Research Article

Prediction of Barrier Option Price Based on Antithetic Monte Carlo and Machine Learning Methods

Ying Li[1] , Keyue Yan[2*]

[1]College of Global Talents, Beijing Institute of Technology-Zhuhai Campus, Zhuhai, China
[2]Department of Computer and Information Science, University of Macau, Macau, China
 E-mail: conneryan@foxmail.com

Abstract: Option pricing has become a popular topic in the fields of finance and mathematics with the rapid development of stock and option markets. Now, more and more academics, financial companies, and investors are attracted to study and conduct research on it. The theory of option pricing can also be used to price financial instruments with structures similar to options and contribute to risk control and management. The Black-Scholes model is the basic and well-known method applied for different options pricing with modifications and adjustments, and the results can be solved by some traditional numerical methods such as the binomial model, finite difference method, Monte Carlo method, and so on. Machine learning has risen recently and has begun to replace some complex work in traditional methods with the evolution of computers and computing power. How to use machine learning methods to predict the option price is a problem worthy of being solved. In this research, the antithetic Monte Carlo method is used to generate the prices of the up-and-out barrier options without rebate based on the Black-Scholes model. The generated dataset is divided into a training set and a test set for support vector regression, random forest, adaptive boosting, and artificial neural networks. We compare the fitting and performance of all machine learning methods and find that random forest and artificial neural network methods fit better than others with fewer errors in predictions.

Keywords: option pricing, barrier option price, antithetic monte carlo, machine learning

1. Introduction

With the improvement of the financial market system, different kinds of options play an important role with the advantages of low transaction costs and diverse strategies, which are more suitable for individual investors to buy and sell [1]. Meanwhile, new options are constantly being constructed to trade in markets as underlying stock indices, metals, energy, and other technologies emerge [2]. Barrier options are quite popular with cheaper prices among all options.

Zheng and Wei [3] constructed a hedging strategy of barrier option prices using the Black-Scholes model and Monte Carlo method to analyze the results of the Greek letters and observed that this strategy can reduce the cost of hedges and perform well. Nouri and Abbasi [4] used random variables from a uniform distribution for pricing down-and-out call options and double barrier options to modify the Monte Carlo algorithm. They focused on calculating the

first hitting time and concluded that the modified one can reduce the errors of predictions. Umeorah and Mashele [5] calculated the barrier option prices by Crank-Nicolson finite difference method. They observed the convergence of results by increasing the steps of discretization and the effect of changing the rebates on option values, and the values of the Crank-Nicolson finite difference method and antithetic Monte Carlo are more accurate than the value of Monte Carlo method. Gao et al. [6] deduced the formulas for pricing American barrier options under uncertain environments by uncertain differential equations and may consider other options in this way in the future. Cao et al. [7] selected several stocks as the basket options with barriers and used the Monte Carlo method to simulate the price and test the performance of the options. Shrivastava and Verma [8] applied bagging methods that contained support vector machine, nearest neighbor classification and decision tree to forecast the option prices for Indian stock market to increase the accuracy of predictions. They also combined cascaded neural network model with optimization algorithms to predict the price of options and obtained better results compared to deep neural network [9]. Gan et al. [10] used traditional numerical methods for pricing geometric and arithmetic Asian options and trained Back Propagation neural network of which the run time is reduced significantly. Chowdhury et al. [11] changed the strike price, time and volatility parameters to modify the Monte Carlo method and forecasted the values of call and put options with a decision tree, neural network and ensemble learning method, and then they discussed the relationship between stock price and volatility based on experimental results.

Most papers on pricing the options above are concentrated on deriving the formulas and improving the numerical methods to reduce the errors, and the others focus on the applications of machine learning methods with historical data. The goal of this research is present the methods of barrier option pricing and provided suggestions to budding researchers. Based on the improvement and combination of models above, this research is comparing the performance of different machine learning methods with the data simulated by the improved Monte Carlo method and determinines which method is more capable of handling the task for up-and-out barrier option pricing.

The remainder of this research is summarized as follows. Section 2 introduces the concepts of barrier option, the Black-Scholes model and the antithetic Monte Carlo method of option pricing in detail. Section 3 describes the basic contents of four machine learning methods: support vector regression, random forest, adaptive boosting and artificial neural network. Section 4 prepares the data for the experiment and summarizes the performance of machine learning methods by analyzing the results of the experiment. Finally, we conclude this research in Section 5.

2. Option pricing models

Option, a type of financial derivative, is a contract that gives the holders the right to buy or sell an asset at a certain date or any time before the date [12]. The certain date of the option is called the maturity date. There are two types of the options which are call options and put options. The holder of a call option has the right to buy an asset at a certain price defined as the strike price, while that of a put option has the right to sell. According to the time of exercise, there are European options and American options. A European option can only be exercised at the maturity date, while an American option can be exercised at any time before or at the maturity date.

Barrier options are path-dependent exotic options with a more complex structure than basic European call or put options. A barrier option can only be exercised or expired before or at the maturity date if the price of the underlying asset reaches a predetermined level, which is called the barrier level [12]. Generally, barrier options are classified as knock-in and knock-out options.

A knock-in barrier option will be existence till maturity if the underlying asset price reaches or exceeds the barrier level, otherwise, it will expire. There are two types of knock-in options: up-and-in and down-and-in options. For the existence of an up-and-in option, the price of the underlying asset should be above the barrier level that is set above the initial price. Similarly, a down-and-in option only comes into existence if the asset price is below the barrier level, which is set below the initial price.

Contrary to knock-in options, a knock-out option is expired once the asset price crosses the barrier level before the maturity date. If the underlying asset price of a knock-out option does not cross the barrier level during the whole option's period, this option can be considered a normal European call or put option. If a barrier option is expired because of knocking out, a rebate may sometimes be provided to the holder of the contract. There are also two types which are

up-and-out and down-and-out options. An up-and-out barrier option will expire worthless when the underlying asset price rises above the barrier level before or at the maturity date. For a down-and-out option, the option will be knocked out if the asset price decreases below the barrier level.

2.1 *Black-Scholes model*

The Black-Scholes model is used for computing the price of a European call or put options in continuous time, which was proposed by Myron Scholes and Fisher Black and modified by Robert C. Merton [12]. The Black-Scholes model assumes that there is one risk-free asset and one risky asset in the financial market. It is assumed that r is the risk-free interest rate during the period and is obtained from national debt, q and σ are the dividend yield and the volatility of a stock, respectively. Let $W(t)$ be a Wiener process under the risk-neutral probability measure. The process of the stock price $S(t)$ follows the stochastic differential equation below.

$$dS(t) = (r - q)S(t)dt + \sigma S(t)dW(t) \tag{1}$$

By Ito's lemma, the value of option V should be satisfied the following partial differential equation.

$$\frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S^2 + (r - q)\frac{\partial V}{\partial S}S - rV = 0 \tag{2}$$

Considering an up-and-out option with strike price K, time to maturity T, barrier level B and no rebate, the barrier option value is given by

$$V(0) = e^{-rT}V(T) \tag{3}$$

with the payoff as below.

$$V(T) = \begin{cases} \max(S(T) - K, \ 0), & \text{if } \max S(t) < B \text{ with } 0 \le t \le T \\ 0, & otherwise \end{cases} \tag{4}$$

2.2 *Antithetic Monte Carlo*

The Monte Carlo method is a statistical method for solving computational problems by generating random samples. It can simulate several paths of a target variable and get the results for each path. The solution is the mean of the results or the frequency of certain values [12]. Compared with other numerical methods, there are not too many deductions of formulas and algorithms in the Monte Carlo method [13]. It is easier to understand and get the simulation of processes by computer programming. Monte Carlo methods with antithetic sampling considered antithetic Monte Carlo can reduce the variance by generating new samples which are negatively correlated with the original samples [14].

For calculating an option price, the first step is to set the number of paths and steps for each path. According to the formula of stock price, the process of price can be generated randomly step by step for each path. The option price is the current value of the expected value of the payoffs. Solving Equation 1 by Ito's lemma, the stock price processes in antithetic Monte Carlo follow the function below:

$$S_1(t + \Delta t) = S_1(t)e^{\left(r - q - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t}\varepsilon(t)} \tag{5}$$

$$S_2(t + \Delta t) = S_2(t)e^{\left(r - q - \frac{1}{2}\sigma^2\right)\Delta t - \sigma\sqrt{\Delta t}\varepsilon(t)} \tag{6}$$

where $\Delta t = \dfrac{T}{M}$ is the time interval, M is the number of steps and ε is a random number of the standard normal distribution.

Using antithetic Monte Carlo to simulate the stock prices for one path will obtain two values $S_1(T)$ and $S_2(T)$ based on Equation 5 and Equation 6 step by step. Each $S(T)$ in a path is corresponding to a value of $V(T)$ satisfying Equation 4. The mean of these two values of $V(T)$ is the payoff in this path. The option payoff is the expected value of all payoffs with N paths, and the formula of the option value is shown below:

$$V(0) = \frac{e^{-rT}}{N} \sum_{i=1}^{N} \frac{V_{1i}(T) + V_{2i}(T)}{2} \tag{7}$$
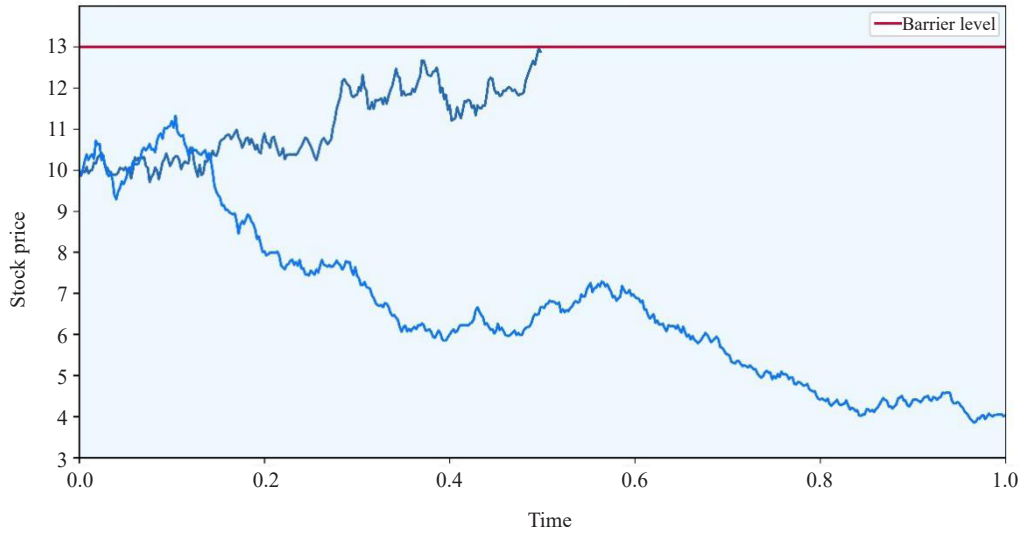
Figure 1. Antithetic Monte Carlo simulation

Set the parameters $S(0) = 10$; $K = 11$; $B = 13$; $r = 0.03$; $q = 0.04$; $T = 1$; $M = 500$; $\sigma = 0.3$. The stock price processes of barrier options can be simulated by the antithetic Monte Carlo method as Figure 1 shows. There are two possibilities for the stock price shown in the Figure 1. One is knocked out for hitting the barrier level before the maturity date, the other one is regarded as a normal European option. The computer programming runs in 0.0180 seconds for these two processes.

3. Machine learning methods

With the development of computer science, machine learning is a study of computer algorithms widely used for processing data and simulating to obtain new information or knowledge from raw and messy data. In this research, there are four machine learning methods such as support vector regression, random forest, adaptive boosting and artificial neural network for predicting the prices of barrier options.

3.1 *Support vector regression*

Developed by Vladimir N. Vapnik and Alexander Y. Lemer, a Support Vector Machine (SVM) is a generalized supervised learning classifier used for regression and classification [15]. The objective of an SVM algorithm is to find out a hyperplane by maximizing the margin between the hyperplane and the nearest sample point to separate two classes

so that most sample points are outside the decision boundaries. Different from SVM, a Support Vector Regression (SVR) is minimizing the margin between the hyperplane and the farthest sample point to make as many points as possible inside the decision boundaries, and it is easy to understand and operate.

3.2 *Random forest*

Random Forest (RF), a decision-making method, contains multiple decision trees unconnected between each one, and the result is the mode or the mean of the individual trees' classes. This method can speed up the training speed, reduce the risk of overfitting problems and get a more accurate result than an individual decision tree model [16]. The algorithm of RF is selecting a certain amount of samples from the training set based on sampling with the replacement for training each decision tree. For each decision tree, the root node represents the object, each internal node represents an attribute, and each leaf node represents a class [15]. Determining which attribute to satisfy and choosing the branch to enter until reaching the leaf node can get the result in a tree.

3.3 *Adaptive boosting*

Adaptive Boosting (AdaBoost) is one of the Boosting algorithms for reducing bias in machine learning. It obtains a weaker learner by training a certain amount of samples at first and then combines the misclassified samples with new data to get a new training set for a second weaker learner [17]. The third training set is constituted by the misclassified samples of the first two sets and new data, and so on until all samples are classified correctly. The main advantage of AdaBoost is considering the weight of each weaker learner. A strong learner is built when the number of iterations or error rate reaches the specified value.

3.4 *Artificial neural network*

Artificial Neural Network (ANN or NN) is a key deep learning algorithm of artificial intelligence for imitating the features of biological neural networks and processing the information. An NN model simulates the nervous system of brains by modeling artificial neurons so that it can learn based on the data set and store the results in the edges which keep the signals to connected neurons [18]. Each neuron with weight and threshold adjusted with learning will send a signal if the output is over the threshold, otherwise, there is no message transmitted to the next layer of the networks. To build an NN model, we can not only use the sklearn package but also some famous commercial packages such as tensorflow, keras and pytorch [19]. In this research, keras has been used to build a NN regressor with different numbers of layers and neurons.

4. Experiments
4.1 *Data preparation*

The data of the up-and-out barrier options without rebate of which the prices are generated by antithetic Monte Carlo is considered as the dataset in the research. Using the antithetic Monte Carlo method simulates 10,000 up-and-out option prices with 5,000 paths for each option and 500 time steps for each path. The parameters with fixed values for computing the prices are the initial value of the stock price $S(0)$ and the risk free interest rate r which are set as 100 and 0.03, respectively. The other parameters used to compute the option prices are the dividend yield q and the volatility σ of the stock, time to maturity T, strike price K and barrier level B which are generated randomly from normal distribution based on certain conditions and reference possible values [20]. The ratios of strike price and barrier level to initial value replace strike price and barrier level in the data set in order to concern the relationship between these three parameters. There are 6 parameters totally in the data set, and the data range of them is as follows.

Table 1. Data range of parameters

	mean	std	min	max
$V(0)$	0.9539	1.1529	0.0000	6.9777
R_K ($K/S(0)$)	1.1203	0.0906	1.0000	1.6000
R_B ($B/S(0)$)	1.3208	0.1320	1.0200	2.1000
q	0.0501	0.0099	0.0130	0.0887
σ	0.2000	0.0202	0.1330	0.2801
T	0.5008	0.2863	0.0001	0.9999

In Table 1 above, the minimum value of up-and-out barrier option prices is 0, which means that the option is knocked out or the price of option does not exceed the strike price during the maturity, so it is meaningful and cannot be deleted from the dataset. The programming takes around 28 seconds for calculating one barrier option price and more than 48 hours for all option prices in dataset by the antithetic Monte Carlo method.

The data set is divided into the training set and test set in the ratio of 7:3 for training and testing machine learning methods discussed in the research. The option value is considered the dependent variable, and the other parameters in the data set are independent variables. The programming uses the packages of sklearn and keras to complete the parameter tuning of these machine learning methods and forecast the option values.

4.2 *Model performance*

Evaluating the performance of machine learning methods needs analyzing errors between the predicted results and the actual values in the training set and the test set. The smaller the error of predictions, the better the performance of model. The following indicators are commonly used for model evaluations.

R-squared also called the coefficient of determination is a basic indicator used for evaluating the fitting of prediction models, and its value is usually between 0 and 1. The closer the result is to 1, the higher the fitting degrees. Conversely, the model fits worse if the value is 0. The formula of *R-squared* is as below:

$$R - squared = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2} \tag{8}$$

where y_i, $\hat{y}_i$ and $\overline{y}$ are the actual value, predicted value and the mean of variable in models, respectively.

Mean Absolute Error (*MAE*) measures the average of residuals between the predictions and actual data. Due to taking the absolute value of residuals, the residuals cannot cancel each other out. The smaller the result, the better the fitting. The formula of *MAE* is as follows

$$MAE = \frac{\sum | y_i - \hat{y}_i |}{n} \tag{9}$$

where n is the number of the training set or test set for models.

Mean Squared Error (*MSE*) is an indicator for measuring bias from the predictions. Taking the square root of *MSE* obtains Root Mean Square Error (*RMSE*). The smaller the values of both, the better the model fits. The formulas are shown below.

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n} \qquad (10)$$

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}} \qquad (11)$$

By improving the process of parameter tuning of the NN method, the numbers of layers and neurons of the best model are 2 and 30, respectively. The results of evaluating the machine learning methods used in this research are shown as the following.

Table 2. Evaluation results of machine learning methods in training set

	Running time	R-squared	MAE	MSE	RMSE
SVR	4.7986	0.6632	0.4709	0.4503	0.6710
RF	0.9619	0.9983	0.0271	0.0022	0.0473
AdaBoost	0.3630	0.8111	0.4445	0.2525	0.5025
NN	5.8461	0.9966	0.0465	0.0046	0.0677

Table 3. Evaluation results of machine learning methods in test set

	Running time	R-squared	MAE	MSE	RMSE
SVR	0.4169	0.6740	0.4568	0.4271	0.6535
RF	0.0312	0.9881	0.0728	0.0156	0.1247
AdaBoost	0.0156	0.7994	0.4490	0.2628	0.5126
NN	1.5981	0.9962	0.0486	0.0050	0.0707

As Table 2 and Table 3 show, the values of *R-squared* for RF and NN methods are over 0.98 and significantly higher than the values for other methods both in the training set and test set, and all the other indicators' values of these two methods are below 0.2. It indicates that the methods of RF and NN fit better with the lower error of predictions. For the AdaBoost method in these two sets, although the values of *R-squared* are around 0.8 which means the performance of this method performs well enough, the values of *MAE* and *RMSE* are about 0.5 and higher compared with the results of the NN method. For the SVR method in training and test sets, the values of *R-squared* are around 0.67 which are the lowest of all methods, and the values of *MAE*, *MSE* and *RMSE* are higher than others, especially values of *MSE*. The fitting result for SVR is not satisfying. Compared to the running time in the antithetic Monte Carlo mentioned above, the computational cost of machine learning methods is reduced significantly.

The fitting efficiency of four machine learning methods for predicting the up-and-out option prices while testing is shown in Figure 2. The more the points are on the diagonal, the less the error of prediction is, which indicates the model has good performance and perfect efficiency. In Figure 2, the predicted values for SVR are generally gathered around 1 to 2, although the option prices are from 0 to 7. Most options are overvalued when the price is less than 2 and undervalued when the price exceeds, which states the SVR method with poor effects is not recommended to forecast prices based on the dataset in this research. For the AdaBoost method, the overall trend of points is clear and rises as the price increases, but the points are scattered on both sides of the diagonal. The performance of AdaBoost is not good

enough. For RF and NN methods, points are concentrated on the diagonal, especially two extreme values, which means these two methods fit well and the efficiency of the NN method is better.

(a) SVR Method

(b) RF Method

(c) AdaBoost Method

(d) NN Method
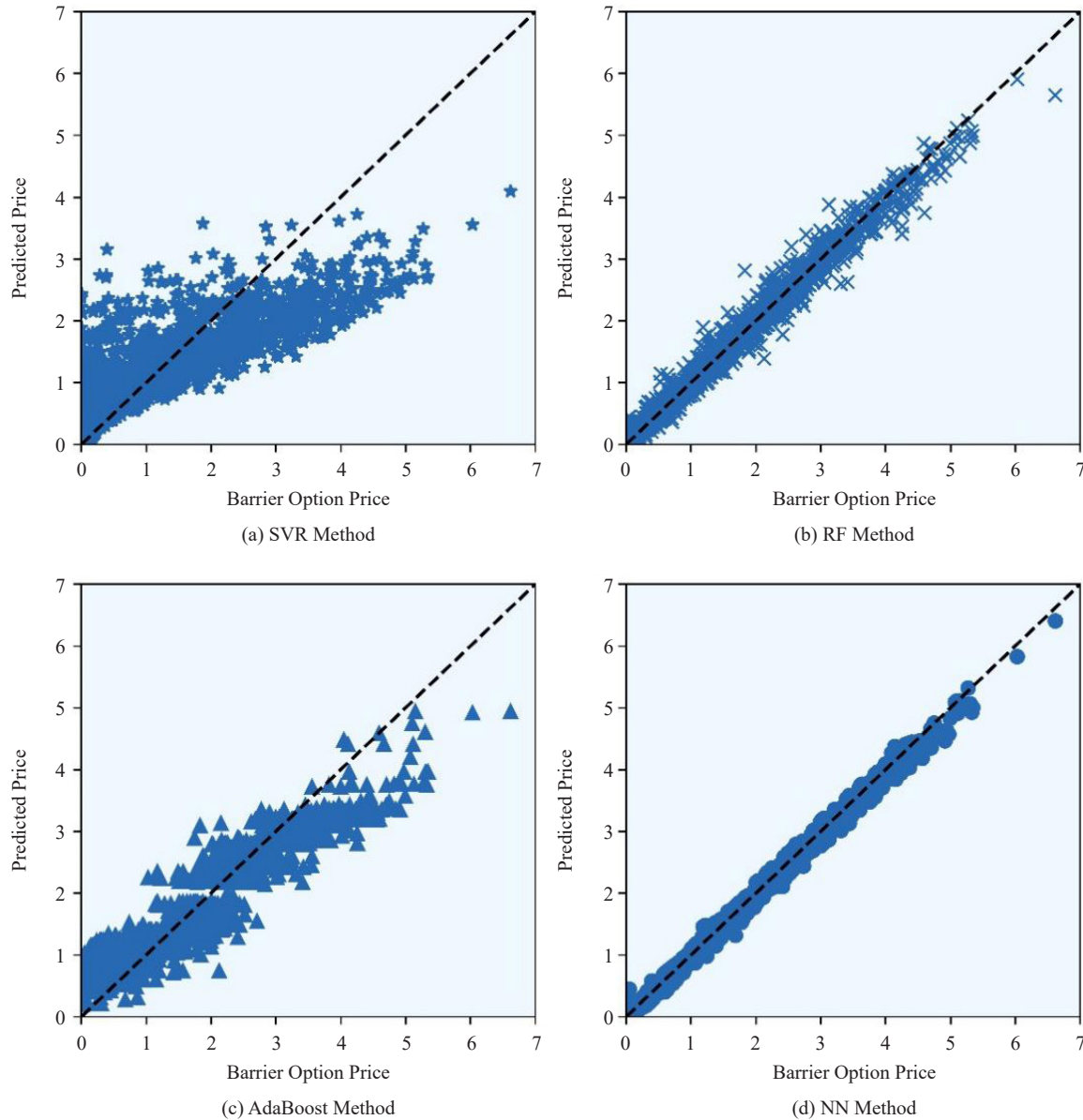
Figure 2. Prediction of prices in testing machine learning methods

5. Conclusion

The research proposes machine learning methods to forecast barrier option prices with the example of up-and-out options without rebates. Based on the Black-Scholes model, the data used for training and testing machine learning methods is generated by antithetic Monte Carlo method which reduces the variance during the simulations and shortens the convergence time. The experiential results of the model evaluation show that the predictions of RF and NN fit the generated data better than the performances of SVR and AdaBoost methods, and they also give references for investors to trade. This research provides a framework for pricing various options with different properties.

In further research, using machine learning methods to price other kinds of barrier options with stochastic volatility and rebates and how to modify the method with poor performance should be studied.

Conflict of interest

The authors declare there is no personal or organizational conflict of interest with this work.

References

[1] Yang S, Wang C. Trends of global exchange-traded stock index futures and options market and its implication. *Studies of International Finance*. 2006; 11: 36-41.

[2] Zhang X. The new trend of international futures and options market and its enlightenment to China. *Globalization*. 2018; 7: 50-61. Available from: doi: 10.16845/j.cnki.ccieeqqh.2018.07.006.

[3] Zheng X, Wei, Y. Pricing and hedging barrier options based on Merton model and Monte Carlo simulation. *Journal of University of Science and Technology of China*. 2018; 48(11): 906-922.

[4] Nouri K, Abbasi B. Implementation of the modified Monte Carlo simulation for evaluate the barrier option prices. *Journal of Taibah University for Science*. 2017; 11(2): 233-240. Available from: doi: 10.1016/j.jtusci.2015.02.010.

[5] Umeorah N, Mashele P. A Crank-Nicolson finite difference approach on the numerical estimation of rebate barrier option prices. *Cogent Economics & Finance*. 2019; 7(1): 1598835. Available from: doi: 10.1080/23322039.2019.1598835.

[6] Gao R, Liu K, Li Z, Lang L. American barrier option pricing formulas for currency model in uncertain environment. *Journal of Systems Science & Complexity*. 2021; 35(1): 283-312. Available from: doi: 10.1007/s11424-021-0039-y.

[7] Cao N, Mai H, Zhang S. A monte carlo simulation scheme for basket options with barriers. *Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)*. 2022. Available from: doi: 10.2991/aebmr.k.220307.485.

[8] Shrivastava P, Verma CK. Prediction of option price using ensemble of machine learning algorithms for Indian stock market. *International Journal of Recent Technology and Engineering (IJRTE)*. 2019; 8(2): 3231-3241. Available from: doi: 10.35940/ijrte.b2683.078219.

[9] Shrivastava P, Verma CK. A hybrid forecasting model for option price prediction using machine learning technique. *International Journal of Innovative Technology and Exploring Engineering*. 2019; 9(1): 3444-3450. Available from: doi: 10.35940/ijitee.a4622.119119.

[10] Gan L, Wang H, Yang Z. Machine learning solutions to challenges in finance: An application to the pricing of financial products. *Technological Forecasting and Social Change*. 2020; 153: 119928. Available from: doi: 10.1016/j.techfore.2020.119928.

[11] Chowdhury R, Mahdy MRC, Alam TN, Al Quaderi GD, Arifur Rahman M. Predicting the stock price of frontier markets using machine learning and modified Black-Scholes Option pricing model. *Physica A: Statistical Mechanics and its Applications*. 2020; 555: 124444. Available from: doi: 10.1016/j.physa.2020.124444.

[12] Hull J. *Options, Futures, and Other Derivatives*. 11th ed. New York: Pearson; 2022.

[13] Graf S, Korn R. A guide to Monte Carlo simulation concepts for assessment of risk-return profiles for regulatory purposes. *European Actuarial Journal*. 2020; 10(2): 273-293. Available from: doi: 10.1007/s13385-020-00232-3.

[14] Wen X, Deng G. Monte Carlo Simulations with dual variables pricing of barrier options in a stochastic volatility model. *Journal of Guangxi Normal University (Natural Science Edition)*. 2016; 34(2): 90-97. Available from: doi: 10.16088/j.issn.1001-6600.2016.02.013.

[15] Mohammed M. *Machine Learning: Algorithms and Applications*. S.L.: Crc Press; 2020.

[16] Song S, He R, Shi Z, Zhang W. Variable importance measure system based on Advanced Random Forest. *Computer Modeling in Engineering & Sciences*. 2021; 128(1): 65-85. Available from: doi: 10.32604/cmes.2021.015378.

[17] Ogunseye EO, Adenusi CA, Nwanakwaugwu AC, Ajagbe SA, Akinola SO. Predictive analysis of mental health conditions using AdaBoost algorithm. *ParadigmPlus*. 2022; 3(2): 11-26. Available from: doi: 10.55969/paradigmplus.v3n2a2.

[18] Panagiotou DK, Dounis AI. Comparison of hospital building's energy consumption prediction using artificial neural networks, ANFIS, and LSTM Network. *Energies*. 2022; 15(17): 6453. Available from: doi: 10.3390/

en15176453.

[19] Müller AC, Guido S. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. Sebastopol, CA: O'Reilly Media, Inc.; 2018.

[20] Culkin R, Das SR. Machine learning in finance: The case of deep learning for option pricing. *Journal of Investment management*. 2017; 15(4): 92-100.