Lecture 4: More on The Smile: Arbitrage Bounds, Valuation Problems, Models

Recapitulation of Lecture 3:

Transactions Costs

Hedge more frequently: replicate better, greater cost.
Hedge less frequently: replicate worse, less cost.

No transactions cost: hedging error ~ $(\Delta t^{0.5})$, converges to zero.

Regular rehedging: hedging error ~ $(\Delta t^{-0.5})$, diverges.

Some characteristics of the equity index implied volatility smile

- Volatilities are steepest for small expirations as a function of strike, shallower for longer expirations.
- The minimum volatility as a function of strike occurs near the at-the-money strikes.
- Low strike volatilities are usually higher than at-the-money volatilities, but high strike volatilities can also rise above at-the-money levels.
- The term structure is upward sloping but can change depending on views of the future. After large sudden market declines, the term structure may slope downwards, reflecting an expectation of diminishing future realized volatility, and the implied volatility of high-strike options may temporarily rise above at-may levels, indicating an expectation of a market recovery.
- The *volatility* of implied volatility is greatest for short maturities, as it is with Treasury rates.
- There is a negative correlation between changes in implied at-the-money volatility and changes in the underlying asset itself. [Fengler: $\rho = -0.32$ for three-month options on DAX in the late 1990s.]
- Implied volatility, like interest rates, is a parameter in a valuation model, and appears to be mean reverting with a life of about 60 days.
- Implied volatility tends to rise fast and decline slowly. Index markets tend to rise slowly and fall fast.

- Shocks across the implied volatility surface are highly correlated. There are a small number of principal components or driving factors. The surface has three main modes of movement: changes in overall level, changes in term structure, and changes in the skew, especially short term. We'll study these effects more closely later in the course.

- Implied volatility is usually greater than recent historical volatility.

Equity index smiles tend to be skewed to the downside. The large painful move for an index is a downward move, and needs the most protection. Upward moves hurt almost no-one. An option on index vs. cash is very different and much more asymmetric than an option on JPY vs. USD.

Single-stock smiles tend to be more symmetric than index smiles, since single stock prices can move dramatically up or down.

Interest-rate or swaption volatility, which we will not consider much in this course, tend to be more skewed and less symmetric, with higher implied volatilities corresponding to lower interest rate strikes. This can be partially understood by the tendency of interest rates to move normally rather than lognormally as rates get low (the Federal Reserve Banks makes more or less constant basis-point changes in interest rates), as well as the need of some investors to buy protection against low rates.

One must be very careful in speaking about volatility because there are so many different kinds: realized volatility σ, at-the-money volatility, implied volatility for a *definite strike*, $\Sigma = \Sigma(S, t;K, T)$, and at-the-money implied volatility $\Sigma_{atm} = \Sigma(S, t;S, T)$. When you talk about the change in volatility, which volatility are you referring to?

One often-used measure of the skew is a *risk reversal*, the difference in volatility between an out-of-the-money call option with a 25% Δ and an out-of-the-money put with a –25% Δ.

The dominant problems caused by the smile and its conflict with the Black-Scholes model are the hedging of standard options and the valuation and hedging of exotic options.

This Lecture:

Parametrizing options prices: delta, strike and their relationship
Estimating the effects of the smile on delta and on exotic options
Reasons for the existence of a smile
No-riskless-arbitrage bounds on the size of the smile
Fitting the smile

4.1 How to Graph the Smile

We observe implied volatility as a function of strike at a given time t_0 when the stock or index is at S_0; that is we are given only the snapshot $\Sigma(S_0, t_0, K, T)$. Our problem is similar to that of yield curve modeling: you see the yield curve at one instant and wonder what will happen to it later. Similarly, if we are interested in volatility, what we would like to know is its dynamic behavior as a function of S and t, namely $\Sigma(S,t;K,T)$, assuming implicitly that the Black-Scholes Σ is the customary (and appropriate?) way to indicate value.

We can plot $\Sigma(\,)$ against strike K, moneyness K/S, forward moneyness K/S_F, $(\ln K/S_F)/(\sigma\sqrt{\tau})$, or even more generally against $\Delta = N(d_1)$, which depends on stock price, strike, time to expiration and *implied volatility*, the function we are plotting, itself.

Traders usually like to plot the smile against Δ because they believe that the shape of the smile changes less with time and stock price level when expressed in that functional form. There are some other good practical reasons for preference too:

- Plotting implied volatilities against Δ immediately indicates the hedge for an option at that strike.
- Since Δ depends on both strike and expiration, you can compare the implied volatilities of differing expirations and strikes as a function of single variable.
- Finally, Δ is approximately equal to the risk-neutral probability $N(d_2)$ that a standard option will expire in the money – d_2 is roughly the number of standard deviations the stock price must move to expire in the money – and therefore seems to be a sensible measure of moneyness. Plotting implied volatilities against Δ embodies the notion that what matters for an option's price is how likely it is to move into the money from its current value.

Using the wrong quoting convention can distort the simplicity of the underlying dynamics. Perhaps the Black-Scholes model uses the wrong dynamics for stocks and therefore the smile looks peculiar in that quoting convention. That's the underlying hope behind advanced models of the smile.

An example is the case of a stock price that undergoes arithmetic (rather than geometric) Brownian motion. A constant arithmetic volatility then corresponds to a variable lognormal volatility that varies inversely with the level of the stock price. Plotting lognormal volatility against stock price would obscure the simplicity of the underlying evolution.

4.2 Δ and the Smile

4.2.1 The meaning of delta

Suppose that

$$\frac{dS}{S} = \mu dt + \sigma dZ$$

$$d\ln(S) = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma dZ$$

where $dZ^2 = dt$.

Then

$$\ln\frac{S_t}{S_0} = \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma\sqrt{t}Z(0, 1) \qquad \text{Eq.4.1}$$

where $Z(0, 1)$ is a normally distributed random variable with mean 0 and standard deviation 1.

The risk-neutral ($\mu = r$) probability of $S_t > K$ is $P(S_t > K)$ given by

$$\begin{aligned} P(\ln S_t > \ln K) &= P\left(\ln\frac{S_t}{S_0} > \ln\frac{K}{S_0}\right) \\ &= P\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma\sqrt{t}Z > \ln\frac{K}{S_0}\right] \\ &= P\left[Z > \frac{\ln K/S_0 - \left(\mu - \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}}\right] \\ &= P[Z > -d_2] = P[Z < d_2] \\ &= N(d_2) \end{aligned}$$

For small $\sigma\sqrt{t}$ this is approximately equal to $N(d_1)$ or the delta of a call, which is therefore approximately the risk-neutral probability of the option finishing in the money at expiration.

4.3 The Relationship between Δ and Strike

The most popular and liquid option is an at-the-money option, with $\Delta \sim 0.5$. Why? On the day it's sold it's a bet that could go either way, with roughly equal odds, and therefore attractive and relevant to speculating on or hedging the current positions in the stock. Far out-of-the-money options are also popular for buyers, because they're like cheap lottery tickets with small probabilities of success. Trading desks don't like to sell them; they are subsequently illiquid because they have a small (and hard to estimate) probability of expiring in the money. They embody tail risk which is very difficult to price.

A standard measure of the skew is the difference in volatility between an out-of-the-money call option with a 25% Δ and an out-of-the-money put with a –25% Δ. This trade – long the call and short the put at these deltas – is called a *risk reversal*.

What's the relation between the strike or moneyness of an option and its delta? Traders think about moneyness rather than strike, because moneyness measures the strike relative to where spot is today, which is what matters when you are trading. What percentage of moneyness corresponds to a given Δ? It's convenient to measure moneyness in units of standard deviation of the log return for the relevant expiration.

For simplicity set $r = 0$. Then $C = SN(d_1) - KN(d_2)$ where

$$d_1 = \frac{\ln\frac{S}{K}}{\Sigma\sqrt{\tau}} + \frac{\Sigma\sqrt{\tau}}{2}$$

At the money, $K = S$ and $d_1 = 0.5\Sigma\sqrt{T}$, assumed small, so that

$$\begin{aligned}\Delta = N(d_1) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_1} \exp\left(-\frac{y^2}{2}\right)dy \\ &= \int_{-\infty}^{0} \exp\left(-\frac{y^2}{2}\right)dy + \int_0^{d_1} \exp\left(-\frac{y^2}{2}\right)dy \\ &\approx \frac{1}{2} + \frac{d_1}{\sqrt{2\pi}}\end{aligned}$$

So, at the money,

$$\Delta = \frac{1}{2} + \frac{\Sigma\sqrt{\tau}}{2\sqrt{2\pi}}$$

As an example, for a typical annual volatility of 0.2 (20%) and an expiration of one year, we have $\Delta \approx 0.5 + 0.04 = 0.54$. (Check for yourself on a Black-Scholes calculator that this is approximately the correct delta for an at-the-money option at this volatility.)

Now suppose we move slightly out of the money, so that $K = S + \delta S$ where δS is small. Then $\ln\left(\frac{S}{S+\delta S}\right) = -\ln(1+\delta S/S) \approx -\frac{\delta S}{S}$ and so

$$d_1 \approx -\frac{(\delta S)/S}{\Sigma\sqrt{\tau}} + \frac{\Sigma\sqrt{\tau}}{2} = -\frac{J}{\Sigma\sqrt{\tau}} + \frac{\Sigma\sqrt{\tau}}{2}$$

where J is the fractional move in the strike away from the money, and $\Sigma\sqrt{\tau}$ is the square root of the variance over the life of the option.

Then for a slightly out-of-the-money option, a fraction J away from the at-the-money level,

$$\Delta \approx \frac{1}{2} + \frac{d_1}{\sqrt{2\pi}} \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\left(\frac{\Sigma\sqrt{\tau}}{2} - \frac{J}{\Sigma\sqrt{\tau}}\right)$$

Let's look at a real example. Suppose J = 0.01, a 1% move away from the at-the-money. And also assume $T = 1$ year and $\Sigma = 0.2$.

Then $\Delta \approx 0.54 - \frac{(0.4)(0.01)}{0.2} = 0.54 - 0.02 = 0.52$

Thus, Δ decreases by two basis points for every 1% that the strike moves out of the money for a one-year 20%-volatility call option.

The difference between a 50-delta and a 25-delta option therefore corresponds to about a 12% or 13% move in the strike price.

The move J to higher index levels necessary to decrease the delta of an initially at-the-money call by 0.29 from 0.54 to 0.25 is approximately given by

$$\frac{1}{\sqrt{2\pi}}\frac{J}{\Sigma\sqrt{\tau}} \approx 0.29 \quad \text{or} \quad J = 0.29\sqrt{2\pi}\Sigma\sqrt{\tau} \approx 0.29 \times 2.5 \times 0.2 \approx 0.15$$

Thus the strike of the 25-delta call is about 115. Actually it's about 117 if you use the exact Black-Scholes formula to compute deltas.

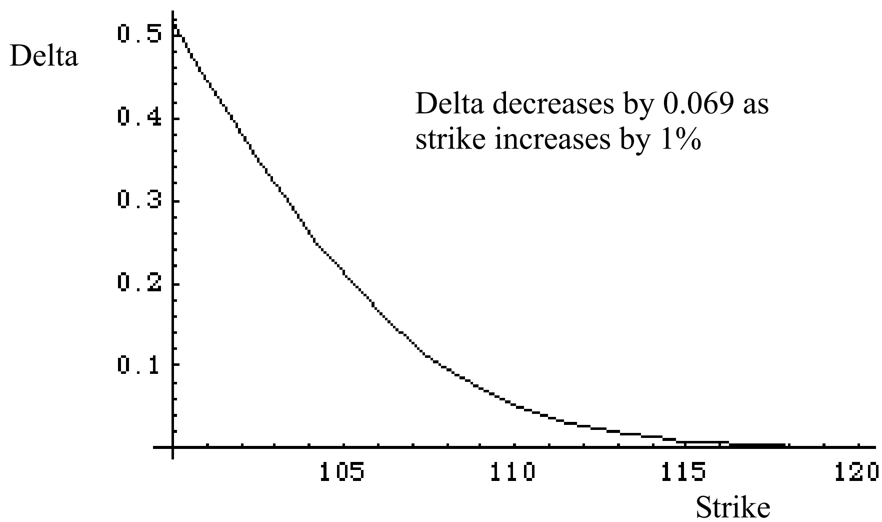
More generally, the change in the Δ of a call for a change in J is approximately given by

$$\frac{1}{\sqrt{2\pi}}\left(-\frac{J}{\Sigma\sqrt{\tau}}\right)$$

and a one-basis point change in Δ corresponds to a change in J of about $0.025\Sigma\sqrt{\tau}$.

The key variable here is the fractional shift *J* divided by the square root of the annual variance. For a greater volatility or time to expiration, the distribution of the stock is broader, and you need a bigger move in the strike to get to the same Δ.

Example for a 1-month call with zero interest rates, 20% volatility



4.4 No-Arbitrage Bounds on the Smile

You should think of Σ as the parameter that determines options prices via the model, in the same way as you think of yield to maturity as the parameter that determines quoted bond prices.

There are no-arbitrage bounds on bond yields. For example, consider zero coupon bonds of maturity T whose price is given by $B_T = 100\exp(-y_T T)$. Nos suppose $B_1 = 90$ and $B_2 = 91$. Then you can create a portfolio long the one year and short the two year, $\pi = \frac{91}{90}B_1 - B_2$, for zero cost. After one year the long position is worth more than \$100, so if you wait for B_2 to mature and pay off the face value you have a riskless profit. The necessary absence of riskless profits puts constraints on the yields.

There are similarly constraints on the prices of options which lead to subsequent constraints on the smile.

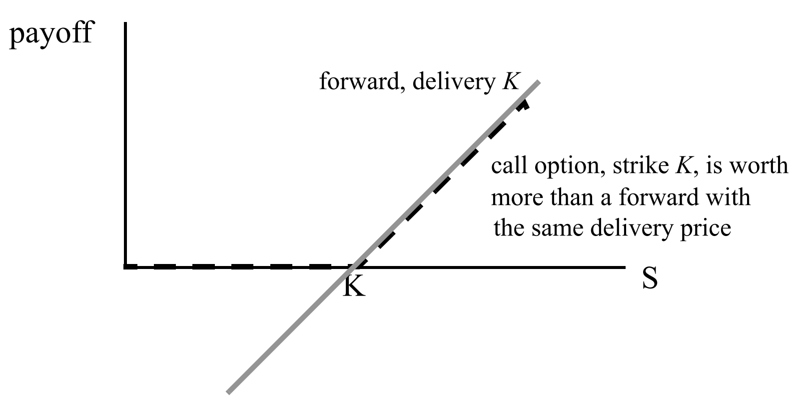
4.4.1 Some of the Merton Inequalities for Strike

Assume zero dividends, European calls.

1. $C \ge S - Ke^{-r(T-t)}$

   Proof: A forward F is worth $S - K$ at expiration, and therefore, assuming zero dividends and a riskless rate r, it worth $S - Ke^{-r(T-t)}$ now. An option is always worth more than a forward, because it has the same payoff when $S_T > K$, and is worth more when $S_T < K$.

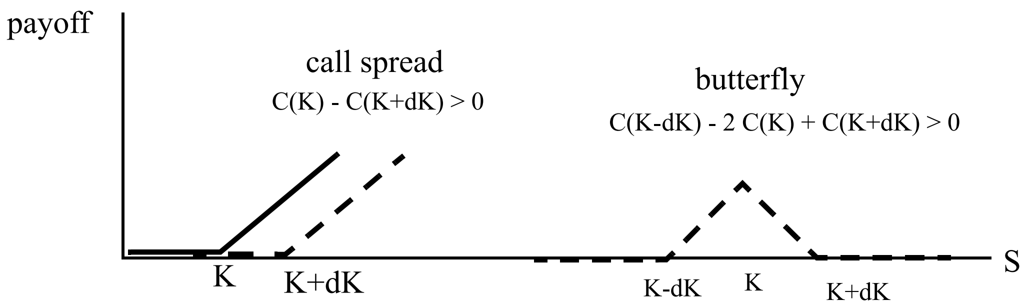
   Diagrammatically:



2. For the same expiration, options prices satisfy two constraints on their derivatives:

$$\frac{\partial C}{\partial K} < 0 \text{ and } \frac{\partial^2 C}{\partial K^2} > 0,$$

Proof: Look at payoff of a *call spread* and a *butterfly*.



The values of these portfolios are always positive, since they have positive payoff, and are respectively proportional to the first and second derivatives of the call price w/r/t strike in the limit as the $dK \to 0$. Therefore these derivatives must be positive, for actual call prices, independent of a model.

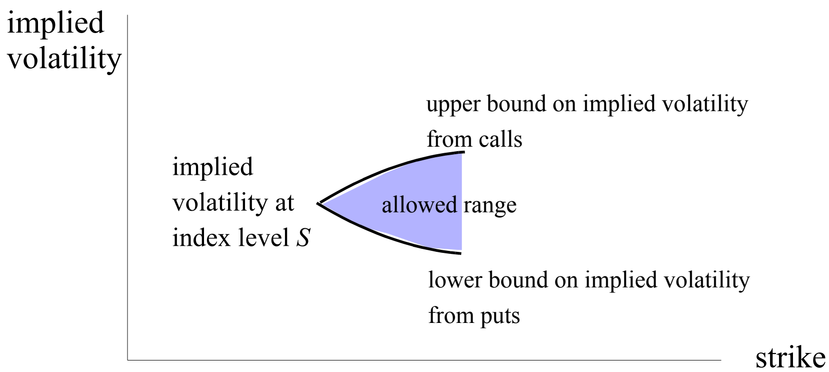
There are similar constraints on European put prices:

$$\frac{\partial P}{\partial K} > 0 \text{ and } \frac{\partial^2 P}{\partial K^2} > 0$$

4.4.2 Inequalities for the slope of the smile

The constraints on $\frac{\partial C}{\partial K} < 0$ and $\frac{\partial P}{\partial K} > 0$ put limits on the slope of the smile.

These constraint are true in the Black-Scholes formula with strike-independent volatility. Now suppose you parameterize actual market call and put prices in terms of the Black-Scholes formula, and so allow the implied volatility to vary with strike (and expiration). Then, if volatility were to increase (decrease) with strike level in the Black-Scholes formula, a too rapid increase (decrease) in volatility could offset the natural decrease (increase) with strike for a call (put) and so cause the call (put) price actually increase (decrease) with strike level. These limits on call- and put-price slopes sets respective limits on the positive and negative slope of the skew, as illustrated schematically below for call and put prices at some fixed index level S.



Now let's develop this idea more quantitatively.

$$C = C_{BS}(S, t, K, T, r, \Sigma)$$

$$\frac{\partial C}{\partial K} = \frac{\partial C_{BS}}{\partial K} + \frac{\partial C_{BS}}{\partial \Sigma}\frac{\partial \Sigma}{\partial K} < 0$$

$$\frac{\partial C_{BS}}{\partial \Sigma} = S\sqrt{\tau}N'(d_1) \equiv Ke^{-r\tau}\sqrt{\tau}N'(d_2) \qquad \text{Eq.4.2}$$

$$\frac{\partial \Sigma}{\partial K} \le -\frac{\frac{\partial C_{BS}}{\partial K}}{\frac{\partial C_{BS}}{\partial \Sigma}} = \frac{e^{-r\tau}N(d_2)}{Ke^{-r\tau}\sqrt{\tau}N'(d_2)} = \frac{N(d_2)}{K\sqrt{\tau}N'(d_2)}$$

Assume that volatility is small and strike price is at-the-money forward so that $S_F = K$. Then $d_2 \approx 0$, $N(d_2) \approx 0.5$ and $N'(d_2) \approx \frac{1}{\sqrt{2\pi}}$, so that

$$\frac{\partial\Sigma}{\partial K} \le \sqrt{\frac{\pi}{2}}\frac{1}{K\sqrt{\tau}} \qquad \text{Eq.4.3}$$

or

$$\frac{d\Sigma}{dK} \le \frac{(1.25)}{K\sqrt{\tau}} \qquad \text{Eq.4.4}$$

Recall that annualized volatility is measured as volatility of returns per year, so that 20% volatility means $\Sigma = 0.20$

For 1-month options, the bound is $\frac{d\Sigma}{dK} < 0.0043$. Equivalently, the maximum amount that volatility can change for a 1% change in strike without violating the principle of riskless arbitrage is 4.3 percentage points. For comparison, the S&P skew slope for one-month options in the figure we showed earlier was 0.001, or 1 volatility point for a 1% change in the strike, only a factor of 4 below the arbitrage limit.

We can also examine the limits for asymptotically short and long expirations.

From Equation 4.2 we see that

$$\frac{\partial\Sigma}{\partial K} \le \frac{N(d_2)}{K\sqrt{\tau}N'(d_2)}$$

When the strike is at-the-money forward, then as $\tau \to 0$

$$d_2 \to -\frac{\Sigma\sqrt{\tau}}{2} \to 0$$

$$N(d_2) \to \frac{1}{2}$$

$$N'(d_2) \to \frac{1}{\sqrt{2\pi}}$$

and so

$$\frac{\partial\Sigma}{\partial K} \le o(\tau^{-1/2})$$

as $\tau \to 0$.

As the time to expiration $\tau \to 0$, the slope can diverge no faster than $o(\tau^{-1/2})$.

S

At the other extreme, as $\tau \to \infty$, $d_2 \to -\infty$, and therefore

$$\frac{\partial \Sigma}{\partial K} \le \frac{1}{K\sqrt{\tau}}\frac{N(d_2)}{N'(d_2)} \sim o\left(\frac{1}{\sqrt{\tau}}\frac{1}{\sqrt{\tau}}\right) \sim o\left(\frac{1}{\tau}\right)$$

as the time to expiration gets large. (To prove the line above we have made use of the asymptotic relation $(N(d_2))/N'(d_2) \sim o(\tau^{-0.5})$ as $\tau \to \infty$.)

Thus, the slope of the smile can decrease with time to expiration no more slowly than $o(\tau^{-1})$.

Reference: *Arbitrage Bounds on the Implied Volatility Strike and Term Structures of European-Style Options*. Hardy M. Hodges, Journal of Derivatives, Summer 1996, pp. 23-35.

4.5 Problems Caused By The Smile

The Black-Scholes model assumes constant volatility and cannot account for the pattern of options prices at a given time. It attributes a different underlying stock implied volatility to *each* option with a different strike, *but* that implied volatility is not really the volatility of the option; it is the volatility of the stock, and in the model, a stock must have one lognormal volatility of returns which cannot know about the option's strike. There cannot be many GBM volatilities for the same stock.

Therefore, Black-Scholes is often simply being used as a quoting mechanism, rather than a valuation mechanism, similar to the way in which yield to maturity is used in quoting rather than calculating bond or mortgage prices.

What problems does this cause.

4.5.1 Fluctuations in the P&L from incorrect hedging of standard options

If we have the wrong model, then, even if liquid vanilla options prices are forced to be correct by calibrating the implied volatility parameter to the price, one is using the wrong model formula to obtain the right (market) price. Therefore the derivatives of the formula are wrong and so are hedge ratios. The price *should* reflect the riskless hedging strategy, but it won't.

How large a fluctuation in the P&L will be induced by using the wrong hedge ratio? Let's make a rough estimate.

Using the chain rule,

$$\Delta = \frac{dC_{BS}(S, t, K, T, \Sigma)}{dS} = \frac{\partial C_{BS}}{\partial S} + \frac{\partial C \partial \Sigma}{\partial \Sigma \partial S} \qquad \text{Eq.4.5}$$

We make the following approximations and assumptions. At the money, the vega for the S&P 500 index assuming $S \sim 1000$ and $T = 1$ year is given by

$$\frac{\partial C}{\partial \Sigma} \sim \frac{S\sqrt{T}}{\sqrt{2\pi}} \sim 400$$

We can plausibly estimate $\frac{\partial \Sigma}{\partial S}$ on dimensional grounds by guessing that it is of order $\frac{\partial \Sigma}{\partial K}$, so that

$$\frac{\partial \Sigma}{\partial S} \sim \frac{\partial \Sigma}{\partial K} \sim \frac{0.02}{100} \sim 0.0002$$

where, as we saw earlier, the right hand side of the equation above is the typical order of the magnitude for the S&P 500 skew.

Then, the mismatch between the true Δ and the Black-Scholes Δ owing to the skew, using Equation 4.5, is approximately

$$\Delta - \frac{\partial C_{BS}}{\partial S} = \frac{\partial C \partial \Sigma}{\partial \Sigma \partial S} \sim 400 \times 0.0002 = 0.08$$

For a not uncommon daily index move of 1%, or about 10 S&P 500 index points, we can compare the mismatch in the P&L from using the wrong delta to the incremental P&L expected in the Black-Scholes model with perfect hedging.

The mismatch in the P&L from being long or short 0.08 of a share when the index moves 10 points is about 0.8 index points.

The incremental P&L from the curvature in a position in a perfectly hedged at-the-money one-year option in a Black-Scholes world with a not atypical volatility of 0.2 when the index moves 10 points is of order

$$\Gamma \times \frac{\delta S^2}{2} \sim \frac{1}{S\Sigma\sqrt{T}} \frac{\delta S^2}{2} \sim \frac{1}{200}(50) = 0.25 \quad \text{points}$$

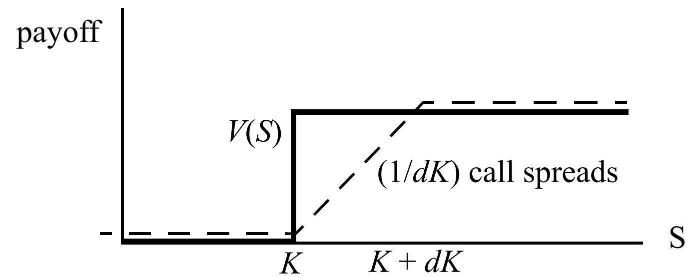
The mismatch in Δ can cause a large distortion in the incremental P&L from hedging at each step.

4.5.2 Errors in the Valuation of Exotic Options

Even if the prices of standard calls and puts are known without the need for a model, the value of exotic options will be incorrect if we calibrate these standard options to the wrong model.

Look for example at a rather simple exotic European-style option V at time t which pays \$1 if $S \geq K$ at time T, and zero otherwise. This serves as insurance against a fixed loss above the strike K, but not against a proportional loss as in the case of a vanilla call. It is very hard to hedge this because the payoff oscillates between 0 and 1.

A graph of the payoff at expiration is shown below.



One can approximately replicate V with a call spread, a long position in a call with strike K and expiration T and a short position in a call with strike $K + dK$ and the same expiration. In the limit as $dK \to 0$ the call spread's payoff converges to that of the exotic option.

The value of the call spread at stock price S and time to time t is

$$\frac{-C_{BS}(S, K+dK, t, T, \Sigma(K)) + C_{BS}(K, S, t, T, \Sigma(K))}{dK} \approx -\frac{d}{dK}C_{BS}(S, K, t, T, \Sigma(K))$$

where $C_{BS}(S, K, t, T, \Sigma(K))$ is the market price of the call quoted in Black-Scholes terms, with $\Sigma(K)$ incorporating the dependence of implied volatility on K. The total derivative with respect to K includes the change of all variables with K, including that of the implied volatility.

We can estimate the current value $V(S, K, t, T, \Sigma(K))$ if we know how call prices vary with strike K:

$$V(S, K, t, T) = -\frac{d}{dK}C_{BS}(S, K, t, T) = -\frac{\partial C_{BS}}{\partial K} - \frac{\partial C_{BS}}{\partial \Sigma} \times \frac{\partial \Sigma}{\partial K}$$

For $r = 0$, $\Sigma = 20\%$, $T - t = 1$ year, $K = S = 1000$, and a skew slope 0.0002,

$$-\frac{\partial C_{BS}}{\partial K} = N(d_2) = N\left(-\frac{\Sigma}{2}\right) = \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\frac{\Sigma}{2} = 0.46$$

$$\frac{\partial C_{BS}}{\partial \Sigma} \sim \frac{S\sqrt{T}}{\sqrt{2\pi}} \sim 400$$

$$\begin{aligned} V(100, 100, 0, 1, 0.2) &\approx (0.46) + (400 \times 0.0002) \\ &= 0.46 + 0.08 = 0.54 \end{aligned}$$

The non-zero slope of the skew adds about 16% to the value of the option. This is a significant difference.

Why does the skew *add* to the value of the derivative V? Because it's a call spread, and a negative skew means there is less volatility at higher strikes, and therefore "less risk-neutral probability" of the stock price moving upwards; therefore the second, higher leg of the call spread is worth less than when there is no skew.

How can we "fix" it or extend Black-Scholes to match the skew and allow us to calculate all these quantities correctly? What changes can we make? Or, how, as we did in the above example, can we tread carefully and so avoid our lack of knowledge about the right model and still get reasonable estimates of value? Those are the questions we will tackle later in the course.

4.6 Some Behavioral Reasons for an Implied Volatility Skew

Think of options trading as the trading of volatility or variance as an asset. Then there are a number of reasons you can think of that might cause a skew or smile.

- Knowledge of past behavior in options markets suggests a skew in options would be wise. (How much, though? What's the fair value?) Out-of-the-money puts provide crash protection cheaply, which may have driven up implied volatilities after 1987. Also, realized volatility will increase in a crash, therefore this is self-consistent.

- Expectation of future changes in volatility naturally gives rise to a term structure.

- Expectation of changes in volatility as the market approaches certain significant levels can give rise to skew structure. For example, investors' perception of support or resistance levels in currencies and interest rates suggests that realized volatility will decrease as those levels are approached.

- Expectation of an increase in the correlation between individual returns as the market drops can cause an increase in the volatility of the entire index. (Some volatility arbitrage dispersion strategies are based on this.)

- An aversion to downward market moves/jumps can produce negative skew. Dealers' tend to be short options because they sell zero-cost collars (short otm call-long otm put) to investors who want protection against a decline. This fear of being short options in a crash can lead them to charge more to protect themselves against gamma losses and increases in volatility after a downward jump.

4.7 An Overview of Smile-Consistent Models

We want to model the process that produces a smile. One consideration is whether to model the stochastic evolution of the underlying asset S and its realized volatility, and then deduce the implicit form of the implied volatility surface $\Sigma(S, t, K, T)$ produced by the model, or whether to directly model the dynamics of the parametric surface $\Sigma(S, t, K, T)$. Though the former is more common, the latter is also possible.

Modeling the stochastic evolution of the stock price and its volatility is more fundamental, and arbitrage violations are more easily avoidable. But it is difficult to figure out the process that accurately describes stock evolution. As I mentioned in the initial lecture, financial models are often better aimed at modeling observable variables rather than dropping down several levels to model unobservable processes and then deducing the behavior of observables from them. We have to make the difficult choice between modeling implied volatility (a parameter in the Black-Scholes model that incorrectly describes options values) or dropping down a level and modeling instantaneous volatility, a quantity whose evolution we know little about.

Traders think in terms of Black-Scholes implied volatility, which they observe as they make markets every day, so for them it is natural to describe the dynamics of $\Sigma(S, t, K, T)$. In addition, changes in implied volatility are relatively easy to observe, so that a model of implied volatility can be calibrated without too much difficulty. But one has to be careful in modeling the stochastic evolution of Σ directly, because changing Σ changes all options prices, and we do not want to violate the constraints imposed by no riskless arbitrage. Analogously, in the interest rate world (and most smile models are inspired by interest-rate models) you can't just write down any stochastic process you like for moving the yield curve around because you might generate negative forward rates which violate the no-arbitrage constraints on bond prices. One can develop Heath-Jarrow-Morton-style models for implied volatilities directly, and we'll review these later in the course. In these lectures we will predominantly concentrate on more fundamental lower-level models of the stock price evolution.

One other comment: from the data we've seen on the smile so far: different markets have such different smiles that we should probably recognize that it is unlikely that one grand replacement for Black-Scholes will cover all smiles in all markets. What may well be important is choosing the right model for the right market.

Since Black-Scholes is inadequate, the most common models of the smile extend Black-Scholes to accommodate a wider range of asset evolution.

4.7.1 Local Volatility Models

Local volatility models were the earliest models of the smile. These models depart just enough from the Black-Scholes model to become consistent with the smile.

In the Black-Scholes model the stock's volatility σ is a constant, independent of stock price and future time, and in consequence $\Sigma(S, t, K, T) = \sigma$ is independent of strike and expiration.

In local volatility models, the stock's realized volatility is allowed to vary *deterministically* as a function of future time *t* and the future (*random*) stock price *S*, so that it takes the functional form $\sigma(S, t)$. This function is called the local volatility function, and leads to an implied volatility that can vary with strike and expiration. In these models the evolution of the stock price is given by

$$\frac{dS}{S} = \mu(S, t)dt + \sigma(S, t)dZ \qquad \text{Eq.4.6}$$

Note that $\sigma(S, t)$ is a *deterministic* function of a *stochastic variable* S, and so this *is* a stochastic volatility model, but of a limited kind.

Local volatility models are one-factor models – only the stock price is stochastic – and so most of the standard Black-Scholes scheme of perfect replication in terms of the riskless bond and the stock still works; we can use risk-neutral valuation methods to obtain unique arbitrage-free option values in a similar way. This is very attractive from a theoretical point of view. But, is it the right model, or perhaps putting it better, for which asset is it the right model?

In using any model the first problem is that of calibration: how to choose $\sigma(S, t)$ to match market values of $\Sigma(S, t, K, T)$. We'll show later how this can be done in principle. But one must be careful: just because you can fit the diffusive process of Equation 4.6 to match the smile doesn't necessarily mean that the model is an accurate description. The right model is presumably the one that (most closely – this is imperfect finance, not almost perfect physics) matches the behavior of the underlying asset.

Nevertheless, right or wrong, there is lots to learn about local volatility models and we're going to spend an appreciable amount of time on them later.

What might account for local volatility being a function $\sigma(S, t)$? One example is the so-called *leverage effect* for stocks which models the fact that if a company is leveraged, its volatility should increase as the stock price moves lower and closer to the level of debt. Here is a simplistic illustration:

$$S = A - B \qquad \text{assets - liabilities}$$

$$\frac{dA}{A} = \sigma dZ$$

$$\frac{dS}{S} = \frac{dA}{S} = \frac{A\sigma dZ}{S} = \frac{(S+B)}{S}\sigma dZ$$

$$\sigma_S = \sigma(1 + B/S)$$

So, stock volatility increases as stock price decreases.

Constant Elasticity of Variance (CEV) models, developed by Cox and Ross soon after the Black-Scholes model appeared, are the earliest local volatility models:

$$dS = \mu(S, t)dt + \sigma S^{\beta} dZ$$

Here $\beta = 1$ corresponds to the usual lognormal case, and $\beta = 0$ to normal evolution. You need β to be negative and quite large in magnitude to account for the magnitude of the negatively sloped equity index skew.

The conceptual difference between these parametric models and local volatility models is that the CEV and leverage models are parametric models that have only a few parameters and cannot fit an arbitrary smile, whereas local volatility models are non-parametric, so that $\sigma(S, t)$ must be numerically calibrated to the observed smile at time *t*.

4.7.2 Stochastic Volatility Models

One of the Black-Scholes assumptions that is certainly violated is the supposition that volatility is constant. In stochastic volatility models, the volatility or variance *V* of the stock itself is an independent random variable whose evolution is correlated with that of the stock price *S*.

$$\begin{aligned} dS &= \mu_S(S, V, t)dt + \sigma_S(S, V, t)dZ_t \\ dV &= \mu_V(S, V, t)dt + \sigma_V(S, V, t)dW_t \\ V &= \sigma^2 \\ E[dWdZ] &= \rho dt \end{aligned}$$

If you are allowed to replicate options through dynamic trading *only* in the stock and the bond markets, and volatility itself is stochastic, then options markets are not complete and perfect replication of the option's payoff isn't possible. The principle of no riskless arbitrage will not lead to a unique price and you need to know the market price of risk or a utility function for risk and reward to price options; that's not a preference-free method, and less reliable than either static or dynamic replication, and we'd like to avoid it here. (Of course, just because we prefer to avoid it from a theoretical point of view doesn't mean that the market itself doesn't operate that way.)

If you can trade options, and if you know (or rather *assume* that you know) the stochastic process for option prices or volatility as well as the stochastic process stock prices, then you can hedge one option's exposure to volatility with another option, you can derive an arbitrage-free formula for options values. But assuming that you know all this is, as Rebonato says, "a tall order."

An additional objection to stochastic volatility models is that they assume that the correlation ρ is constant. In fact, correlations are stochastic too, with perhaps a greater variance than volatility. Choosing ρ constant may be too extreme an assumption.

But, you must start somewhere if you want to get anywhere. So, model we must. Later in the course we'll study stochastic volatility models.

4.7.3 Jump-Diffusion Models

Another element that Black-Scholes ignores is the discontinuous movement (a.k.a. jumps) of stock prices. Jump-diffusion models were first invented by Merton shortly after Black-Scholes. These models allow the stock to make an arbitrary number of jumps in addition to undergoing diffusion.

With a finite number of jumps of known size in the model, one can replicate any payoff perfectly with by dynamic trading in a finite number of options, the stock and the bond, and so achieve risk-neutral pricing. But if there is an infinite number of jumps of different sizes possible, then perfect replication is impossible, and though people use risk-neutral pricing, it's not strictly correct.

4.7.4 A Plenitude of Other Models

There are many other smile models too, which we may discuss later: mixing models, variance gamma models, stochastic volatility models of other types, stochastic implied volatility models …

In practise, one has to see which model best describes the market one is working in. You could argue reasonably convincingly that in the real world there is indeed diffusion, jumps and stochastic volatility! Then, however, you have so many different ways of fitting the observed smile that the model is non-parsimonious and offers too many choices. In the end, you want to model the market with reasonable (but not perfect) accuracy via a fairly simple model that captures most of the important behavior of the asset. A model is only a model, not the real thing.