Dean&Francis

ISSN 2959-6157

Improving Binomial Tree Option Pricing: Estimating Volatility via Linear Regression

Hang Zhang

School of Engineering, Nanfang College Guangzhou, Guangzhou, 510000, China
Corresponding author: 250025zhangh@stu.nfu.edu.cn

Abstract:

The binomial tree option pricing model, proposed by Cox, Ross and Rubinstein, is widely used due to its intuitive discrete-time structure and flexibility in handling various option types. However, the traditional model often assumes constant volatility, which deviates from actual market conditions. This paper proposes an improvement to relax the constant volatility assumption. Specifically, the study first reviews the basic knowledge of the binomial tree model and transforms it into a semi-parametric model. Unlike traditional methods that assume constant volatility, the author employs linear regression to estimate volatility from real option market data. Under the risk-neutral pricing principle, the fitted volatility is embedded into each node of the binomial tree, allowing the model to reflect market-implied volatility characteristics. Using SSE 50 ETF call option data as the sample, empirical analysis shows that the proposed semi-parametric model significantly reduces pricing errors compared to the traditional binomial tree model. These results indicate that incorporating market implied volatility through information regression enhances the practicality of the binomial tree model. Future research could extend this approach to more complex volatility structures or other types of options and underlying assets.

Keywords: Binomial Tree Option Pricing Model; Semi-parametric Model; Linear Regression

1. Introduction

In the early development of option markets, effective pricing tools were not available. As a result, option trading activities were largely based on the experience and intuition of market participants [1]. It was not until the 1970s that Black, Scholes, and Merton pioneered the Black-Scholes model. People could use a mathematical model to caught the virtual data fluctuations. Although the bs model helped trader making options more reliable through mathematical derivation, in real-world financial markets, these assumptions can hardly be fully satisfied. In 1979 Cox, Ross and Rubinstein proposed the Binomial Tree model, it used a more intuitive to present the change of option price. It employs a more intuitive and discrete method to characterize the price dynamics of the underlying asset. Time is split into numerous small

ISSN 2959-6157

periods. In each period, the asset price can only move in two possible directions: up or down, which resembles the branches of a binary tree. Compared with the bs model, it is obviously more flexible.

In the present, the current research mainly improves the binomial tree model from three aspects: the first of among them is Li Jiaxin's thesis [2]. This article points out that the risk-free is difficult to achieve in reality. The second one is used fuzzy mathematics to deal with the uncertainty of stock price fluctuations, built a fuzzy binary tree model, and gave a reasonable range of option prices [3]. The last one is to implement model applications with the help of software [4].

In fact, this model could indeed give people an understanding of ever-changing prices, especially for European options. But both models have a biggest problem that they need fixed parameters to calculate the price of the option. It will lead to a huge problem, which the traditional binomial tree model is built on the assumption of constant volatility. Nevertheless, volatility in realworld option markets is not fixed but dynamic. As a result, applying this model tends to generate considerable errors between theoretical prices and actual market values. So, in order to solve this problem, the author will provide empirical evidence on the application of semi-parametric methods in the binomial tree model.

The purpose of this paper is to estimate the real market volatility using linear regression and optimize the traditional binomial tree model to calculate more accurate option prices. Section 2 will theoretically introduce the application of Risk-Neutral Pricing Principle in the Model and how to improve the model into a semi-parametric model in addition use real market data to obtain option prices with smaller errors. In the third section, experiments will be carried out to compare data generated by the traditional and improved binomial tree models. The last Section will devote to the conclusion.

2. Basic Theory and Model Construction:

2.1 Traditional binomial tree as well as its limitations

To clearly illustrate the fundamental theory of the binomial tree model, this paper begins with the simplest form: single-period case. Suppose the current time is S_0.The option has strike price K and expires after one time interval of length Δt. Within the binomial framework, the asset price is assumed to follow a simple two-state process at expiration: it can either increase to S_0u (where $u>1$) or decrease to S_od (where $d<1$). Consequently, the option's earnings take two possible values: C_u = max (s_0u-K, 0) in the up state and C_d = max (s_0d-K, 0) in the down [5]. According to the no-arbitrage principle, one can construct a replicating portfolio consisting of Δ units of the underlying asset and a risk-free borrowing of B, such that its cash flows at expiration exactly match those of the option. Equating the portfolio's expiration values to C_u and C_d in the two states yields the following system:

$$\begin{cases} \Delta S_0 u - B(1+r\Delta \mathrm{t}) = C_u \\ \Delta S_0 d - B(1+r\Delta \mathrm{t}) = C_d \end{cases} \tag{1}$$

Solving this system gives:

$$\begin{cases} \Delta = \dfrac{C_{u-C_d}}{S_0(u-d)} \\ B = \dfrac{\Delta S_0 u - C_u}{1+r\Delta \mathrm{t}} \end{cases} \tag{2}$$

By the no-arbitrage principle, the current option price C_0 must equal the current value of the replicating portfolio:

$$C_0 = \Delta S_0 - B = \frac{1}{1+r\Delta \mathrm{t}}\left(\frac{(1+r\Delta \mathrm{t})-d}{u-d}C_u + \frac{u-(1+r\Delta \mathrm{t})}{u-d}C_d\right) \tag{3}$$

Among them $p=\dfrac{1+r\Delta t\text{-}d}{u\text{-}d}$, it is called risk-neutral probability. The risk-neutral probability is a key concept in option pricing. It is not an actual probability, but rather a virtual probability introduced under the no-arbitrage assumption to facilitate calculation. In a risk-neutral world, the expected return of all assets equals the risk-free rate, which makes option pricing more concise.

The multi-period binomial model is a natural extension of the single-period model. Its pricing follows a recursive backward induction approach: first, calculate the option values at each node on the expiration date (for a call option, $\max(S_T\text{-K},0)$; for a put option, $\max(\text{K-}S_T,0)$). Then, working backwards from the final period to the present, at each node, use the single-period pricing formula to compute the option value as the risk-neutral probability-weighted average of the two succeeding nodes' option values, discounted at the risk-free rate. Repeating this process until reaching the current time yields the option price. In an n-period model, the i period has $i+1$ price nodes, corresponding to different combinations of upward and downward paths [6,7].

2.2 Volatility and Its Market Characteristics

First of all, one needs to clarify the concept of volatility.

Volatility is an indicator used to measure the magnitude of price fluctuations and the severity of fluctuations. In short, it means how unstable price changes are. From this, one can summarize 2 Conclusion: First, price fluctuations will be more violent in a high volatility environment. Second, volatility becoming greater will increase the value of options because the price is more likely to rise to the strike price. The traditional binomial tree option pricing model used historical volatility to predict future value changes. This means that when one calculates the future price, the user's past data will lead to a large error. So, in comparison, the author prefers to choose the implicit volatility instead of the historical volatility. This implicit volatility reflects the invisible conditions in the current situation, so one can use this improved model to make some predictions about future risks. Different from historical volatility, implied volatility has a "forward - looking" nature and takes into account the market's assessment of future risks.

Volatility smile is an important empirical phenomenon in option markets. It describes the pattern that implied volatility varies with the strike price. Specifically, at-the-money options usually have the lowest implied volatility, while in-the-money and out-of-the-money options have relatively higher implied volatility, forming a shape similar to a smile. This phenomenon was first found in the foreign exchange option market. After the 1987 stock market crash, similar patterns known as volatility skew also appeared in stock option markets. Its causes involve many factors, such as the leverage effect, panic premium, fat-tailed distribution of asset returns, and imbalance between supply and demand. The existence of the volatility smile indicates that the assumption of constant volatility in the traditional binomial tree model is unreasonable, providing empirical evidence for the introduction of a semi-parametric improvement method in this paper.

2.3 Principle of risk-neutral pricing

The risk-neutral pricing is one of the core concepts in option pricing theory. This concept was proposed by Cox and Ross in the context of no arbitrage in 1976.This principle states that in a risk-neutral world, the expected rate of return on all assets is equal to the risk-free interest rate, and investors no longer demand any risk premium. Under this setting, the current value of an option is equal to the discounted value of its future expected discounted value. This is exactly the theoretical basis for the pricing formula of the binary tree model. So, one has a formula to express at the present moment, the stock price S_0 is equal to the discounted value of its expected future price in a risk-neutral world. If this formula holds, it means that the p people defined is indeed the risk-neutral probability—under this probability, the expected rate of return on the stock is exactly equal to the risk-free interest rate

$$S_0 = e^{-r\Delta t}\left[p\left(S_0 u\right) + \left(1-p\right)\left(S_0 d\right)\right] \tag{4}$$

Solving this equation, it is found that

$$p = \frac{e^{r\Delta t} - d}{u-d} \tag{5}$$

The P derived from this equation is exactly the same as the P derived under no arbitrage conditions. The significance of risk-neutral pricing lies in the fact that it makes option pricing no dependent on investors' risk preferences and real probability. Instead, it allows for the pricing of options by constructing risk-neutral probabilities through no-arbitrage conditions. This principle provides a rigorous theoretical for the binomial tree model. One can use the expected discount formula to set the price with confidence [8].

2.4 Summary of model assumptions

According to the previous content, one can understand clearly that the traditional binomial tree base on the several relatively important assumption: The first is no arbitrage opportunities. This assumption is the base of financial pricing, it can keep all products in a balanced state. The second is Constant volatility. The traditional assumption is that σ remains constant throughout the option's period. The third is discrete time: Asset prices change only at discrete time intervals. Among the assumptions, the constant volatility is the most problem with actual market conditions. As describes in 2.2 above, in real markets, implied volatility varies with time and other data changes. It exhibits a volatility smile pattern. This disparity represents the core issue that this article aims to address. By introducing a semi-parametric approach, volatility can reflect market structural characteristics, thereby improving the traditional binomial tree option pricing model [9].

3. Experimental Design and Research

The data source of this article is based on the option data which one gets from the Shanghai Stock Exchange (SSE). The author collects option prices from the SSE website. The sample focuses on SSE 50 ETF call options traded on March 20, 2026. The underlying asset is the SSE 50 ETF, with a closing price of RMB 2.957 on that day. These options expire in March 2026, and the author chooses five strike prices between RMB 2.800 and RMB 3.000. This range includes options that are near the current asset price, which helps us observe how option prices changes.

For each option contract, the author collects the following basic parameter in Table 1.

ISSN 2959-6157

Table 1. Brief introduction of parameters

Parameter	Symbol	Description
Asset price	S	Closing price of the SSE 50 ETF on the selected day
Strike price	K	The price at which the option can be exercised
Market price	C	Observed closing price of the option
Time to maturity	T	Remaining days until expiration, converted to years
Risk-free rate	r	Government bond yield with a similar maturity

Using these parameters, the author computes the important variable: moneyness, defined as the ratio of asset price to strike price, which indicates whether an option is in-the-money, at-the-money, or out-of-the-money. Table 2 presents the sample data of Shanghai Stock Exchange 50ETF call options, obtained from the Shanghai Stock Exchange on March 20, 2026; however, obtaining a small sample size may cause errors. Note that market implied volatility is calculated using the Black-Scholes model.

Table 2. The values of each parameter

Strike price K(yuan)	Market price C(yuan)	Moneyness S/K	Implied volatility σ(%)
2.800	0.1609	1.0561	18.2
2.850	0.0115	1.0375	21.5
2.900	0.0640	1.0197	19.8
2.950	0.0275	1.0024	17.3
3.000	0.0071	0.9857	16.1

The author uses BS model to compute the real market data, after that one got the "market-implied volatility" (as the Y-value) for each option. Use linear regression to find the relationship between "market-implied volatility" and "moneyness (S/K)". Apply this relationship to the binomial tree--each node calculates the volatility using the regression formula based on its S/K. To estimate the implied volatility for each node in the binomial tree, the author specifies the following regression model:

$$\sigma_{\text{implied}}=\beta_0+\beta_1(\text{S/K})+\epsilon \tag{6}$$

This model can show the linear dependence of implied volatility on the moneyness ratio. These two coefficients β_0 and β_1 will be used to predict volatility for each node in the semi-parametric binomial tree. The Regression formula is given by

$$Y=-26.0390+43.6585\left(S/K\right) \tag{7}$$

In this formula, one can see the β_0 value is -26.0390 and the β_1 is 43.6585. The R^2 is 0.348.

So now the author uses the traditional binomial tree option pricing model to run again and compare the results for error comparison, see Table 3. The author first computes the theoretical option prices using the traditional binomial model (with constant volatility) and the semi-parametric binomial model (with volatility predicted by the regression). The results are then compared against market prices.

Table 3. The values of each parameter

Strike Price K (Yuan)	Market Price (Yuan)	Traditional Binomial Tree Model (Yuan)	Semi-parametric Binomial Tree Model (Yuan)
2.800	0.1609	0.1690	0.1725
2.850	0.0115	0.1157	0.1074
2.900	0.0640	0.0681	0.0612
2.950	0.0275	0.0280	0.0271
3.000	0.0071	0.0058	0.0056

So now one uses the traditional binary tree option pricing model to run again and compare the results for error comparison. The mean absolute error (MAE) and root mean squared error (RMSE) are important bases for judgment. After calculation, one can know that the values of MAE and RMSE of the traditional model are 0.0272 and 0.0419 respectively. The two values of the semi-parametric model are 0.0223 and 0.352 respectively. From them, one can see that they are indeed different. The performance of the semi-parametric model reflected on these two data is better than that of the traditional model. The mean absolute error (MAE) has been reduced by 18%, and the root mean squared error (RMSE) has been reduced by 16%, which shows that the volatility structure of market banks can be estimated through linear regression to improve the accuracy of pricing [10].

Conclusion

This paper mainly solves the problem that the volatility assumption is constant which is different with the actually market. In this study, linear regression is used to estimate the relationship between implied volatility and moneyness. Then the author adds the structure of volatility to the original binomial tree model. So, one gets a semi-parametric pricing model. Based on empirical data from Shanghai 50ETF call options, the results show that linear regression can effectively characterize the connection between volatility and moneyness. People can see that it has a positive relationship and R^2 is 0.348. which explains roughly 35% of volatility variations. At the same time, the semi-parametric model significantly improves the accuracy of pricing. Compared with the traditional binomial tree model, the MAE of the proposed model is 0.0223 and the RMSE is 0.0352, representing an 18% reduction in MAE and a 16% reduction in RMSE. Experimental results confirm that adding a volatility structure that based on market data to the binomial tree model can effectively reduce pricing errors and demonstrate the effectiveness of the semi-parametric improvement method proposed in this paper.

In terms of innovation, this study integrates linear regression into the binomial tree to better fit real market volatility and validates the model using Chinese 50ETF option data while retaining the simplicity. In the meantime, it retains the flexibility and intuitiveness of the traditional binary tree model. However, the small sample size the author selected seriously led to the error of people's experiment and the non-universality of his conclusion. Future research may further explore the convergence properties of the model and compare its efficiency with other improved binomial tree structures.

References

[1] Cox J C, Ross S A, Rubinstein M. Option pricing: A simplified approach. Journal of Financial Economics, 1979, 7(3): 229-263.

[2] Li J X. Exploring the limitations of option pricing models. Economic Horizons, 2023(3): 44-52.

[3] He Y. Developing an American option binary tree pricing model using EXCEL2007. Accounting Monthly Journal, 2012(36): 57-59.

[4] Lu L J, Hu Y J. Pricing of American put options based on a fuzzy binary tree model. Journal of Beijing University of Chemical Technology (Natural Sciences Edition), 2012, 39(3): 114-118.

[5] Ren F, Liu L. Binary tree option pricing model with q-Gaussian distribution of return and empirical analysis. Journal of Henan Science, 2024, 42(8): 1093-1101.

[6] Cui J, Yin X, Huo Q. Valuation of convertible bonds: Application of binomial trees. China Appraisal Journal, 2022(1): 42-50.

[7] Muroi Y, Suda S. Binomial tree method for American option pricing: Discrete cosine transform approach. Computational Economics, 2025.

[8] Zhao C, Zhang F. Internet enterprise data asset value assessment: Based on binary tree option pricing model. China Appraisal Journal, 2023(9): 51-60.

[9] Lian Y Y, Zhang T. Construction of a novel binary tree parameter model for option pricing. Mathematical Practice and Understanding, 2010, 40(2): 15-19.

[10] Josheski D, Apostolov M. A review of the binomial and trinomial models for option pricing and their convergence to the Black-Scholes model determined option prices. Econometrics. Advances in Applied Data Analysis, 2020, 24(2): 53-85.