Research and Analysis of Real-World Asset Token Pricing Model Based on Machine Learning

Zekai Mo

School of Business, Hong Kong Baptist University, Hong Kong SAR, China

25429426@life.hkbu.edu.hk

Abstract. This article aims at providing a methodical review and brief of the application of machine learning. Especially reviewed tokens RWA assets pricing on blockchain, based on systematic methodology. First, this paper clarifies both core definitions and scenarios of RWA tokenization, asset pricing, and machine learning models in this research. Next, this paper review current research findings in blockchain finance, traditional pricing models, and machine learning pricing methods, identifying the strengths and weaknesses of existing work. This paper summarizes the main model labels in existing research, particularly exploring supervised learning, unsupervised learning, and deep learning in area of asset pricing of tokenization. Then this paper identifies weakness in existing articles, provides a conceptual framework, and suggests future research questions and directions. Finally, discuss the practical implications and potential risks of these models for market participants, platform operators, and regulators, providing a methodological framework and theoretical foundation for subsequent empirical and model development research.

Keywords: Blockchain, Tokenization, Asset pricing, Machine learning.

1. Introduction

As a major force behind the digital transformation of financial markets, blockchain technology, which has recently disrupted banking and data infrastructure, has gradually moved into the tokenization of physical assets. The basic concept is to create rights of ownership for assets such as equities, real estate, commodities, and artwork. After that use smart contracts to map those rights to the blockchain so that they may be traded and distributed as digital tokens [1]. However, the precise and constantly changing costs of tokenized assets has emerged as a major concern in both academic research and industry as RWA tokenization quickly. Blockchain transactions are immediate, transparent, and worldwide, yet there is a great deal of confusion surrounding the value of real-world assets. It is challenging to apply standard financial asset pricing models to the decentralized, very unpredictable blockchain market since they often depend on stable market assumptions [2]. Machine learning has been a popular tool to forecast stock, derivative, and cryptocurrency prices in recent years [3]. Using methods like deep learning and ensemble learning, traditional financial markets is successfully pricing stocks and bonds. However, there is a dearth of systematic research and model transfer in the context of RWA tokenization [4].

From an academic perspective, various machine learning methods for asset price tokenization may enhance the knowledge base of blockchain finance in the interdisciplinary field of financial technology and information systems.

In real life, accurate pricing can enhance the trust of investors, make the market more unified and open, and make the integration of real estate on the blockchain easier. It also provides investors, platform operators and regulators with technical tools to manage risks, check compliance and evaluate assets, thus making the market more efficient and open [5]. This study comprehensively evaluates the application of machine learning method in the pricing of blockchain tokenized real world assets (RWA). The main ideas and applications of machine learning modeling, asset pricing and risk weighted asset tokenization are given. This paper also introduces the latest research on blockchain, machine learning and classical pricing methods. The advantages and disadvantages of each method are discussed. The limitations of previous studies are analyzed, and the conceptual framework and research approaches for further exploration are proposed. As the basis for further empirical research and model improvement. The end of these publications focuses on the significance of the model in real life and the concerns it may bring to regulators and market participants.

2. Definition of key concepts

To establish a clear conceptual framework and analysis variables for the study of machine learning driven markup RWA pricing model. This will provide a solid foundation for further research. This will be achieved by systematically outlining the basic concepts, theoretical framework and research methods.

2.1. Core Concepts of RWA Tokenization

The act of converting ownership rights in digital or physical assets into digital tokens on a blockchain is referred to tokenization of real-world assets. "The conversion of tangible and intangible assets into digital tokens that can be traded on decentralized platforms" is the broad definition of tokenization given by current research [1]. In this process, the asset must be recognized, a digital representation is required, a legal framework is needed, and trading capabilities on a digital platform must be permitted.

Utility tokens, security tokens, and hybrid tokens are the three categories into which tokens may be divided due to their practical characteristics. According to utility token pricing research, a token's worth is defined by its capacity to deliver platform services; it only increases in value when the platform services demand instant access [6]. Security tokens are created by security token offerings (STOs), which need to be backed by financial assets and adhere to securities laws. Equity token issuance has reduced agency costs, according to ICO-related studies, but it necessitates a reliable profit-sharing implementation mechanism [3].

2.2. Asset Pricing Theoretical Framework

Capital asset pricing model (CAPM) and arbitrage pricing theory (APT) are the core ideas of classical asset pricing theory. The actual data show that the beta value will change with time. However, CAPM research shows that the standard static CAPM assumes that the beta value remains unchanged. When considering the reality such as weight limit and short selling ban, the effect of market portfolio is not ideal. The perfect linear relationship between expected asset return and beta value will also fail [7]. Risk return tradeoff is the main idea of modern portfolio theory (MPT), which is the most popular theory at present. Research on deep portfolio optimization shows that traditional portfolio management mainly depends on linear models, while MPT does not consider transaction costs and relies on accurate market price forecasts and limited constraints [8].

In the blockchain environment, asset pricing faces new challenges. Alcassa pointed out in his research on quantum machine learning that the traditional parameter model has limitations in terms of computational power and economic statistics assumptions [9]. Data driven nonparametric technology based on machine learning algorithm conceptualizes option as a functional relationship between contract terms (input) and premiums (output). Scholars need to consider a lot when pricing token assets. Issues such as technical risk, regulatory uncertainty and market liquidity. The latest research on option pricing shows that machine learning algorithms are far better than traditional models in processing nonlinear data, because they do not need to make statistical assumptions [10].

2.3. Classification of Machine Learning Methods

Supervised learning is widely used in asset pricing, including regression and classification methods. In his research on mutual fund selection, Huang adopted technologies including support vector regression (SVR), extreme gradient lifting (xgboost) and lightweight gradient lifting machine (lightgbm), which proved that the machine learning model can effectively identify funds with growth potential [11]. In the research of bitcoin forecasting, Chen proposed a sample dimension engineering method for price, using models with different complexity to predict data with different frequencies:

the statistical method is suitable for daily price forecasting with high-dimensional characteristics, while the machine learning model is suitable for high-frequency price forecasting with less characteristics [5].

3. Overview of mainstream technologies:

This part selects the core thesis from the literature of machine learning method and asset pricing model, and analyzes its basic concept, structure process and characteristics. By comparing and analyzing the advantages and disadvantages of different methods, this paper provides a theoretical basis for subsequent research.

Table 1. Overview of selected literature in this article

literature	Core Concepts	Main Methods	Application Areas
Deep Learning in Asset Pricing	Deep Neural Network Asset Pricing	GAN+LSTM+FFN	Stock Return Forecast
Autoencoder asset pricing models	Nonlinear factor models	Conditional Autoencoder	Factor Exposure Modeling
Comparing Asset Pricing Models	Bayesian model comparison	F statistic + model probability	Factor model selection
Machine Learning and Portfolio Optimization	Performance-based regularization	PBR + cross validation	Portfolio Construction

3.1. Deep Learning in Asset Pricing

The author proposes an asset pricing model based on deep neural network. The model takes the no arbitrage condition as the constraint function, and combines three neural network structures to estimate the stochastic discount factor (SDF). The core innovation of the model is to incorporate economic constraints into machine learning algorithms, which significantly improves the ability to identify risk premium signals [5].

The model adopts a theoretical framework that integrates three layers of neural networks: the feedforward neural network (FFN) is responsible for describing the overall function of the stochastic discount factor (SDF), the long-term and short-term memory network (LSTM) extracts the dynamic macroeconomic state variables, and generates the confrontation network (GAN) to construct the test asset portfolio with the highest information content; These three factors form a collaborative estimation framework through no arbitrage conditions, so as to effectively distinguish the role of systematic signals and random noise in the risk premium.

In general, the practice of asset pricing using deep learning discussed in this paper shows that this method has many technical advantages and practical application value. At the same time, the model has strong explanatory power, which can explain the fluctuation of 8% in individual stock returns and 23% of expected returns. This fully demonstrates how powerful deep learning is in capturing complex nonlinear correlations and interactions.

However, this progress is accompanied by corresponding challenges: the complexity of the model is extremely high, which requires a lot of computing resources; The parameter adjustment process is cumbersome and requires a large number of validation data sets; At the same time, there are strict requirements for the quality and integrity of data.

3.2. Autoencoder Asset Pricing Models

Gu et al. Developed an innovative implicit factor conditional asset pricing model [4]. The model uses automatic encoder neural network to explore the nonlinear relationship between asset characteristics and factor risk exposure. Compared with the typical linear model, this method can more flexibly and truly describe the changes of asset prices over time.

The model architecture includes two main modules: conditional beta network (which helps to map asset characteristics nonlinearly to factor risk exposure) and factor network (which compresses yield data into a low dimensional factor set). These modules are integrated into the collaborative modeling framework to improve the structural identification of systemic and unique risks.

Mathematically, this model can be expressed as $r\_(i,t) = \beta\_(i,t-1)^\wedge{}' f\_t + u\_(i,t),$ where β is modeled via a neural network, enabling a deep characterization of complex market mechanisms.

This method correctly evaluates a series of portfolio anomalies, and there are few statistically insignificant pricing errors. It solves the problem of abnormal interpretation of traditional technology. The model also has some disadvantages in practical application, such as requiring a large number of historical data training, difficult to master, difficult to calculate and so on. In order to avoid over fitting, a large number of regularization algorithms are needed.

3.3. Comparing Asset Pricing Models

Barillas shanken proposed a Bayesian asset pricing test method using F statistics. This test facilitates the calculation of model probabilities and enables the comparison of all alternative price models based on a specific set of factors. They found that the five factor and six factor models including momentum factors were better than the newly introduced hxz and FF models [12].

This methodological framework includes three core steps: first, the joint alpha constraint is introduced into the Bayesian setting based on economic motivation to construct a priori belief. Then, the posterior probability calculation function is performed through the F statistics calculation module. Finally, we use the model comparison method to systematically evaluate the probabilities of all possible pricing models, so as to form a complete Bayesian model selection process.

This method provides you with a systematic method to compare models, so the research can compare nested and non nested models at the same time. It also uses prior information based on economic theory to help people better understand the results. However, its disadvantages include that it is more sensitive to the choice of prior distribution, the computational complexity will increase exponentially with the increase of the number of components, and it is vulnerable to the influence of model setting errors. These characteristics may make it unsuitable for some complex application situations.

3.4. Machine Learning and Portfolio Optimization

Ban et al. introduced a performance-based regularization (PBR) approach to improve the performance of the sample average approximation (SAA) method by constraining the sample variance of estimated portfolio risk and returns [13]. This approach aims to guide the solution toward a direction that is less associated with performance estimation errors.

The core steps of this method include three key links: constructing an optimization framework by introducing a quartic polynomial constraint into the PBR model for the mean-variance problem; transforming the PBR model for the mean-CVaR problem into a combinatorial optimization problem for solution; and finally, using performance-based K-fold cross-validation to calibrate the right-hand side constraints in the model, thereby systematically improving the robustness and applicability of the model configuration.

The authors' performance-based regularization (PBR) method shows many benefits for machine learning-based portfolio optimization. The main benefit of this strategy is that it can directly use the performance of portfolio decisions that are not in the sample to improve the optimization process. It can build stronger and more predictive portfolio strategies by carefully managing the estimation errors of risk and return.

There are certain problems with PBR. It can be hard to use with high-dimensional data, and it needs very accurate setup of constraint bounds, which makes it rather hard to compute.

4. Application

4.1. Basic Research on Blockchain and RWA Tokenization

Blockchain technology offers a decentralized, transparent, and secure framework for the tokenization of tangible goods. This system uses Ethereum staking and Layer-2 scaling methods to fix problems with scalability, security, and decentralized governance in RWA tokenization. This system uniquely combines Ethereum staking, Layer-2 scaling (Optimistic Rollups), decentralized autonomous organizations (DAOs), and strong oracle integration (Chainlink) to solve these problems at the same time [14].

The distributed ledger feature of blockchain technology architecture is what makes it the best technical base for the tokenization of RWAs. According to Tian et al., blockchain is a decentralized and trusted ledger system that stores data in a way that can't be changed. This quality makes it great for tokenizing infrastructure assets, which are very important for the economy and frequently need regular income flows over a long period of time. They also need to be well-structured throughout the project lifecycle [15].

According to Tian et al., blockchain is a decentralized and trusted ledger system, which stores data in an unchangeable way. This quality makes the tokenization of infrastructure assets very important, because infrastructure assets are very important to the economy and often require a long-term fixed income stream. They also need to be well structured throughout the project lifecycle.

A multi case study analyzes the asset tokenization business cases of four different industries, and the results show that blockchain technology has significantly improved the transaction efficiency of real estate, gold, game assets, carbon credit and other industries [16]. Research shows that tokenization has greatly modified the traditional transaction process by reducing or reducing the number of intermediaries. Token in real estate allows people to buy and sell real estate directly on the platform, so brokers and brokers do not have to intervene. The licensed and unlicensed architectures of blockchain provide adaptive solutions for various RWAs. Systems that require licenses are usually faster and can handle more users, but they are also more centralized than systems that do not require licenses. Licensed blockchain is a better way to track and manage infrastructure assets that must comply with rules and regulations.

4.2. Application of Traditional Asset Pricing Models in Blockchain Environment

The adoption of standard asset pricing methods in the blockchain sector has brought new problems and opportunities. Blockchain technology has fundamentally changed the basic characteristics of assets, such as liquidity, transparency and transaction process, and has a far-reaching impact on the traditional pricing model.

The traditional CAPM model faces new challenges in the blockchain environment. Wang and Chen pointed out that the traditional static CAPM assumes that β is constant, but empirical evidence shows that the actual β is unstable due to various factors [17]. These factors include changes in corporate strategy, capital structure and earnings, as well as microeconomic factors such as dividend policy and financial leverage. Therefore, the change of enterprise level characteristics and external macroeconomic environment leads to the change of beta coefficient over time.

Prat et al. point out that the return of a token can be decomposed into two components: capital return and convenience return [18]. Convenience return is unique to utility tokens. The marginal token provides a service quantity $p_{(t+1)}$ and its utility is equal to $z_{(t+1)} u^{\prime} (m_t p_{(t+1)})$. From a pricing theory perspective, this is the key difference between stocks and tokens. Unlike stocks, which offer dividends without the need for exchange, the fundamental value of a token is determined by the discounted surplus from the next transaction.

4.3. Overview of Machine Learning Applications in Asset Pricing

In the field of asset pricing, the application of machine learning is developing rapidly. Now, it can provide a new perspective and tool for the traditional pricing method. These tools can process multidimensional data, discover nonlinear relationships, and adjust according to the complexity of financial markets.

4.3.1 Supervised Learning Methods

Supervised learning methods have demonstrated strong predictive power in asset pricing. Paiva et al. proposed a unique model for intraday stock market trading decisions that combines a classifier based on support vector machines (SVM) with a mean-variance (MV) portfolio selection approach [19]. Experimental results show that Paiva et al.'s SVM + MV model achieved 3809.90% profitability during the analysis period, significantly outperforming other benchmark models [19].

Chen et al. Used the sample dimension engineering in the research of predicting the value of bitcoin, and classified the data into daily and high-frequency (5-minute interval) prices [4]. They used statistical methods including logistic regression and linear discriminant analysis to predict daily prices. In the case of high-dimensional features, the accuracy rate of this method is 66%, which is better than the complex machine learning system. On the other hand, machine learning models include random forest XGBoost, Twice discriminant analysis, support vector machine and long and short memory network, the accuracy rate of predicting the price every five minutes is only 67.2%.

4.3.2 Deep Learning Methods

Baek and Kim proposed the ModAugNet framework, a novel stock market index prediction framework designed to predict stock market index values while mitigating overfitting risk [20]. The framework consists of two key modules: an overfitting prevention LSTM module and a prediction LSTM module. ModAugNet-c achieved test mean squared error (MSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) on the S&P 500, respectively, reducing them to 54.1%, 35.5%, and 32.7% of the corresponding errors of SingleNet [20].

DPO not only uses the advantages of machine learning to make investment decisions, but also retains the main ideas of contemporary portfolio theory and optimizes the portfolio. The framework uses both convolutional neural network (CNN) and recurrent neural network (RNN) modules. CNN module learns short-term time characteristics to analyze asset correlation, while RNN module learns long-term time characteristics to analyze sequence correlation [8].

5. Conclusion

This review systematically discusses the application of machine learning in the pricing of heterogeneous real-world assets (rwas). It aims to build a theoretical framework that combines technology and finance, and provide practical guidance. This research reveals significant progress in theoretical innovation and practical application in this field. Finally, it also points out the various challenges it faces and the future development direction.

This paper combines the characteristics of blockchain technology with traditional pricing theory to build a framework of technology and theory integration, which provides a new pricing basis for token assets. The deep learning model proposed by Gu et al. And the implementation of automatic encoder reflect the major research achievements in this aspect of fusion. In addition, for various risks associated with tagged assets, it is also necessary to establish a complete pricing system. The study of financial contagion within the network economy provides a novel perspective for understanding systemic risk. Machine learning technologies, including supervised learning and stable learning frameworks, have shown potential in capturing high-dimensional nonlinear relationships and mitigating distributed changes. At the same time, it provides some methods to solve the inherent limitations of traditional pricing models in managing complex dynamic structures.

This assessment has important reference value for investors, platform operators and regulators. Investors can improve their strategies by using machine learning, and the platform can optimize its architecture by studying past successful cases. On the other hand, regulators need to find a way to encourage new ideas and reduce risks. They should consider flexible solutions such as "regulatory sandbox", and strive to strengthen international regulatory cooperation and unify technical standards.

Future research should focus on the following areas: Research on the prospective application of complex technologies (including federated learning and quantum machine learning); Theoretically, behavioral finance is combined with the complex system framework to develop a pricing theory more in line with the characteristics of token assets. It is suggested to expand the scope of research from a single asset to cover cross market and cross chain portfolios, focusing on the solution of practical problems such as data scarcity, lack of cross chain interoperability standards and blurred regulatory prospects.

References

[1] Tian, Y., Adriaens, P., Minchin, R. E., Chang, C., Lu, Z., & Qi, C. (2020). Asset Tokenization: A blockchain Solution to Financing Infrastructure in Emerging Markets and Developing Economies. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3837703.

[2] Gan, J., Tsoukalas, G., & Netessine, S. (2020). Initial coin offerings, speculation, and asset tokenization. Management Science, 67 (2), 914–931. https://doi.org/10.1287/mnsc.2020.3796.

[3] Gu, S., Kelly, B., & Xiu, D. (2020). Autoencoder asset pricing models. Journal of Econometrics, 222 (1), 429–450. https://doi.org/10.1016/j.jeconom.2020.07.009.

[4] Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. Journal of Computational and Applied Mathematics, 365, 112395. https://doi.org/10.1016/j.cam.2019.112395.

[5] Zhang, Y., Gong, B., & Zhou, P. (2024). Centralized use of decentralized technology: Tokenization of currencies and assets. Structural Change and Economic Dynamics, 71, 15–25. https://doi.org/10.1016/j.strueco.2024.06.006.

[6] Prat, J., Danos, V., & Marcassa, S. (2024). Fundamental pricing of utility tokens. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4771372.

[7] Simaan, Y. (2024). Markowitz and the CAPM. Annals of Operations Research. https://doi.org/10.1007/s10479-024-06404-8.

[8] Yan, R., Jin, J., & Han, K. (2024). Reinforcement learning for deep portfolio optimization. Electronic Research Archive, 32 (9), 5176–5200. https://doi.org/10.3934/era.2024239.

[9] Alcazar, J., Leyton-Ortega, V., & Perdomo-Ortiz, A. (2020). Classical versus quantum models in machine learning: insights from a finance application. Machine Learning Science and Technology, 1 (3), 035003. https://doi.org/10.1088/2632-2153/ab9009.

[10] Ivașcu, C. (2020). Option pricing using Machine Learning. Expert Systems With Applications, 163, 113799. https://doi.org/10.1016/j.eswa.2020.113799.

[11] Huang, C. S. J., & Huang, Y. (2024). Mutual fund selection strategies based on machine learning. Computational Economics. https://doi.org/10.1007/s10614-024-10766-3.

[12] Barillas, F., & Shanken, J. (2018). Comparing asset pricing models. The Journal of Finance, 73 (2), 715–754. https://doi.org/10.1111/jofi.12607.

[13] Ban, G., Karoui, N. E., & Lim, A. E. B. (2016). Machine learning and portfolio optimization. Management Science, 64 (3), 1136–1154. https://doi.org/10.1287/mnsc.2016.2644.

[14] Zhao, X., Ding, J., Su, Y., Wang, H., Guo, F., Zhang, Q., & Mu, M. (2025). Scalable & secure real-world asset tokenization using ethereum staking & layer-2 solutions. Peer-to-Peer Networking and Applications, 18 (5). https://doi.org/10.1007/s12083-025-02032-6.

[15] Tian, Y., Lu, Z., Adriaens, P., Minchin, R. E., Caithness, A., & Woo, J. (2020). Finance infrastructure through blockchain-based tokenization. Frontiers of Engineering Management, 7 (4), 485–499. https://doi.org/10.1007/s42524-020-0140-2.

[16] Tanveer, U., Ishaq, S., & Hoang, T. G. (2025). Tokenized assets in a decentralized economy: balancing efficiency, value, and risks. International Journal of Production Economics, 109554. https://doi.org/10.1016/j.ijpe.2025.109554.

[17] Wong, M. C. S., Chan, E. K. H., & Yousaf, I. (2024). CBDCs, regulated stablecoins and tokenized traditional assets under the Basel Committee rules on cryptoassets. Journal of Financial Regulation and Compliance. https://doi.org/10.1108/jfrc-03-2024-0050.

[18] Prat, J., Danos, V., & Marcassa, S. (2024). Fundamental pricing of utility tokens. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4771372.

[19] Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2018). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. Expert Systems With Applications, 115, 635–655. https://doi.org/10.1016/j.eswa.2018.08.003.

[20] Baek, Y., & Kim, H. Y. (2018). ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. Expert Systems With Applications, 113, 457–480. https://doi.org/10.1016/j.eswa.2018.07.019.