A Binomial Tree-Based Empirical Study on the Biases of American Call Option

Yiyuan Miao*

Department of Foreign Languages and Literatures, Tsinghua University, Beijing, China

* Corresponding Author Email: miaoyy21@mails.tsinghua.edu.cn

Abstract. The application of financial derivatives, particularly options, has become pervasive in modern finance, offering effective tools for risk management. Among these, American options, allowing flexibility in exercising rights prior to expiration, dominate the derivatives market. Accurate pricing of American options is crucial for informed investment decisions and risk assessments. While various pricing models exist, the binomial tree model is welcomed for its simplicity and accessibility. However, inherent biases in pricing models raise questions about their efficacy across different sectors of the market. Based on the binomial tree model, this study empirically examines the pricing accuracy of American call options in both the technology sector and the industrial sector, by collecting data from top five largest companies in each sector. For the technology sector, samples are mainly from semiconductor manufacturers, software companies, etc., which are all situated in thriving fields; while those five companies from the industrial sector mainly belong to traditional industries, such as energy and engine suppliers. By analyzing prediction biases between two sectors, this research tries to give out reasons behind it and grasps a deeper understanding of option pricing nuances between sectors, and also the limitations of the binomial tree model.

Keywords: Option pricing, American option, binomial tree, pricing bias.

1. Introduction

The application of financial derivatives can be dated back to approximately 50 years ago, when companies started to hedge against insurable and uninsurable risk that unprecedentedly increased after the collapse of Bretton Woods [1]. After several decades' development, derivatives have been widely used around the world and evidence have shown that they tend to enhance firm value conspicuously [2]. Among various financial derivatives, options, mostly American options rather than European options in today's market, stand out as effective instruments that attract both individual investors and the broader societal economic structure. Whereas European options must be exercised at a precise date, these American ones endow buyers right to buy or sell an underlying asset at a specified price at any time before a certain date, thus providing a highly powerful tool for hedging, speculation, and other kinds of financial managements. This flexibility, catering to a wide array of strategic needs and risk profiles, makes American options the majority of the market of the derivatives [3].

To price these options as accurate as possible is essential to their effective use, which directly influences investment decisions, risk assessments, and the overall efficiency of financial markets. Although formulas might be devised to predict the price of European options, American options should be priced usually numerically, since they tend to be exercised in advance with no fixed time [4]. Among the various methodologies developed, binomial tree model is known for its accessibility and adaptability, which is firstly proposed by Cox et al., 1979 [5]. This method, by breaking down the option's life into finite discrete time spans and postulating the prices on every node, demonstrates a clear framework of option pricing, making it one of the most basic and commonly used approaches in the financial industry.

Despite its widespread application and fundamental role in the valuation of American options, binomial tree model is not perfect. Every pricing method may lead to biases, since the market is manipulated by people who have distinct judgement, experiences, and interests, and might be subjective. Some previous research and empirical studies examine the accordance between the pricing outcomes generated by this model and the actual market prices, such as an essay which takes Alibaba Stock as an example [6]. Moreover, lots of studies focus on how to improve the efficacy of this model. For example, in Hu and Cao's work, a randomized binomial tree is developed, in which the stock prices' going up or down is no longer depicted by fixed percentiles, but rather random variables [7]; also, many studies are fixated on implied binomial trees [8-9].

Nonetheless, seldom studies probe into the difference of prediction errors between different sectors of the stock market when using the same model. A related empirical study has been done by Jiang and Feng, which is placed in Chinese ETF market and examines the factors influencing the biases caused by the binomial tree, but still, it doesn't mention the biases of different sectors [10].

With rapid blooming of the technology sector and the eclipse of the industrial sector, the conditions of their stock options are doomed to be dramatically distinct. Thus, it's of high significance to examine whether binomial tree method has different efficacy and accuracy in pricing those options, whose underlying assets belonging to these two sectors, prompting a need for investigation.

Based on the traditional binomial tree model, this essay embarks on an empirical study to price 10 American call options, half of which coming from technology sector and the other half industrial sector, with a focus on prediction biases of two distinct sectors. By scrutinizing the accuracy of this pricing method across different market sectors, this essay aims to shed light on the nuances of option pricing and grasp a deeper understanding of the binomial tree model's strengths and limitations.

2. Pricing Method and Sample Collection

2.1. Binomial Tree Method

The binomial tree model assumes that the price of the underlying asset can only rise or fall in each time interval. Suppose that the starting stock price is S0, the probability of rising is p and the probability of falling is (1-p), the price after rising is S0u, and the price after falling is S0d, where u=1/d, with u and d representing the rising factor and the descending factor respectively. The probability of rising, p, is a calculated from u, d, days to maturity and risk-free rate, by a certain formula. After m times of simulations (that is the number of intervals we divided the whole period into), we can simulate m+1 different ST at the maturity date, and then reverse to derive the option value at the beginning of the period based on the option value at the maturity date (See Fig. 1).

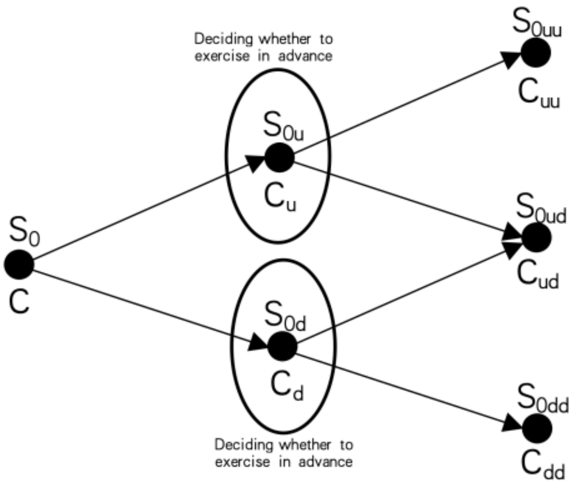

Fig. 1 The process of a two-step binomial tree

However, American option pricing has an additional process, in that they tend to be exercised in advance due to investors' judgement that they might gain more by doing so. The specific steps are listed below:

Calculate the stock price on every node from the beginning to the end;

Subtract each of the ending stock prices by K (the strike price), and the ending option value should be Max(ST-K,0);

Start from the last strata to calculate the theoretical option prices of the previous nodes by using the weight p, which is, precisely, to multiply the option value derived from the rising the stock price by p and multiply the option value of the decreased stock price by (1-p), and then add them together; compare them with the option prices one would get by directly exercising on that node, which should be Max(ST-1-K, 0); if the latter value is larger than the former, then take the latter as the actual option price on that node, otherwise the former;

Repeat step 3 until gaining the option price of the first node.

These steps are done by codes, especially when the number of intervals is large. This essay applies Python to construct the process.

2.2. Option Selection

This study collects options of the companies with top-5 market weight in either technology sector or industrial sector, so there are in all 10 stock options analyzed. Since they rank top in each of their sectors and have absorbed most of investors' concentrations, it is reasonable to state that they are typical enough to represent the conditions of options in their own sectors. The ten stocks are attached below in Table 1.

Table 1. Sample companies chosen

Technology Sector	Industrial Sector
MSFT: Microsoft Corporation	GE: General Electric Company
AAPL: Apple Inc.	CAT: Caterpillar Inc.
NVDA: NVIDIA Corporation	UNP: Union Pacific Corporation
AVGO: Broadcom Inc.	UPS: United Parcel Service, Inc.
ORCL: Oracle Corporation	HON: Honeywell International Inc.

With current date set on March 12, 2024, this essay aggregates from nasdaq.com all of the call option contracts maturing on March 15, March 22 and March 28, 2024, respectively, to gain a set of panel data consisting of over 650 rows.

2.3. Parameters

To realize the binomial tree calculation process in Python, some parameters need to be stated in clarity. "m", the number of iterations, should be a relatively large number, since the more iterations, the higher the accuracy. It is valued 100 here, which means that time to mature is divided into 100 intervals and there will be 100 layers in the "tree". "S0" stands for the stock prices at the start of the time span, which, in this essay, is the closing prices on March 12. "T" represents days to maturity, which could only be 3, 10 and 16, since there are only 3 distinct maturing date, March 15, March 22 and March 28, among data collected.

The volatility rate, "sigma", is represented by the historical volatility of the stock, calculated from daily closing prices in the previous year. For each day, calculate the natural logarithm of the ratio of the stock price at the end of that day to that at the end of the previous day; and then, calculate the standard deviation of these logarithmic values and multiply it by the square root of the number of trading days in a year, which is usually 250:

$$\mathrm{STD}[\ln(\frac{S_t}{S_{t-1}})] * \sqrt{250} \tag{1}$$

The resulting value is the historical volatility. This study calculates volatility belonging to 10 stocks respectively from the historical quotes downloaded from nasdaq.com. It can be figured out from the following Table 2 of volatility that the technology sector is more unstable, consistent with the current fever of investors in that field, technology companies' intense internal competition and unprecedented market dynamism.

Table 2. Volatility of ten stocks

Volatility (Technology)					Volatility (Industrial)				
MSDT	AAPL	NVDA	AVGO	ORCL	GE	CAT	UNP	UPS	HON
0.23	0.19	0.45	0.34	0.33	0.22	0.22	0.19	0.25	0.17

The strike price, commonly denoted as “K”, is also an essential parameter applied. The risk-free rate “r” is collected from the return rate of the American Treasury Bond maturing on March 11, which is 0.0409. Finally, “b” is the holding cost equal to r, since it is assumed that during such short period of time to mature all those companies would not give out dividends. After aggregating m, S0, T, sigma, K, r and b, feed them into Python so that pricing results will be generated.

3. Regression Design

The difference between option market prices and the pricing results, biases, may be influenced by lots of factors, which can be roughly divided into 2 categories: the attributes of the option contract itself and the characteristics of its underlying asset.

Option value is basically separated into 2 sections: intrinsic value and time value. When it comes to the accuracy of option pricing, the most unstable part baffling scholars for decades is the time value of option. The time value of the option is related to the remaining term of the option and the historical volatility of the stock price: the farther the date to maturity, the greater the price change and the higher the time value of the option; the higher the volatility of stocks, the higher the time value of options. The certainty is quite expanded, since American call options might be exercised on an uncertain date largely at the discretion of the investors. These factors extend the possible range of future option prices, which might contribute to the inaccuracy of pricing. Meanwhile, the specific strike price of the option under examination might also be taken into consideration, in that for a call option, the higher the strike price, the lower the option price, which might affect the denominator of the bias rate. Furthermore, which sector its underlying stock belongs to might also have impact on investor’s judgement of the option’s prospect. Thus, two attributes of the option contract itself (days to mature and strike price), and two attributes of the the underlying stock (stock volatility and sector) are embraced in the construction of the model.

To examine whether the type of sectors affect the rate of bias, two equations are structured as below:

$$\mathrm{ABSDIFFRATE} = \beta_0 + \beta_1\mathrm{VOLA} + \beta_2\mathrm{DATETOMATURE} + \beta_3\mathrm{DISTOHIGH} \tag{2}$$

$$\begin{aligned}\mathrm{ABSDIFFRATE} &\\ &= \beta_0 + \beta_1\mathrm{VOLA} + \beta_2\mathrm{DATETOMATURE} + \beta_3\mathrm{DISTOHIGH} + \beta_4\mathrm{TYPE} \\ &+ \beta_5\mathrm{TYPE} * \mathrm{VOLA} + \beta_6\mathrm{TYPE} * \mathrm{DATETOMATURE} + \beta_7\mathrm{TYPE} \\ &* \mathrm{DISTOHIGH}\end{aligned} \tag{3}$$

The clarifications of parameters are listed below in Table 3.

Table 3. Illustrations of parameters

Parameter Name	Illustration
ABSDIFFRATE	Absolute value of the bias rate (binomial-tree pricing minus actual market price, timing 100, divided by actual market price)
VOLA	Stock volatility, estimated by last year’s historical volatility
DATETOMATURE	The number of days to maturity
DISTOHIGH	The difference between the specific strike price and the maximum strike price of that stock option, divided by the maximum strike price
TYPE	Dummy variable, 0 for tech company and 1 for traditional industrial company

If there are significant differences between technology and industrial sector, there must be at least one non-zero coefficient among β_4 to β_7. Thus,the experiment process of this study is to firstly regress both models to prove their significance and reliability, and then to do an F-test on coefficients β_4 to β_7, with the hypotheses that:

H0: $\beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

H1: At least one of β_4, β_5, β_6 and β_7 is not 0.

4. Results and Analysis

4.1. Descriptive Statistics

According to the summaries by Stata, there are in general 664 samples collected from a total of 10 companies, with the absolute difference rate ranging from 0.028% to 109.101%. Those variable names starting with “type...” stands for the product of type and the corresponding variable. There are 458 samples from tech sector and 206 from industrial sector, and both their bias rate and absolute bias rate do not differ conspicuously (See Table 4).

Table 4. Descriptive Statistics

Variable	Obs	Mean	Std. dev.	Min	Max
DIFFRATE	664	-23.42267	38.71167	-99.99996	109.1005
ABSDIFFRATE	664	34.18787	29.62211	0.0284949	109.1005
DATETOMATURE	664	9.051205	5.211696	3	16
DISTOHIGH	664	0.0843537	0.0509938	0	0.1795775
VOLA	664	0.300512	0.0903273	0.17	0.45
TYPE	664	0.310241	0.4629409	0	1
TYPEDISTOHIGH	664	0.0262529	0.048309	0	0.1780822
TYPEDATETOMATURE	664	2.593373	4.783741	0	16
TYPEVOLA	664	0.0660392	0.0996667	0	0.25
DIFFRATE (type=0)	458	-24.58909	37.06157	-99.99887	74.89545
DIFFRATE (type=1)	206	-20.82938	42.12884	-99.99996	109.1005
ABSDIFFRATE (type=0)	458	34.85519	27.60369	0.1585261	99.99887
ABSDIFFRATE (type=1)	206	32.70422	33.70512	0.0284949	109.1005

4.2. Regression Result

Both two models give out satisfying results, with R^2 of model (1) being 0.5392 and R^2 of model (2) being 0.5761, implying that the models roughly fit the curve, with the latter having a higher degree of goodness of fit. The p value (Prob>F) is less than 0.05 in both models (0.000), indicating that the model is significant at the significance level of 0.05, which, in other words, means that the coefficients of the variables in the model, looked from the overall, have significant influence on the dependent variable (See Table 5).

Table 5. Regression results

Model (1)			Model (2)		
R-squared = 0.5392			R-squared = 0.5761		
Prob > F = 0.0000			Prob > F = 0.0000		
Variable	Coefficient	P>|t|	Variable	Coefficient	P>|t|
DATETOMATURE	-0.7765839	0.000	DATETOMATURE	-0.2894331	0.094
DISTOHIGH	-415.9337	0.000	DISTOHIGH	-410.0409	0.000
VOLA	75.48558	0.000	VOLA	116.6441	0.000
/	/	/	TYPE	38.99619	0.001
/	/	/	TYPEDISTOHIGH	-44.66776	0.174
/	/	/	TYPEDATETOMATURE	-1.536048	0.000
/	/	/	TYPEVOLA	-46.21042	0.386
_cons	53.61812	0.000	_cons	32.45279	0.000

4.3. F-test

Both models have the same dependent variable and are estimated on the same sample. Model (2) is called "unrestricted model" while model (1) is called "restricted", since it restricts the number of variables, making itself a special case of the unrestricted model in which some of the coefficients are constrained to be zero. Under null hypothesis " $\beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$", we have:

$$\mathrm{f} = \frac{(R_U^2 - R_R^2)/q}{(1 - R_U^2)/DF_U} \sim F(q, DF_U) \tag{4}$$

in which R_U^2 and R_R^2 stand for the R-square of unrestricted and restricted model respectively; q is the number of restrictions, that is, the number of coefficients deemed as zero in model (1), which is 4; DF_U stands for the degree of freedom of the unrestricted model, which is the number of observations minus the number of coefficients (including the constant term), equaling to 656. Plugging in those numbers so that f equals to 14.28. From the F-table, one can get F (4, 656) = F (4, ∞) =1.94, which is far less than f, showing that the F-test is significant, and the null hypothesis should be rejected. Thus, at least one of β_4, β_5, β_6 and β_7 is not zero, strongly suggesting that model (2) is a necessary improvement against model (1). Since those four coefficients are of four cross terms, each of them multiplied by the dummy variable "type", this test strongly supports the idea that the some of the coefficients gained from regressing on one of the sectors alone is distinct between sectors.

4.4. Analysis

From the regression results of those two models, it can be clearly figured out that for all samples, days to mature and difference to highest strike price negatively correlates with the absolute bias rate, and stock volatility positively correlates with it. It's easy to understand the latter, since the higher the stock volatility is, the more uncertainty there is of people's transactions on options due to difficulties to make precise judgement, leading to the inaccuracy of model prediction that situates in a theoretically perfect world where everyone behaves perfectly on managing their American call options. As for differences to the highest strike price, it's reasonable that with the increase of strike prices (the decrease of distances to the highest strike price) and the decrease of the intrinsic value of options, the denominator of the dependent variable, the actual market option price, is prone to be reduced, thus rendering the rate of bias to be larger. Another explanation might be that for call options, investors are mostly inclined to lower strike prices and very few people buy those contracts that are of extremely high strike prices, making those contracts less fluid and thus there tend to be some irrational transactions which are significantly different from the ideal estimation of binomial tree models. Nevertheless, it's counter-intuitive that the farther the maturity date is, the lower the bias becomes, indicating that the binomial tree model is more suitable for estimating options with farther maturity dates, which is left to be further proved.

Among those four extra terms existing in model (2), two are significant on the level of 0.001: "TYPE" and "TYPEDATETOMATRUE". TYPE, as a dummy variable regarding technology as 0 and industrial as 1, is clear evidence that industrial sector's option prices are more inaccurately predicted than those of technology sector, with its coefficient as high as 38.99619. Moreover, the maturity dates' negative effect on biases is more significant for industrial-sector options. The nuance between predictions on industrial-sector and technology-sector options might be attributed to the fact that the inner distinctions between those two sectors. The flourishing of technology enterprises underscores the promising prospect of this sector, making its options more welcomed by investors, and as more investors pile in, the regulating role of the "invisible hand" of the market becomes more obvious. Through frequent trading and gaming, the market prices of those options gradually get close to the theoretical prices estimated by the model. On the contrary, there are fewer investors in the market of industrial-sector options, so there might be fewer transaction contracts (this is obvious when collecting samples). Furthermore, there might be in this market a larger proportion of large-size fund companies and venture capitals, etc., who tend to have certain degree of control against the option prices, making prices harder to predict simply by the traditional binomial tree model.

5. Conclusion

This study mainly focuses on the factors affecting the biases of American call option pricing based on traditional binomial tree model by collecting samples of top 5 largest company's stock options in both technology sector and industrial sector. In all 10 company's stock options are collected into the data table with a total of 664 samples. They are analyzed and regressed into two models, along with an F-test, which finally generates the conclusion that there are significant differences in biases of technology-sector and industrial-sector option pricing. In other words, pricing technology-sector call options is of higher accuracy than pricing industrial-sector ones, the reasons behind which are left to be further investigated in the future. Moreover, the intrinsic biases of the binomial tree method have been proved to be difficult to thoroughly avoid due to the highly randomized behavior of market transactions. Thus, it's necessary to ameliorate the traditional model and adapt it to the highly dynamic and volatile modern American option market.

References

[1] Till, H. The History of Financial Derivatives: A 2-Part Feature - Part 1: The Emergence and Development of Financial Derivatives Post-Bretton Woods. Finance Educator: Courses, 2014.

[2] Bartram, S. M., Brown, G. W., Fehle, F. R. International evidence on financial derivatives usage. Financial management, 2009, 38(1): 185-206.

[3] AitSahlia, F., Carr, P. American Options: A Comparison. Numerical methods in finance, 1997, 13: 67.

[4] Toivanen, J. Numerical valuation of European and American options under Kou's jump-diffusion model. SIAM Journal on Scientific Computing, 2008, 30(4): 1949-1970.

[5] Cox, J., Ross, S.A., Rubinstein, M. Option pricing: a simplified approach. Journal of Financial Economics, 1979, 7: 229-263.

[6] Yang, X., Fei, Y. An Empirical Study on Option Pricing Based on binary Tree and Black-scholes Model —Taking Alibaba Stock as an example. Journal of Hunan Industry Polytechnic, 2019, 19(03): 36-39.

[7] Xiaoping, H., Jie, C. Randomized binomial tree and pricing of American-style options. Mathematical Problems in Engineering, 2014.

[8] Rubinstein, Mark. Implied binomial trees. The journal of finance, 1994, 49(3): 771-818.

[9] Wang, Tianyang, James S. Dyer. Valuing multifactor real options using an implied binomial tree. Decision Analysis, 2010, 7(2): 185-195.

[10] Jiang, F., Feng, L. Pricing of Convertible Bonds based on Binary Tree Model: Analysis of Influencing Factors of Pricing Deviation. Research of Finance and Education, 2021, 34(01): 21-30.