Chapter 14

Volatility Trading

This chapter is attributed to the theory and practice of volatility trading, with primary focus on the variance risk premiuim. Necessarily, the theory of option pricing will be involved. Although volatility may as well be traded with the use of linear products (such as UVXY), or with other derivatives (such as variance swaps), the accessibility and depth of the option markets give us the most commonplace access to trading volatility. Option pricing is discussed in our chapter on stochastic calculus (Chapter 13), exploring option pricing through the lenses of the Black-Scholes-Merton model and various extensions. Both the pricing of European and American (and Asian) options were given treatment in those sections - here we will give more focus to the pricing of European options, since they are less unwieldy and produce better intuition - we will find out that as long as interest rates are low or time horizon is small, or both - then the European option is satisfactory approximation for American one. In most cases, such approximation should suffice. After all, long term memory in volatility is regime dependent and most variance premiums are harvested over short horizons.

Options are the right but not the obligation to buy or sell some underlying asset S at (European) or up to (American) some specified time T at some specified price K. The right to do so is valued as the option premium, or the value of the option. The seller is in turn obligated to fulfill the terms of this contract if the buyer chooses to exercise his right. The parameters defining an option contract is given by the option type (call/put/barrier), underlying asset S, strike K, expiry T, exercise style (typically specified by a region). A call is the right to buy, and a put is the right to sell. The underlying asset may affect the form of settlement, and therefore the obligation the seller of an option contract is subjected to. For instance, the underlying asset of a stock option are shares (often 100 units) in the stock, and some multiple of the cash value of the index in index options.

The specifications of the option contract in relation to market structure, clearing and margin concerns are left up to the reader to find out and clearly depends on both product and brokerage being traded with.

14.1   Arbitrage Bounds

We begin by presenting bounds for option pricing that do not require any models. The arguments rely on the principle of no-arbitrage. The formal definition is given in Definition 453, and the existence of a risk-neutral measure asserts the absence of arbitrage (Theorem 420). We would not need such formality though - simply said, we should not be able to construct a portfolio that begins with zero capital and at

some point in the future has positive probability of profit and zero probability of loss. No free money, that is.

We begin by denoting the value of the European call, European put, American call, American put as c, p, C, P respectively. Then we have

$$c \leq C, p \leq P. \tag{5860}$$

The American option is always worth at least as much as the European option, since the American option becomes a European option if we simply do not choose to exercise before expiry. All we have on top is choice. If the inequalities do not hold, the arbitrage portfolio is short the European and long the American at same strike and expiry.

A call option cannot cost more than the underlying. We have

$$c \leq S. \tag{5861}$$

Otherwise the arbitrage portfolio is short the call and long stock. We use the short proceeds to purchase stock, and invest the remaining. If the call expires worthless, we are done. If the call is exercised, we fulfill our obligations with the stock held. The money market investment is our free lunch. See that we can similarly fulfill our obligations at any time prior to expiry, so the argument holds for American calls

$$C \leq S. \tag{5862}$$

An American put cannot be more than the strike price.

$$P \leq K. \tag{5863}$$

Otherwise, he can sell the put, receive P and his maximum payout is $(0 - K) = -K$, and he makes risk-free profit of $P - K$. For the European option, since the exercise can only occur at expiry, we have

$$p \leq K \exp(-rt). \tag{5864}$$

The arbitrage portfolio if this inequality does not hold sells put for p, invests in the money market for $p \exp(rt) > K$ at expiry, which is greater than the maximum loss K.

The arguments for put-call parity relationship between a European call and European put is given in Equation 3727. There we presented that the payoff of a long call, short put portfolio is replicating portfolio for a forward contract and therefore must satisfy

$$c - p = S - K \exp(-rt). \tag{5865}$$

If the underlying pays dividends, we would have to modify the discounting and use $c - p = S - D - K \exp(-rt)$, where D is the present value of any dividends paid.

We may easily derive from this put-call parity relationship the following bounds:

$$c \geq S - K \exp(-rt), \tag{5866}$$
$$p \geq K \exp(-rt) - S. \tag{5867}$$

The put-call argument does not hold when working with American options, since there is no guarantee that the short American put is not exercised to expiry. We know, however, that the value of the American options must be at least as great as their intrinsic value, else the arbitrage portfolio purchases and immediately exercises the right:

$$C \geq \max(0, S - K), \tag{5868}$$
$$P \geq \max(0, K - S). \tag{5869}$$

In addition the American call is at least as valuable as the European call, and since we have a lower bound already for the European call

$$C \geq \max(0, S - K \exp(-rt)). \tag{5870}$$

This tighter bound implies that the American call option on a non-dividend paying stock should never be exerised early. An in-the-money call option may be *sold* for $S - K \exp(-rt)$, greater than the intrinsic value $S - K$ obtained from early *exercise*. In this particular scenario, the early exercise option is not useful and both American and European call options on non-dividend stocks are priced the same.

Now consider two strikes $K_2 > K_1$. We want to construct bounds for the relationship between calls and puts of different strike price. Consider a portfolio that is long one call at K_1 strike and short one call at K_2 strike. For the European option, at maturity, if $S < K_1$, then both calls are worthless, if $K_1 < S < K_2$, then only the long call position is in-the-money with payout $S - K_1$, else if $S > K_2$, the long call position is worth $S - K_1$, short call position is worth $-(S - K_2)$ and the portfolio is worth $K_2 - K_1$. In any case the portfolio is at least worth zero and we can write

$$c(K_1) > c(K_2). \tag{5871}$$

In all three scenarios the portfolio value at expiration is $\leq K_2 - K_1$, so we must have

$$c(K_1) - c(K_2) \leq (K_2 - K_1) \exp(-rt). \tag{5872}$$

But the same argument holds if we were to allow exercise prior to expiry, with the exclusion of the discount factor -

$$C(K_1) - C(K_2) \leq (K_2 - K_1). \tag{5873}$$

Now consider a portfolio that is short one put at K_1 strike and long one put at strike K_2. We can run through the same scenarios and determine that the value of the final portfolio is $K_2 - K_1$, $K_2 - S$, 0 in the three states of the world $S < K_1$, $K_1 < S < K_2$ and $S > K_2$ respectively. In all states the portfolio value at expiry is positive and $\leq K_2 - K_1$, so

$$p(K_2) \geq p(K_1) \tag{5874}$$

and

$$p(K_2) - p(K_1) \leq (X_2 - X_1) \exp(-rt) \tag{5875}$$

and again if we were to consider the possibility of early exercise

$$P(K_2) - P(K_1) \leq (X_2 - X_1). \tag{5876}$$

For the next relationship, we would not outline the arguments, but it should be relatively straightforward algebra to verify. For three strikes given $K_1 < K_2 < K_3$ and $\alpha := \frac{K_3 - K_2}{K_3 - K_1}$, we have

$$\alpha c(K_1) + (1 - \alpha)c(K_3) \geq c(K_2), \tag{5877}$$
$$\alpha C(K_1) + (1 - \alpha)C(K_3) \geq C(K_2), \tag{5878}$$
$$\alpha p(K_1) + (1 - \alpha)p(K_3) \geq p(K_2), \tag{5879}$$
$$\alpha P(K_1) + (1 - \alpha)P(K_3) \geq P(K_2). \tag{5880}$$

Consider two portfolios. The first portfolio is long one American call, strike K with K in cash. The second portfolio is long one American put and one stock. For the first portfolio, we assume the underlying pays no dividends, so there is no early exericse - or rather we choose not to. At expiry, if $S < K$ then our portfolio is worth $K \exp(rt)$ and otherwise we choose to exercise, and our portfolio is worth $S - K + K \exp(rt) = S + K(\exp(rt) - 1)$. Consider the second portfolio - if we do not exercise early, we have K if the put expires in the money and we exercise the right to sell our stock, otherwise we hold on to the stock and our portfolio is worth S. If we do exercise early, it is only so when the put is in the money, at which point we receive K at t^* and invest in the money market to get $K \exp(r(t - t^*))$ at expiry. See that the first portfolio is worth at least as much as the second portfolio, so $C + K \geq P + S$, or

$$C - P \geq S - K. \tag{5881}$$

Again recall the put-call parity relationship $c - p = S - K \exp(-rt)$. Then $p = c - S + K \exp(-rT)$. Since the $P > p$, we have

$$P \geq c - S + K \exp(-rT). \tag{5882}$$

We said when early exercise feature is worthless then the American option pricing is priced as the European one, so in the absence of dividends we have

$$C - P \leq S - K \exp(-rT). \tag{5883}$$

Using the lower bound found earlier - we have

$$S - K \leq C - P \leq S - K \exp(-rT). \tag{5884}$$

Here the put-call parity bounds on American options demonstrate that if the time to expiry is small, or rates are low, then the European one approximates American ones pretty well. This has to do with how favorable it is for the buyer of a put to do early exercise on a deep in the money option and invest his proceeds in the money market. When the stock pays dividends then our bound is modified:

$$S - D - K \leq C - P \leq S - K \exp(-rT), \tag{5885}$$

where D is the present value of dividends paid.

An interesting option structure is the box spread. Consider two strikes $K_1 < K_2$, and then construct a portfolio $c(K_1), -c(K_2), -p(K_1), p(K_2)$, and go through the mental exercise to verify that in all states of the world, the terminal value of the portfolio is $K_2 - K_1$. So

$$c(K_1) - c(K_2) - p(K_1) + p(K_2) = (K_2 - K_1)exp(-rT). \tag{5886}$$

The sole risk is in the interest rate movements. Note that only the European options perform this function.

14.2 Pricing Model

For those who have already taken the stochastic calculus pill (Chapter 13), we walk through the central arguments that lead to the pricing of a European put and call option. For those who have not done so, we will outline the main points such that our arguments remain tractable in this section, but the next page or so will probably not make any sense - fret not.

It all begins with the symmetric random walk (Definition 429), which is scaled (Definiton 431) and shown to be normal in distribution as the number of time steps $n \to \infty$ (Theorem 391), the continuous version of which is the Brownian motion (Definition 432). We saw the Brownian motion is not continuously differentiable, and there is non-trivial second order quadratic variation in the Brownian motion, a fact we summarise as $dW(t)dW(t) = dt$ (Theorem 397). Furthermore it was shown that the cross-variation of the Brownian motion with time and the quadratic variation of time itself is given by $dW(t)dt = dtdt = 0$ (Theorem 398). We showed how an exponentiated Brownian motion gives rise to the geometric Brownian motion which we typically assume as asset price dynamics, and how this leads to the (sampled) volatility of an asset price process (Definition 435). These are the central results of the first part in financial stochastic calculus.

The Ito integral is then introduced - we are not able to differentiate paths of the Brownian motion w.r.t to time, so we defined an integral directly w.r.t the Brownian motion such that we may specify an integral $I(t) = \int_0^t \Delta(u)dW(u)$, where $\Delta(u)$ is a simple process - we began with simple integrands (Definiton 439), and we observe that Ito integrals w.r.t to martingales are also martingales (Theorem 404). An important result is Ito isometry (Theorem 405) and the quadratic variation of the Ito integral, facts we write as $\mathbb{E}I^2(t) = \mathbb{E}\int_0^t Delta^2(u)du$ and $dI(t)dI(t) = \Delta^2(t)dt$ respectively. It is later extended to general integrands as a limit of the simple integrands, which now allows us to include portfolio processes (Definition 441). Most of this is to arrive at the important Ito-Lemma.

The above points are lost on many a traders, and they instead choose to begin with a memorization of the Ito-Lemma, which would mostly suffice in following the pricing arguments in the Black-Scholes Merton model. There are many variants, and we summarize: the Ito Doeblin formula for a differentiable $f(x)$ is written (Definition 442)

$$df(W(t)) = f'(W(t))dW(t) + \frac{1}{2}f''(W(t))dt. \tag{5887}$$

We can let f further depend on time t and so (Theorem 408)

$$df(t, W(t)) = f_t(t, W(t))dt + f_x(t, W(t))dW(t) + \frac{1}{2}f_{xx}(t, W(t))dt. \tag{5888}$$

An Ito process (Definition 443) is a stochastic process that evolves both w.r.t time and a Brownian motion, given

$$X(t) - X(0) = \int_0^t \Delta(u)dW(u) + \int_0^t \Theta(u)du. \tag{5889}$$

All the other important results can be obtained from the above results about cross and quadratic variation of Brownian motions. For instance the differential form of the Ito process is written $dX(t) = \Delta(t)dW(t) + \Theta(t)dt$. One may easily derive that the quadratic variation of the Ito process:

$$dX(t)dX(t) = (\Delta(t)dW(t) + \Theta(t)dt)(\Delta(t)dW(t) + \Theta(t)dt) = \Delta^2(t)dt \tag{5890}$$

using $dW(t)dW(t) = dt, dW(t)dt = dtdt = 0$ without going through the arduous proof (Theorem 409). Now we may define an Ito integral w.r.t to an Ito process (Definition 444), we write:

$$\int_0^t \Gamma(u)dX(u) = \int_0^t \Gamma(u)\Delta(u)dW(u) + \int_0^t \Gamma(u)\Theta(u)du. \tag{5891}$$

and the Ito Doeblin formula for this is (as we may expect) shown in Theorem 410

$$df(t, X(t)) = f_t(t, X(t))dt + f_x(t, X(t))dX(t) + \frac{1}{2}f_{xx}(t, X(t))dX(t)dX(t). \tag{5892}$$

It won't be difficult to express this in terms of dt and $dW(t)$ by expanding $dX(t)$ which we already know how to do.

We see how the geometric Brownian motion can be generalized in Example 650. There we see how the Ito Lemma can be used to relate the asset price process $S(t) = S(0) \exp \left\{ \int_0^t \sigma(s)dW(s) + \int_0^t \left(\alpha(s) - \frac{1}{2}\sigma^2(s) \right) ds \right\}$ to the asset price dynamics

$$dS(t) = \alpha(t)S(t)dt + \sigma(t)S(t)dW(t). \tag{5893}$$

The reverse derivation - obtaining the asset price process from asset price dynamics is given in Exercise 651.

We shall move on to the Black-Scholes-Merton derivation. We begin with some $X(t)$ cash at the outset, and follow its evolution across time holding some $\Delta(t)$ units of stock worth $S(t)$ at time t dynamically across time (Section 13.2.3.1). We then see how a portfolio of a single European call option might evolve (Section 13.2.3.2) - using Ito's Lemma. Now we want to construct a hedge portfolio for a short European call option - then naturally all we need to do is match their evolution dynamics by matching the drift and diffusion terms. It turns out that the hedge units required $\Delta(t)$ is precisely $c_x(t, S(t))$, and to match the drift terms, we arrive at the Black Scholes Merton partial differential equation (Equation 3631):

$$c_t(t, x) + rxc_x(t, x) + \frac{1}{2}\sigma^2 x^2 c_{xx}(t, x) = rc(t, x) \qquad \forall t \in [0, T), x \geq 0. \tag{5894}$$

In this discussion we did not raise the concern of the payoff related to the European call option at all - in fact this partial differentiation equation needs to be satisfied by all sorts of derivatives. It is a solution to the partial differentiation equation, parameterized by the boundary and terminal conditions relating to the specific derivative that determines the asset pricing formula. The specific solution to the PDE for European call option is verified in Section 13.2.4 and in Exercise 659, where we assert it is given by Equation 3633

$$c(t, x) = x\Phi(d_+(\tau, x)) - K \exp(-r\tau)\Phi(d_-(\tau, x)) \qquad 0 \leq t < T, x > 0, \tag{5895}$$

where $\tau = T - t$, Φ is normal c.d.f of $\Phi(0, 1)$ (see Section 6.17.6) and

$$d_\pm(\tau, x) = \frac{1}{\sigma\sqrt{\tau}} \left[\log \frac{x}{K} + \left(r \pm \frac{\sigma^2}{2} \right) \tau \right]. \tag{5896}$$

In Exercise 659 we work out the arduous calculus steps in deriving the option greeks from the closed form solution for European calls, giving

1. call delta: $\Phi(d_+(\tau, x))$,

2. call theta: $-rK \exp(-r\tau)\Phi(d_-(\tau, x)) - \frac{\sigma x}{2\sqrt{\tau}}\Phi'(d_+(\tau, x))$,

3. call gamma: $\Phi'(d_+(\tau, x))\frac{1}{x\sigma\sqrt{\tau}}$.

The other two first-order Greeks such as rho and vega were not explored there - particularly because rho effects are secondary to option price changes in the real world, and in the purest form the Black-Scholes-Merton assumes constant volatility. However, the vega term *is* in fact important to option price dynamics and we will explore them here, in addition to more second-order Greeks for completeness.

The mathematics of gamma scalping imposed by the curvature of the option pricing is explored in Example 660. An alternative way of arriving at the Black-Scholes-Merton PDE is presented in Exercise

661. One constructs a portfolio of a call option and short $\Delta(t)$ of stock and realises this portfolio is risk-free - hence it must earn the risk-free rate. The arguments in the proof are precise, but use Ito-Lemma. Again, it is not assumed one is familiar with stochastic calculus, so we will present the same argument but in an approximated, discrete form in this section, but whether you are familiar with one or the other, it is the motivation at which we arrive the PDE that is important. At this stage we only verified the solution to the BSM - now in Exercise 662 we derive it, and show that it is the discounted expected payoff. The reward from correctly predicting volatility distinct from the market implied volatility levels are highlighted by Example 663.

Most of the arguments outlined above should do the trick; for good measure we will go abit further. Section 13.3 introduces the concept of risk-neutrality via the change of measure, the important result being that derivatives are priced as risk-neutral expected payoffs, and is given the risk-neutral pricing formula (Equation 3976). Explaining exactly what this means would be too mathematical at this point, but the intuition of why risk-neutral probabilities arise in the first place is encapsulated well by the examples given in Exercise 724, which shows that actuarial value need not equal to the market price even when economic agents are rational investors. Actually we have covered the European option call pricing for time-varying rates and volatility using an extended-form of the Black-Scholes-Merton model, but we will leave this as reference (Exercise 679) and not use it here. The more important extension is when the stock presents dividends. Here it is the portfolio with the re-investment of dividends in money-market that is risk-neutral martingale (Section 13.3.7.1). The European option price hence takes an adjusted close-form solution (Equation 4181):

$$S(t)\exp(-a\tau)\Phi(d_+(\tau, x)) - \exp(-r\tau)K\Phi(d_-(\tau, x)). \tag{5897}$$

where τ is time to maturity and a is the continuous dividend yield. This is appropriate if the underlying is an index, for instance. For a stock option it is more likely that the dividends are lump-sum (Definition 13.3.7.3), and the pricing is abit more difficult (Equation 13.3.7.4):

$$c(t, x) = S(0)\Pi_{j=0}^{n-1}(1 - a_{j+1})\Phi(d_+) - \exp(-rT)K\Phi(d_-) \tag{5898}$$

with $d_\pm = \frac{1}{\sigma\sqrt{T}}\left[\log\frac{S(0)}{K} + \sum_{j=0}^{n-1}\log(1 - a_{j+1}) + \left(r \pm \frac{1}{2}\sigma^2\right)T\right]$, where $a_{j\in[n]}$ are the dividend rates prior to option expiry.

If one were to carefully go through each of the references above, then understanding the precise continuous time arguments should be no issue, albeit abit challenging. Either way, here we give an outline of the arguments that should be understandable for any reader, regardless of whether she has gone through referenced material. The binomial one-step option pricing model is given before in Exercise 724 to highlight the existence of a risk-neutral world - here we give some more details. Assume a stock trades at S today, and we short Δ units of stock and long one call option struck at $K = S$ (at the money). In the next time step, with probability p it has gross return u and with probability $1 - p$ it has gross return d. On the up move our portfolio is worth $C_u = \max(S_0 u - 100, 0) - \Delta S_0 u$, on the down move it is worth $C_d = \max(S_0 d - 100, 0) - \Delta S_0 d$, to be hedged we want:

$$C_u - \Delta S_0 u = C_d - \Delta S_0 d. \tag{5899}$$

Obviously the hedge ratio we require is

$$\Delta = \frac{C_u - C_d}{S_u - S_d}. \tag{5900}$$

Then we are indifferent to the stock price move and we are perfectly hedged in this two-state economy. This also means that our portfolio is risk-free, and should therefore earn the risk-free rate. Then the

present value of the portfolio is given by $\exp(-rT)(C_u - \Delta S_u)$. The current value of the portfolio is given by $C - \Delta S$. So the value of the call option today is given by

$$
\begin{align}
C &= \Delta S + \exp(-rT)(C_u - \Delta S_u) \tag{5901} \\
&= \exp(-rT)\left[\exp(rT)\Delta S + C_u - \Delta S_u\right]. \tag{5902}
\end{align}
$$

Define $p := \frac{\exp(rt)-d}{u-d}$, and we have

$$
\begin{align}
C &= \exp(-rT)\left[\exp(rT)\frac{C_u - C_d}{S_u - S_d}S + C_u - \frac{C_u - C_d}{S_u - S_d}S_u\right] \tag{5903} \\
&= \exp(-rT)\left[\exp(rT)\frac{C_u - C_d}{u - d} + C_u - \frac{C_u - C_d}{u - d}u\right] \tag{5904} \\
&= \exp(-rT)\left[\frac{\exp(rT) + u - d - u}{u - d}C_u + \frac{\exp(rT)(-1) + u}{u - d}\right] \tag{5905} \\
&= \exp(-rT)\left[pC_u + (1-p)C_d\right]. \tag{5906}
\end{align}
$$

Interestingly p does not involve any probability estimations - but rather depend on the components related to the magnitude of up and down moves; it is the volatility that matters. If we interpret p to be some probability, then one may see that

$$
pS_u + (1-p)S_d = \frac{\exp(rt) - d}{u - d}S_u + \frac{u - \exp(rt)}{u - d}S_d = \frac{\exp(rt)Su - \exp(rt)Sd - Sdu + Sdu}{u - d} = \exp(rt)S.
$$

This is an expectation of the stock price in the next time step *in the risk-neutral world*. The risk-neutral pricing assumes that the stock grows at the risk free rate. It is the limit of the binomial model that gives rise to the geometric Brownian motion assumed for asset prices (Theorem 392). One may use this to price paths to obtain the pricing of different kinds of options (such as put options) by changing the payoff, as well as different exercise styles such as that of an American option - however, we will not explore this option here. It should be enough to keep in mind that as the number of time steps is allowed to reach infinity, the binomial tree results converges to our BSM result.

We derive the BSM model using some discrete time arguments, parallel to what we did in Exercise 661. We start by assuming a delta-hedged position holding a call option and short Δ stock units, so portfolio is worth $C - \Delta S_t$. The change in the portfolio value is given by

$$
C(S_{t+1}) - C(S_t) - \Delta(S_{t+1} - S_t) - r(C - \Delta S_t), \tag{5907}
$$

corresponding to change in the call price, change in the underlying stock price and interest accrued. We may approximate this first term $C(S_{t+1}) - C(S_t)$ by second-order Taylor expansion w.r.t to asset price change and change in call value w.r.t to time passing, so we may write

$$
\Delta(S_{t+1} - S_t) + \frac{1}{2}\Gamma(S_{t+1} - S_t)^2 + \theta - \Delta(S_{t+1} - S_t) - r(C - \Delta S_t) \tag{5908}
$$

$$
\frac{1}{2}\Gamma(S_{t+1} - S_t)^2 + \theta - r(C - \Delta S_t) \tag{5909}
$$

where $\Delta = \frac{\delta c}{\delta S}, \Gamma = \frac{\delta^2 c}{\delta S^2}, \theta = -\frac{\delta C}{\delta t}$. On average we have $\left(\frac{S_{t+1} - S_t}{S}\right)^2 \approx \sigma^2$, so we let $(S_{t+1} - S_t)^2$ take $\sigma^2 S^2$, and substituting we obtain

$$
\frac{1}{2}\Gamma\sigma^2 S^2 + \theta - r(C - \Delta S_t) \stackrel{!}{=} 0. \tag{5910}
$$

and this should equal zero since the portfolio is financed with zero initial capital. We can refer to the above pargraphs and see this is just the Black-Scholes-Merton PDE. We assumed that the underlying

can be traded in any size, shortable and that it is deeply liquid, with a common interest for borrowing and investing, and that there were no dividends. We also assumed that the price is continuous and that asset prices are log-normal. We assumed that volatility is constant, and that we live in a tax-free world. Clearly this world does not exist, and it is these assumptions that cause the inaccuracy of BSM models in pricing options in the real world. Even though it might not be that accurate in pricing a single option, the BSM model is still a great tool in allowing us to evaluate the relative pricing between different options, including those of different underlying, strike or expiry dates. It is likely that the errors in the BSM model for an option correlates to the errors in the estimation relating to a similar option, and the estimate of the spread value may be fairly robust. This allows us to make statements such as 'option A looks to be more rich in volatility compared to option B', and as a volatility trader this should be critically helpful.

14.3   Option Greeks

Our discussion so far was mostly on call options, but we know what the put-call parity relationship is, so we do also know how to price a put option (Equaiton 3731). The European options give closed form solutions, and the American ones detract only mildly from the European ones, so we will stick to our analysis on European options. We will also discard with the convention of c, C, p, P and treat them indiscriminately.

So let's repeat some of the more salient points:

$$c(t, x) = x\Phi(d_+(\tau, x)) - K\exp(-r\tau)\Phi(d_-(\tau, x)) \tag{5911}$$

$$p(t, x) = K\exp(-r\tau)\Phi(-d_-(\tau, x)) - x\Phi(-d_+(\tau, x)) \tag{5912}$$

with

$$d_\pm(\tau, x) = \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} + \left(r \pm \frac{\sigma^2}{2}\right)\tau\right]. \tag{5913}$$

For graphical interpretation of these formulas, use the Internet. As mentioned, we have already done the hard work to compute option greeks in Exercise 659 on European calls, and the same for European puts by the put-call parity relationship (Exercise 664) :

1. $\frac{\delta C}{\delta S} = \Delta_c = \Phi(d_+(\tau, x))$,

2. $\frac{\delta P}{\delta S} = \Delta_p = \Phi(d_+(\tau, x)) - 1$,

3. $-\frac{\delta C}{\delta t} = \theta_c = -rK\exp(-r\tau)\Phi(d_-(\tau, x)) - \frac{\sigma x}{2\sqrt{\tau}}\Phi'(d_+(\tau, x))$,

4. $-\frac{\delta P}{\delta t} = \theta_p = rK\exp(-r\tau)\Phi(-d_-(\tau, x)) - \frac{\sigma x}{2\sqrt{\tau}}\Phi'(d_+(\tau, x))$.

5. $\frac{\delta^2 C}{\delta S^2} = \Gamma_c = \frac{\delta^2 P}{\delta S^2} = \Gamma_p = \Phi'(d_+(\tau, x))\frac{1}{x\sigma\sqrt{\tau}}$.

A rough rule of thumb for an ATM call option delta is about 0.50, and this approximation is more accurate the closer the option is to expiry. Gamma is the measure of the convexity of the curve interpolating between a deep-in-the money option and a far-out-of-the money option. As such, one may think of Gamma as a measure of how option-like an option is. When option is near expiry, Gamma is high for ATM options and is roughly bell-shaped w.r.t to the strike, although the peak lies slightly to the left of the strike. The maximum is actually at

$$S^* = K\exp\left(-(r + \frac{3\sigma^2}{2})t\right). \tag{5914}$$

To make comparisons across different underlying stock levels, we might take $S \cdot \Gamma$ to normalize the stock price component that contributes to Gamma absolute levels. The Gamma scalping effects are explored in Exercise 660. Theta is the negative partial derivative of the option value w.r.t time. It is usually presented as a normalized value over a day, so we may take $\tilde{\theta} = \frac{\theta}{365}$. Some argue that θ effects should not be fully accounted over the weekends. We won't bother with the philosophical discussion. Theta is the minimum at $\hat{S} = K \exp\left((r + \frac{3\sigma^2}{2})t\right)$.

Consider Equation 5909 that describes our PnL approximately:

$$\frac{1}{2}\Gamma\sigma^2 S^2 + \theta - r(C - \Delta S_t) = 0. \tag{5915}$$

We are interested in PnL movements over short periods of time - we may ignore the third term. The two Greeks that dominate our PnL is written

$$\frac{1}{2}\Gamma(S_{t+1} - S_t)^2 + \theta = \frac{1}{2}\Gamma(\Delta_S)^2 + \theta = 0. \tag{5916}$$

Then $\theta = -\frac{1}{2}S^2(\frac{\Delta_S}{S})^2\Gamma \approx -\frac{1}{2}S^2\sigma_{imp}^2\Gamma$ where the approximation holds on average. The gamma and theta effects are priced to cancel each other out. The PnL for a delta hedged position can approximately be written as such:

$$PnL = \frac{1}{2}\Gamma(\Delta_S)^2 + \theta(\Delta_t) \tag{5917}$$

$$= \frac{1}{2}\Gamma S^2\left[\left(\frac{\Delta_S}{S}\right)^2 - \sigma_{imp}^2(\Delta t)\right]. \tag{5918}$$

The breakeven price movement in the underlying for our delta-hedged portfolio to have net zero PnL is the level where the squared daily return equals implied variance. We may also relate this to the profits from correctly calibrating a mispriced market implied volatility - see Exercise 663, where we argue that we may earn at instantaneous rate of $\frac{1}{2}(\sigma_{rlz}^2 - \sigma_{imp}^2)S_t^2\Gamma$.

So far we have introduced delta, gamma and theta. To get vega and rho, we need to do some more mathematics. We know from Equation 3636 that

$$d_-(\tau, x) = d_+ - \sigma\sqrt{\tau}, \tag{5919}$$

$$d_+(\tau, x) = d_- + \sigma\sqrt{\tau}. \tag{5920}$$

So we have

$$\frac{\delta d_-}{\delta\sigma} = \frac{\delta d_+}{\delta\sigma} - \sqrt{\tau}, \qquad \frac{\delta d_+}{\delta r} = \frac{\delta d_-}{\delta r}. \tag{5921}$$

We assert an additional relation.

Lemma 34. *For d_+, d_- given by $d_\pm(\tau, x) = \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} + \left(r \pm \frac{\sigma^2}{2}\right)\tau\right]$ we have*

$$S_0\Phi'(d_+) = K\exp(-rt)\Phi'(d_-). \tag{5922}$$

Proof. First see that

$$d_-^2 - d_+^2 = (d_- - d_+)(d_- + d_+) \tag{5923}$$

$$= (-\sigma\sqrt{\tau})(2d_+ - \sigma\sqrt{\tau}) \tag{5924}$$

$$= (-\sigma\sqrt{\tau})\left[\frac{2\ln\frac{S_0}{K} + 2(r + \frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}} - \sigma\sqrt{\tau}\right] \tag{5925}$$

$$= -2\left[\ln\frac{S_0}{K} + r\tau\right]. \tag{5926}$$

Then since $\Phi'(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$, see that

$$\ln \frac{\Phi'(d_+)}{\Phi'(d_-)} = \frac{1}{2} \left(d_-^2 - d_+^2 \right) = - \left(\ln \frac{S_0}{K} + r\tau \right). \tag{5927}$$

The result follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Let's continue by computing vega, the partial derivative of an option value w.r.t implied volatility. See that

$$\begin{aligned}
\frac{\delta C}{\delta \sigma} &= S_0 \Phi'(d_+) \frac{\delta d_+}{\delta \sigma} - K \exp(-r\tau) \Phi'(d_-) \frac{\delta d_-}{\delta \sigma} & &\text{(5928)} \\
&= S_0 \Phi'(d_+) \frac{\delta d_+}{\delta \sigma} - K \exp(-r\tau) \Phi'(d_-) (\frac{\delta d_+}{\delta \sigma} - \sqrt{\tau}) & &\text{Equation 5921} & &\text{(5929)} \\
&= (S_0 \Phi'(d_+) - K \exp(-r\tau) \Phi'(d_-)) \frac{\delta d_+}{\delta \sigma} + \sqrt{\tau} K \exp(-r\tau) \Phi'(d_-) & &\text{(5930)} \\
&= \sqrt{\tau} S_0 \Phi'(d_+). \quad \text{Lemma 34.} & &\text{(5931)}
\end{aligned}$$

Clearly the put call parity asserts that the vega of a European call and European put are identical. So we have

$$\nu_C = \nu_P = \sqrt{\tau} S_0 \Phi'(d_+). \tag{5932}$$

Vega is typically scaled so that it represents the dollar change in option value per percentage point change in implied volatility. It can be shown that $\nu = \Gamma \sigma S^2 \tau$. As volatility acts on the underlying process over time, the sensitivity of the call price to volatility changes, which is precisely what vega measures, increases w.r.t to time. An ATM option that is close to expiry has large gamma and small vega, and vice versa. This is the reason why purchasing short dated options are referred to as buying gamma, and buying long dated options are referred to as buying vega.

Often traders are more interested in the total profit from correctly predicting mispriced volatility rather than the instantaneous profit given $\frac{1}{2}(\sigma_{rlz}^2 - \sigma_{imp}^2) S_t^2 \Gamma$. We may approximate it in vega terms as follows: suppose we have some call priced at σ_{imp} but σ_{rlz} was realized, then for $\delta = \sigma_{rlz}^2 - \sigma_{imp}^2$ we have

$$C(\sigma_{imp}^2 + \delta) = C(\sigma_{imp}^2) + \delta \frac{\delta C}{\delta \sigma_{rlz}^2}. \tag{5933}$$

All else equal, the profit is given by $\delta \frac{\delta C}{\delta \sigma_{rlz}^2}$ and we have

$$\frac{\delta C}{\delta \sigma_{rlz}^2} = \frac{\delta C}{\delta \sigma_{rlz}} \frac{\delta \sigma_{rlz}}{\delta \sigma^2} = \nu \cdot \frac{1}{2\sigma_{rlz}}. \tag{5934}$$

So the PnL is approximated

$$\begin{aligned}
\delta \cdot \nu \frac{1}{2\sigma_{rlz}} &= \frac{\nu}{2\sigma_{rlz}} \left(\sigma_{rlz}^2 - \sigma_{imp}^2 \right) & &\text{(5935)} \\
&= \frac{\nu}{2\sigma_{rlz}} \left(\sigma_{rlz} - \sigma_{imp} \right) \left(\sigma_{rlz} + \sigma_{imp} \right) & &\text{(5936)} \\
&\approx \nu \left(\sigma_{rlz} - \sigma_{imp} \right), & &\text{(5937)}
\end{aligned}$$

the approximate, or average total profit from accurate prediction of volatility over implied levels.

To compute rho we take $\rho_c =$

$$\begin{aligned}
\frac{\delta C}{\delta r} &= S_0 \Phi'(d_+) \frac{\delta d_+}{\delta r} - K \exp(-r\tau) \left[\Phi'(d_-) \frac{\delta d_-}{\delta r} - \tau \Phi(d_-) \right] & &\text{(5938)} \\
&= \frac{\delta d_+}{\delta r} \left[S_0 \Phi'(d_+) - K \exp(-r\tau) \Phi'(d_-) \right] + \tau K \exp(-r\tau) \Phi(d_-) & &\text{(5939)} \\
&= \tau K \exp(-r\tau) \Phi(d_-). \quad \text{Lemma 34} & &\text{(5940)}
\end{aligned}$$

By the put-call parity we have $\frac{\delta C}{\delta r} - \frac{\delta P}{\delta r} = \tau K \exp(-r\tau)$ so $\rho_P =$

$$
\begin{align}
\frac{\delta P}{\delta r} &= \frac{\delta C}{\delta r} - \tau K \exp(-r\tau) \tag{5941} \\
&= \tau K \exp(-r\tau) \left(\Phi(d_-) - 1 \right) \tag{5942} \\
&= \tau K \exp(-r\tau) \left(1 - \Phi(-d_-) - 1 \right) \tag{5943} \\
&= -\tau K \exp(-r\tau) \Phi(-d_-). \tag{5944}
\end{align}
$$

Rho is typically scaled to be the dollar change in option value per percentage point change in interest rates.

14.4   Volatility Measurement and Forecasting

If the mathematics of the previous section were lost on traders, the primary realization should be that volatility is an important, if not *the* most important input into option pricing. The estimation of volatility that is subsequently more accurate than implied by the market should, on avarege, lead to monetary profits - but not much has been said so far about volatility measurement or forecasting it. This section shall be dedicated to that front.

14.4.1   Measurements of Volatility

Volatility is a concept, but it is not directly measurable. We may assume the true price processes are continuous and unknown, whereas the data we obtain are samples of this price process measured at discrete time intervals. Our computation for the estimate of population volatility from sample data use statistical estimators. There are many statistical estimators, differing in bias (see Definition 371) and efficiency (see Definition 336). The bias of an estimator is related to the systematic over/under estimation of the population parameter, while the efficiency indicates the convergence properties of our statistical estimator - which is related to the sampling error, or the average amount of noise in our statistical estimator. The *forecasting* of volatility requires a model that captures our notion of the *behavior* of volatility processes as measured by our choice of statistical *estimator*.

The definition for volatility is given as the square root of variance of logarithmic returns. The sample variance is given as

$$
s^2 = \frac{1}{N} \sum_{i=1}^{n} (x_i - \bar{x})^2, \tag{5945}
$$

where s is the volatility and x_i are logarithmic returns, typically measured by close-to-close prices, and $\bar{x}$ is the sample mean for the logarithmic returns. n is the size of the sample used for volatility estimation. Normally, volatility is presented in annualized terms, and variance is (assumed) additive, so we need a scalarization factor. If s^2 in Equation 5945 was computed with daily returns, then the annualized *variance* would be $252 \cdot s^2$ (because there are typically 252 trading days in a year) and the annualized *volatility* would be $\sqrt{252} \cdot s$. The sample variance is an example of a *statistical estimator* for the population variance parameter σ^2. Given logarithmic returns r, we know that $\sigma^2 = \mathbb{E}[r^2] - \mathbb{E}[r]^2$, but often $\mathbb{E}[r]$ is very close to zero. It is even smaller when we square it, so $\sigma^2 \approx \mathbb{E}[r^2]$ and we often treat the mean return as zero - this is rather convenient, and also standard practice, since $\bar{x}$, the sample mean of return in Equation 5945 is very noisy. We can typically improve the accuracy of measurement

by removing the noisy parameter; so that the estimator we use becomes

$$s^2 = \frac{1}{n} \sum_i^n x_i^2. \tag{5946}$$

But the sample variance is biased (Exercise 343) - note that this does not require normality assumptions, although in the derivation of the Black-Scholes model we assumed that returns are normal. To get an unbiased estimate we take

$$\hat{\sigma}^2 = \frac{n}{n-1} s^2. \tag{5947}$$

$\hat{\sigma}^2$ is unbiased estimator for the population variance, so $\mathbb{E}\hat{\sigma}^2 = \sigma^2$. One should note that taking the square root of this does *not* yield an unbiased estimator of the population volatility, due to the concavity imposed by $\sqrt{\cdot}$. In particular, by Jensen's inequality (Theorem 336) we have

$$\mathbb{E}\hat{\sigma} = \mathbb{E}\sqrt{\hat{\sigma}^2} \le \sqrt{\mathbb{E}\hat{\sigma}^2} = \sqrt{\sigma^2} = \sigma. \tag{5948}$$

If the returns are assumed normal then it turns out that

$$\mathbb{E}\hat{\sigma} = \sqrt{\frac{2}{n}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \sigma, \tag{5949}$$

so that $\frac{\hat{\sigma}}{k}$ is unbiased estimator of population standard deviation, where $k = \sqrt{\frac{2}{n}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$ is the bias correction factor. A rough value of n commonly used in practice is about 30, and for this value k is slightly greater than 0.96. In volatility trading, it is often also customary to set n to the dte of the option. Although unbiased, our close-to-close estimator is an inefficient estimator for portfolio volatility - so the convergence is rather slow. We need more samples, so we can increase n - but we are not measuring volatility over a stationary process. As n increases, we are capturing market dynamics that are increasingly further from the current regime - and hence the tradeoffs between convergence and sampling relevant data to prevailing market regimes. The variance of our estimator is roughly given

$$Var(\hat{\sigma}) \approx \frac{\sigma^2}{2n}. \tag{5950}$$

The tradeoff may be alleviated in two directions - one is sampling at higher frequencies, or using a different statistical estimator all together.

An alternative volatility estimator is the Parkinson estimator, given

$$\hat{\sigma} = \sqrt{\frac{1}{4n \ln 2} \sum_i^n \left(\ln \frac{h_i}{l_i} \right)^2} \tag{5951}$$

where h_i, l_i are the high and low prices in the data bar. Here the volatility is measured by the bar range, which captures a different aspect of price volatility. By sampling two data points per OHLC bar, for a fixed window of time n, we have more data points, and hence better convergence properties. Assuming a driftless geometric brownian motion, the Parkinson estimator is shown to be about five times more efficient than the close to close estimator. However, when the GBM assumptions are violated, this efficiency benefit is muted. Furthermore, Parkinson estimator assumes continuous prices, but since high and low prices are obtained from discretely sampled market data during regular trading hours, it is highly probably that the true, underlying extreme prices were not recorded, especially during overnight trading - hence Parkinson estimator tends to underestimate true volatility. The Parkinson estimator should be

divided by correction factors to reduce the impact of bias, which can be rather significant. For $n = 30$, the correction factor is roughly 0.75. Another issue other than the bias is the tendency for jumps in the opening prices in markets.

Another popular volatility estimator combines the information from both the bar range (like the Parkinson estimator) and close-to-close and is named the Garman-Klass estimator

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_i^n \frac{1}{2}\left(\ln\frac{h_i}{l_i}\right)^2 - \frac{1}{n}\sum_i^n (2\ln 2 - 1)\left(\ln\frac{c_i}{c_{i-1}}\right)^2}, \tag{5952}$$

where c_i is the closing price sampled at period i. To relax the assumptions on a driftless GBM process, the Rogers-Satchel estimator gives

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_i^n \left(\ln\frac{h_i}{c_i}\right)\left(\ln\frac{h_i}{o_i}\right) + \left(\ln\frac{l_i}{c_i}\right)\left(\ln\frac{l_i}{o_i}\right)} \tag{5953}$$

which allow price trends. To further account for possibility of jumps in opening prices, the Yang-Zhang estimator gives

$$\hat{\sigma} = \sqrt{\hat{\sigma}_o^2 + k\hat{\sigma}_c^2 + (1-k)\hat{\sigma}_{rs}^2} \tag{5954}$$

where

$$\hat{\sigma}_o^2 = \frac{1}{n-1}\sum_i^n \left(\ln\frac{o_i}{c_{i-1}} - \frac{1}{n}\sum_i^n \ln\frac{o_i}{c_{i-1}}\right)^2, \tag{5955}$$

$$\hat{\sigma}_c^2 = \frac{1}{n-1}\sum_i^n \left(\ln\frac{c_i}{o_i} - \frac{1}{n}\sum_i^n \ln\frac{c_i}{o_i}\right)^2, \tag{5956}$$

$$\hat{\sigma}_{rs}^2 = \frac{1}{n}\sum_i^n \left(\ln\frac{h_i}{c_i}\right)\left(\ln\frac{h_i}{o_i}\right) + \left(\ln\frac{l_i}{c_i}\right)\left(\ln\frac{l_i}{o_i}\right), \tag{5957}$$

$$k = \frac{0.34}{1.34 + \frac{n+1}{n-1}}. \tag{5958}$$

We can see that Yang-Zhang estimator is a mixture of the overnight volatility, intraday volatility and Rogers-Satchel estimator. The Garman-Klass estimator, Rogers-Satchel estimator and Yang Zhang estimator are all biased downwards, since they use price extremes on assumptions of continuous prices, but in reality the prices are sampled discretely only at market hours. Although each successove estimator is constructed to address a shortcoming of the last, in practice, there is no 'best' estimator. The efficency gains and improvements assume a geometric Brownian motion, but when tested on simulated market data behaving closer to real market data, their differences are less distinct. Furthermore it is impossible to benchmark their performance on real market data, as the true value (and hence errors) are unknown. If they were known - we would not need an estimator in the first place. A fine estimator would be the mean of all of these estimators. It is more important that we know what these estimators are telling us. For instance, if the Parkinson volatility is higher than the close-to-close volatility, we might suspect that a large proportion of volatility is driven by intraday movements.