

Volatility Trading

Vladislav Zakatov & Euan Sinclair

Contents

1	Option Pricing	2
1.1	The Black-Scholes-Merton model	2
2	Volatility Measurement and Forecasting	8
2.1	Defining and Measuring Volatility	9
2.2	Alternative Volatility Estimators	13

Chapter 1

Option Pricing

Black-Scholes model is a framework we can use to compare options of different maturities, underlyings, and strikes. We do not insist that it is in any sense true or even a particularly accurate reflection of the real world. As options are highly leveraged, nonlinear, time-dependent bets on the underlying, their prices change very quickly. The major goal of a pricing model is to translate these prices into a more slowly moving system. A model that perfectly captures all aspects of a financial market is probably unobtainable. Further, even if it existed it would be too complex to calibrate and use. So we need to somewhat simplify the world in order to model it. Still, with any model we must be aware of the simplifying assumptions that are being used and the range of applicability. The specific choice of model isn't as important as developing this level of understanding.

1.1 The Black-Scholes-Merton model

We start from the assumption that a trader holds a delta hedged portfolio consisting of a call option and Δ units of short stock. We then apply our knowledge of option dynamics to derive the BSM equation. Even before we make any assumptions about the distribution of underlying returns, we can state a number of the properties that an option must possess:

- A call (put) becomes more valuable as the underlying rises (falls), as it has more chance of becoming intrinsically valuable.
- An option loses value as time passes, as it has less time to become intrinsically valuable.

- An option loses value as rates increase. Since we have to borrow money to pay for options, as rates increase our financing costs increase, ignoring for now any rate effects on the underlying.
- The value of a call (put) can never be more than the value of the underlying (strike).

Let's assume we hold the delta hedged option position,

$$C - \Delta S_t \quad (1.1)$$

where C is the value of the option, S_t is the underlying price at time t , Δ is the number of shares we are short.

Over the next time step the underlying changes to S_{t+1} . The change in the value of the portfolio is given by the change in the option and stock positions together with any financing charges we incur by borrowing money to pay for the position.

$$C(S_{t+1}) - C(S_t) - \Delta(S_{t+1} - S_t) + r(C - \Delta S_t) \quad (1.2)$$

The change in the option value due to the underlying price change can be approximated by a second-order Taylor expansion. Also, we know that when other things are held constant, the option will decrease due to the passing of time by an amount denoted by θ . So we get

$$\Delta(S_{t+1} - S_t) + \frac{1}{2}(S_{t+1} - S_t)^2 \frac{\delta^2 C}{\delta S^2} + \theta - \Delta(S_{t+1} - S_t) + r(C - \Delta S_t) \quad (1.3)$$

Or

$$\frac{1}{2}(S_{t+1} - S_t)^2 \Gamma + \theta + r(C - \Delta S_t) \quad (1.4)$$

where Γ is the second derivative of the option price with respect to the underlying.

Expression (1.4) gives the change in value of the portfolio, or the profit the trader makes when the stock price changes by a small amount. It has three separate components:

1. The first term gives the effect of gamma. Since gamma is positive, the option holder makes money. The return is proportional to half the square of the underlying price change.
2. The second term gives the effect of theta. The option holder loses money due to the passing of time.

3. The third term gives the effect of financing. Holding a hedged long option portfolio is equivalent to lending money.

Further, we see in Chapter 2 that on average

$$(S_{t+1} - S_t)^2 \cong \sigma^2 S^2 \quad (1.5)$$

where σ is the standard deviation of the underlyings returns, generally known as *volatility*. So we can rewrite expression (1.4) as

$$\frac{1}{2}\sigma^2 S^2 \Gamma + \theta + r(C - \Delta S_t) \quad (1.6)$$

If we accept that this position should not earn any abnormal profits because it is riskless and financed with borrowed money, the expression can be set equal to zero. Therefore the equation for the fair value of the option is

$$\frac{1}{2}\sigma^2 S^2 \Gamma + \theta + r(C - \Delta S_t) = 0 \quad (1.7)$$

Before continuing, we need to make explicit some of the assumptions that this informal derivation has hidden.

- In order to write down expression (1.1) we needed to assume the existence of a tradable underlying asset. In fact, we assume that it can be **shorted** and the underlying can be **traded in any size** necessary without incurring transaction costs.
- Expression (1.2) has assumed that the proceeds from the short sale can be reinvested **at the same interest rate** at which we have borrowed to finance the purchase of the call. We have also taken this rate to be constant.
- Expression (1.3) has assumed that the underlying changes are **continuous** and **smooth**. Further, we have considered second-order derivatives with respect to price but only first-order with respect to time.

The effect of drift can be negated by combining the option with the share in the correct proportion. As the drift can be hedged away, the holder of the option is not compensated for it. When we consider hedging later in Chapter 4, we find that in the real world, where the assumptions about continuity fail, directional dependence will reemerge.

If our estimate of volatility differs significantly from that implied by the option market, then we can trade the option accordingly. If we forecast

volatility to be higher than that implied by the option, we would buy the option and hedge in the underlying market. Our expected profit would depend on the difference between implied volatility and realized volatility. Equation (1.7) says that instantaneously this profit would be proportional to

$$\frac{1}{2}S^2\Gamma(\sigma^2 - \sigma_{implied}^2) \quad (1.8)$$

A complementary way to think of the expected profit of a hedged option is by considering *vega*. Vega is defined as the change in value of an option if implied volatility changes by one point. This means that if we buy an option at $\sigma_{implied}$ and volatility immediately increases to σ we would make a profit of

$$vega(\sigma - \sigma_{implied}) \quad (1.9)$$

If we have to hold the option to expiration and realized volatility averages σ we will also make this amount, but only on average. The vega profit is realized as the sum of the hedges as we rebalance our delta. This can be formalized by noting the relationship between vega and gamma,

$$vega = \sigma T S^2 \Gamma \quad (1.10)$$

So expression (1.8) can also be written as

$$\frac{vega}{\sigma T}(\sigma^2 - \sigma_{implied}^2) \quad (1.11)$$

It is perfectly acceptable to make simplifying assumptions when developing a model. It is totally unacceptable to make assumptions that are so egregiously incorrect that the model is useless, even as a basic guide. So before we go any further we look at how limiting our assumptions really are.

- We assumed that the underlying was a *tradable* asset. While the BSM formalism has been extended to cases where this is not true, notably in the pricing of real options, we are primarily concerned with options on equities and futures, so this assumption is not restrictive. However, on many optionable underlyings liquidity is an issue, so tradable is not always a clearly defined quality.
- We assumed that the underlying pays no *dividends* or any other income. This changes the equations slightly, but the same principles still apply.

- We also assumed this asset was able to be *shorted*. This is not a problem where the underlying is a future but when it is a stock, shorting is often more difficult. For example, in the United States, stocks can only be shorted on an uptick. Further, even when shorting is achievable, the short seller rarely receives the full proceeds of the sale for investment, as fees must be paid to borrow the stock. This can be accounted for synthetically by assuming an extra dividend yield on the underlying, equal to the penalty cost associated with shorting the stock.
- Interest rates have a *bid/ask spread*. We cannot invest the proceeds of a sale at the same rate at which we borrow. Further, rates are not constant. However, the BSM is often used to price options on bonds and money market rates which would have no volatility if this assumption was valid. We can get away with this because the risk due to interest charges (*rho*) is insubstantial in comparison to other risks, at least for short-dated options.
- We have assumed that volatility is a constant, neither a function of time nor of the underlying price. Not only is this untrue but we will be actively trying to trade these changes. There are models that explicitly take into account volatility changes. However, we choose to recognize this limitation and learn to use the BSM model anyway. This is consistent with our philosophy of the model as a framework for organizing our thoughts rather than as an accurate depiction of reality.
- We assumed that volatility is the only parameter needed to specify the distribution of the underlying returns. The mean can be hedged away and we have ignored higher-order moments. This is the same as assuming a normal return distribution or a lognormal price distribution. The fact that this is incorrect leads to the well-known phenomenon of the *volatility smile*, where implied volatility is a function of strike. In essence, implied volatility is the wrong number we put into the wrong formula to get the correct option price. This can be rectified in several ways. In Chapter 3 we present methods of quantifying the implied skewness and kurtosis.
- We have assumed that the underlying's changes are continuous so we can continually adjust our hedge. This is not true. Sometimes the underlying has vast jumps. For example, it isn't uncommon for the shares of a biotech company to jump by 70 to 80 percent in one day.

Modifications have been made to the BSM formalism to price options in these circumstances (Merton 1976), but this isn't really the point. These jumps cannot be hedged and the replication strategy fails utterly. We have to learn to hedge this risk with other options. This is the concept of semistatic hedging that traders need to use in practice.

The BSM model is remarkably robust. Most of the assumptions that we need to derive the equation can be loosened without destroying the model's utility. But note that we will only use the BSM paradigm as a pricing method, not a risk control method. It is useful to translate the fast-moving option prices into a slow-moving parameter, implied volatility, which can be compared to the estimated realized volatility. But risk control must be handled separately. Traders should never think about extreme risk in terms of the moments of the Gaussian distribution. Tail risk can often be capped by trading far out-of-the-money options, and keeping individual positions to a small proportion of the total portfolio can also help. But generally we get paid for taking risks. Just try to be aware of the risks you have an edge in and those you don't. And never estimate the magnitude of risks from within the same model that you priced them with.

Chapter 2

Volatility Measurement and Forecasting

To find an edge in option trading we need an estimate of future realized volatility to trade against that implied by the options. But before we can forecast future volatility, we need to be able to measure what it has been in the past. Here we look at methods of historical volatility measurement including:

- Close-to-close volatility;
- Parkinson volatility;
- Rogers-Satchell volatility;
- Garman-Klass volatility;
- Yang-Zhang volatility.

We discuss the **efficiency** (how quickly our estimated value converges to the true value) and **bias** (whether our method systematically estimates above or below the true value) of each estimator and also how each is perturbed by different aspects of real markets such as **fat tails** in the return distribution, **trends**, and **microstructure noise**. We discuss different **frequencies** of measurement. Once we understand what is meant by historical volatility, we can look at its properties. We demonstrate **mean reversion**, **volatility clustering**, and **seasonality effects**.

2.1 Defining and Measuring Volatility

For trading we need more than a point estimate of future volatility; we need some estimate of possible range of volatilities¹. To find this we examine the construction and sampling properties of **volatility cones**. Measuring volatility and having a forecast of its distribution are essential for successfully trading options, but it is not sufficient. Buying volatility because it is cheap or selling because it is rich is seldom a good idea. Often things will be cheap for a reason. Any forecast we make has to be supplemented by our fundamental analysis. The markets are very complex and interrelated, and all measurements and forecasts must be placed in the context of the current trading environment.

The standard definition of volatility is the square root of the variance. And variance is defined as

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (2.1)$$

where x_i are the logarithmic returns, \bar{x} is the mean return in the sample, N is the sample size.

To express the variance in annualized terms, we would need to multiply the raw variance by the annualization factor given by N , the number of trading periods in a year (for example, 252 if using daily data).

If the historical price series includes the payment of a dividend (or a stock split), we must adjust the price series. The effect of a stock going ex-dividend makes it look like there was volatility even though there was none. If this adjustment is not done, our volatility estimate may well be wrong by several percentage points.

There are several different ways of making this adjustment. The first is simply to subtract the dividend from the price before the ex-dividend date. This leaves the absolute values of the day-to-day changes before the ex-dividend date unchanged, but if we have enough dividends in the series this process can lead to apparently negative stock prices.

A better method is to multiply by an adjustment factor that leaves the percentage changes unaffected. This factor is

$$1 - \frac{\text{dividend}}{\text{price}} \quad (2.2)$$

Prices before the ex-dividend date are multiplied by this factor. This is *backward adjustment*. Alternatively, it is possible to *forward-adjust* prices,

¹Poon (2005) lists over 100 references on volatility forecasting.

which would mean that the current price won't be the same as the adjusted price.

In finance it is very difficult to distinguish mean returns (*drift*) from variance (this is a central problem in many arguments about trading methods and results), and estimates of the mean return are notoriously noisy, especially for small samples. So we generally set the mean return in equation (2.1) to zero. This increases accuracy of measurement by removing a source of noise.

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i)^2 \quad (2.3)$$

Equations (2.1) and (2.3) make no assumptions about the distribution, other than that the sum converges. All finite samples have variance. However, to use volatility for pricing options, we need to make assumptions about the process generating the returns. As mentioned in Chapter 1, the BSM model assumed that the returns were normally distributed. In this case variance completely characterizes the shape of the distribution. We know this isn't true, but we still expect variance (and hence volatility) to be a very important parameter, indeed the dominant parameter, for describing the width of the return distribution.

To estimate the population variance from this sample variance we need to make the conversion (Kenny and Keeping 1951):

$$\sigma^2 = \frac{N}{N-1} s^2 \quad (2.4)$$

Equations (2.3) and (2.4) give unbiased estimates of variance, but simply taking the square root will give an estimate of volatility that is biased low. This is because of Jensen's inequality, which states that the average of a square root is always less than the square root of the average. Specifically

$$E(s) = E(\sqrt{s^2}) < \sqrt{E(s^2)} = \sqrt{s^2} = \sigma \quad (2.5)$$

If we assume that the true process is a normal distribution of returns, we can use the fact that the distribution of the sample standard deviation as a function of the length of the sample is given by

$$f_N(s) = 2 \frac{\left(\frac{N}{2\sigma^2}\right)^{\frac{N-1}{2}}}{\Gamma\left(\frac{N-1}{2}\right)} \exp\left(\frac{-Ns^2}{2\sigma^2}\right) s^{N-2} \quad (2.6)$$

where s is the sample standard deviation, σ is the population standard deviation, $\Gamma(x)$ is the gamma function defined by $\Gamma(n) = (n-1)!$.

The extent of the bias can be exactly quantified by the relationship

$$\bar{s} = b(N)\sigma \quad (2.7)$$

where

$$b(N) = \sqrt{\frac{2}{N}} \frac{\Gamma(\frac{N}{2})}{\Gamma(\frac{N-1}{2})} \quad (2.8)$$

s/b is an unbiased estimator of the population standard deviation.

This corrects the bias. This means that it will not systematically over- or underestimate the true volatility. However, this estimator converges to the true volatility slowly, which is technically referred to as being *inefficient*.

The variance of the estimator is given by

$$\text{var}(s) = \frac{1}{N} \left(N - 1 - 2 \frac{\Gamma^2(\frac{N}{2})}{\Gamma^2(\frac{N-1}{2})} \right) \sigma^2 \quad (2.9)$$

Mental arithmetic involving the gamma function can be tricky. A simpler approximation to this equation would be more useful. First we note that

$$\frac{\Gamma(k + \frac{1}{2})}{\Gamma(k)} = \sqrt{k} \left(1 - \frac{1}{8k} + \frac{1}{128k^2} + \dots \right) \quad (2.10)$$

So to leading order in N we eventually get

$$b(N)^2 \approx 1 - \frac{3}{2N} \quad (2.11)$$

Therefore equations (2.8) and (2.9) give us

$$\text{var}(s) \approx \frac{\sigma^2}{2N} \quad (2.12)$$

This gives a simpler expression for levels of the confidence interval of the measured volatility.

Using more data would get us closer to the true result in this case. While this is no problem if we are measuring the volatility of an unchanging process, it is problematic for financial markets. If we use too little data we will have a noisy measurement of volatility that, due to sampling error, might not be close to the true volatility; but if we use too much data we will be using information that is no longer relevant to the current state of the market. Choosing the right compromise is something of an art, and the most appropriate solution will be dependent on current market conditions. However, it is obvious that the commonly used method of measuring volatility

from the last 30 closing prices gives an unacceptably large sampling error. The 95 percent confidence interval of two standard deviations means that we could be off by as much as 25 percent of the true value!

Sampling error is not the same as measurement error. There is no uncertainty due to measurement. But there is uncertainty over whether the measure number is truly representative of the underlying reality. Volatility is an unobservable quantity that we can only estimate.

Before addressing this, let's note that another reason the close-to-close estimator is useful is because it can be rewritten in a form that allows us to easily relate typical average stock moves to volatility, which is very useful for traders.

The definition of the standard deviation involves the square root of an average of squares, and we typically don't have good intuition about how these behave. So instead we will look at an estimator based on the typical stock move. This is

$$\sigma = 19.896 \left(\frac{1}{N} \sum_{t=1}^N |R_t| \right) \quad (2.13)$$

This is because

$$E[|R_t|] = \sqrt{\frac{2}{\pi}} \sigma \quad (2.14)$$

This means that

$$\text{average move} = 0.04986 \sigma S \approx \frac{\sigma S}{20} \quad (2.15)$$

This allows a simple translation between daily returns and annualized volatility. Multiplying the daily return by 20 gives a useful quick-and-dirty estimate of annualized volatility.

There are two basic ways of addressing the problem of the large sampling error. We can use the close-to-close estimator with higher-frequency data, or we can use another estimator that doesn't throw away all data points other than closing prices. Each has limitations. First we will try to develop better estimators, an approach that is also more generally applicable. If we can find a better estimator we could always apply it to higher-frequency data.

2.2 Alternative Volatility Estimators

The first such estimator was developed by Parkinson (Parkinson 1980). His estimator is

$$\sigma = \sqrt{\frac{1}{4N \ln 2} \sum_{i=1}^N \left(\ln \frac{h_i}{l_i} \right)^2} \quad (2.16)$$

where h_i is the high price in the trading period, l_i is the low price.

As before, this would need to be annualized by multiplying it by the square root of the number of trading periods in a year. This estimate needs fewer time periods to converge to the true volatility as it uses two prices from each period, instead of just one as with the close-to-close estimator. The Parkinson estimator is about five times more efficient at estimating volatility than the close-to-close estimator when it is tested on an artificially generated geometric Brownian motion (GBM). (*Efficiency* is defined as the ratio of the variance of the close-to-close estimator to the range-based estimator).

If prices are continuous, the Parkinson estimate of variance is *unbiased* (but remember that there is a bias introduced by Jensen's inequality when we convert any variance estimate to a volatility estimate). However, prices are only sampled discretely. This is true both because markets only trade in discrete units and, more important, because markets are only open for part of the day. This means that the unobservable true price may not make a high or a low when we can actually measure it. So we will systematically underestimate volatility by using an estimator based on the observed range.

Garman and Klass (Garman and Klass 1980) showed by simulation the underestimation due to discrete sampling as a function of sample size.

The fact that this biases the estimates of volatility low comes as a surprise to some. There is a pervasive misunderstanding that the Parkinson estimator is biased high, as it is impossible to actually trade at the extremes. This is true but irrelevant. Parkinson makes no claims about being able to trade at the extremes of the range, just that the range is related to the volatility. It is an estimator of volatility, not of tradability.

This bias is clearly a significant issue. As a practical matter, the variance estimate can be unbiased by dividing by these correction factors as we did in the case of the close-to-close estimator. But this does not address the fact that opening jumps exist in the price series.

The other well-known volatility estimator was developed by Garman and

Klass. It is

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{1}{2} \left(\ln \frac{h_i}{l_i} \right)^2 - \frac{1}{N} \sum_{i=1}^N (2\ln 2 - 1) \left(\ln \frac{c_i}{c_{i-1}} \right)^2} \quad (2.17)$$

This estimator is up to eight times as efficient as the close-to-close estimator (the exact efficiency improvement is dependent on the sample size) but is also biased due to the discrete sampling leading to a low estimate of the range. Its bias is actually worse than the Parkinson estimator, as shown in Table 2.4 (also from Garman and Klass 1980).

If we know what the bias is, it can be corrected. More problematically, the studies that show improved efficiency of these estimators rely on assumptions that do not apply to real markets. In particular they assume that the underlying follows a driftless GBM and trades continuously. Rogers, Satchell, and Yoon (Rogers and Satchell 1991; Rogers, Satchell, and Yoon 1994) relax this restriction and introduce an estimator that outperforms the others when a drift term is introduced: