Option pricing and profitability: A comprehensive examination of machine learning, Black-Scholes, and Monte Carlo method

Sojin Kim[a], Jimin Kim[a], Jongwoo Song[1,a]

[a]Department of Statistics, Ewha Womans University, Korea

Abstract

Options pricing remains a critical aspect of finance, dominated by traditional models such as Black-Scholes and binomial tree. However, as market dynamics become more complex, numerical methods such as Monte Carlo simulation are accommodating uncertainty and offering promising alternatives. In this paper, we examine how effective different options pricing methods, from traditional models to machine learning algorithms, are at predicting KOSPI200 option prices and maximizing investment returns. Using a dataset of 2023, we compare the performance of models over different time frames and highlight the strengths and limitations of each model. In particular, we find that machine learning models are not as good at predicting prices as traditional models but are adept at identifying undervalued options and producing significant returns. Our findings challenge existing assumptions about the relationship between forecast accuracy and investment profitability and highlight the potential of advanced methods in exploring dynamic financial environments.

Keywords: option pricing, Black-Scholes model, MonteCarlo simulation, variance reduction, machine learning

1. Introduction

Option pricing is one of the main topics in finance, closely related to the calculation of integrals. Traditional option pricing theory includes the Black-Scholes model, binomial tree, and the finite difference method. Most of these methods can be valued analytically.

As the path of the underlying asset becomes more complex and dynamic, it might be better to use numerical methods such as Monte Carlo simulations to remove the assumptions of the analytical method. Simulation methods have proven to be valuable and flexible computational tools. Monte Carlo simulation, in particular, performs well in a variety of environments due to its ability to cope with uncertainty. However, one of the drawbacks of this method is that the standard error of the estimate increases with the number of simulation runs. As an alternative, Boyle (1976) and Larsson (2020) found that reducing variance is an effective way to reduce errors and obtain more accurate results.

Data-driven models also draw attention to pricing financial derivatives. Studies using machine learning have shown that future stock prices can be predicted with near-accuracy based on past stock

---

prices. Chowdhury *et al.* (2020) concluded that data mining results are similar to those of the Black-Scholes equation. Furthermore, Ivașcu (2021) confirms that decision tree algorithms such as Random Forest, XGBoost, and LightGBM are effective for option valuation. More recently, Ke and Yang (2019) found that deep learning models such as LSTM outperform Black-Scholes.

In this paper, we predict the price of KOSPI200 options and show the results of investment simulations. We use traditional methods including the Black-Scholes model and binomial tree, and simulation methods such as Monte Carlo and importance sampling. We also apply machine learning and deep learning models. Besides, we want to see how well these methods work for different timeframes. By examining these models closely, we aim to offer valuable insights that can help investors and financial professionals make better decisions in the dynamic world of finance.

In Section 2, we describes the concept of options. We briefly explain the concepts of option positions, call and put options, and their payoffs. Section 3 describes various option pricing methods including the traditional option pricing models such as the Black-Scholes model, and the binomial tree, and also the non-traditional option pricing methods. Simulation methods, such as Monte Carlo simulation, antithetic sampling, control variables, and importance sampling are used to improve the limitations of traditional methods. We also briefly discuss machine learning and deep learning methods which use historical data to predict option prices. Section 4 shows the results of our experiments. We compare the actual and predicted prices of KOSPI200 options. We also demonstrate an investment simulation that shows the expected return rate of option pricing methods. Finally, Section 5 summarizes and concludes our paper.

2. Options

An option is a type of derivative, a contract that exchanges the right to buy or sell at a predetermined price within a predetermined period of time. A person who buys an option is called an option buyer, and a person who sells an option is called an option seller. The right to purchase a specific product at a set price on the option expiration date is called a call option, and the right to sell it is called a put option. Options have value because they are a type of complex transaction, and are affected by the price of the underlying asset, the period until maturity, market volatility, risk-free interest rate, and the exercise price of the option.

The type of option is primarily defined by the date on which the option is exercised. Most options are European or American options. European options are contracts that can be exercised only at expiration, while American options can be exercised at any time before. Options that are calculated similarly to these two options are called "vanilla options." In contrast to regular vanilla options, "exotic options" have relatively complex and unusual features. Exotic options may include products developed for specific markets, and representative examples include Asian options and binary options.

There are three terms that describe where an option is traded in relation to the price of the underlying asset. These terms are called "at the money," "in the money," and "out of the money." If there is intrinsic value by comparing the exercise price of the option and the price of the underlying asset, it is called an in the money state. If there is no intrinsic value, it is called an out of the money state. When the spot price and futures price are the same, the state is at the money (Larsson, 2020).

If S_t is the underlying stock price and K is the strike price, the call option can be defined as follows:

- In the money status : $S_T > K$

- At the money status : $S_T = K$

- Out of the money status : $S_T < K$

On the other hand, the put option is as follows:

- In the money status : $S_T < K$

- At the money status : $S_T = K$

- Out of the money status : $S_T > K$

　　In options trading, "long position" and "short position" are terms that refer to holding and selling an option contract, respectively. Long position refers to purchasing and holding a specific type of option contract. Options held by the buyer can only be exercised during a certain period of time in the future, and if there is profit, profit can be earned by exercising the option contract. Short position refers to selling a specific type of option contract. Sellers make a profit by writing option contracts and selling them to other investors. In other words, a long position is when the stock price is expected to rise, and a short position is when the stock price is expected to fall and the person is trying to make a profit.

　　The position of the call option can be expressed as:

- the payoff from a long position in a call option : $\max(S_T - K, 0)$

- the payoff from a short position in a call option : $\min(K - S_T, 0)$

The Put option is expressed as:

- the payoff from a long position in a put option : $\max(K - S_T, 0)$

- the payoff from a short position in a put option : $\min(S_T - K, 0)$

3. Option pricing methods

Determining option prices is one of the most important aspects of derivatives trading. Among them, the most traditional and famous methods are the Black-Scholes model and binomial tree model. The Black-Scholes model was designed by Fisher Black and Myron Scholes in 1973. With this model, they presented a closed-form mathematical formula for measuring the price of European call options. However, even after the Black-Scholes model was introduced, limitations were revealed in the market, making it difficult to apply this model in some special situations. In particular, because American options can be exercised at any time, a model that considers discrete times was needed. Accordingly, the binomial tree model developed by Cox-Ross-Rubinstein appeared in 1979. This model considers stock price movements at discrete time steps and is useful when evaluating more complex options.

3.1. Black-Scholes model

The Black-Scholes model calculates the theoretical price of an option by considering the stock price, exercise price, interest rate, investment period, and volatility of the underlying asset. At the time of its introduction, this model was recognized as a simple and effective tool for calculating accurate prices. However, since the assumption about the random movement of stock prices is made using a stochastic process called geometric brownian motion, there are limitations in assuming that this assumption always accurately represents stock movements in the real world. The model also makes

several assumptions: Arbitrage opportunities do not exist in the market, there are no taxes, and the risk-free interest rate is constant and volatile.

Basically, the Black-Scholes model shows how to determine the price of an option contract using a simple formula. The calculation formulas for European call option c and put option p are as follows (Chowdhury *et al.*, 2020).

$$c = S e^{-qT} N(d_1) - K e^{-rT} N(d_2) \tag{3.1}$$

$$p = K e^{-rT} N(-d_2) - S e^{-qT} N(-d_1) \tag{3.2}$$

$$d_1 = \frac{\ln(S/K) + (r - q + (\sigma^2/2))T}{\sigma \sqrt{T}} \tag{3.3}$$

$$d_2 = d_1 - \sigma \sqrt{T}, \tag{3.4}$$

where, S is the stock price; K is the exercise price; r is the risk free interest rate; T is the time to expiration; σ : volatility of stock price; q is the dividends; $N(x)$ is the cumulative normal distribution function (MacBeth and Merville, 1979).

3.2. Binomial tree model

The binomial tree model is a model that considers option prices at discrete time steps. This model calculates the price of an option by dividing it into several time stages until the expiration date of the option and creating a binomial tree in which the stock price rises or falls at each stage (Rubinstein, 1994). Each node represents the stock price at a specific point in time and the option price at that point in time. The model calculates option value by considering possible future price movements at each node of the tree, which includes calculating the intrinsic value of the option and the expected value of the option in the future. The binomial tree model is recognized to converge to the Black-Scholes model as the number of time steps increases. This will be confirmed by the experimental part using real KOSPI 200 options data.

Consider the binomial tree model of the first step. This model starts at time $t = 0$ when the underlying asset price is S_t. The price may rise by u or fall by d at each step. At this time, if the probability of rising is p, the probability of falling is $1 - p$. Therefore, the price at time $t = 1$ is as follows (Dar and Anuradha, 2018).

$$S_1 = \begin{cases} S_0 u, & \text{when the price of the underlying asset rises,} \\ S_0 d, & \text{when the price of the underlying asset falls.} \end{cases}$$

As a result, according to Benninga and Wiener (1997), if the number of price increases during t period is j, the stock value at time t becomes

$$S_t = S_0 u^j d^{t-j}. \tag{3.5}$$

3.3. Monte-Carlo method and variance reduction techniques

3.3.1. Monte-Carlo method

The Monte Carlo method is one of the most essential tools in option pricing in the financial field. This method generates a random sample of a stochastic process to calculate the option value and uses this to evaluate the option's payoff. This is then used to estimate the expected value of the entire option. The Monte Carlo method becomes increasingly competitive compared to other numerical integration

methods as the dimensionality increases (Jia, 2009). This is because the strong law of large numbers allows convergence to an accurate integral value even if the number of samples becomes infinite. Additionally, this method has the advantage of being easy to implement and modify.

Stock prices are influenced by the following geometric Brownian motion:

$$S_T = S_0 \cdot e^{(r-q-\frac{\sigma^2}{2})T+\sigma\sqrt{T}Z_i}, \tag{3.6}$$

when Z_i are independent samples from the standard normal distribution. Then, the price of the call option is determined as follows.

$$c = e^{-rT} \frac{1}{n} \sum_{i=1}^{n} \max\left(S_T^{(i)} - K, 0\right). \tag{3.7}$$

However, the Monte Carlo method requires a large number of simulations and can be computationally expensive. To solve these problems and increase precision, various variance reduction techniques have been introduced. Boyle *et al.* (1997) Afterwards, the Monte Carlo approach was applied. Representative examples of these methods include antithetic variates, control variates, and importance sampling.

3.3.2. Antithetic sampling

One of the most common variance reduction techniques is the antithetic variates technique. This technique reduces variance by generating paths in opposite directions when running a simulation. For example, when creating a path where the stock price increases, you can also create a path where the stock price decreases and use the correlation between the two paths to reduce variance.

For example, it can be generated using Monte Carlo methods.

$$Y_i = e^{-rT} \max\left(S_T^{(i)} - K, 0\right). \tag{3.8}$$

In this context, the antithetic sampling method is based on the fact that when Z_i follows a standard normal distribution, $-Z_i$ also follows a standard normal distribution (Boyle *et al.*, 1997). Therefore, Z_i can be replaced with $-Z_i$ and expressed as follows.

$$\hat{Y}_i = e^{-rT} \max\left(S_T^{\hat{(i)}} - K, 0\right). \tag{3.9}$$

Therefore,

$$\hat{Y_{AV}} = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i + \hat{Y}_i}{2}. \tag{3.10}$$

Here Y_i and $\hat{Y}_i$ are balanced and thus the variance is reduced.

3.3.3. Control variates

The control variates mathod is a technique to reduce variance by using other variables that affect option prices. For example, variables such as the interest rate or volatility of options are used to adjust the simulation results of the stock price path to reduce variance. In Monte Carlo simulation, the value

of a given option is used as a control variable to calculate the estimated value. This can furthermore be described in a mathematical way

$$P_A = \hat{P_A} + \beta \left(P_B - \hat{P_B} \right). \tag{3.11}$$

Here, P_A is an estimate of the option price for A, $\hat{P_A}$ is the option price estimated through Monte Carlo simulation, P_B is the known price for a similar option B, $\hat{P_B}$ is the expected price of similar option, and β is the parameter that needs to be calculated. The goal is to minimize price fluctuations, so fluctuations appear like this.

$$\text{Var}(P_A) = \text{Var}\left(\hat{P_A}\right) + \beta^2 \text{Var}\left(\hat{P_B}\right) - 2\beta \text{Cov}\left(\hat{P_A}\hat{P_B}\right). \tag{3.12}$$

Then, when the derivative of the above equation with respect to B is set to 0, variance-minimizing B is as follows.

$$\beta^* = \frac{\text{Cov}(\hat{P_A}\hat{P_B})}{\text{Var}(\hat{P_B})}. \tag{3.13}$$

It can be seen that a high correlation between A and B provides high variance reduction.

3.3.4. Importance sampling

This method improves the accuracy of the simulation by generating more samples in cases where the probability of a particular event occurring is low (Bolia and Juneja, 2005). For example, if the stock price rarely exceeds the option strike price, more accurate estimates can be obtained by sampling these events more frequently. This is done by adjusting the distribution, or path. Consider the problem of estimating

$$V = E_f[h(x)] = \int h(x)f(x)dx. \tag{3.14}$$

Here X is a random vector in R^n with probability density f. h is the payoff according to the generated stock price path, and f is a new random probability distribution for sampling. The ordinary Monte Carlo estimator is

$$\hat{V}_f = \frac{1}{n} \sum_{i=1}^{n} h(X_i), \tag{3.15}$$

with $X_1, \ldots, X_n$ independent draws from f.

Let g be any pdf satisfying $f(x) > 0 \Rightarrow g(x) > 0$ for all $x \in R^n$. Then You can then change the measurements to get

$$V = \int h(x)\frac{f(x)}{g(x)}g(x)dx = E_g\left[h(x)\frac{f(x)}{g(x)}\right]. \tag{3.16}$$

If $X_1, \ldots, X_n$ are independent from g, the importance sampling estimator associated with g is

$$\hat{V}_g = \frac{1}{n} \sum_{i=1}^{n} h(X_i)\frac{f(X_i)}{g(X_i)}. \tag{3.17}$$

The weight $f(X_i)/g(X_i)$ will be the likelihood ratio evaluated at X_i. If $V \neq 0$, estimate is thus obtained by choosing

$$g(X) = \frac{1}{V}h(x)f(x), \tag{3.18}$$

which we call the optimal density function. Unfortunately, there is no general way to find the optimal g for an ordinary function f. So, we should find something relatively optimal to make the variance as small as possible.

4. Experiments

In the experimental part, we predict the price of KOSPI200 options and calculate the profit/loss of an investment simulation. Accordingly, we find out which method is the most predictive and which is the most profitable. The price prediction is performed for three cases: Weekly, monthly, and 50 days to maturity referring to Martinkute - Kauliene *et al.* (2013).

The methods used to predict option prices and calculate profit/loss include traditional methods, and models in Monte Carlo simulations with variance reduction techniques. We also make a comprehensive comparison using machine learning methods such as XGBoost, random forest, and multilayer perceptron.

The machine learning methods include inputs such as S (stock price), K (strike price), r (risk-free rate), T (time to expiration), σ (volatility), and q (dividend rate), similar to other methods for fair comparison. However, machine learning methods necessitate training data, unlike traditional methods where input values can be directly inserted into a predefined formula. Consequently, in a machine learning method for predicting option prices, data from up to one day prior is employed as the training data.

We use XGBoost (Chen and Guestrin, 2016) without parameter tuning and for random forest (Breiman, 2001), we set the number of estimators to 500, maximum depth to 8, and maximum features to 5 to optimize performance. The MLP (Haykin, 1994) model was built based on Culkin and Das (2017). The first layer consisted of 120 nodes, and the model consisted of a total of 7 hidden layers. The output layer comprised a single node with a linear activation function to predict option prices. For training the MLP, we employed the mean squared error (MSE) loss function and RMSProp optimizer. We trained the model using a batch size of 32 and iterated over 50 epochs to optimize performance.

4.1. KOSPI 200 data description

The KOSPI 200 index includes 200 stocks that best represent the market capitalization, liquidity, and industry composition of the KRX (Ahn *et al.*, 2008). The KOSPI 200 option, introduced in July 1997, is an option based on the KOSPI 200 index as an underlying asset. Secondary derivatives exist such as the KOSPI 200 Weekly option, which has a one-week expiry, and the Mini KOSPI 200 option, which enables small investments by reducing the contract size by 1/5. The KOSPI 200 option is a European option that can only be exercised at expiration, and the expiration date is fixed to the second Thursday of every month. The trading history of KOSPI 200 options, including S (stock price), K (strike price), T (time to expiration), σ (implied volatility), and q (dividend rate) was obtained from data provided by KRX market data system. We collected the data by crawling the option transaction history. The r (risk-free rate) corresponds to the KOFR (Korea overnight financing repo rate). The price of the KOSPI 200 option is expressed in premium points. The multiplier is 250,000(KRW), which means that the actual price equals the quoted price multiplied by 250,000(KRW).

Table 1: Summary of KOSPI200 option

Call/Put	Count	Mean	Min	Max	Quantile		
					25%	50%	75%
Call option	3.5K	6.76	0.01	176.65	0.09	0.83	5.40
Put option	4.8K	4.43	0.01	134.9	0.08	0.635	3.275
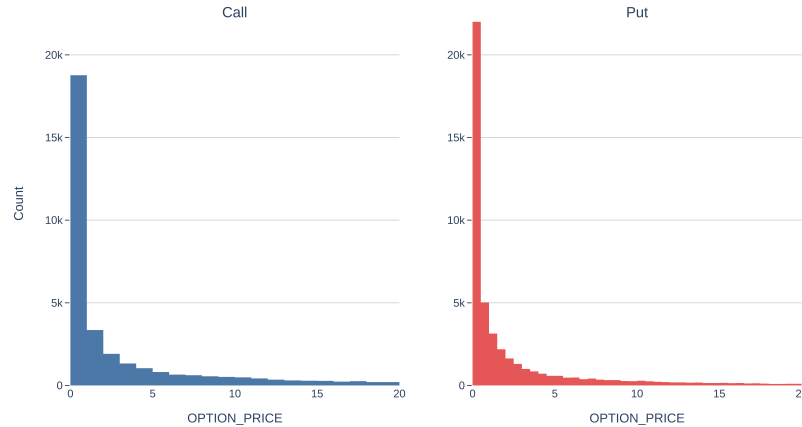
Figure 1: *Histogram of call/put option price.*

We use one year of data from January to December 2023. However, we need earlier data to apply the machine learning method in predicting Jan 2023 option price. Therefore, we use data from November 2022 for our analysis.

Table 1 presents a summary of call and put options within the dataset. The analysis reveals a higher transaction volume for put options as opposed to call options. Furthermore, it is observed that put options exhibit lower maximum and quantile values compared to their call counterparts. This can also be seen in the histogram in Figure 1, where the number of options close to 0 is much higher for puts than for calls.

4.2. Results

4.2.1. Option price prediction

Table 2 shows the RMSE of each method based on call options in three time frames: Week, month, and 50days. The table is organized into three sections, each corresponding to different time frames (week, month, and 50 days) for option pricing.

Highlighted in the table are the two smallest RMSEs for each time frame. The Black-Scholes (BS) model and the binomial tree (BinT) model have the smallest values in every time frame, suggesting the potential applicability of the Black-Scholes (BS) model in real-world option pricing. Also, the binomial tree (BinT) model and the Black-Scholes (BS) model show similar results. This supports the fact that the binomial tree (BinT) converges with Black-Scholes (BS) even with moderate levels of time steps(100). Furthermore, simulation methods with variance reduction techniques demonstrate comparable performance. As we can see in Figure 2, machine learning methods are more widespread than other methods with larger RMSE. Among the machine learning techniques, XGBoost and random forest models outperform MLP models, particularly in predicting call options over put options.

Table 2: RMSE of each method in call option

Timeframe	Traditional		Simulation				Machine Learning		
	BS	BinT	MC	AS	CV	IS	XGB	RF	MLP
Week	**0.650**	**0.650**	0.684	0.685	0.685	0.686	6.307	5.097	21.351
Month	**0.641**	**0.641**	0.664	0.664	0.664	0.656	8.863	8.557	25.638
50days	**0.386**	**0.385**	0.485	0.485	0.486	0.495	4.498	3.935	8.366

BS (Black-Scholes), BinT (Binomial Tree), MC (Monte-Carlo), AS (Antithetic Sampling), CV (Control Variates), IS (Importance Sampling), XGB (XGBoost), RF (Random Forest), MLP (Multilayer Perceptron)

(a) *Call option*   (b) *Put option*
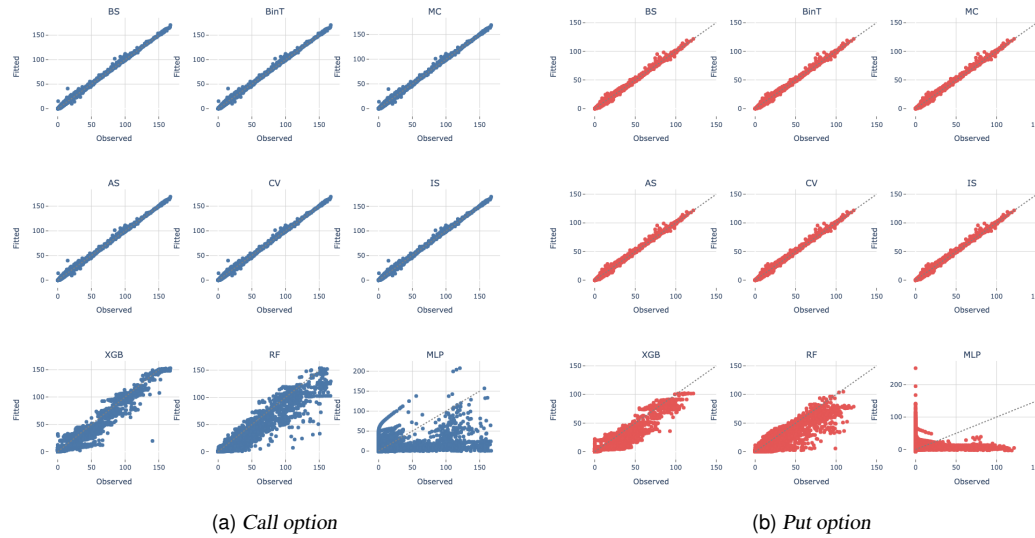
Figure 2: *Observed vs fitted graph.*

A sample table of the prices of the options predicted by each method and the actual options is attached in the Appendix A.1. A table of the results for put options is also available in the Appendix B.1.

4.2.2. Profit/loss in investment simulation

The investment simulation is a calculation of the profit/loss associated with buying options in the real options market. An option is considered undervalued if its predicted price is higher than its actual market price, thereby presenting an opportunity for profit. Consequently, a greater disparity between the predicted and actual option prices increases the profit potential. We assume a monthly investment and acquire options on the first of each month. The predicted option prices are derived through the methods used earlier. Subsequently, the 10 options are selected for investment which have the most significant differences between actual and predicted prices. For diversified investment across 10 options, two scenarios are considered: One where we invest the same amount in each chosen option and another where we invest proportionally with differences between their actual and predicted prices.

The outcomes of these simulations are presented in Table 3 and Table 4, with the return rate. The last column indicates the average returns for each month, with bold formatting applied to the three

Table 3: Return rate of call option

(a) Equally Invest

Method	202301	202302	202303	202304	202305	202306	202307	202308	202309	202310	202311	202312	Average
BS	0.056	0.008	0.026	0.009	−0.017	−0.012	−0.016	−0.045	−0.041	0.016	0.082	0.009	0.006
BinT	0.051	0.008	0.026	0.001	−0.017	−0.012	−0.016	−0.045	−0.041	0.016	0.082	0.009	0.005
MC	0.066	0.011	0.007	−0.1	−0.001	0.015	−0.014	−0.036	−0.03	0.007	0.075	0.005	0.000
AS	0.066	0.011	0.004	−0.1	−0.001	0.015	−0.014	−0.036	−0.03	0.007	0.075	0.005	0.000
CV	0.066	0.011	0.004	−0.1	−0.001	0.015	−0.014	−0.036	−0.03	0.007	0.075	0.005	0.000
IS	0.066	0.011	0.004	−0.1	−0.001	0.014	−0.014	−0.036	−0.03	0.007	0.075	0	0.000
XGB	0.219	0.001	−0.011	0.015	−0.1	0.037	−0.1	−0.1	0.013	0.048	0.72	−0.1	**0.054**
RF	1.325	0.002	−0.026	0.245	0.003	0.031	−0.006	−0.021	0.001	0.036	0.542	0.026	**0.180**
MLP	0.123	0.011	−0.012	0.019	−0.007	0.008	−0.002	−0.03	0.027	0.027	0.015	0.005	0.015

(b) Proportionally Invest

Method	202301	202302	202303	202304	202305	202306	202307	202308	202309	202310	202311	202312	Average
BS	0.253	0.01	0.025	0.006	−0.007	−0.012	−0.013	−0.045	−0.04	0.03	0.093	0.032	0.028
BinT	0.239	0.01	0.025	0.002	−0.008	−0.014	−0.013	−0.045	−0.04	0.031	0.093	0.032	0.026
MC	0.662	0.012	0.015	−1	0.004	0.02	−0.011	−0.036	−0.028	0.038	0.094	0.048	−0.015
AS	0.662	0.012	0.014	−1	0.004	0.02	−0.011	−0.036	−0.028	0.038	0.094	0.048	−0.015
CV	0.662	0.012	0.014	−1	0.004	0.02	−0.011	−0.036	−0.028	0.038	0.094	0.048	−0.015
IS	0.645	0.012	0.014	−0.415	0.004	0.02	−0.011	−0.036	−0.028	0.037	0.094	0	0.028
XGB	0.234	0.001	−0.013	0.02	−0.095	0.052	−0.099	−0.099	0.026	0.057	0.704	−0.087	**0.058**
RF	0.694	0.006	−0.028	0.22	−0.001	0.046	−0.005	−0.023	0.012	0.034	0.473	0.029	**0.121**
MLP	0.125	0.013	−0.006	0.026	−0.008	0.012	−0.015	−0.029	0.027	0.068	0.05	0.014	0.023

*BS (Black-Scholes), BinT (Binomial Tree), MC (Monte Carlo), AS (Antithetic Sampling), CV (Control Variates), IS (Importance Sampling), XGB (XGBoost), RF (Random Forest), MLP (Multilayer Perceptron)

Table 4: Return rate of put option

(a) Equally Invest

Method	202301	202302	202303	202304	202305	202306	202307	202308	202309	202310	202311	202312	Average
BS	−0.053	−0.1	−0.01	−0.035	−0.1	−0.045	−0.001	−0.1	−0.014	−0.053	−0.039	−0.026	−0.048
BinT	−0.053	−0.1	−0.01	−0.035	−0.1	−0.045	0	−0.1	−0.014	−0.053	−0.039	−0.026	−0.048
MC	−0.058	−0.069	−0.054	−0.033	−0.1	−0.062	−0.025	0.034	0.012	−0.077	−0.057	−0.025	−0.043
AS	−0.058	−0.069	−0.054	−0.033	−0.1	−0.062	−0.025	0.043	0.012	−0.077	−0.057	−0.025	−0.042
CV	−0.058	−0.069	−0.054	−0.033	−0.1	−0.062	−0.025	0.034	0.012	−0.077	−0.057	−0.025	−0.043
IS	−0.058	−0.069	−0.056	−0.034	−0.1	−0.067	−0.025	0.034	0.006	−0.077	−0.057	−0.023	−0.044
XGB	−0.1	−0.035	0.004	−0.1	−0.1	−0.1	−0.001	0.06	−0.1	−0.045	−0.1	−0.1	−0.060
RF	−0.048	−0.021	−0.1	−0.1	−0.1	−0.072	−0.027	0.026	−0.1	−0.054	−0.1	−0.012	−0.059
MLP	−0.1	−0.1	−0.076	−0.1	−0.1	−0.1	−0.1	−0.1	−0.012	−0.1	−0.1	−0.006	−0.083

(b) Proportionally Invest

Method	202301	202302	202303	202304	202305	202306	202307	202308	202309	202310	202311	202312	Average
BS	−0.054	−0.113	−0.004	−0.035	−0.102	−0.064	0.023	−0.117	−0.113	−0.051	−0.129	−0.026	−0.065
BinT	−0.054	−0.573	−0.004	−0.035	−0.25	−0.064	0.033	−0.165	−0.117	−0.051	−0.13	−0.026	−0.120
MC	−0.058	−0.087	−0.055	−0.036	−0.143	−0.07	−0.024	0.043	0.006	−0.077	−0.067	−0.024	−0.049
AS	−0.058	−0.086	−0.055	−0.036	−0.143	−0.07	−0.024	0.047	0.006	−0.077	−0.067	−0.024	−0.049
CV	−0.058	−0.087	−0.055	−0.036	−0.143	−0.07	−0.024	0.043	0.006	−0.077	−0.067	−0.024	−0.049
IS	−0.058	−0.087	−0.055	−0.036	−0.135	−0.076	−0.024	0.037	0.005	−0.077	−0.071	−0.022	−0.050
XGB	−0.099	−0.046	0.007	−0.1	−0.094	−0.094	0.002	0.066	−0.099	−0.062	−0.091	−0.073	−0.057
RF	−0.06	−0.027	−0.096	−0.1	−0.099	−0.074	−0.025	0.018	−0.099	−0.096	−0.096	−0.014	−0.064
MLP	−0.176	−0.1	−0.122	−0.143	−0.084	−0.091	−0.1	−0.095	−0.012	−0.095	−0.092	0	−0.093

*BS (Black-Scholes), BinT (Binomial Tree), MC (Monte Carlo), AS (Antithetic Sampling), CV (Control Variates), IS (Importance Sampling), XGB (XGBoost), RF (Random Forest), MLP (Multilayer Perceptron)

highest positive returns. In call options, positive return is observed across the Black-Scholes model, binomial tree model, and three machine learning models, under both equal and proportional investment strategies. Interestingly, machine learning methods tend to yield higher profits than traditional methods. This implies that despite less precise predictions, the volatility in machine learning forecasts facilitates investment in undervalued options.

Table 5: Random forest profit table

No	DATE	CALL/PUT	Expiry	K	S0 (a)	St (b)	Option price (c)	RF (d)	Difference (e) = (d) − (c)	Payoff *(f) = ((b) − (a))+	Profit (g) = (f) − (c)	Return_rate (h) = (g)/(c)
1	2023.1.2	C	38	397.5	289.79	324.9	0.01	17.11	17.10	0	−0.01	−1
2	2023.1.2	C	10	342.5	289.79	310.7	0.01	15.28	15.27	0	−0.01	−1
3	2023.1.2	C	10	350	289.79	310.7	0.01	13.97	13.96	0	−0.01	−1
4	2023.1.2	C	10	325	289.79	310.7	0.015	8.25	8.23	0	−0.015	−1
5	2023.1.2	C	10	322.5	289.79	310.7	0.015	7.51	7.49	0	−0.015	−1
6	2023.1.2	C	10	327.5	289.79	310.7	0.01	5.65	5.64	0	−0.01	−1
7	2023.1.2	C	10	330	289.79	310.7	0.01	5.54	5.53	0	−0.01	−1
8	2023.1.2	C	10	320	289.79	310.7	0.02	5.29	5.27	0	−0.02	−1
9	**2023.1.2**	**C**	**10**	**305**	**289.79**	**310.7**	**0.255**	**4.94**	**4.68**	**5.7**	**5.445**	**21.4**
10	2023.1.2	C	10	312.5	289.79	310.7	0.045	4.73	4.68	0	−0.045	−1

$^{*}(X)^{+} = \text{Max}(X, 0)$

Random forest and Xgboost are the most effective methods for both equal and proportional investment, especially in January. The profit outcomes for random forest in January 2023 are detailed in Table 5. It shows the 10 call options with the largest difference between random forest's predicted price and the actual price. Looking closely at the table, we can see that the actual option prices vary from 0.015 to 0.255 depending on the expiry and strike price. Notably, the option ranked 9th for its return rate with a strike price of 305, stands out for its relatively lower strike price in comparison to the others. For a call option to be profitable, the price of the underlying asset must rise above the strike price. The low strike price of 305 made this option significantly more expensive than others, yet the substantial increase in the asset's price resulted in a notable profit.

Contrary to the initial expectation that the size of the difference would be related to the expected return, The option with the largest return was the ninth option, not the first. Therefore, it is unlikely that the difference between the actual and expected option price is related to the return.

5. Conclusion

This paper demonstrates a comprehensive examination of various option pricing methods ranging from traditional models like the Black-Scholes and binomial tree to advanced techniques involving machine learning and simulation methods. The primary objective was to discern the most effective methods for predicting KOSPI200 option prices and to evaluate the profitability of these predictions through investment simulations.

Our findings reveal that traditional models such as the Black-Scholes and the binomial tree models have shown superior performance in predicting option prices. This allows us to make reasonable inferences about using these models in actual option pricing.

Machine learning models, especially random forest and XGBoost, fell short in accuracy when predicting option prices compared to traditional methods, but performed well in maximizing returns on investment. This outcome was remarkable in January 2023, leveraging the difference between predicted and actual prices to find undervalued options, generating significant profits.

This study suggests that while traditional models are efficient in terms of prediction accuracy, machine learning methods can provide investors with a competitive edge by finding opportunities that traditional models might overlook. The adaptability of these methods to various market conditions and their ability to handle complex, dynamic scenarios prove invaluable in the constantly evolving finance fields.

It's important to note that these results are specific to a particular option and time period, so generalization is not advisable. Future research could explore a more thorough study with a wider variety and longer period of options data. By continuing to refine these models and strategies, the financial

community can better explore the intricacies of market dynamics, enhance investment strategies, and ultimately achieve more robust financial outcomes.

Appendix A: Predicted option prices

Table A.1: Call option - Week

DATE	REAL	BS	BinT	MC	AS	CV	IS	XGB	RF	MLP
Jan	0.620	0.342	0.339	0.327	0.327	0.326	0.330	1.786	1.168	0.036
Feb	0.800	0.558	0.561	0.542	0.539	0.538	0.534	1.903	1.071	0.938
Mar	0.650	0.263	0.261	0.249	0.250	0.251	0.244	4.631	3.692	0.131
Apr	1.185	0.421	0.422	0.406	0.404	0.403	0.412	2.547	1.573	0.281
May	0.820	0.659	0.658	0.631	0.628	0.628	0.625	1.772	1.646	−0.204
Jun	0.900	0.560	0.563	0.531	0.533	0.532	0.532	1.096	0.632	1.059
Jul	0.660	0.334	0.332	0.317	0.317	0.318	0.330	1.487	1.677	−0.112
Aug	0.685	0.337	0.338	0.323	0.323	0.322	0.337	1.083	0.613	0.843
Sep	0.175	0.095	0.095	0.089	0.088	0.089	0.082	1.555	1.064	0.007
Oct	0.220	0.056	0.054	0.052	0.052	0.052	0.050	1.965	1.940	0.919
Nov	1.190	0.871	0.872	0.841	0.840	0.843	0.825	2.113	2.462	0.545
Dec	1.120	0.929	0.933	0.900	0.897	0.899	0.919	2.696	2.260	0.248

Table A.2: Call option - Month

DATE	REAL	BS	BinT	MC	AS	CV	IS	XGB	RF	MLP
Jan	1.775	2.711	2.720	2.550	2.551	2.538	2.580	4.399	4.421	0.590
Feb	0.555	0.760	0.765	0.707	0.702	0.699	0.703	0.278	0.296	1.448
Mar	0.715	0.776	0.781	0.711	0.714	0.713	0.699	2.980	1.124	0.972
Apr	1.485	1.843	1.841	1.709	1.712	1.714	1.693	4.821	4.185	1.956
May	1.065	1.158	1.161	1.068	1.071	1.065	1.081	0.140	0.222	1.228
Jun	1.155	1.037	1.040	0.943	0.942	0.939	0.981	4.147	2.417	1.182
Jul	1.025	0.865	0.860	0.791	0.789	0.787	0.765	0.296	0.166	1.667
Aug	0.770	0.932	0.939	0.850	0.847	0.847	0.825	3.595	2.690	0.987
Sep	0.530	0.415	0.418	0.371	0.372	0.372	0.375	2.843	1.835	3.237
Oct	0.675	0.515	0.516	0.459	0.458	0.459	0.498	3.061	1.431	0.994
Nov	0.535	0.303	0.305	0.273	0.272	0.271	0.282	1.674	1.180	1.651
Dec	1.160	0.882	0.883	0.819	0.817	0.812	0.790	1.140	0.959	1.544

Table A.3: Call option - 50days

DATE	REAL	BS	BinT	MC	AS	CV	IS	XGB	RF	MLP
Jan	0.575	0.855	0.854	0.769	0.769	0.761	0.816	-	-	-
Feb	0.710	0.981	0.988	0.869	0.868	0.868	0.869	3.595	2.120	1.803
Mar	0.775	0.694	0.696	0.621	0.616	0.619	0.592	0.271	0.504	1.565
Apr	0.860	0.774	0.775	0.690	0.697	0.692	0.709	1.926	1.408	1.081
May	0.580	0.652	0.657	0.583	0.579	0.579	0.556	0.885	0.839	0.885
Jun	0.635	0.561	0.565	0.492	0.491	0.490	0.510	0.117	0.164	2.947
Jul	0.560	0.607	0.612	0.533	0.537	0.531	0.529	0.075	0.181	1.562
Aug	0.570	0.550	0.553	0.480	0.482	0.482	0.504	1.121	0.606	1.333
Sep	0.475	0.330	0.329	0.288	0.287	0.288	0.280	0.212	0.545	1.399
Oct	0.670	0.720	0.718	0.622	0.628	0.628	0.634	4.354	2.212	1.711
Nov	0.660	0.717	0.713	0.620	0.622	0.620	0.616	1.263	0.605	2.951
Dec	0.715	0.559	0.559	0.489	0.494	0.494	0.485	1.962	1.214	1.384

XGB, RF, and MLP for January are not calculable due to insufficient training data preceding 50 days.

Table A.4: Put option - Week

DATE	REAL	BS	BinT	MC	AS	CV	IS	XGB	RF	MLP
Jan	0.660	0.542	0.542	0.562	0.561	0.559	0.570	1.786	1.168	0.927
Feb	0.585	0.480	0.482	0.498	0.498	0.499	0.497	1.903	1.071	1.675
Mar	1.260	1.082	1.085	1.106	1.111	1.112	1.087	4.631	3.692	0.857
Apr	1.090	1.211	1.215	1.254	1.256	1.253	1.264	2.547	1.573	1.007
May	0.915	0.658	0.661	0.678	0.682	0.679	0.703	1.772	1.646	2.987
Jun	1.060	0.925	0.926	0.963	0.964	0.964	0.972	1.096	0.632	1.051
Jul	1.310	1.565	1.563	1.621	1.620	1.619	1.641	1.487	1.677	1.487
Aug	0.900	0.799	0.801	0.824	0.827	0.823	0.808	1.083	0.613	1.549
Sep	0.510	0.403	0.405	0.421	0.422	0.420	0.431	1.555	1.064	1.886
Oct	1.190	1.226	1.231	1.270	1.266	1.267	1.247	1.965	1.940	6.166
Nov	1.085	0.973	0.978	1.005	1.005	1.002	1.004	2.113	2.462	1.166
Dec	1.090	0.924	0.928	0.953	0.949	0.952	0.917	2.696	2.260	3.109

Table A.5: Put option - Month

DATE	REAL	BS	BinT	MC	AS	CV	IS	XGB	RF	MLP
Jan	1.540	0.833	0.825	0.903	0.901	0.899	0.867	4.399	4.421	0.341
Feb	0.450	0.355	0.352	0.385	0.382	0.383	0.399	0.278	0.296	0.847
Mar	0.920	0.754	0.754	0.814	0.812	0.811	0.794	2.980	1.124	1.696
Apr	1.800	1.879	1.883	1.978	1.986	1.983	1.964	4.821	4.185	1.318
May	0.410	0.334	0.328	0.366	0.362	0.362	0.379	0.140	0.222	1.793
Jun	0.165	0.141	0.140	0.154	0.153	0.152	0.165	4.147	2.417	2.631
Jul	0.845	0.798	0.805	0.879	0.880	0.877	0.851	0.296	0.166	0.882
Aug	0.860	0.953	0.959	1.028	1.029	1.028	1.016	3.595	2.690	1.701
Sep	1.425	1.415	1.413	1.530	1.531	1.530	1.562	2.843	1.835	1.166
Oct	0.530	0.524	0.527	0.581	0.585	0.586	0.576	3.061	1.431	0.931
Nov	0.045	0.060	0.058	0.064	0.063	0.064	0.062	1.674	1.180	1.098
Dec	1.010	0.998	1.005	1.064	1.070	1.071	1.046	1.140	0.959	1.636

Table A.6: Put option - 50days

DATE	REAL	BS	BinT	MC	AS	CV	IS	XGB	RF	MLP
Jan	0.985	0.596	0.600	0.657	0.654	0.656	0.658	-	-	-
Feb	0.040	0.026	0.025	0.028	0.028	0.029	0.026	3.595	2.120	2.563
Mar	0.455	0.393	0.393	0.436	0.439	0.435	0.428	0.271	0.504	2.970
Apr	0.030	0.031	0.030	0.032	0.032	0.033	0.030	1.926	1.408	0.999
May	0.580	0.498	0.492	0.544	0.540	0.538	0.515	0.885	0.839	1.020
Jun	0.410	0.436	0.432	0.488	0.488	0.488	0.479	0.117	0.164	1.240
Jul	0.875	0.789	0.796	0.877	0.875	0.875	0.886	0.075	0.181	0.989
Aug	0.565	0.525	0.524	0.592	0.592	0.590	0.582	1.121	0.606	1.523
Sep	0.465	0.520	0.524	0.578	0.577	0.578	0.600	0.212	0.545	1.009
Oct	0.775	0.597	0.594	0.663	0.663	0.664	0.677	4.354	2.212	2.695
Nov	1.260	1.106	1.114	1.240	1.239	1.234	1.242	1.263	0.605	2.188
Dec	0.400	0.439	0.441	0.483	0.479	0.484	0.489	1.962	1.214	0.570

XGB, RF, and MLP for January are not calculable due to insufficient training data preceding 50 days.

Appendix B:

Table B.1: RMSE for each method in put option

Timeframe	Traditional			Simulation			Data mining		
	BS	BinT	MC	AS	CV	IS	XGB	RF	MLP
Week	0.592	0.591	0.637	0.636	0.635	0.637	6.816	7.425	20.00
Month	0.627	0.627	0.711	0.711	0.711	0.708	6.270	6.102	19.717
50days	0.328	0.326	0.383	0.383	0.383	0.383	3.653	2.782	11.189

BS (Black-Scholes), BinT (Binomial Tree), MC (Monte-Carlo), AS (Antithetic Sampling), CV (Control Variates), IS (Importance Sampling), XGB (XGBoost), RF (Random Forest), MLP (Multilayer Perceptron)

References

Ahn H-J, Kang J, and Ryu D (2008). Informed trading in the index option market: The case of KOSPI 200 options, *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, **28**, 1118–1146.

Benninga S and Wiener Z (1997). The binomial option pricing model, *Mathematica in Education and Research*, **6**, 27–34.

Bolia N and Juneja S (2005). Monte Carlo methods for pricing financial options, *Sadhana*, **30**, 347–385.

Boyle PP (1977). Options: A monte carlo approach, *Journal of Financial Economics*, **4**, 323–338.

Boyle P, Broadie M, and Glasserman P (1997). Monte Carlo methods for security pricing, *Journal of Economic Dynamics and Control*, **21**, 1267–1321.

Breiman L (2001). Random forests, *Machine Learning*, **45**, 5–32.

Chen T and Guestrin C (2016). XGBoost: A scalable tree boosting system, In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, New York, NY, 785–794.

Chowdhury R, Mahdy MRC, Alam TN, Al Quaderi GD, and Rahman MA (2020). Predicting the stock price of frontier markets using machine learning and modified Black-Scholes option pricing model, In *Physica A: Statistical Mechanics and Its Applications*, **555**, (pp. 124444), Elsevier, Amsterdam.

Culkin R and Das SR (2017). Machine learning in finance: The case of deep learning for option pricing, *Journal of Investment Management*, **15**, 92–100.

Dar AA and Anuradha N (2018). Comparison: Binomial model and Black Scholes model, *Quantitative Finance and Economics*, **2**, 230–245.

Haykin S (1994). *Neural Networks: A Comprehensive Foundation*, Prentice Hall PTR, United States.

Ivașcu C-F (2021). Option pricing using machine learning, *Expert Systems with Applications*, **163**, 113799.

Jia Q (2009). *Pricing American Options Using Monte Carlo Methods*, Unpublished manuscript, Uppsala University, Uppsala.

Ke A and Yang A (2019). Option pricing with deep learning, *Department of Computer Science, Standford University, In CS230: Deep Learning*, **8**, 1–8.

Larsson J (2020). Optimization of option pricing: Variance reduction and low-discrepancy techniques (Unpublished bachelor's thesis), Umeå University, Umeå

Martinkutė-Kaulienė R, Stankevičienė J, and Venslavienė, Santautė (2013). Option pricing using Monte Carlo simulation, Generolo Jono Žemaičio Lietuvos karo akademija, *Journal of Secu-*

rity and Sustainability Issues www.lka.lt/index.php/lt/217049/ ISSN 2029-7017/ISSN 2029-7025 online, **2**, 65–79, Available from: http://dx.doi.org/10.9770/jssi.2013.2.4(7)

MacBeth JD and Merville LJ (1979). An empirical examination of the Black-Scholes call option pricing model, *The Journal of Finance*, **34**, 1173–1186.

Rubinstein M (1994). Implied binomial trees, *The Journal of Finance*, **49**, 771–818.