Deep Reinforcement Learning Agents in Algorithmic Trading

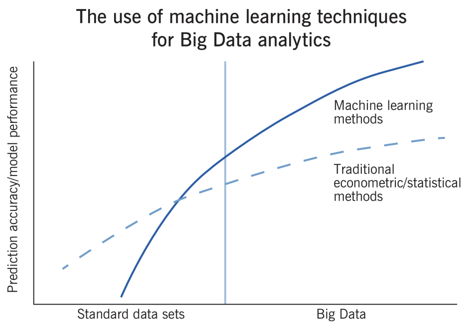
Michał Maria Andrzejewski

15 November 2024

Introduction

Traditionally, quantitative finance has relied heavily on statistical and econometric models in order to forecast market fluctuations. These models were built upon a set of mathematical methods, grounded in advanced probability theory, such as time-series models, continuous-time stochastic processes, or Monte Carlo simulations to optimize the returns on investments or trades in an ever-changing financial landscape. However, these models, which all fundamentally resemble multivariate linear regression models[1], are inherently limited by their assumption of linear relationships—a notion dating back to the 18th-century work of Carl Friedrich Gauss on geodesic and astronomical datasets[2], incomparable in scale to the magnitude of information that stimulates financial decision-making these days.

With the rise of machine learning, a new influx of powerful methods has entered the field, offering the ability to capture much more complex, nonlinear relationships in vast datasets, thereby addressing some limitations of traditional models[3]. This shift reflects both the technological progress and the growing demand for methods capable of handling today's dynamic, intricate financial systems that require a larger number of predictors than the purely econometric approaches were able to analyse. The machine learning algorithms allow predictions based on diverse data sources beyond classic financial statements including patterns found in alternative sources like the news, credit card transactions, or social media. As a result, these algorithms are exposed to a broader range of inputs to draw from, offering richer insights into market sentiment and behavior[4], often outperforming traditional methods, when trained on sufficiently large datasets, as shown in Figure 1 - making them attractive tools for modern trading.



r

Figure 1: Performance Comparison of Machine Learning and Econometric Methods. Source: Harding, M., Hersh, J. Big Data in economics. *IZA World of Labor* 2018: 451 doi: 10.15185/izawol.451

[1] For instance, a multivariate linear regression model can be represented as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon,$$

where y is the dependent variable, $x_1, x_2, \ldots, x_n$ are independent variables, β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are coefficients, and ϵ is the error term.

[2] Lopez de Prado, M. (2018, October 20). *Advances in Financial Machine Learning: Lecture 1/10 (seminar slides)* [Presentation slides]. Cornell University ORIE 5256. Retrieved from `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3252423`

[3] López de Prado, Marcos. Machine Learning for Asset Managers. United States, Cambridge University Press, 2020.

[4] Zhang, Zihao, Stefan Zohren, and Stephen Roberts. *Deep Reinforcement Learning for Trading.* Department of Engineering Science, Oxford-Man Institute of Quantitative Finance, University of Oxford, 2020.

Machine Learning in Algorithmic Trading

Since algorithmic trading has become so highly prevalent, accounting for around 75% of trading volume in the U.S. stock exchange[5], the interest in finding trading models that produce greater alpha[6] has surged. Leveraging the availability of large amounts of financial data, the recent rapid advances in artificial intelligence have "emerged as a revolutionary element in algorithmic trading"[7], incorporating the unique advantages that the machine learning paradigms bring to the table, as presented in the Figure 2.

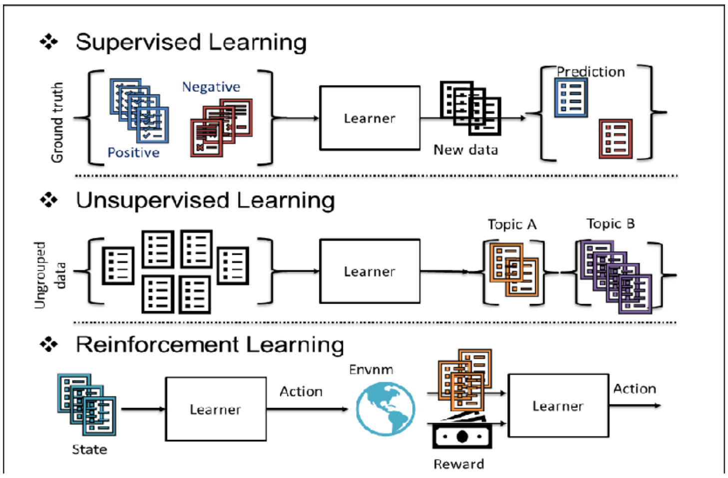


Figure 2: Machine Learning Paradigms in Quantitative Finance. *Source:Koshiyama, Adriano Firoozye, Nick Treleaven, Philip. (2020). Algorithms in Future Capital Markets. SSRN Electronic Journal. 10.2139/ssrn.3527511.*

Among these three, reinforcement learning (RL) has become particularly promising by allowing an agent (the trading algorithm) to learn optimal strategies through continuous interaction with the environment (the financial market), without a priori assumptions about its dynamics, leading to rapid progress in adapting RL techniques in algorithmic trading [8].

Methods of Reinforcement Learning

In markets such as FX, cryptocurrency, or stocks, RL algorithms can be policy-based, hybrid (e.g., actor-critic), or value-based. Due to the contemporary importance of value-based methods, which we'll elaborate on later, we will illustrate their construction, which require defining a Markov Decision Process (MDP) aligned with the trading objective and involve two main approaches[9]:
Equation 1.1: Infinite Time Horizon with Discounted Rewards used usually for Portfolio Optimization

$$V(s) = \sup_{\pi} E^{\pi} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s \right] \tag{1}$$

.

Equation 1.2: Finite Time Horizon with Terminal Reward, typically for High-Frequency Trading.

$$V_t(s) = \sup_{\pi} E^{\pi} \left[\sum_{u=t}^{T-1} r(s_u, a_u) + r_T(s_T) \mid s_t = s \right] \tag{2}$$

,where: V and V_t are the value functions of the strategy π dependant on the state $s \in S$, the action $a \in A$, and the discount factor, $\gamma \in [0, 1]$, corresponding to the respective 5-tuple (S, A, P, R, γ) of the

[5] Perez, Eugenia & Parra-Dominguez, Javier & Omatu, Sigeru & Herrera-Viedma, Enrique & Corchado Rodríguez, Juan. (2021). *Machine Learning and Traditional Econometric Models: A Systematic Mapping Study. Journal of Artificial Intelligence and Soft Computing Research*, 12, 79-100. doi:10.2478/jaiscr-2022-0006

[6] An investment strategy's ability to beat the market

[7] Team DigitalDefynd. *10 Ways AI Is Being Used in Algorithmic Trading.* 2024. DigitalDefynd, Retrieved from `https://digitaldefynd.com/IQ/ai-in-algorithmic-trading/`

[8] Hambly, B., Xu, R., Yang, H. (2022). *Recent Advances in Reinforcement Learning in Finance.* February 3, 2022.

[9] Ibid.

MDP, with P and R representing the transition probabilities and the reward function. [10]

With the MDP established, we can apply value-based methods such as Q-learning, which aims to identify an optimal policy by estimating the Q-function, $Q(s,a)$, an extension of the value function $V(s)$ (Equation 1) or $V_t(s)$ (Equation 2) that also incorporates the reward for action a in state s and then following the optimal policy thereafter[11].

$$Q(s,a) = E\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, a_0 = a\right] \tag{3}$$

To iteratively update $Q(s,a)$, Q-learning relies on the Bellman equation [12]:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[r(s,a) + \gamma \max_{a'} Q(s',a') - Q(s,a)\right] \tag{4}$$

Here, α is the learning rate.

The Q-learning algorithm follows these steps[13]::

1. **Initialize** $Q(s,a)$ for all $s \in S$ and $a \in A$.
2. **Loop until convergence**:
   - Choose a for s using a policy derived from Q (e.g., ϵ-greedy).
   - **Observe** the reward r and next state s'.
   - **Update** $Q(s,a)$ according to the Bellman equation.
   - **Set** $s = s'$.

Through repeated application, this algorithm approximates Q^*, the optimal Q-function, converging towards an optimal trading policy. However, it has limitations in financial markets where the action space evolves every second [14], necessitating manual adjustments of MDP parameters and limiting full autonomy in real-time trading.

The Promise of Deep Reinforcement Learning

Following the introduction of the Deep Q-network (DQN)[15], which efficiently handles high-dimensional inputs using deep neural networks, researchers have focused on parameterizing value functions and policies, including approximating the transition and reward functions within the Markov Decision Process (MDP) framework.[16]

In DQN, the Q-value function $Q(s,a;\theta)$ is approximated by a neural network with parameters θ, where the loss function used to train the network is the mean-squared error between predicted and target Q-values[17]:

$$L(\theta) = E_{(s_t,a_t,r_{t+1},s_{t+1})}\left[(y_t - Q(s_t,a_t;\theta))^2\right]$$

with the target y_t computed using periodically updated target network parameters θ^-:

$$y_t = r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-)$$

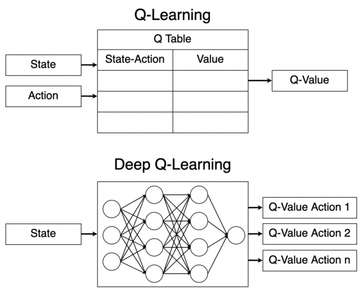


Figure 3: Comparison of Q-learning vs Deep Q-learning. Source: Sebastianelli, A., Tipaldi, M., Ullo, S., Glielmo, L. (2021). "A Deep Q-Learning Based Approach Applied to the Snake Game."

[10] Y. Li, "Deep reinforcement learning: An overview," arXiv preprint arXiv:1701.07274, 201

[11] Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3-4), 279-292.

[12] Bellman, R. (1957). *The Theory of Dynamic Programming.*

[13] Hambly, B., Xu, R., Yang, H. (2022). *Recent Advances in Reinforcement Learning in Finance*

[14] Bacoyannis, V., et al. *Idiosyncrasies and Challenges of Data-Driven Learning in Electronic Trading.* arXiv preprint, 2018.

[15] Mnih, V., Kavukcuoglu, K., Silver, D. et al. Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015). `https://doi.org/10.1038/nature14236`

[16] Hambly, B., Xu, R., Yang, H. (2022). *Recent Advances in Reinforcement Learning in Finance.* February 3, 2022.

[17] Alvaro Cartea, Sebastian Jaimungal, and Leandro Sánchez-Betancourt, *Deep Reinforcement Learning for Algorithmic Trading*

In recent years, the increasing application of neural networks, ranging from simple fully-connected neural networks (FNN) to more sophisticated recurrent convolutional neural networks (RCNN), in algorithmic trading has garnered attention. Many studies suggest that variations of DQN, which replace the conventional Q-table with neural networks to approximate Q-values (as depicted in Figure 3), offer considerable advantages in financial markets due to their ability to generalize complex patterns in financial data and autonomously adapt to changes occurring in them.

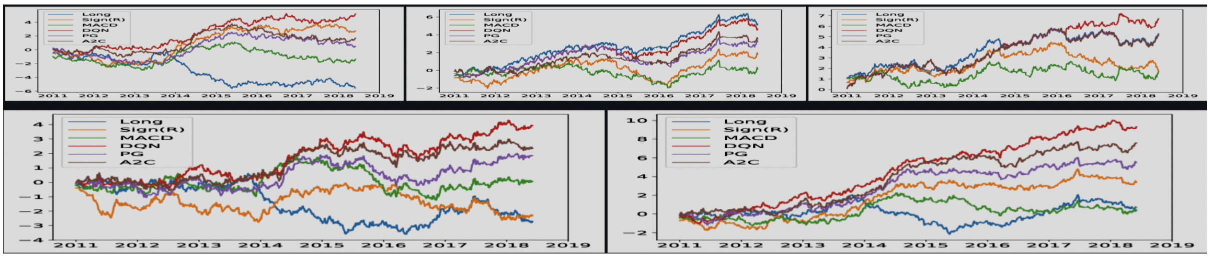


Figure 4: Cumulative trade returns from the paper "Deep Reinforcement Learning for Trading" by Zhang et al. First row: commodity, equity index, and fixed income; second row: FX and portfolio of all contracts.

Papers such as *Deep Reinforcement Learning for Trading*[18] present in their experiments how DQN can outperform other methods in cumulative trade returns for various markets (commodity, equity index, fixed income, FX, and portfolio optimization) . In other papers, which specify slightly different environments and objectives, we can see its' cousin algorithm, Deep Recursive Q-learning (DRQN) showing superior performance[19], as has Proximal Policy Optimization (PPO) in high-frequency trading[20]. In sentiment analysis applications, models combining Deep Deterministic Policy Gradient (DDPG) with RCNN have also shown promising results.[21] Yet, while these studies outline impressive gains, it's crucial to avoid idealizing DRL agents, as their actual effectiveness may vary in live, unpredictable market settings.

Conclusions and Concerns

There is no denying that Deep Reinforcement Learning algorithms are powerful tools with immense potential in algorithmic trading. Their high performance and flexibility, grounded in strong capabilities for recognizing patterns and adapting dynamically to data from diverse sources, offer promising pathways for financial applications as the technology advances. Nonetheless, the growing popularity of AI has fostered a layer of perceived mysticism and infallibility, which sometimes overshadows critical flaws in current research.

While impressive in backtesting environments, these algorithms often struggle to remain as consistent in live market conditions[22].This discrepancy is often rooted in the overfitting, instability, and failure to generalize that DRL models exhibit in real-time trading scenarios[23]. While favorable results in historical backtests are commonly presented, the models often falter under unpredictable market conditions. This trend of focusing on idealised backtest results, while downplaying real-world failures, is a consistent critique of DRL applications in finance. Furthermore, effective deployment in actual trading requires continuous retraining, adaptability to market dynamics, and rigorous handling of noisy financial data.

While DRL-based algorithms hold "revolutionary" potential for trading strategies, it is important to recognize that the true factors enabling success in algorithmic trading are access to high-quality

[18]Zhang, Zihao, Zohren, Stefan, Roberts, Stephen. (2020). Deep Reinforcement Learning for Trading. *The Journal of Financial Data Science*, 2(2), 25-40. `https://doi.org/10.3905/jfds.2020.1.030`

[19]Chen, L., and Gao, Q. "Application of Deep Reinforcement Learning on Automated Stock Trading." *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*, IEEE, 2019, pp. 29-33. `https://doi.org/10.1109/ICSESS47205.2019.9040728`.

[20]Briola, Antonio, et al. "Deep Reinforcement Learning for Active High Frequency Trading." *arXiv*, 2021. `https://doi.org/10.48550/arXiv.2101.07107`.

[21]Azhikodan, Akhil, et al. "Stock Trading Bot Using Deep Reinforcement Learning." *Lecture Notes in Electrical Engineering*, vol. 505, 2019. `https://doi.org/10.1007/978-981-10-8201-6_5`.

[22]Pricope, Tidor-Vlad. "Deep Reinforcement Learning in Quantitative Algorithmic Trading: A Review." 2021. arXiv, doi:10.48550/arXiv.2106.00123.

[23]Millea, A. "Deep Reinforcement Learning for Trading—A Critical Survey." Data, vol. 6, no. 11, 2021, p. 119, doi:10.3390/data6110119.

market data, computational power, and the structured preprocessing of this data to train algorithms effectively. As these algorithms assume greater control over substantial asset values, their transparency becomes essential to mitigate the risks associated with "black-box" models operating on a global scale. A malfunction or misinterpretation could have far-reaching consequences, potentially destabilizing markets in times of economic distress. Consequently, it is equally vital to consider the regulatory framework in which these algorithms operate, ensuring compliance to minimize risks of market manipulation and regulatory failures[24].

References

Azhikodan, Akhil et al. (2019). *Stock Trading Bot Using Deep Reinforcement Learning.* Lecture Notes in Electrical Engineering, vol. 505, `https://doi.org/10.1007/978-981-10-8201-6_5`.

Azzutti, Alessio et al. (2021). "Machine Learning, Market Manipulation and Collusion on Capital Markets: Why the 'Black Box' Matters". In: *European Banking Institute Working Paper Series 2021, no. 84.* University of Pennsylvania Journal of International Law, vol. 43, no. 1, SSRN, `doi:10.2139/ssrn.3788872`.

Bacoyannis, V. et al. (2018). "Idiosyncrasies and Challenges of Data-Driven Learning in Electronic Trading". In: arXiv preprint.

Bellman, R. (1957). *The Theory of Dynamic Programming.*

Briola, Antonio et al. (2021). "Deep Reinforcement Learning for Active High Frequency Trading". In: arXiv, `https://doi.org/10.48550/arXiv.2101.07107`.

Cartea, A., S. Jaimungal, and L. Sánchez-Betancourt (2022). *Deep Reinforcement Learning for Algorithmic Trading.*

Chen, L. and Q. Gao (2019). "Application of Deep Reinforcement Learning on Automated Stock Trading". In: *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*, pp. 29–33. DOI: `10.1109/ICSESS47205.2019.9040728`.

DigitalDefynd, Team (2024). *10 Ways AI Is Being Used in Algorithmic Trading.* DigitalDefynd. Retrieved from `https://digitaldefynd.com/IQ/ai-in-algorithmic-trading/`.

Hambly, B., R. Xu, and H. Yang (2022). "Recent Advances in Reinforcement Learning in Finance". In: February 3, 2022.

Li, Y. (2017). *Deep reinforcement learning: An overview.* arXiv preprint arXiv:1701.07274.

Lopez de Prado, M. (Oct. 2018). *Advances in Financial Machine Learning: Lecture 1/10 (seminar slides).* Cornell University ORIE 5256. Retrieved from `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3252423`.

Lopez de Prado, Marcos (2020). *Machine Learning for Asset Managers.* Cambridge University Press.

Millea, A. (2021). "Deep Reinforcement Learning for Trading—A Critical Survey". In: *Data* 6.11, p. 119. DOI: `10.3390/data6110119`.

Mnih, V., K. Kavukcuoglu, D. Silver, et al. (2015). "Human-level control through deep reinforcement learning". In: *Nature* 518, pp. 529–533. DOI: `10.1038/nature14236`.

Perez, Eugenia et al. (2021). "Machine Learning and Traditional Econometric Models: A Systematic Mapping Study". In: *Journal of Artificial Intelligence and Soft Computing Research* 12, pp. 79–100. DOI: `10.2478/jaiscr-2022-0006`.

Pricope, Tidor-Vlad (2021). "Deep Reinforcement Learning in Quantitative Algorithmic Trading: A Review". In: arXiv, `doi:10.48550/arXiv.2106.00123`.

Watkins, C. J. C. H. and P. Dayan (1992). "Q-learning". In: *Machine Learning* 8.3-4, pp. 279–292.

Zhang, Zihao, Stefan Zohren, and Stephen Roberts (2020). *Deep Reinforcement Learning for Trading.* Department of Engineering Science, Oxford-Man Institute of Quantitative Finance, University of Oxford.

[24] Azzutti, Alessio, Ringe, Wolf-Georg, and Stiehl, H. Siegfried, "Machine Learning, Market Manipulation and Collusion on Capital Markets: Why the 'Black Box' Matters," *European Banking Institute Working Paper Series* 2021, no. 84, *University of Pennsylvania Journal of International Law*, vol. 43, no. 1, 2021, SSRN, doi:10.2139/ssrn.3788872.