How we came up with the option formula

Like many great inventions, it started with tinkering and ended with delayed recognition.

Fischer Black

My paper with Myron Scholes giving the derivation of our option formula appeared in the spring of 1973. We had published a paper on the results of some empirical tests of the formula, however, in the spring of 1972. The work that led to the formula started in the spring of 1969, and the background research started in 1965. Here is the story of how the formula and the papers describing it came to be.

THE SHORT STORY

Before I describe the events surrounding our discovery of the formula, here is the idea behind the formula.

Suppose there is a formula that tells how the value of a call option depends on the price of the underlying stock, the volatility of the stock, the exercise price and maturity of the option, and the interest rate.

Such a formula will tell us, among other things, how much the option value will change when the stock price changes by a small amount within a short time. Suppose that the option goes up about $.50 when the stock goes up $1.00, and down about $.50 when the stock goes down $1.00. Then you can create a hedged position by going short two option contracts and long one round lot of stock.

Such a position will be close to riskless. For small moves in the stock in the short run, your losses on one side will be mostly offset by gains on the other side. If the stock goes up, you will lose on the option but make it up on the stock. If the stock goes down, you will lose on the stock but make it up on the option.

At first, you create a hedged position by going short two options and long one stock. As the stock price changes, and as the option approaches maturity, the ratio of option to stock needed to maintain a close-to-riskless hedge will change. To maintain a neutral hedge, you will have to change your position in the stock, your position in the option, or both.

As the hedged position will be close to riskless, it should return an amount equal to the short-term interest rate on close-to-riskless securities. This one principle gives us the option formula. It turns out that there is only one formula for the value of an option that has the property that the return on a hedged position of option and stock is always equal to the short-term interest rate.

The same argument works for a "reverse hedge," if you assume that you can sell stock short and invest the proceeds of the short sale for your benefit. A short position in the stock combined with a long position in the option (in the right ratio) will be close to riskless. Your equity in that position will be negative, but there is only one formula such that the return on that position is the interest rate — the same formula that we derive from the direct hedging argument.

We can even get the formula by assuming that a neutral spread must earn the interest rate. If you are short one option and long another option on the same stock in the right ratio, you will have a neutral spread. The argument is plausible even for a spread where you take in money, because you are probably in a position to invest the proceeds of a sale of options for your own benefit.

FISCHER BLACK is a Partner and Director of The Quantitative Strategies Group at Goldman Sachs in New York (NY 10004). A short version of this article appeared in *Current Contents/Social & Behavioral Sciences*, Vol. 19, No. 33, 1987, published by the Institute for Scientific Information, Inc., Philadelphia, PA 19104.

In fact, we can get the formula without assuming any hedging or spreading at all. We just compare a long stock position with a long option position that has the same action as the stock. In our example, the comparable positions would be long one round lot of stock and long two option contracts. These two positions have the same movements for small changes in stock price in the short run, so their returns should differ only by an amount equal to the interest rate times the difference in the total values of the two positions. We can have equilibrium only if investors are indifferent between the two positions. This gives us the same formula as the hedging and spreading arguments.

THE DIFFERENTIAL EQUATION

Jack Treynor was at Arthur D. Little, Inc. when I started work there in 1965. He had developed, starting in 1961, a model for the pricing of securities and other assets that is now called the "capital asset pricing model." William Sharpe, John Lintner, and Jan Mossin developed more or less independent versions of the same model, and their versions began to be published in 1965. Jack's papers were never published, in part because they never quite satisfied the perfectionist in him, and in part (I believe) because he did not have an academic job.

In any case, Jack sparked my interest in finance, and I began to spend more and more time studying the capital asset pricing model and other theories of finance. The notion of equilibrium in the market for risky assets had great beauty for me. It implies that riskier securities must have higher expected returns, or investors will not hold them — except that investors do not count the part of the risk that they can diversify away.

I started trying to apply the capital asset pricing model to assets other than common stock. I looked at bonds, cash flows within a company, and even monetary assets. One of Treynor's papers was on the valuation of cash flows within a company, and he had derived a differential equation to help in figuring this value. His equation had an error because he had omitted some terms involving second derivatives, but we found out how to put in the missing terms and correct the equation.

With this background, I started working on a formula for valuing a warrant. At that time, we thought about warrants more than about options, because the over-the-counter options market was such an imperfect market. I'm not sure when I started work on the warrant problem, but it was probably in 1968 or 1969. I have notes containing the differential equation that are dated June 1969.

Back then, most of the best papers about warrants tried to find the value of a warrant by taking the expected value of the warrant at expiration and discounting it to the present. That method has two problems: you have to know the stock's expected return to find the warrant's expected value at expiration, and you have to choose a discount rate for the warrant. No single discount rate will do, however, because the risk of the warrant depends on the stock price and time. Hence, the discount rate depends on the stock price and time too. None of the papers had dealt with this problem.

One key step in solving the problem is to write the warrant value as a formula that depends on the stock price and other factors. As Treynor had used this approach with his "value equation," I tried it too. And about the same time that I was using this approach, Samuelson and Merton were using it in a paper that appeared in 1969 (although they didn't come up with the same formula).

Another thing that made it possible to solve the problem was to assume away all kinds of complications. I assumed that trading costs are zero, that both borrowing and lending can be done at a single short-term interest rate that is constant over time, and that the volatility of a stock is constant, which means that the future price of the stock follows a lognormal distribution. I made a few other simplifying assumptions, some of which turned out to be unnecessary.

The equation I wrote down said simply that the expected return on a warrant should depend on the risk of the warrant in the same way that a common stock's expected return depends on its risk. I applied the capital asset pricing model to every moment in a warrant's life, for every possible stock price and warrant value. To put it another way, I used the capital asset pricing model to write down how the discount rate for a warrant varies with time and the stock price.

This gave me a differential equation. It was an equation for the warrant formula. It has just one solution, if we use the known value of the warrant at expiration and another condition that I didn't know about at the time.

I spent many, many days trying to find the solution to that equation. I have a Ph.D. in applied mathematics, but had never spent much time on differential equations, so I didn't know the standard methods used to solve problems like that. I have an A.B. in physics, but I didn't recognize the equation as a version of the "heat equation," which has well-known solutions.

I did notice that some of the factors in the original equation were not in the final equation. The warrant value did not seem to depend on how the risk of the stock was divided between risk that could be diversified away and risk that could not be diversified away. It depended only on the total risk of the stock (as measured, for example, by the standard deviation of the return on the stock). The warrant value did not depend on the stock's expected return, or on any other asset's expected return. That fascinated me.

But I was still unable to come up with the formula. So I put the problem aside and worked on other things.

In 1969, Myron Scholes was at MIT, and I had my office near Boston, where I did both research and consulting. Myron invited me to join him in some of the research activities at MIT. We started working together on the option problem, and made rapid progress.

THE FORMULA

First, we concentrated on the fact that the option formula was going to depend on the underlying stock's volatility — not on its expected return. That meant that we could solve the problem using any expected return for the stock.

We decided to try assuming that the stock's expected return was equal to the interest rate. (We were assuming a constant interest rate, so short-term and long-term rates were equal.) In other words, we assumed that the stock's beta was zero; all of its risk could be diversified away.

As we also assumed that the stock's volatility was constant (when expressed in percentage terms) it was easy to figure the likelihood of each possible value of an investment in the stock at the time the option expired. We knew that the stock's terminal value (including reinvested dividends) would have to fit a lognormal distribution.

Other writers on options had made the same sort of assumption about the underlying stock, but they had not assumed an expected return equal to the interest rate. They had, however, assumed a constant expected return, which means a lognormal distribution for the terminal value of a stock that pays no dividends.

If you know the distribution for the stock's terminal value, you can cut it off at the option's exercise price and have the distribution for the option's terminal value. The expected value of that cutoff distribution gives you the expected terminal value of the option.

An article by Case Sprenkle presented a formula for the expected terminal value of an option with these same assumptions, except that Sprenkle allowed the stock to have any constant expected return. By putting the interest rate for the expected stock return into his formula, we got the expected terminal value of the option under our assumptions.

But we didn't want the expected terminal value of the option. We wanted the present value of the option: the value at some time before maturity. So we had to find some way to discount the option's expected terminal value to the present.

Rather suddenly, it came to us. We were looking for a formula relating the option value to the stock price. If the stock had an expected return equal to the interest rate, so would the option. After all, if all the stock's risk could be diversified away, so could all the option's risk. If the beta of the stock were zero, the beta of the option would have to be zero too.

If the option always had an expected return equal to the interest rate, then the discount rate that would take us from the option's expected future value to its present value would always be the interest rate. The discount rate would not depend on time or on the stock price, as it would if the stock had an expected return other than the interest rate.

So we discounted the expected terminal value of the option at the constant interest rate to get the present value of the option. Then we took Sprenkle's formula, put in the interest rate for the expected return on the stock, and put in the interest rate again for the discount rate for the option. We had our option formula.

We checked the formula against the differential equation, and sure enough, it fit. We knew it was right. A few changes, and we had a formula for puts, too.

WORKING TOWARD PUBLICATION

Our first thought was to publish a paper describing the formula. (Later, we thought also about trying to use the formula to make money trading in options and warrants.) As we worked on the paper, we had long discussions with Robert Merton, who was also working on option valuation.

Merton made a number of suggestions that improved our paper. In particular, he pointed out that if you assume continuous trading in the option or the stock, you can maintain a hedged position between them that is literally riskless. In the final version of the paper, we derived the formula that way, because it seemed to be the most general derivation.

Merton started working on a paper on aspects of the option formula. He was able to prove, along with other important points, that if you don't want a constant interest rate in the formula, you should use the interest rate on a discount bond that matures when the option expires.

Scholes and I started thinking about applying the formula to figuring the values of risky corporate bonds and common stock. Merton began thinking about that too, but neither of us told the other. We were both working on papers about the formula, so there was a mixture of rivalry and cooperation. Scholes and I gave an early version of our paper at a conference on capital market theory sponsored by Wells Fargo Bank in the summer of 1970. We talked then about the application to corporate finance. Merton attended the conference, but he overslept on the morning of our talk, so it was only later that all of us discovered we were working on the corporate finance applications.

The first surviving draft of our paper describing the option formula (dated October 1970) was called "A Theoretical Valuation Formula for Options, Warrants, and Other Securities." I sent it to the *Journal of Political Economy* and promptly got back a rejection letter. They said that it was too specialized for them, and that it would be better in the *Journal of Finance*. I then sent it to *The Review of Economics and Statistics* and promptly got back another rejection letter. They said they could publish only a few of the papers they received. Neither journal had the paper reviewed.

I suspected that one reason these journals didn't take the paper seriously was my non-academic return address. In any case, we rewrote the paper to emphasize the economics behind the formula's derivation. The next draft (dated January 1971) was called "Capital Market Equilibrium and the Pricing of Corporate Liabilities."

Merton Miller and Eugene Fama at the University of Chicago took an interest in the paper. They gave us extensive comments on this draft, and suggested to the *Journal of Political Economy* (which is published there) that perhaps the paper was worth more serious consideration. In August 1971, the *Journal* accepted the paper, conditional on further revisions suggested by the referees.

The final draft of the paper (dated May 1972) was called "The Pricing of Options and Corporate Liabilities." It appeared in the May/June 1973 issue of the *Journal of Political Economy*. Meanwhile, we had written a paper on the results of some empirical tests of the formula, which appeared in the May 1972 *Journal of Finance*.

TESTING THE FORMULA

While we were working on our paper telling about the formula, we began to look for ways to test it on real securities. We started with warrants.

We estimated the volatility of the stock of each of a group of companies with warrants outstanding. We applied the formula in a simple way to these warrants, ignoring some of the ways in which warrants differ from options. We noticed that several warrants looked like very good buys. The best buy of all seemed to be National General new warrants.

Scholes, Merton, and I and others jumped right in and bought a bunch of these warrants. For a while, it looked as if we had done just the right thing. Then a company called American Financial announced a tender offer for National General shares. The original terms of the tender offer had the effect of sharply reducing the value of the warrants.

In other words, the market knew something that our formula didn't know. The market knew that such a tender offer was likely or possible, and that's why the warrants seemed so low in price. Although our trading didn't turn out very well, this event helped validate the formula. The market price was out of line for a very good reason.

It also illustrates a general rule. The formula and the volatility estimates we put into the formula are always based on the information at hand. The market will always have some kinds of information affecting the values of options and warrants that we don't have. Sometimes the values given by the formula will be better than market prices; at other times the market prices will be better than the formula values.

We learned that rule again in our next set of tests. One of Scholes's students managed to get data on the premiums received by a broker's option-writing customers in the over-the-counter options market. The data covered a period of several years.

We used the formula, with some simple volatility estimates, to test trading rules. We wanted to find out how much money we could have made if we had bought the options whose prices seemed lower than our formula's values, and sold the options whose prices seemed higher than our values.

Ignoring transaction costs, our profits seemed to be substantial. As these were over-the-counter options, we assumed the positions were held to maturity. To highlight the profits and losses, we combined each option position with a continuously changing stock position that created a close-to-riskless hedge all the time. The profits were consistent at around fifty cents per day per contract. Nevertheless, transaction costs in the over-the-counter options market could easily wipe out those profits.

We also tried assuming that we bought the underpriced options and sold the overpriced options at the values given by our formula, rather than at the market prices. Then we had losses of around fifty cents per day per contract. In other words, the formula seemed to have some information the market didn't have, but the market had just as much information that the formula didn't have.

Our findings do not mean that you lose if you use the formula for trading. If you trade at market prices, you get the benefit of what the market knows. But it is not a good idea to insist on trading at the values given by the formula. The market may want to trade at prices away from those values for good reasons that the formula cannot consider.

Later, after the CBOE started trading in listed options, Dan Galai wrote a Ph.D. thesis at the University of Chicago in which he tested trading rules based on the formula. Ignoring trading costs, the profits he found in trading listed options were much larger than the profits we found in over-the-counter options, because he assumed that an option position would be changed every time it became underpriced or overpriced.

For example, he tested the profitability of spreads that are kept neutral continuously. A neutral spread is a long position in one option combined with a short position in another option on the same stock. The position is close to riskless. To maintain a neutral spread, you need to change either your long position or your short position (or both) as the stock price and time-to-maturity change.

Galai figures option values using simple volatility estimates, based on past daily data on stock prices. He has only closing prices for the options, but he tries to take out some of their distortions. He assumes that you decide what to do by comparing option values and closing option prices one day, but you execute the trades at closing prices the next day. If closing prices are distorted in the same direction two days in a row, they may still overstate your prices. But if it's possible to trade only at a favorable price, and not at just any next-day's price, then this will understate your profits. This method also ignores profits that market makers can make by opening and closing positions within a single day.

The spreads that Galai looked at involve buying one contract of the underpriced option and selling either more or less than one contract of the overpriced option: whatever is needed to create and maintain a neutral spread. Ignoring transaction costs, the average spread gives a consistent profit of $4.00 or $5.00 per day.

That sounds like a fast way to make money. But it does ignore trading costs, which are especially high for people who have to pay retail commissions. And it does assume trading at the next day's closing prices — a conservative assumption, but one that still may cause profits to be overstated. Finally, the period Galai studied was July 1973 to April 1974. Opportunities like this are harder to come by these days.

One reason for the change is that traders now use the formula and its variants extensively. They use it so much that market prices are usually close to formula values even in situations where there should be a large difference: situations, for example, where a cash takeover is likely to end the life of the option or warrant.