The Black-Scholes paper: a personal perspective

Anthony Neuberger

Bayes Business School, City, University of London, 106, Bunhill Row, London, EC1Y 8TZ, United Kingdom.

Contributing authors: anthony.neuberger.1@city.ac.uk;

Abstract

This is a personal assessment of the intellectual contribution of the Black-Scholes model of option pricing. I argue that the real contribution of the paper is to show that European options can be replicated exactly if the future variability of the path of transaction prices is known. The continuous rebalancing and the probabilistic setting of the original paper mask this insight.

Keywords: option pricing, dynamic hedging, realized variance

1 Introduction

It is fifty years since Fischer Black and Myron Scholes published the paper (Black and Scholes, 1973) that revolutionized modern finance, prompting numerous conferences and publications. This note is a personal reflection on the theoretical contribution of their paper. It does not attempt to cover the impact of the paper on academics and practitioners; for a brief and accessible account of this the reader may want to look at MacKenzie (2023).

But first an issue of attribution. As Black himself makes clear in his account of how they came up with the Black-Scholes formula (Black, 1989), Robert Merton also made a substantial contribution to the work, and indeed published his own paper on rational option pricing in the same year (Merton, 1973). In recognition of their work on option pricing, Scholes and Merton shared the Nobel prize for economics in 1997. Fischer Black would doubtless have won it with them but for having died in 1995. While some people rightly refer to the model as the Black-Scholes-Merton model, I will follow the common usage (popularised indeed by Merton himself) of referring to the model as Black-Scholes.

It was just ten years after the famous paper was published that I first heard of Black and Scholes. I was doing an MBA at London Business School. It was my first exposure to financial economics. I have been obsessed with their paper ever since. As with any theory that has radically transformed a field of thought, it is different things to different people: "a pioneering formula for the valuation of stock options . . . [it] has also generated new types of financial instruments and facilitated more efficient risk management in society" (as in the Nobel prize citation), or "defective and unscientific" and its use by traders a "myth" (Haug and Taleb, 2011).

My own questions about it ranged from the severely practical (should the interest rate used in the formula be the short-term rate or the rate for the maturity of the option? in calculating historic volatility should one work with returns or log returns, actual returns or demeaned returns?) to the fundamental (if the replication strategy works path by path, how could it derive from an assumption about the probability distribution of paths?).

This is an account of my own grappling with the paper. I start with the formula itself and its meaning, and go on to examine the dynamic arbitrage strategy that underpins it. Dynamic arbitrage is radically different from the static arbitrages that has long been familiar to practitioners. The theory of continuous time stochastic processes which underpins dynamic arbitrage was entirely new to me. It is subtle and beautiful, but I found it hard to believe that the elaborate machinery of equivalent martingale measures was really necessary to deliver the option pricing results. I came to believe that the real insights of option pricing are independent of probability theory, and the function of the machinery is to disguise the assumption on which the Black-Scholes result is based.

While my take on the paper is no doubt idiosyncratic, I do not lay claim to originality, and have not attempted to give credit to those who reached similar conclusions well ahead of me. But the people to whom I should give credit are the many students, and in particular practitioners, I have taught. For in looking back at my own intellectual journey I realize how much my thinking has been influenced by teaching option pricing theory and responding to my students' questions.

2 The formula

The Black-Scholes formula, as set out in the original paper, is

$$w(x,t) = xN(d_1) - ce^{r(t-t^*)}N(d_2), \tag{1}$$

where

$$\begin{aligned} d_1 =& \{\ln x/c + (r + v^2/2)(t^* - t)\}/v\sqrt{t^* - t}, \\ d_2 =& \{\ln x/c + (r - v^2/2)(t^* - t)\}/v\sqrt{t^* - t}. \end{aligned} \tag{2}$$

w is the price of the option, x the spot price of the asset, c the strike price, r the risk free interest rate, t the current time, t^* the time of expiry, v the volatility and $N(.)$ the cumulative standard normal distribution.

It can be argued that the formula is not very original; it does not go beyond what a bright graduate student might have come up with at the time. Picture such a student trying to value a call option on a non-dividend paying stock. They would naturally start from the traditional approach to the valuation of risky cash flows, and value the option as the discounted expectation of its pay-off. In algebra, using Black-Scholes' notation,

$$w(x_t, t) = PV(\mathbb{E}_t[w(x_{t^*}, t^*)]) \text{ where } w(x, t^*) = \max\{x - c, 0\}. \tag{3}$$

x_t is the spot price of the asset at the valuation date. x_{t^*} is the (currently unknown) spot price at the option expiration date t^*. At expiration, the value of the option is $x_{t^*} - c$ if the spot price at maturity exceeds the strike price, and zero otherwise. $\mathbb{E}$ is the conventional expectation operator – our student is no probabilist and knows nothing about different measures - and PV is the present value operator.

To go further down this line of argument requires a choice of probability distribution for x_{t^*}. The normal distribution is one possibility but the student, being a bright student, might worry that a normal distribution gives positive probability to negative stock prices, and hence makes the zero strike call on the stock worth more than the stock itself. The lognormal is a better choice. Like the normal, it has two parameters – a growth rate g and a volatility v. Integrating the lognormal is not particularly hard, so the student would have little trouble in writing down the formula for the expected terminal value of the option.

The final step is the choice of discount rate to get the current price of the option. With no very good theory, the student might simply leave the rate d to be chosen by the user.

The student would finish with a model that looks like

$$w(x, t) = xe^{(d-g)(t-t^*)}N(d_1) - ce^{d(t-t^*)}N(d_2), \tag{4}$$

where

$$\begin{aligned} d_1 =& \frac{\ln x/c + (g + v^2/2)(t^* - t)}{v\sqrt{t^* - t}}, \\ d_2 =& \frac{\ln x/c + (g - v^2/2)(t^* - t)}{v\sqrt{t^* - t}}. \end{aligned} \tag{5}$$

This looks very similar to the Black-Scholes formula, but in place of the risk-free rate r, it requires the asset growth rate g and the discount rate d; if $d = g = r$, the two formulae are identical. The student's formula would not only look quite similar to Black Scholes; it would also give option valuations which are quite similar, at least for short-dated options which are not very sensitive to the choice of rates. Our student would deserve a respectable grade for their answer, but not a Nobel Prize.

Doing this kind of thought experiment is obviously open to the accusation that most truths are obvious in retrospect. But in this case, it is not just hindsight. Black and Scholes themselves refer to no fewer than six papers published since 1960 (and there are several earlier papers) which contained similar formulae to theirs, but criticises them because they (like our graduate student's formula) "were not complete, since they all involved one or more arbitrary parameters." From the purely practical perspective of valuing an individual option, the removal of the growth and discount rate parameters and their replacement by the risk free rate is an improvement, but still leaves the much bigger problem of estimating the volatility.

To my mind though, this entire line of criticism of the paper and the formula itself misses the point. What is special and new about the formula is not the right hand side of the equation, but the equal sign. The traditional valuation formula, $w_t = PV\left(\mathbb{E}_t\left[w_{t^*}\right]\right)$ is a near tautology (I say "near" because it does have the falsifiable implication that two securities with identical cash flows trade at the same price). The Black-Scholes valuation formula on the other hand is a valuation based on replication. To make the point in more colloquial language, Black and Scholes show what the price of the option must be, while all their predecessors merely showed what it ought to be from some subjective viewpoint. So it is not the formula but the way that dynamic replication works that is at the heart of the Black-Scholes theory.[1]

3 Simplifying the formula

But before looking at replication it is worth spending a little more time with the algebra to see what intellectual juice one can extract from it.

Even if the formula is not the major contribution of the paper, it is the aspect which immediately appeals to the MBA, to the consultant and to the practitioner. The formula turns the follower of Black and Scholes into a kind of priest, able to give a valuation that is both mysterious and authoritative. At a time in the 1970s when few people could actually use the formula with ease,[2] its very complexity, and its derivation using exotic tools from stochastic calculus made it appear particularly powerful. Academics loved it.

In conjunction with the Miller-Modigliani theorems in corporate finance (1958 and 1963), and the Capital Asset Pricing Model (1961-66), the Black-Scholes Formula gave Business School academics an intellectual dominance over practitioners which was potent in securing funding, students and respect.

I found the vision of academic as priest disturbing. The function of academics is to reveal not to hide, to question not to instruct. The Oxford English Dictionary notes that the word formula is applied "to rules unintelligently or slavishly followed, to fettering conventionalities of usage, to beliefs held or professed out of mere acquiescence in tradition", the very opposite of critical enlightenment. I wanted to unpack, simplify and demystify the formula to better understand the underlying theory.

[1] Harrison and Pliska (1981) put the point rather more elegantly when they say "It can be argued that the important and interesting, feature of the model ... is its completeness, not the fact that it yields the explicit valuation formula ... for call options. ... (In the end, however, it is the explicit calculation which give the subject its vitality.)"

[2] Calculators at the time were quite rudimentary. Most users of the formula had to look up tables of the cumulative normal distribution by hand. More geeky people used nomograms, which are exotic looking charts from which option values could be read off (Dimson, 1977)

By the standards of other equations in finance the formula is both complex and opaque. It has many (six) inputs; it involves a function (N) that has to be approximated numerically and looked up in tables; it is too complex to describe in words; and its structure does not obviously relate to the logic on which it is founded.

The formula can be simplified. Self-evidently, the calendar date is irrelevant to the option price. It must hold in exactly the same way when it was published in 1973 as 2500 years earlier when Thales of Miletus took out his famous call options on olive presses[3] . So we can drop t and t^*, and replace them with the length of time to maturity, so reducing the number of inputs by one.

Less obviously, but much more significantly, the riskless interest rate r can be dropped. It is only needed because Black and Scholes choose to do all their transactions in the spot market and so need a bank account to carry forward cash flows to maturity. But it is much simpler to do all the transactions in the forward market – where all transactions (including the purchase of the call itself) are settled on one date - and dispense with cash.

It is worth at this stage quoting Black and Scholes' seven assumptions as stated in their paper:

(a) The short-term interest rate is known and is constant through time.
(b) The stock price follows a random walk in continuous time with a variance rate proportional to the square of the stock price. Thus the distribution of possible stock prices at the end of any finite interval is lognormal. The variance rate of the return on the stock is constant.
(c) The stock pays no dividends or other distributions.
(d) The option is 'European,' that is, it can only be exercised at maturity.
(e) There are no transaction costs in buying or selling the stock or the option.
(f) It is possible to borrow any fraction of the price of a security to buy it or to hold it, at the short-term interest rate.
(g) There are no penalties to shortselling. A seller who does not own a security will simply accept the price of the security from a buyer, and will agree to settle with the buyer on some future date by paying him an amount equal to the price of the security on that date.

Doing everything in the forward market not only reduces the number of inputs to the formula, it also dispenses with assumptions (a) about the constancy of interest rates and (c) about the absence of dividends. The diffusion assumption (b) needs to be amended to refer to the dynamics of the forward price rather than the spot price. The last three assumptions can be rolled up into the assumption of a frictionless forward market. Assumption (d) is not really an assumption at all. So we see that the theory

[3]The source of the story of Thales and the olive presses – the impoverished philosopher knew from his skills in astronomy that the next olive harvest would be very large, so he speculated by placing deposits on many olive presses - comes from Aristotle's Politics (Book 1, chapter XI in Aristotle and Lane (2016)). Aristotle is writing some 200 years after Thales' death, and describes the story as having been "attributed to him on account of his reputation for wisdom". Although Aristotle describes the transaction as "his financial scheme, which involves a principle of universal application", the principle he is referring to is not in fact the use of options, but rather the exercise of monopoly power. The interpretation of the contract as a call option assumes that Thales, having made the deposit on the olive presses, had no further obligation if he chose not to use them.

is really based on just two assumptions: a frictionless forward market, and a forward price that follows a diffusion process with constant volatility.

One possible objection to specializing Black Scholes to forward prices is that in many cases forward markets do not exist. But the original Black-Scholes assumptions (costless shorting, absence of transactions costs, zero dividends, borrowing and lending at the same rate) themselves ensure both that forward contracts can be replicated perfectly and that the dynamics of the synthetic forward price and the spot price are identical. So nothing is lost and much is gained by transposing Black and Scholes from the spot market to the forward market. In the rest of this paper, I will use price to mean the forward price; when I need to refer to the spot price, I will do so explicitly.

The payoff from working in the forward market goes beyond the aesthetic pleasure of simplifying the assumptions and reducing the number of variables in the formula. It makes the wider application of the theory much more obvious and straightforward. The extension of Black-Scholes to options on futures (Black, 1976), to rights to exchange one risky asset for another (Margrabe, 1978) and to currency options (Garman and Kohlhagen, 1983) all become trivial. It becomes self-evident that the right way to price options on commodities (where spot and forward prices are not perfectly correlated) is to start from the dynamics of the forward price. Given that one is forced to make simplifying assumptions about the dynamics of prices in order to price options, much better to do it about the object of interest (the forward price) than to derive the dynamics by making more complex assumptions about the joint dynamics of spot prices, storage costs and convenience yields.

I learnt much of this from lecturing. It would have been some time in the late 1980s when I was teaching a group of practitioners about option pricing. I remember (with some embarrassment) fervently denouncing the common practice of using Black-Scholes to value fixed income and interest rate options. It is logically inconsistent, I argued to the class, to use a model that assumes constant interest rates to value options on bonds whose volatility derives entirely from unexpected variation in interest rates. It is absurd, I would continue, to use a Brownian diffusion process to model the price of a default-free bond, when the bond's future price at maturity is known with certainty. It makes no sense, I would conclude, to treat the interest rate in a cap contract (where the writer of the cap pays the other party the difference between the market rate and some fixed rate, if positive) as if it is a price. My arguments fell on deaf ears; while the practitioners were prepared to accept that I might be right, they would continue to use this discredited methodology which served them well in practice. "I am sure you are right in theory, Professor, but . . . ".

Their scepticism made me rethink my own ideas, and understand why the inconsistencies between theory and practice that I had identified were illusory – the constancy of interest rates is irrelevant to Black-Scholes, and it is only the dynamics of the forward price that matter.

Having got rid of the calendar time t^*, and the interest rate r, it is worth noting that the time to maturity now only occurs in conjunction with the volatility. The formula can be written in terms of the total variance - the square of the volatility multiplied by the period length – reducing the number of inputs to three. There is a reason for doing this that goes beyond an obsession with reducing the number of variables in the equation. It highlights the role of time in Black-Scholes. With interest rates out of the way, the passage of time is measured only by volatility. The algebra hints that the crucial assumption is not the constancy of volatility, but rather that the total volatility of the price path is known. This turns out to be key to understanding why the theory works.

We can now write a simplified version of the Black-Scholes formula (with refreshed notation) as

$$C = N(d_1)F - N(d_2)K, \tag{6}$$

where

$$\begin{aligned} d_1 =& \frac{\ln F/K + V/2}{\sqrt{V}}, \\ d_2 =& d_1 - \sqrt{V}. \end{aligned} \tag{7}$$

C is the price of the option to buy the asset for an amount K at maturity. F is the price of the asset, V is the total variance (that is the variance of the log return of the asset over the period) and N the cumulative standard normal distribution. Since all prices in this equation are forward prices, there is no need for discounting.

The formula (it is effectively Black's Formula) is still not simple, but it is less cluttered than before. Interpreting it as a static equation, it hints that the option resembles a leveraged position in the asset: the call option is like a holding of $N(d_1)$ units of the first asset financed in part by borrowing $N(d_2)K$. This is a valuable insight when risk managing an option position. But it is the dynamic interpretation which is more significant.

For the valuation formula can be seen as a *recipe* for replicating and hence pricing the option. A trader who starts with wealth C raises further cash by borrowing $N(d_2)K$, and uses the money to buy $N(d_1)$ units of the asset. As time passes, as remaining variance V reduces and the price F of the asset changes, $N(d_1)$, otherwise known as the *delta* of the option, also changes. The recipe requires the trader to trade continually to keep a long position in the asset equal to the current delta, buying the asset as the price goes up, selling as it goes down. If the assumptions of the model hold, this strategy - called delta hedging - will lead to the trader holding a position at maturity t^* whose value exactly matches the call option: whenever the asset is worth more than the strike, so $F_{t^*} > K$ and the call option is worth exercising ("in the money"), the trader is long one unit of the asset, and short K. When the asset is below the strike, and the option is not worth exercising (the call option is "out of the money"), the portfolio is empty. So for a price c the trader can replicate the call option perfectly.

The step from replication to arbitrage based pricing is simple. If the price of the option in the market is C^+ where $C^+ > C$, then the strategy of selling the option and replicating it guarantees a profit of $C^+ - C$. This is an arbitrage. If $C^+ < C$, buying the option and reversing of the delta hedging strategy (swapping buying and selling) locks in a profit of $C - C^+$. To avoid arbitrage possibilities the price of the call option must be exactly C.

4 The replication strategy

But before accepting too readily that Black and Scholes' achievement is pricing by replication, it is important to recognise that their dynamic replication only exists as a theoretical construct and is not a practical reality.

Arbitrage or replication based pricing was of course familiar way before Black and Scholes – traders have long understood triangular arbitrage in the currency markets (trading euro-yen rates against euro-dollar and yen-dollar rates), cash and carry arbitrage (trading the forward euro-yen rate against the the spot rate), and put-call arbitrage in the options market (trading the option to sell a euro for 100 yen - a put option - against the corresponding call option, the right to buy a euro for 100 yen). But these trades – call them traditional arbitrages - are very different from the dynamic strategy in Black and Scholes:

- In a traditional arbitrage, the trade is well specified: the quantities to be bought or sold are well-defined. In Black-Scholes the trading strategy is not a single strategy, but a family of strategies which depend on the level of volatility. The arbitrage only works if you use the right volatility parameter, but the level of volatility is not known to the arbitrageur.
- A traditional arbitrage is practically feasible. It involves a finite number of discrete trades - typically, setting up a portfolio initially and liquidating it at maturity - and trading volumes are bounded. The Black-Scholes strategy requires continuous trading and infinite trading volumes and is impossible to execute.
- In a traditional arbitrage, any shortfall between the theoretical profit from an arbitrage and the profits actually achieved can readily be quantified and explained. All trades carry risk. Actual transaction prices differ from quoted prices; fees and commissions need to be paid; counter-parties default. These risks mean that arbitrageurs rarely capture the full difference between theoretical value and market price. But in a traditional arbitrage, the shortfall can be attributed to specific causes and properly quantified. In the case of option pricing, Black and Scholes theory provides no obvious way of explaining, and therefore of understanding, any shortfall.

Black-Scholes option replication requires continuous rebalancing. In the presence of any market friction - bid-ask spreads for example - this is impossible since trading volume is unbounded. If the theory is to have any practical application, it cannot be restricted to continuous trading but must also work with discrete rebalancing. But Black Scholes arbitrage with discrete rebalancing is quite problematic as the following thought experiment illustrates.

Consider a world where trading costs in the forward market are negligible, and the price follows a constant volatility diffusion process. The volatility (known to the arbitrageur) is 30%. The fair (Black-Scholes) price of a one year at the money call is $12.00. The call is actually trading at $10.00. The arbitrageur buys the call and delta hedges following the formula, rebalancing every day. At the end of the year, all positions are liquidated and the arbitrageur finds that instead of making a profit of $2, she has actually lost $2. She is understandably upset.

She consults a friendly academic in the hope that he can explain why an arbitrage trade lost money. He asks for her trading records. He calculates the realized volatility

– that is the volatility calculated from the prices she actually traded at - and finds that it was only 20%. The academic concludes that there is no great mystery here; the arbitrageur had thought that volatility would be 30%, the market had believed that it would be 25% and it actually turned out to be 20%. The option she bought was in fact over-priced, not under-priced.

The arbitrageur does not accept this. She knows (divine revelation?) that she was right, and that the data generating process was Brownian with volatility of 30%. The fact that the realized volatility of the observed prices on the path was 20% does not prove that she was wrong. The probability of a process with volatility of 30% generating daily returns over a year with a realized volatility of 20% is small, but it is not zero. She concludes that the theory is sound, her beliefs about volatility were correct, and that the reason she lost money was that she failed to rebalance continuously.

The arbitrageur is of course correct. The prices at which she traded could have been generated by a Brownian process with volatility of 30% (think for example of successive daily prices being linked to each other by Brownian bridges with a volatility of 30%). If that were the case, and had she rebalanced continuously (as the theory provided), she would have made her $2 (ignoring transaction costs).

This thought experiment suggests that the Black Scholes replication strategy is immune to refutation even when it does not work. The strategy never fails because it is never properly implemented, and it is never properly implemented because it requires continuous trading, and that is impossible. On this reading, the replication strategy is just a theoretical construct with limited practical application.

If that were a fair summary, it is hard to see how the paper could have had such an impact on the practice of finance. We need to look at the arbitrage strategy in more detail.

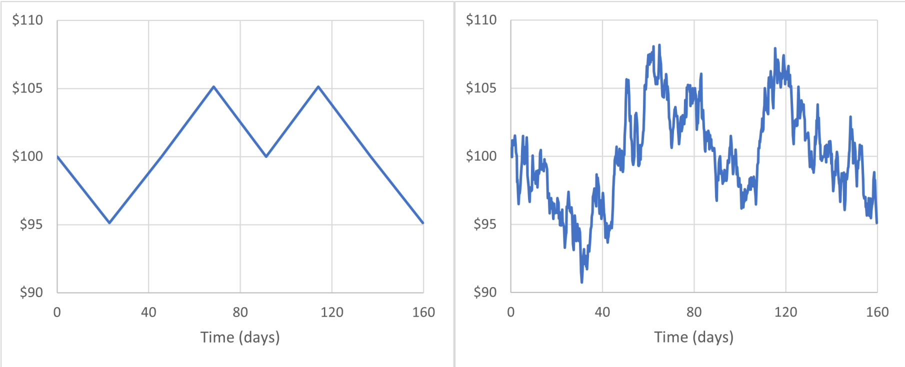
5 The binomial model

The key to understanding why Black and Scholes does work in the real world comes from a somewhat unexpected source – not an article in an academic journal but rather from a chapter in a textbook. It happens to be one written by another Nobel Prize winning economist, Bill Sharpe. In the first edition of Investments (Sharpe, 1978), published in 1978, just five years after the Black and Scholes paper, he sets out the multi-period binomial model in order, as he puts it, "to make concrete some further principles of option valuation". He then describes "the somewhat formidable but highly practical option valuation formula developed by Black and Scholes" but does not attempt to derive it. The binomial model which Sharpe originated and Cox, Ross and Rubinstein developed, is generally seen as a convenient way of explaining the ideas behind the Black and Scholes model in a way that "requires only elementary mathematics" (Cox et al., 1979).

As a model it appears highly artificial. Geometric Brownian paths look like real price paths; binomial paths, with their clumsy sawtooth shape, look like cartoon representations of price paths. Figure 1 illustrates the contrast.[4] Sharpe's stock starts at

[4]There is actually a bit more to Figure 1. The binomial path has an annualized volatility of 30%, while the Brownian path has a volatility of 20%. Despite this, the Brownian path is constructed to coincide with

Fig. 1 Sample price paths under binomial and Geometric Brownian motion



The left hand graph shows a path from a seven period symmetric binomial tree with an absolute log return of 5% at each node. The right hand side shows a path from a Geometric Brownian process with an annualized volatility of 20%, using a mesh of 0.2 days.

$100 and goes up or down 5% each period. At the end of the first period the stock is worth either $95 or $105; it is not allowed to be $101 or $120. This is a handicap for a model that attempts to represent the real world.

The standard interpretation of the binomial model is that is a fairy tale or parable to explain a complex mathematical theory to poets. If students complain that the binomial assumption is artificial, they are told that economic theories must be judged by their predictions not by their assumptions. But this reply is not satisfactory. Option pricing theory is not really economics – there are no agents, no equilibria – it is mathematics. It says that a particular trading strategy will have a particular pay-off. If we do not understand the assumptions, we can not know under what conditions the Black-Scholes replication strategy works.

Furthermore, the insights of the binomial model are rather fragile. When I first started thinking about these issues, I was prepared to go along with modelling the price path as a process in discrete time, knowing that one can always make the time divisions finer and, in the limit, (if limits are taken in the right way) obtain geometric Brownian motion. But I did not like the binomial. I preferred the trinomial. I thought it more pleasing aesthetically. As the binomial tree becomes finer, the size of price change goes to zero, and the probability of a price change per unit time becomes infinite. Most markets have a minimum tick size (for example most stocks on the New York Stock Exchange are quoted to the nearest cent) - and prices move finitely often per unit time. The trinomial can accommodate this. Prices can go up one tick, down one tick or stay the same; as the period length used becomes shorter, the probability of a price change per period becomes smaller, but the jump size remains finite. Whether you share my preferences for the trinomial or not, one might think that it would be harmless to indulge them, but that would be wrong. Arbitrage pricing works in a binomial setting, but it does not work in a trinomial setting - the market is not complete. This immediately arouses suspicion; what am I assenting to when I adopt

the binomial path at the nodes of the tree. It therefore exemplifies the thought experiment at the end of the previous section.

the binomial model? How can I tell whether I am living in a binomial world, where options can be priced by arbitrage, or in a trinomial world where they cannot be?

It turns out that there is a much better way of understanding the binomial model. It does explain when and why option pricing theory works, and it makes the trinomial question irrelevant. The interpretation of the binomial tree as a *recipe* rather than a *process* became clear to me through my teaching.

The binomial model is a gift to teachers. Presenting it to students who have not seen it before has a magic that is rare in teaching financial economics. The teacher sets up the scenario: you, the student, are a trader who has sold a call option on a stock. The stock costs \$100 today. The call gives the holder the right to buy the stock for \$100 in three months' time. Interest rates are zero (fewer pesky calculations) and there are no dividends. The stock goes up or down by 5% each month; it follows the tree shown in panel A. You sold the option for \$3.75 - which is apparently the fair price. In three months you will have to honor the option which may be worth somewhere between nothing and \$15.76.

You need to manage your risk. Fortunately, your grandmother has left you a recipe for doing just that. The recipe is shown in panel B. It sets out the delta - how much of the stock you need to hold - at each node. It fully specifies the trading strategy. How well does the recipe work?

The first thing you need to do is buy 0.51 (as shown in the root node) units of the stock for \$51.00; you use the \$3.75 you received for selling the option, and borrow a further \$47.25 to finance the purchase. A month later panel B shows that you will need to increase the hedge to 0.75 or reduce it to 0.25 depending on whether the stock rises or falls, increasing or reducing your debt accordingly.

Table 1 Binomial replication recipe

Month 0	Month 1	Month 2	Month 3
			\$115.76
		\$110.25	
	\$105.00		\$104.74
\$100.00		\$99.75	
	\$95.00		\$94.76
		\$90.25	
			\$85.74

A. Stock price

Month 0	Month 1	Month 2	Month 3
		1.00	
	0.75		
0.51		0.47	
	0.25		
		0.00	

B. Delta

I tell the students to trust their grandmothers. The strategy is perfect; whatever happens over the next three months, they will be safe. I challenge the students to prove me wrong. They can decide at each node whether the stock goes up or down. As we arrive at each node, we calculate their cash and stock position (to increase the drama I would probably use a four or five period tree). At first the students try to frustrate the strategy by choosing paths which they think will be problematic, but gradually they get convinced that the replication strategy works on every path.

As I used the simulation in class, I would elaborate the story. The student would be a portfolio manager; panel B would be a set of instructions for a trading desk which had responsibility for doing the actual trades. This helpfully emphasises the precommitment to the recipe. I worried about the credibility of the story. The instructions are incomplete: how would/should the desk interpret them if the stock price after a month was \$101 or \$120 rather than \$95 or \$105?

Then I had a moment of revelation. The trading desk does not have a clock or a calendar. The traders observe the price feed. Rather than wait for one month and hope that the price is \$95 or \$105, they trade when the price first hits \$95 or \$105. If the price hits \$105, they then wait until the price first hits \$110.25 or \$99.75 before trading again. Trading in this way, they follow the instructions to the letter, and end up replicating the option perfectly.

This transformed my understanding of the binomial. The binomial is not a Mickey Mouse representation of a price path; it is a representation of a completely practical trading strategy – trade only when the price moves to the next node in the tree. The strategy works exactly as the theory requires; the replication is perfect. By the time the trader reaches the last node, the portfolio consists of a long position in the asset and a short position in the strike if the asset price is above the strike, and is empty if it is below.

6 Realized variance

What could go wrong with the recipe? One problem is jumps. If in the first period the market closes at \$104.90 and then opens at \$105.10, the share purchase required by the recipe would be at a higher than planned price. To avoid the problem, I assume that the price path is continuous.[5]

A more fundamental issue is that the last node of the tree is reached before the call option matures. Suppose for example that the price reaches the last node of the tree after just 2 months. If the price is below the strike, the hedge portfolio is empty, and the trader is short a one month call with a strike of \$100. If the price is above \$100, the hedge portfolio is a long forward contract which, when netted against the short one month call, is a short one month (out of the money) put with a strike of \$100. The trader is left with an unhedged exposure.

It is also possible that the option matures and the price has not reached the last node. In that case too the replication is not perfect, but any error will be in the trader's favour. The trader will be long an out of the money put or call with a strike of \$100 and which expires when the last node is hit.

The replication is imperfect because the maturity of the option is fixed. But now suppose that the option expiry date is not defined as a calendar date, but rather as the time the last node is hit. In that case, the trader can replicate the call perfectly.

[5] Continuity is not quite the right word, though I will continue to use it. Prices that move discretely - tick by tick - present no problem. The problem is jumps, when the price moves several ticks without any opportunity to trade at intermediate prices. It is messy to deal with price processes that have jumps - in a paper I wrote with Mark Britten-Jones (Britten-Jones and Neuberger, 1996) we showed how they can be incorporated in the theory - and our conclusion can crudely be summarized as saying that jumps do matter (replication is imperfect) but they do not matter very much unless the jumps are individually large compared with the overall variance.

By analogy with car rental agreements that charge by time and mileage, I decided to call these options mileage options.

A general mileage call option contract has four parameters (K, u, d, n). K is the strike. The option expires after n periods. The first period starts at the inception of the option. A period ends when the forward price first reaches either uF or dF, where F is the beginning of period price. K, u and d are strictly positive real numbers with $d < 1 < u$, and n is an integer. The mileage option can be replicated perfectly and therefore priced exactly using the binomial model. The only assumptions are price continuity and frictionless markets.

We can simplify matters by restricting attention to symmetric trees where $\ln u = -\ln d = \delta$, where δ is a strictly positive number. The price of a mileage option increases with n and δ. It can readily be shown that the fair price is to all intents and purposes a function of $V^* = n\delta^2$. So the tree can be characterized by the pair (V^*, δ), where the expected time to maturity, and hence the price of the mileage option depends primarily on V^*, and δ measures the granularity of the tree.

It is useful to introduce a bit more notation. Let $t_0, t_1 \ldots$ be the times that the successive nodes of the tree are reached, with t_0 being the root node. For simplicity, restart the clock at the beginning so $t_0 = 0$. Let $F_0, F_1 \ldots$ be the corresponding forward prices. $\tau = t_n$ is the (random) time the last node is reached. V_t is a step function that is initialised at zero at time 0 and which increases by $(\ln F_{j+1}/F_j)^2$ as each node is reached. So V_t is the sum of squared log returns from the inception of the option. Returns are calculated from transaction to transaction, rather than being based, as more conventionally, on prices observed at fixed time intervals. It is natural to refer to V_t as the *realized variance*. Realized variance defined in this way is a property of a path not of a process; it is a well-defined observable quantity, and not an estimate of some parameter; and its value depends on the granularity of the tree.

We can now state what I regard as the fundamental insight of the Black and Scholes paper: in the absence of jumps and market frictions, a European call option can be priced by arbitrage if the realized variance over the life of the option is known.

Note that the concept of arbitrage here is similar to a traditional arbitrage

- the trading strategy is well specified;
- it is practically feasible, as it involves a finite number of discrete trades, and trading volume is bounded;
- any shortfall between the theoretical profit and the profits actually achieved can readily be quantified and explained in terms of price jumps, trading costs, failure of trades, or whatever.

This formulation immediately raises the question: how can one know the realized variance in advance? One answer is: because the option is a mileage option – so the realized variance is known in advance by contractual definition. This is rather unsatisfactory because nobody actually trades mileage options . But I would argue that the interpretation is still useful insofar as mileage options are reasonably closely related to conventional fixed maturity options, and a partial hedge for them.

The other, more standard, answer is that if one knows that the price follows geometric Brownian motion (GBM), then one also knows the realized variance over any fixed horizon. More precisely, given any process for generating price paths, the maturity τ of the mileage option (V^*, δ) is a random variable. If the process is GBM with volatility σ, $\mathbb{E}[\tau] = V^*/\sigma^2$,[6] which is independent of the granularity. τ has sampling error, but the error is proportional to $n^{-1/2}$ and therefore to δ. As δ tends to zero, τ converges to V^*/σ^2 with probability 1 – the mileage option and the time option coincide.

To put the point another way, all the sophisticated infrastructure of continuous time stochastic processes – the semi-martingales, the measure theory, Girsanov, Ito and so on – is being used to hide or justify two propositions: that the price path is continuous, and that the agent knows the quadratic variation of the price path. It is a matter of taste, but it does not seem to me that going from "assuming you know the realized variance" to "assuming the statistical process generating the price path is GBM with known volatility" is a great advance. Indeed, I would go further and say that the first formulation has the advantage that it focuses attention on the issues that matter when applying Black and Scholes.

7 Insights from the mileage option approach

7.1 Implied volatility

The data generating process cannot be known. The mileage option approach highlights the fact that a trader who has sold a conventional option with fixed maturity T for some price C can manage or reduce their risk by using the money received from the sale of the option to replicate a mileage option with uncertain maturity τ, but is left – inevitably – with exposure to realized variance risk as represented by the difference $\tau - T$. The higher C is, the more nodes the trader can have for the mileage option, the greater its maturity.

More specifically, if the trader fixes the granularity parameter δ, then the option price C determines the number of nodes n in the binomial tree, and hence the amount of realized variance V^* the trader can buy. Option traders normally think in terms of volatility - the square root of variance per unit time - so one way of describing the transaction is that the trader sells the conventional option on an *implied volatility* of $\sigma^{imp} = \sqrt{V^*/T}$ and constructs a mileage option which has a *realized volatility* of $\sigma^{real} = \sqrt{V^*/\tau}$. If the realized volatility is greater than the implied volatility, the mileage option expires before the conventional option and the trader is exposed to loss; conversely if the realized volatility is less than the implied volatility, the trader gains.

This definition of implied volatility is slightly difference from the standard Black-Scholes implied volatility, but converges to it as δ tends to zero.

7.2 Conditional volatility forecasts

The trader who sells a conventional option and creates a mileage option is subject to volatility risk. But the risk depends also on where the price ends up. Suppose that

[6] The distribution of τ also depends on the drift of the diffusion process; the simple formulation here is correct when the log price is a martingale - see Karlin and Taylor (1975), ch 7, theorem 5.1; choosing some other drift parameter complicates the algebra but does not change the underlying argument.

realized volatility is higher than implied volatility, so $\tau < T$. Then if $F_\tau < K$ the mileage option pays nothing, and the trader is fully exposed to the conventional option, with payoff $[F_T - K]^+$. If F_τ is far below K then the risk to the trader is small, but if it is close to K, the trader's exposure may be quite large. If $F_\tau > K$ the mileage option pays $F_\tau - K$, and the trader's net exposure is $[K - F_T]^+$. Again, it can be seen that the exposure is greatest when F_τ is close to K.

Similar analysis applies when realized volatility is greater than implied volatility, with the trader's risk exposure being greatest when F_T is close to K, except that now the risk is entirely on the upside.

This analysis suggests the trader will be most concerned about the realized volatility on paths that end up close to the strike. This provides a possible explanation for the variation in implied volatility across exercise prices - the so-called smile - which of course does not exist in a Black-Scholes world.

7.3 Transaction costs

The assumption of zero transaction costs is standard in arbitrage models. It is obviously unrealistic; in general, there is a spread between bid and ask prices. In traditional arbitrage the presence of frictions is not a problem. It is straightforward – if somewhat tedious - to do the analysis using different bid and ask prices rather than assuming a single price. It means that the price of the replicated asset must lie in a range rather than a single number if the market is to be arbitrage free.

As already observed, transaction costs are a serious problem for the Black-Scholes model because the volume of trading is infinite. One advantage of the mileage option approach is that transaction costs can be directly incorporated in the model. Instead of there being just one price for the asset, there are two: a *bid* price at which the asset can be sold and an *ask* price at which the asset can be bought. We assume the bid is no higher than the ask, and both prices are continuous. Now consider the case of a trader replicating a long call position.[7] The trader buys when the price hits an up-node and sells when it hits a down-node. The instructions to the trading desk are modified as follows: if the most recent transaction was done at price F then the portfolio is to be rebalanced when the bid price reaches $Fe^{-\delta}$ or the ask price reaches $Fe^{+\delta}$. Provided that the bid-ask spread is smaller than δ, the instructions are coherent, all transactions take place at node prices and the replication of the mileage option is perfect.

It does not mean that transaction costs have no impact. The effect of transaction costs is to shorten the life of the mileage option. To quantify the effect, assume that the mid price follows a GBM with volatility σ, and that log bid and ask prices are lower or higher than the log mid-price by an amount s. The price change (that is the absolute log return) between successive portfolio rebalancing as measured by the actual transaction cost is still δ, But the change in the mid-price is either δ (when the trade at the beginning and end of the period have the same sign) or $\delta - 2s$ (when a buy is followed by a sell, or vice versa). The average time between successive nodes goes down from δ^2/σ^2 to $\delta(\delta - 2s)/\sigma^2$. The effect of transaction costs is to reduce

[7] The analysis assumes that the claim being replicated is convex. The replication of a concave payoff - for example, a short option position - is discussed later.

the expected maturity of the mileage option from V^*/σ^2 to $(V^*/\sigma^2)(1-2s/\delta)$, which corresponds to increasing the realized volatility from σ to $\sigma/\sqrt{1-2s/\delta}$.

Transaction costs limit how fine a tree the trader will want to use. In a world with no transaction costs, the maturity of a mileage option is subject to two types of risk: volatility risk, because volatility is not known in advance; and sampling risk, because the price path is observed a finite number of times. The sampling risk can be reduced by using a finer tree (lower δ, greater n). In a GBM world there is no volatility risk, so all risk can be removed by using an infinitely fine tree. But in reality volatility risk is substantial (as shown by the difference between implied and realized volatility, for example Christensen and Prabhala (1998)) and the reduction in risk from having a finer tree are likely to be largely exhausted once n exceeds say 30. In the presence of transaction costs, the finer tree has a shorter life, and the extra cost of hedging a conventional option is large when s is significant relative to δ.

The choice of δ therefore reflects a trade-off. The trader selling an option is exposed to uncertainty about the future level of volatility and the delta hedging strategies we have examined can do nothing about volatility risk. But he is also exposed to sampling risk which he can reduce by using a finer tree. The cost of doing this however is that he faces higher transaction costs which include the probability of losing money.

If the trader is buying an underpriced option and hedging by replicating a short position in a mileage option, strategy involves selling when the underlying rises and buying when it falls. The impact of transaction costs is to *reduce* the realized volatility. Thus transaction costs generate a bid-ask spread in volatility that depends both on the level of transaction costs in the underlying and the granularity of the tree used for option replication.

7.4 Short dated options

GBM is self-similar – the process is the same (apart from scale) whatever the horizon. The Black Scholes theory should apply equally whatever the term of the option. But there do seem to be significant difficulties in applying the theory to short-dated options. "A review of the finance literature reveals that options with less than 10 days to maturity are largely ignored (e.g., Bakshi, Cao,and Chen 1997). The apparent reason for this neglect seems to be that the options market becomes structurally unstable for maturities of less than 10 days." (Arnold et al., 2007). The mileage option approach, with its granularity, does provide some insight here.

A crude numerical example may illustrate the point. Take an asset whose volatility is expected to be 20%. Suppose also that the asset has a bid-ask spread of 10 basis points, and that there is a demand for call options on the asset on an implied volatility of 22%. As we have seen, transaction costs increase realized volatility; to give any opportunity for profit, the trader needs to keep realized volatility below 22%. This implies in turn that $\delta \geq 0.6\%$.[8]

To keep sampling error down, the trader wants n, the number of nodes, to be at least 30. These requirements can all be accommodated provided that the maturity of the option is at least $30x0.006^2/0.22^2 = 0.022$ years which is about 8 calendar days.

[8] With $\sigma = 20\%, \delta = 0.6\%$ and $s = 0.05\%, \sigma/\sqrt{1-2s/\delta} = 22\%$.

For shorter maturity options, hedging is either expensive (because of the need to use a finer grid) or risky (because of the small number of nodes).

8 Conclusion

As I suggested at the beginning of this paper, Black-Scholes means different things to different people. My focus has always been quite a narrow one – seeing how the paper manages to solve the specific problem of valuing a single European call option on one asset. I have explained that the central insight, in my view, is the relationship between option value and realized variance. I see the stochastic process assumptions essentially as a device to pretend that the hedger knows the realized variance in advance. I have suggested that the focus on realized variance and the granularity of measurement is helpful in grappling with the application of the theory to practice.

But I do want to acknowledge that the influence of the Black Scholes paper has been much wider than this narrow focus implies. By showing how the mathematics of continuous time stochastic processes can be used to solve the call option valuation problem, it paved the way for the development of financial engineering, and the tools to manage portfolios of derivatives. This does require a reliance on formal modelling which cannot readily be reduced to simple discrete trading strategies, as in the single risky asset case. The extended reach of the theory, which has enabled the enormous growth of activity in the derivatives market, has also led to the accumulation of imperfectly understood assumptions which, from time to time, have caused users to face losses on a scale which the theory failed to predict.

Declarations

- Funding: no grants or other support were received in connection with this work.
- Conflict of interest: the author has no relevant financial or non-financial interests to disclose.

References

Black, F., Scholes, M.: The pricing of options and corporate liabilities. Journal of political economy **81**(3), 637–654 (1973)

MacKenzie, D.: Black-scholes at 50: how a pricing model for options changed finance. Financial Times **26 April 2023** (2023)

Black, F.: How we came up with the option formula. Journal of Portfolio Management **15**(2), 4–8 (1989)

Merton, R.C.: Theory of rational option pricing. The Bell Journal of economics and management science, 141–183 (1973)

Haug, E.G., Taleb, N.N.: Option traders use (very) sophisticated heuristics, never the black–scholes–merton formula. Journal of Economic Behavior & Organization **77**(2), 97–106 (2011)

Harrison, J.M., Pliska, S.R.: Martingales and stochastic integrals in the theory of continuous trading. Stochastic processes and their applications **11**(3), 215–260 (1981)

Dimson, E.: Option valuation nomograms. Financial Analysts Journal **33**(6), 71–74 (1977)

Aristotle, Lane, M.: Aristotle's Politics: Writings from the Complete Works: Politics, Economics, Constitution of Athens. Princeton University Press, Princeton, NJ (2016)

Black, F.: The pricing of commodity contracts. Journal of financial economics **3**(1-2), 167–179 (1976)

Margrabe, W.: The value of an option to exchange one asset for another. The journal of finance **33**(1), 177–186 (1978)

Garman, M.B., Kohlhagen, S.W.: Foreign currency option values. Journal of international money and finance **2**(3), 231–237 (1983)

Sharpe, W.F.: Investments, 1st edn. Prentice-Hall, New Jersey (1978)

Cox, J.C., Ross, S.A., Rubinstein, M.: Option pricing: A simplified approach. Journal of financial Economics **7**(3), 229–263 (1979)

Britten-Jones, M., Neuberger, A.: Arbitrage pricing with incomplete markets. Applied Mathematical Finance **3**(4), 347–363 (1996)

Karlin, S., Taylor, H.: A First Course in Stochastic Processes, 2nd edn. Academic Press, New York (1975)

Christensen, B.J., Prabhala, N.R.: The relation between implied and realized volatility. Journal of financial economics **50**(2), 125–150 (1998)

Arnold, T., Hilliard, J.E., Schwartz, A.: Short-maturity options and jump memory. Journal of Financial Research **30**(3), 437–454 (2007)