (-0,3444)}{0,0046}$$

$$RR = -14,33$$

After these calculations, it can be seen that analyzing risk-return through this approach, the proposed portfolio was much superior to comparative portfolios. As a matter of fact, the fixed weights model actually has negative risk-return. That happens because, even though this study does not focus on risk reduction, but on return on asset maximization, there was low variation between the variances of the proposed portfolio compared to Bitcoin, but a big raise in expected returns.

5. Concluding remarks

In the last few years, the advent of cryptocurrencies like Bitcoin caused a stir among financial market players. Cryptocurrencies are financial assets that might yield high returns, and it is worth to explore the subject from a portfolio optimization return, as envisioned by Markowitz (1952). To find the optimal results of a cryptocurrency portfolio, the chosen method was an agent-based model subject to reinforcement learning, which rewards successes and punishes mistakes in the market's direction. The made-up portfolio started with a net worth of \$ 1.000.000,00, with the study taking place from 01/01/2018 to 19/03/2018.

At the chosen period's end, the portfolio's net worth was bigger than it would be if invested in any of the five individual cryptocurrencies. There were also comparisons to other strategies. Firstly, a portfolio without reinforcement learning, with 40% ranges. For this strategy, the results were worse than the proposed portfolio but still better than

investing all the net worth in Bitcoin. Other model was an allocation with ingenu weights – fixed in 20% per cryptocurrency. This was the worst allocation by all measurements. An interesting fact is that this strategy's returns and daily variation were very similar to Bitcoin's, showing high correlation among the cryptocurrencies.

Table 1 synthesizes the risk and return analysis, utilizing an own ratio (RR) based on the Sharpe ratio.

Table 1 – Stats by portfolio

Portfolio	Biggest drawdown	Returns	Variance	RR
Machine learning	15,74%	45,55%	0,59%	135,58
40% limits	12,84%	-20,66%	0,43%	32,05
20% fixed weights	22,59%	-41,04%	0,46%	-14,33
Only Bitcoin	17,23%	-34,44%	0,41%	-

Own elaboration.

Therefore, it can be observed that the intelligent portfolio which employs reinforcement learning boasts a risk-return ratio much better than the other portfolios (Bitcoin was used as the basis return for the ratio, so it does not have a RR). The portfolio with ingenu weights goes as far as having negative RR. Even though the work was focused on return maximization, it is easy to see that, based on available results, there is no big risk difference among the options, while returns are raised immensely.

This paper demonstrates the utility in employing computational methods such a machine learning in the solving of economic problems, especially in the finance area, allowing the investor to explore uncharted territories much better than more ingenu investors. And the usage of an agent under reinforcement learning dramatically improves his returns against the rest of the market, thus being an excellent option for modeling resource allocation for cryptocurrency investments.

REFERENCES

- ARTHUR, W. B., DURLAUF, S. N. & LANE, D. A. (1997), **The economy as an evolving complex system ii. santafe institute studies in the sciences of complexity**, Medwood City: Addison-Wesley, 1997. 608 p.
- ARTHUR, W.B. Out-of-Equilibrium Economics and Agent-Based Modeling. In: TESFATSION, Leigh; JUDD, K.L.. **Handbook of Computational Economics, Volume 2**. Amsterdam: Elsevier, 2006. Cap. 32, p. 1551-1564.
- ARTHUR, W. B. **Complexity, the Santa Fe approach, and non-equilibrium economy**.

- History of Economic Ideas, Fabrizio Serra Editore, Pisa - Roma, v. 18, n. 2, p. 149-166, 2010.
- CHIU, Jonathan; KOEPL, T.V. **The Economics of Cryptocurrencies – Bitcoin and Beyond**, 2017. Available in: <https://www.chapman.edu/research/institutes-and-centers/economic-science-institute/_files/ifree-papers-and-photos/koeppl-april2017.pdf>. Access in: 01 jun 2018.
- CoinMarketCap: Cryptocurrency Market Capitalizations**. <<https://coinmarketcap.com/>>. Acesso em: 01 jun 2018.
- DIAS JÚNIOR, E.P.F. **Aprendizado por reforço sobre o problema de reavistação de páginas web**, 2012. Dissertação (MSc in Computing) – ComputingGradSchool, Pontifícia Universidade Católica, Rio de Janeiro.
- HONDA, H., FACURE, M., PENG, Y. **The three types of machine learning**, 2017. Available in: <<https://lamfo-unb.github.io/2017/07/27/tres-tipos-am-english>>. Acesso em: 01 jun 2018.
- MARKOWITZ, Harry. **Portfolio Selection**. Journal of Finance, v. 7, p. 77- 91, mar 1952.
- MINSKY, M.L. **Steps toward artificial intelligence**. Proceedings I.R.E., v. 49, p. 8-30, 1961.
- MURPHY, **A brief introduction to reinforcement learning**, 1998. Available in: <<https://www.cs.ubc.ca/~murphyk/Bayes/pomdp.html>>. Acesso em: 01 jun 2018.
- NAKAMOTO, Satoshi. **Bitcoin: A Peer-to-Peer Electronic Cash System**, 2009. Available in: <<https://bitcoin.org/bitcoin.pdf>>. Acesso em: 01 jun 2018.
- PYKA, A.; FAGIOLO, G. **Agent-based Modelling: A Methodology for Neo-Schumpeterian Economics**. Discussion Papers Series, v. 272, Universitaet Augsburg, Institute for Economics, 2005.
- ROTH, A. E.; EREV, I. **Learning in extensive form games: Experimental data and simple dynamic models in the intermediate run**. Games and Economic Behavior, v. 8, p. 164-212, 1995.
- SIMON, Herbert **The Architecture of Complexity**. Proceedings to the American Philosophical Society, v. 106, n. 6, p. 467-482, 1962.
- STODDER, James P. 1995. **The Evolution of Complexity in Primitive Economies: Theory**. Journal of Comparative Economics, v. 20, n.1, p. 1–31, feb. 1995.
- SUTTON, R.S. **Introduction: The Challenge of Reinforcement Learning**. In: Sutton R.S. **Reinforcement Learning**. The Springer International Series in Engineering and Computer Science (Knowledge Representation, Learning and Expert Systems), v. 173. Boston, MA: Springer, 1992.
- WALTZ, M.D., FU, K.S. **A heuristic approach to reinforcement learning control systems**. IEEE Transactions on Automatic Control, v. 10, n. 4, p. 390–398, 1965.
- WATKINS, C.J.C.H. **Learning with delayed rewards**, 1989. Tese (PhD in Psychology), Department of Psychology, Cambridge University.
- WERBOS, P.J. **Building and understanding adaptive systems: A statistical/numerical approach to factory automation and brain research**. IEEE Transactions on Systems, Man and Cybernetics, v. 17, n. 1, p. 7-20, jan-fev 1987.
- YERMACK, David. **Is Bitcoin a Real Currency? Aneconomicappraisal**. NBER WorkingPapers, dez 2013.