In the absence of transaction costs, this option price has a similarity solution. Show that the similarity solution can still be found when the model includes transaction costs of the form \( k_1 + k_2 S \).

3. Solve the Black–Scholes equation for the price of a European call option with a transaction cost of \( k_2 S \) as follows, in the case when

\[
K = \frac{k}{\sigma \sqrt{t}}
\]

is a small parameter. The option price satisfies the equation

\[
\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - K \sigma^2 S^2 \sqrt{\frac{2}{\pi}} \frac{\partial^2 V}{\partial S^2} + r S \frac{\partial V}{\partial S} - r V = 0.
\]

Write

\[ V(S, t) = V_0(S, t) + K V_1(S, t) + \ldots, \]

and solve the problem for \( V_0 \). What problem must we then solve for \( V_1 \)?

### CHAPTER 22
volatility smiles and surfaces

#### In this Chapter...
- volatility smiles and skews
- the volatility surface
- how to determine the volatility surface that gives prices of options that are consistent with the market

#### 22.1 INTRODUCTION

One of the erroneous assumptions of the Black–Scholes world is that the volatility of the underlying is constant. This can be seen in any statistical examination of time-series data for assets, regardless of the sophistication of the analysis. This varying volatility is also observed indirectly through the market prices of traded contracts. In this chapter we are going to examine the relationship between the volatility of the underlying asset and the prices of derivative products. Since the volatility is not directly observable, and is certainly not predictable, we will try to exploit the relationship between prices and volatility to determine the volatility from the market prices. This is the exact inverse of what we have done so far. Previously we modeled the volatility and then found the price, now we take the price and deduce the volatility.

#### 22.2 IMPLIED VOLATILITY

In the Black–Scholes world of constant volatility, the value of a European call option is simply

\[
V(S, t; \sigma, r; E, T) = SN(d_1) - E e^{-r(T-t)} N(d_2),
\]

where

\[
d_1 = \frac{\log(S/E) + (r + \frac{1}{2} \sigma^2)(T-t)}{\sigma \sqrt{T-t}}
\]

and

\[
d_2 = \frac{\log(S/E) + (r - \frac{1}{2} \sigma^2)(T-t)}{\sigma \sqrt{T-t}}.
\]

I have given the function \( V \) six arguments, the first two are the independent variables, the second two are parameters of the asset and the financial world, the last two are specific to the contract in question. All but \( \sigma \) are easy to measure (\( r \) may be a bit inaccurate but the price is typically not too sensitive to this). If we know \( \sigma \) then we can calculate the option price. Conversely, if we know the option price \( V \) then we can calculate \( \sigma \). We can do this because
the value of a call option is monotonic in the volatility. Provided that the market value of the option is greater than \( \max(S - E^{-r(T-t)}, 0) \) and less than \( S \) there is one, and only one, value for \( \sigma \) that makes the theoretical option value and the market price the same. This is called the implied volatility. One is usually taught to think of the implied volatility as the market's view of the future value of volatility. Yes and no. If the 'market' does have a view on the future of volatility then it will be seen in the implied volatility. But the market also has views on the direction of the underlying, and also responds to supply and demand. Let me give examples.

In one month's time there is to be an election, it is not clear who will win. If the right wing party are elected markets will rise, if the left wing are successful, markets will fall. Before the election the market assumes the middle ground, splitting the difference. In fact, little trading occurs and markets have a very low volatility. But option traders know that after the election there will be a lot of movement one way or the other. Prices of both calls and puts are therefore high. If we back out implied volatilities from these option prices we see very high values. Actual and implied volatilities are shown in Figure 22.1 for this scenario.

Traders may increase option prices to reflect the expected sudden moves but if we are only observing implied volatilities then we are getting the underlying story very wrong indeed. This illustrates the fact that if you want to play around with prices there is only one parameter you can fudge, the volatility. As long as it is not too out of line compared with implied volatilities of other products no-one will disbelieve it.

Regardless of a market maker's view of future events, he is at the mercy of supply and demand. If everyone wants calls, then it is only natural for him to increase their prices. As long as he doesn't violate put–call parity (either with himself or with another market maker) who's to know that it is supply and demand driving the volatility?

![Graph](image.png)

**Figure 22.1** Actual and implied volatilities just before and just after an anticipated major news item.

### 22.3 TIME-DEPENDENT VOLATILITY

In Table 22.1 below are the market prices of European call options\(^1\) with one, four and seven months until expiry. All have strike prices of 105 and the underlying asset is currently 106.25. The short-term interest rate over this period is about 5.6%.

As can easily be confirmed by substitution into the Black–Scholes call formula, these prices are consistent with volatilities of 21.2%, 20.5% and 19.4% for the one-, three- and seven-month options respectively. Clearly these prices cannot be correct if the volatility is constant for the whole seven months. What is to be done?

The simplest adjustment we can make to the Black–Scholes world to accommodate these prices (without any serious effect on the theoretical framework) is to assume a time-dependent, deterministic volatility. Let's assume that volatility is a function of time:

\[
\sigma(t) = \sigma(t) e^{-\alpha t}
\]

As explained in Chapter 8 the Black–Scholes formulae are still valid when volatility is time dependent provided we use

\[
\sqrt{\frac{1}{T-t} \int_t^T \sigma^2(\tau) d\tau}
\]

in place of \( \sigma \), i.e. now use

\[
d_1 = \frac{\log(S/E) + r(T-t) + \frac{1}{2} \int_t^T \sigma^2(\tau) d\tau}{\sqrt{\int_t^T \sigma^2(\tau) d\tau}}
\]

and

\[
d_2 = \frac{\log(S/E) + r(T-t) - \frac{1}{2} \int_t^T \sigma^2(\tau) d\tau}{\sqrt{\int_t^T \sigma^2(\tau) d\tau}}
\]

![Table](image1.png)

**Table 22.1** Market prices of European call options; see text for details.

<table>
<thead>
<tr>
<th>Expirel</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 month</td>
<td>3.50</td>
</tr>
<tr>
<td>3 months</td>
<td>5.76</td>
</tr>
<tr>
<td>7 months</td>
<td>7.97</td>
</tr>
</tbody>
</table>

\(^1\)I'm only using call options because put option prices should follow from the prices of the calls by put–call parity, therefore there will not be any more volatility information in the prices of puts than is already present in the prices of calls. Also the following analysis only really works if the options are European. Since most equity options are American we cannot use put prices because the values of American and European puts are different. In the absence of dividends the two call prices are the same.
In our example, all we need to do to ensure consistent pricing is to make
\[ \sqrt{\frac{1}{T-t} \int_t^T \sigma(t')^2 \, dt'} = \text{implied volatilities.} \]

We do this 'fitting' at time \( t^* \), and if I write \( \sigma_{\text{imp}}(t^*, T) \) to mean the implied volatility measured at time \( t^* \) of a European option expiring at time \( T \) then
\[
\sigma(t) = \sqrt{\sigma_{\text{imp}}(t^*, T)^2 + 2(t - t^*) \sigma_{\text{imp}}(t^*, T) \frac{d\sigma_{\text{imp}}(t^*, T)}{dt}}
\]

Practically speaking, we do not have a continuous (and differentiable) implied volatility curve. We have a discrete set of points (three in the above example). We must therefore make some assumption about the term structure of volatility between the data points. Usually one assumes that the function is piecewise constant or linear. If we have implied volatility for expiries \( T_i \) and we assume the volatility curve to be piecewise constant then
\[
\sigma(t) = \sqrt{\frac{(T_i - t^*) \sigma_{\text{imp}}(t^*, T_i)^2 - (T_{i-1} - t^*) \sigma_{\text{imp}}(t^*, T_{i-1})^2}{T_i - T_{i-1}}} \quad \text{for} \quad T_{i-1} < t < T_i
\]

22.4 VOLATILITY SMILES AND SKIWS

Now let me throw the cat among the pigeons. Continuing with the example above, suppose that there is also a European call option struck at 100 with an expiry of seven months and a price of 11.48. This corresponds to a volatility of 20.8% in the Black–Scholes equation. Now we have two conflicting volatilities up to the seven-month expiry, 19.4% and 20.8%. Clearly we cannot adjust the time dependence of the volatility in any way that is consistent with both of these values. What else can we do? Before I answer this, we'll look at a few more examples.

Concentrating on the same example, suppose that there are call options traded with an expiry of seven months and strikes of 90, 92.5, 95, 97.5, 100, 102.5, 105, 107.5 and 110. In Figure 22.2 I plot the implied volatility of these options against the strike price (the actual option prices do not add anything to our insight so I haven’t given them).

The shape of this implied volatility versus strike curve is called the smile. In some markets it shows considerable asymmetry, a skew, and sometimes it is upside down in a brown. The general shape tends to persist for a long time in each underlying.

If we managed to accommodate implied volatility that varied with expiry by making the volatility time dependent perhaps we can accommodate implied volatility that varies with strike by making the volatility asset-price dependent. This is exactly what we'll do. Unfortunately, it's much harder to make analytical progress except in special cases; if we have \( \sigma(S) \) then rarely can we solve the Black–Scholes equation to get nice closed-form solutions for the values of derivative products. In fact, we may as well go all the way and assume that volatility is a function of both the asset and time, \( \sigma(S, t) \).

22.5 VOLATILITY SURFACES

We can show implied volatility against both maturity and strike in a three-dimensional plot. One is shown in Figure 22.3. This implied volatility surface represents the constant value of volatility that gives each traded option a theoretical value equal to the market value.

We saw how the time dependence in implied volatility could be turned into a volatility of the underlying that was time dependent, i.e., we deduced \( \sigma(t) \) from \( \sigma_{\text{imp}}(t^*, T) \). Can we similarly deduce \( \sigma(S, t) \) from \( \sigma_{\text{imp}}(t^*, E, T) \) the implied volatility at time \( t^* \)? If we could, then we might want to call it the local volatility surface \( \sigma(S, t) \). This local volatility surface can be thought of...
as the market’s view of the future value of volatility when the asset price is $S$ at time $t$ (which, of course, may not even be realized). The local volatility is also called the forward volatility or the forward forward volatility.

### 22.6 Backing Out the Local Volatility Surface from European Call Option Prices

To back out the local volatility surface from the prices of market traded instruments I am going to assume that we have a distribution of European call prices of all strikes and maturities. This is not a realistic assumption but it gets the ball rolling. These prices will be denoted by $V(E, T)$. I could use puts but these can be converted to call prices by put–call parity. This notation is vastly different from before. Previously, we had the option value as a function of the underlying and time. Now the asset and time are fixed at $S^*$ and $t^*$, today’s values. I will use the dependence of the market prices on strike and expiry to calculate the volatility structure.

I will assume that the risk-neutral random walk for $S$ is

$$dS = rS dt + \sigma(S, t)S dX.$$  

This is our usual one-factor model for which all the building blocks of delta hedging and arbitrage-free pricing hold. The only novelty is that the volatility is dependent on the level of the asset and time.

In the following, I am going to rely heavily on the transition probability density function $p(S^*, t^*; S, T)$ for the risk-neutral random walk. Note that the backward variables are fixed at today’s values and the forward time variable is $T$. Recalling that the value of an option is the present value of the expected payoff, I can write

$$V(E, T) = e^{-r(T-t)} \int_0^\infty \max(S - E, 0) p(S^*, t^*; S, T) dS \quad (22.2)$$

We are very lucky that the payoff is the maximum function so that after differentiating with respect to $E$ we get

$$\frac{\partial V}{\partial E} = -e^{-r(T-t)} \int_E^\infty p(S^*, t^*; S, T) dS.$$  

And after another differentiation, we arrive at

$$p(S^*, t^*; E, T) = e^{r(T-t)} \frac{\partial^2 V}{\partial E^2} \quad (22.3)$$

Before even calculating volatilities we can find the transition probability density function. In a sense, this is the market’s view of the future distribution. But it’s the market view of the risk-neutral distribution and not the real one. An example is plotted in Figure 22.4.

![Figure 22.4](image)

The next step is to use the forward equation for the transition probability density function, the Fokker–Planck equation,

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial S^2} (\sigma^2 S^2 p) - \frac{\partial}{\partial S} (r S p) . \quad (22.4)$$

Here $\sigma$ is our, still unknown, function of $S$ and $t$. However, in this equation $\sigma(S, t)$ is evaluated at $t = T$.

From (22.2) we have

$$\frac{\partial V}{\partial t} = -rV + e^{-r(T-t)} \int_E^\infty (S - E) \frac{\partial p}{\partial E} dS.$$  

This can be written as

$$\frac{\partial V}{\partial t} = -rV + e^{-r(T-t)} \int_E^\infty \left( \frac{\partial^2}{\partial S^2} (\sigma^2 S^2 p) - \frac{\partial}{\partial S} (r S p) \right) (S - E) dS$$

using the forward equation (22.4). Integrating this by parts twice, assuming that $p$ and its first $S$ derivative tend to zero sufficiently fast as $S$ goes to infinity we get

$$\frac{\partial V}{\partial t} = -rV + \frac{1}{2} e^{-r(T-t)} \sigma^2 E^2 p + r e^{-r(T-t)} \int_E^\infty S p dS . \quad (22.5)$$

In this expression $\sigma(S, t)$ has $S = E$ and $t = T$. Writing

$$\int_E^\infty S p dS = \int_E^\infty (S - E) p dS + E \int_E^\infty p dS$$
and collecting terms, we get

$$\frac{\partial V}{\partial T} = \frac{1}{2} \sigma^2 E^2 \frac{\partial^2 V}{\partial E^2} - rE \frac{\partial V}{\partial E}. $$

Rearranging this we find that

$$\sigma = \frac{\sqrt{\frac{\partial V}{\partial T} + rE \frac{\partial V}{\partial E}}}{\sqrt{\frac{1}{2} \sigma^2 E^2 \frac{\partial^2 V}{\partial E^2}}}.
$$

This gives us $\sigma(E, T)$ and hence, by relabelling the variables, $\sigma(S, t)$.

This calculation of the volatility surface from option prices worked because of the particular form of the payoff, the call payoff, which allowed us to derive the very simple relationship between derivatives of the option price and the transition probability density function.

When there is a constant and continuous dividend yield on the underlying the relationship between call prices and the local volatility is

$$\sigma = \sqrt{\frac{\frac{\partial V}{\partial T} + (r - D)E \frac{\partial V}{\partial E} + DV}{\frac{1}{2} \sigma^2 E^2 \frac{\partial^2 V}{\partial E^2}}}.$$  \hspace{1cm} (22.6)

There is no change in this expression when the interest rate and dividend yield are time dependent, just use the relevant forward rates.

One of the problems with this expression concerns data far in or far out of the money. Unless we are close to at the money both the numerator and denominator of (22.6) are small, leading to inaccuracies when we divide one small number by another. One way of avoiding this is to relate the local volatility surface to the implied volatility surface as I now show.

In the same way that we found a relationship between the local volatility and the implied volatility in the purely time-dependent case, Equation (22.1), we can find a relationship in the general case of asset- and time-dependent local volatility. This relationship is obviously quite complicated and I omit the details of the derivation. The result is

$$\sigma(E, T) = \frac{\sigma_{imp}^2 + 2(T - t)\sigma_{imp} \frac{\partial \sigma_{imp}}{\partial T} + 2(r - D)E(T - t)\sigma_{imp} \frac{\partial \sigma_{imp}}{\partial E}}{\left(1 + Ed_1\sqrt{T - t} \frac{\partial \sigma_{imp}}{\partial E}\right)^2 + E^2(T - t)^2 \sigma_{imp}^2 \frac{\partial^2 \sigma_{imp}}{\partial E^2} - d_1 \left(\frac{\partial \sigma_{imp}}{\partial E}\right)^2 \sqrt{T - t}}.  \hspace{1cm} (22.7)
$$

In terms of the implied volatility the implied risk-neutral probability density function is

$$p(S^*, t^*; E, T) = \frac{1}{E\sigma_{imp} \sqrt{2\pi(T - t^*)}} e^{-\frac{1}{2}(S^* - E)^2 \frac{1}{E\sigma_{imp}^2 \sqrt{T - t^*}}}.$$ 

One of the advantages of writing the local volatility and probability density function in terms of the implied volatility surface is that if you put in a flat implied volatility surface you get out a flat local surface and a lognormal distribution.

In practice there only exist a finite, discretely-spaced set of call prices. To deduce a local volatility surface from this data requires some interpolation and extrapolation. This can be done in a number of ways and there is no right way. One of the problems with these approaches is that the final result depends sensitively on the form of the interpolation. The problem is actually 'ill-posed,' meaning that a small change in the input can lead to a large change in the output. There are many ways to get around this ill-posedness, coming under the general heading of 'regularization.' Several suggestions for further reading in this area are given at the end of the chapter.

![Figure 22.5 Local volatility surface calculated from European call prices.](image)

An example of a local volatility surface is plotted in Figure 22.5.

### 22.7 HOW DO I USE THE LOCAL VOLATILITY SURFACE?

There are two ways to look at the local volatility surface. One is to say that it is the market's view of future volatility and that these predictions will come to pass. We then price other, more
complex, products using this asset- and time-dependent volatility. This is a very naïve belief. Not only do the predictions not come true but even if we come back a few days later to look at the ‘prediction,’ i.e. to re-fit the surface, we see that it has changed.

The other way of using the surface is to acknowledge that it is only a snapshot of the market’s view and that tomorrow it may change. But it can be used to price non-traded contracts in a way that is consistent across all instruments. As long as we price our exotic contract consistently with the vanillas, that is, with the same volatility structure, and simultaneously hedge with these vanillas, then we are reducing our exposure to model error. This approach is readily justifiable, although it is a bit difficult to estimate by how much we have reduced our model exposure. However, if you price using the calibrated volatility surface but only delta hedge, then you are asking for trouble. Suppose you price, and sell, a volatility-sensitive instrument such as an up-and-out call with a fitted volatility surface which increases with stock level. If it turns out that when the volatility is realized it is a downward-sloping function of the asset then you are in big trouble.

As an example, calculate the local volatility surface using vanilla calls. Now price a barrier option using this volatility structure. This means solving the Black–Scholes partial differential equation with the asset- and time-dependent volatility and with the relevant boundary conditions. This must be done numerically, by the methods explained in Part Six, for example. Now statically hedge the barrier by buying and selling the vanilla contracts to mimic as closely as possible the payoff and boundary conditions of the barrier option. This is described more fully in Chapter 30.

22.8 SUMMARY

Whether you believe in them or not, local volatility surfaces have taken the practitioner world by storm. Now that they are commonly used for pricing and hedging exotic contracts there is no way back to the world of constant volatility.

Personally I am in two minds about this issue. But as long as you hedge as much of the cashflows using traded vanillas options then you will be reducing your model exposure anyway. Once you have done this then you have done your best to reduce dependence on volatility. The danger is always going to be there if you never bother to statically hedge, only delta hedge. This is discussed in detail in Chapter 30.

FURTHER READING

- Merton (1973) was the first to find the explicit formulae for European options with time-dependent volatility.
- See Dupire (1993, 1994) and Derman & Kani (1994) for more details of fitting the local volatility surface.
- Rubinstein (1994) constructs an implied tree using an optimization approach, and this has been generalized by Jackwerth & Rubinstein (1996).
- To get around the problem of ill-posedness, Avellaneda, Friedman, Holmes & Samperi (1997) propose calibrating the local volatility surface by entropy minimization. Their article is also a very good source of references to the volatility surface literature.
- For trading strategies involving views on the direction of volatility see Connolly (1997).

- I suspect that there is not much information about future volatility contained in the local volatility surface. This is demonstrated in Dumas, Fleming & Whaley (1998).

EXERCISES

1. Use market data for European call or put options to calculate the implied volatilities for various times to expiry and exercise prices. Now find the risk-neutral probability density function and the local volatility surface.
2. Using market data for European options with the same expiry but different exercise prices, find and plot the implied volatilities. What shape does the plotted curve have?