

Chapter 9

Implied and Local Volatility

In these notes we take seriously the fact that the market does not price options the way chapter 5 said it should. We show that the option prices at one expiry are the same information as the probability distribution of the underlying at that expiry, that there is always a model with a single deterministic volatility function reproducing every one of those prices, and we derive the formula for it. We finish with Gyongyi's theorem, which explains both why that model exists and why it is not the end of the story.

Where the Curve Went

Chapter 8 put the whole curve in motion and priced a swaption from it, and the last thing it did was reduce that swaption to Black's formula: one volatility, one strike, in the annuity measure. That reduction is exact, and it is what the next three chapters stand on. A swap rate is a martingale under its own annuity and a forward rate is one under its own forward measure, so once the numeraire is chosen, a caplet or a swaption is an option on *one* underlying at *one* date — the same object as an equity option, with the curve already integrated into the numeraire.

So these chapters are not ignoring the term structure. They are working in the measure in which it has already been dealt with.

Structure (Two questions, and why they separate). There are two distinct things a rates model has to get right.

The first is the *marginal* question. For one rate, at one date, what is the distribution? This is what a smile is, entirely — the Breeden-Litzenberger theorem below says that the prices at one expiry *are* that distribution, with nothing left over. It needs no curve, because a numeraire has removed it, and it is the subject of this chapter, chapter 10 and chapter 11.

The second is the *joint* question. How do several rates, at several dates, move together? A swaption on the ten-year rate says nothing about how the ten-year moves relative to the two-year, and no collection of single-rate smiles ever will: they are marginals, and marginals do not determine a joint distribution. That is the subject of chapter 12 and chapter 13, and chapter 18 is about the general fact underneath it.

The separation is real rather than expositional. It is why a CMS caplet needs one rate’s smile and no dynamics at all, as chapter 15 shows by replicating it statically, while a Bermudan needs the joint law and cannot be replicated from vanillas at any price. It is also why the two halves are calibrated to different instruments and, on most desks, by different people.

What the separation costs is stated in chapter 10: a smile fixes the marginal and leaves the dynamics free, so several models agreeing on every quoted price today can disagree about tomorrow. Fitting the first question perfectly does not begin to answer the second.

One convention, since it saves writing every statement twice. These three chapters say “spot”, “forward” and “strike” because the results hold for any underlying, and the equity notation is the one they were discovered in. For a rates reader the translation is fixed: read the underlying as a forward rate under its forward measure or a swap rate under its annuity, and read a discounted expectation as one taken under that numeraire. Nothing else changes; chapter 12 picks the curve back up where this leaves it.

9.1 Implied Volatility

Chapter 5 gave us a formula. Given the forward F , a strike K , an expiry T and a volatility σ , it returns a price. Everything on that list except σ can be read off a contract or a screen. So a trader quoting an option is really quoting one number, and rather than quote it as a price, the market quotes it as the volatility that would produce that price.

Definition 9.1 (Implied volatility). The implied volatility of an option with observed price V is the number $\sigma_{\text{imp}}(K, T)$ satisfying

$$V = \text{Black}(F, K, \sigma_{\text{imp}}(K, T), T).$$

Before using this we should check it makes sense.

Lemma 9.2. *For a price strictly between its no-arbitrage bounds, the implied volatility exists and is unique.*

Proof. The Black price is continuous in σ and its derivative, the vega, is

$$\frac{\partial V}{\partial \sigma} = F n(d_1) \sqrt{T} > 0$$

for $F, T > 0$, where n is the standard normal density. A continuous, strictly increasing function is invertible on its range. As $\sigma \rightarrow 0$ the price tends to the intrinsic value $(F - K)^+$ and as $\sigma \rightarrow \infty$ it tends to F , so the range is exactly the interval between the bounds. $\square$

In words: turning the volatility dial always moves the price, and always in the same direction, so from a price you can always work backwards to exactly one volatility. That is all a quoting convention needs. Notice what the lemma does *not* say. It says nothing about whether the underlying is lognormal, or whether any model is right. It says only that this particular formula is a reversible way of writing down a number. The implementation is `implied.vol.black76` in `quant/src/black.rs`, and it does exactly what the proof does: it brackets the answer and squeezes.

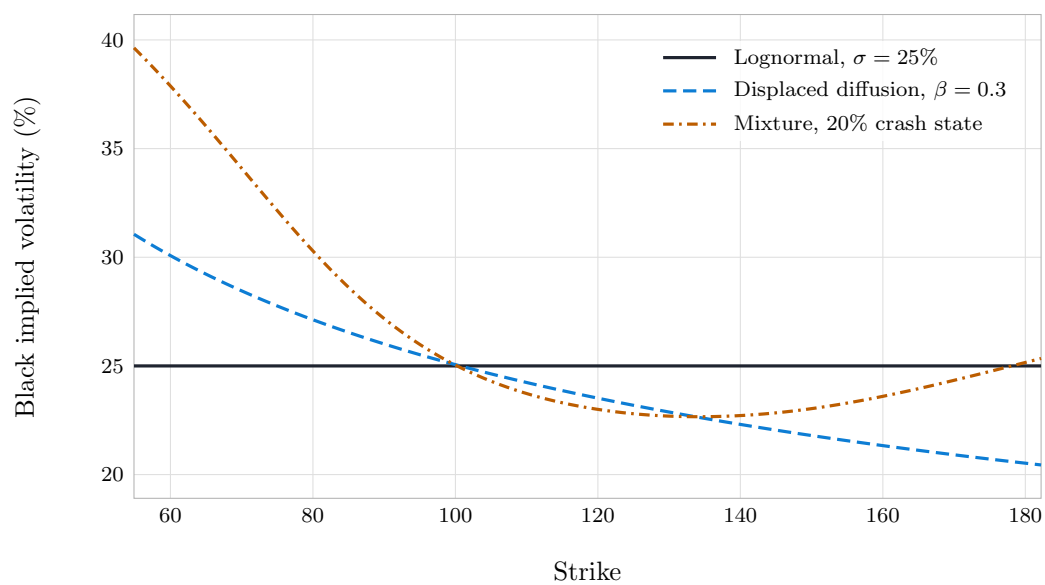


Figure 9.1: The implied volatility of three models, all with the same at-the-money volatility. A lognormal forward gives a flat line at every setting of every parameter, and could not do otherwise: implied volatility is defined by inverting the lognormal formula, so lognormal prices invert to one number. The other two are the two smallest departures from it, and they produce the two features every real market shows — a tilt and a curvature.

Drawn by `Model::price`, `Model::implied_vol` and `implied_vol_black76`.

Now the observation that this whole part of the notes exists to explain. If the Black-Scholes model were literally true, the same σ would appear in every option on the same underlying, whatever its strike. So plotting implied volatility against strike would give a horizontal line. It does not. What one sees instead is called the smile, or, when it slopes, the skew.

Play with the figure for a moment, because it makes a distinction that is easy to lose. The displaced diffusion tilts the smile but leaves it nearly straight; the mixture bends it. These are different statements about the world. Tilt says that a fall in the forward is more violent than a rise. Curvature says that large moves in either direction are likelier than the lognormal model allows. Real markets show both, and a model that produces only one of them will be wrong about half of the options it is asked to price.

9.2 The Prices Are the Distribution

If the smile is a map of something, what exactly is it a map of? The answer is completely precise, and it is the first theorem of this chapter.

Fix an expiry T and work in the T -forward measure of chapter 6, in which the forward F_T is a martingale and a call is worth

$$C(K) = P(0, T) \mathbb{E}^T[(F_T - K)^+]. \quad (9.1)$$

Theorem 9.3 (Breedon-Litzenberger). *Let p be the probability density of F_T in the T -forward measure. Then*

$$\frac{\partial C}{\partial K} = -P(0, T) \mathbb{P}^T(F_T > K), \quad \frac{\partial^2 C}{\partial K^2} = P(0, T) p(K).$$

Proof. Write (9.1) as an integral and differentiate under it:

$$\begin{aligned} C(K) &= P(0, T) \int_0^\infty (x - K)^+ p(x) dx = P(0, T) \int_K^\infty (x - K) p(x) dx \\ \frac{\partial C}{\partial K} &= P(0, T) \left(-[(x - K)p(x)]_{x=K} - \int_K^\infty p(x) dx \right) = -P(0, T) \int_K^\infty p(x) dx, \end{aligned}$$

the boundary term vanishing because $(x - K)$ is zero at $x = K$. That is the first claim. Differentiating once more,

$$\frac{\partial^2 C}{\partial K^2} = -P(0, T) \frac{\partial}{\partial K} \int_K^\infty p(x) dx = P(0, T) p(K). \quad \square$$

Analytically: the second derivative of the call price with respect to strike is the density, discounted. Knowing the price of every strike at one expiry is therefore *exactly* knowing the distribution of the underlying at that date.

Financially: consider what you would have to trade to construct that second derivative. A second derivative is a second difference, so buy a call struck at $K - h$, sell two struck at K , and buy one struck at $K + h$. This is called a butterfly.

Example 9.1 (A butterfly is a bet on where the forward lands). Take $h = \$1$ and $K = \$100$, so we buy the \$99 call, sell two \$100 calls, and buy the \$101 call. Its payoff at expiry, as a function of where F_T lands:

F_T	\$98	\$99.50	\$100	\$102
\$99 call	0	0.50	1.00	3.00
$-2 \times \$100$ call	0	0	0	-4.00
\$101 call	0	0	0	1.00
total	0	0.50	1.00	0

The payoff is zero outside $[\$99, \$101]$ and rises linearly to $\$1$ at the middle: a narrow tent pitched over K . It pays if and only if the forward lands near $\$100$, and it is worthless otherwise.

Now price it. The tent has height h and base $2h$, so its area is h^2 , and for small h the expected payoff is approximately $p(K)h^2$. Its price is therefore about $P(0, T)p(K)h^2$; dividing by h^2 gives the theorem. So $\partial^2 C / \partial K^2$ is not merely a derivative that happens to equal a density. It is the price of the cheapest available bet that the forward finishes at K , and its price is the probability of that happening.

Practically: it tells you what a smile is a map of. It is a map of the market's probability distribution — and since a flat smile corresponds to the lognormal density, the shape of the smile is precisely the shape of the market's disagreement with lognormality.

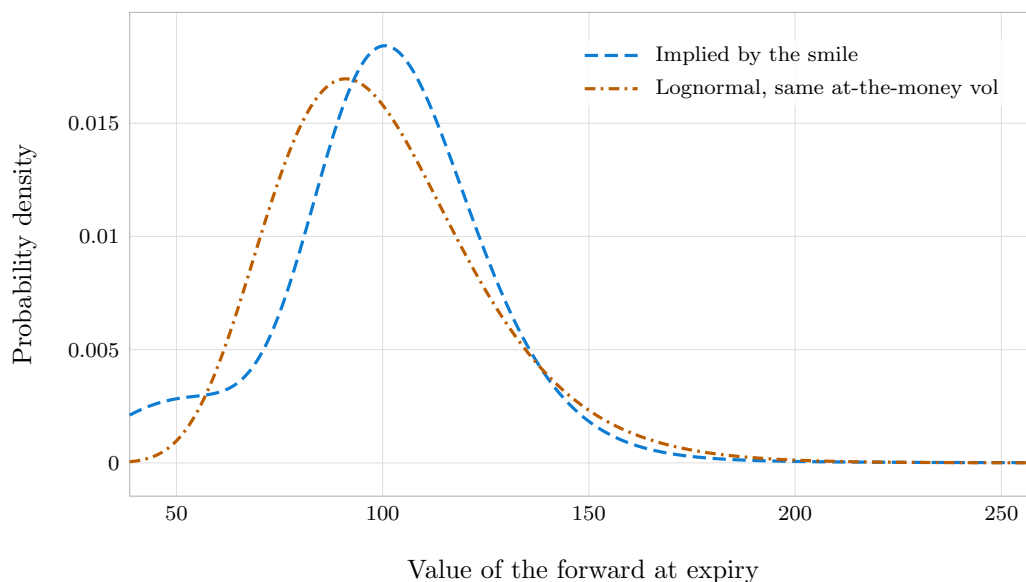


Figure 9.2: The density implied by a smile, next to the lognormal density with the same at-the-money volatility. Both are obtained by differencing option prices twice in the strike, so the picture is the theorem being applied rather than an illustration of it. The market's density has a fatter left tail, a thinner middle, and a slightly fatter right tail: the extra mass in the tails is what the raised wings of the smile are paying for, and it has to come from somewhere.

Drawn by `Model::density` and `lognormal.density`.

The theorem also hands us the arbitrage constraints on a surface, for free. A density is non-negative, so $\partial^2 C / \partial K^2 \geq 0$: the call price must be *convex* in the strike. A probability is at most one, so $-\partial C / \partial K / P(0, T) \leq 1$. A quoted surface violating either of these is not merely ill-behaved, it is arbitrageable, and the trade that exploits it is the butterfly above — sold for a negative price.

Exercise (Finding the arbitrage). Suppose three calls on the same expiry are quoted at $K = 99, 100, 101$ for \$6.10, \$5.30 and \$4.60. Show that the surface is not convex, construct the portfolio that captures the arbitrage, and state what it pays in every state of the world. Then check your answer against the bounds computed by `implied_vol_black76`, which returns no implied volatility at all for prices outside them.

9.3 What the Standard Combinations Price

The butterfly was not chosen for its convenience. It is a trade people actually put on, and so are its relatives, and each of them exists because it isolates one feature of the smile and is insensitive to the others. Setting them out makes the surface easier to read, because a quoted market is usually a market in these rather than in individual options.

Take the smile near the money and write it as a level, a slope and a curvature,

$$\sigma_{\text{imp}}(y) \approx \sigma_0 + s y + \frac{1}{2} c y^2, \quad y = \ln \frac{K}{F},$$

and each combination picks off one coefficient.

- A *straddle* — a call and a put at the same strike — is delta neutral at the money and long vega. It has no exposure to s by symmetry, so it prices σ_0 , the level.
- A *risk reversal* — long an out-of-the-money call, short the equally out-of-the-money put — is antisymmetric about the money and so cancels the level and the curvature, leaving s . It is the market's instrument for the skew, and it is quoted directly as a volatility difference: “the twenty-five delta risk reversal is at minus two” is a statement about s and nothing else.
- A *butterfly* cancels the level and the slope and leaves c . It prices the curvature, and, as the previous section showed, in the limit of narrow wings it prices the density itself.
- A *call spread* is the difference of two calls, so by Theorem 9.3 its price is $-\partial C / \partial K$ times the width: it is a digital, and it prices the *cumulative* distribution rather than the density.
- A *collar* is a risk reversal sized to cost nothing — long a put, short a call, strikes chosen so the premiums cancel. Its usefulness as a hedge is a consequence of the skew: in a market with $s < 0$ the put one wants is expensive and the call one gives up is cheap, so the strikes come out asymmetric, and how asymmetric is a direct reading of s .

Remark. Each combination is a finite-difference operator applied to the price in the strike: a call spread is a first difference, a butterfly a second, a risk reversal an antisymmetric first difference. By Breeden-Litzenberger the price surface *is* the distribution, so differencing it in strike reads off successive features of that distribution — the cumulative, then the density, then its shape. A market that quotes straddles, risk reversals and butterflies is quoting the first three moments of its own risk neutral distribution, in a coordinate system chosen so that each can be traded without disturbing the others.

9.4 How the Density Moves

Dupire's formula needs to know how the risk neutral density changes with maturity, and chapter 3 has already supplied it. The generator's adjoint was computed there, and the forward equation followed from the definition of the adjoint in two lines: the density satisfies $\partial p / \partial t = \mathcal{L}^* p$, with the coefficients inside the derivatives rather than outside.

Written out for the risk neutral price process of this chapter, whose drift is $(r - q)S$ and whose diffusion coefficient is $\sigma_{\text{loc}}(T, S) S$, that reads

$$\frac{\partial p}{\partial T} = -\frac{\partial}{\partial S} [(r - q)S p] + \frac{1}{2} \frac{\partial^2}{\partial S^2} [\sigma_{\text{loc}}^2(T, S) S^2 p]. \quad (9.2)$$

That is the only new thing the next section needs.

Remark (It is a conservation law). One property of (9.2) matters before it is used, because it is what makes the equation safe to solve numerically. It can be written

$$\frac{\partial p}{\partial T} = -\frac{\partial J}{\partial S}, \quad J = (r - q)S p - \frac{1}{2} \frac{\partial}{\partial S} [\sigma_{\text{loc}}^2 S^2 p], \quad (9.3)$$

with J a probability flux. Written this way the content is plain: probability is not created or destroyed, only moved. The first term carries it along with the drift and the second spreads it down its own gradient.

This is more than elegance: it is the property a numerical solution must have, and it fails loudly when the scheme is wrong. Chapter 12 solves a forward equation on a grid, and the scheme in `QuasiGaussian::density` is written in the flux form (9.3) for exactly this reason: it moves mass between cells rather than evaluating a second derivative, so the total is conserved by construction rather than by hope. Another version evaluated the derivative directly, and returned a “density” integrating to 1.42.

9.5 Local Volatility and Dupire's Formula

We now know that at each expiry the market has told us a distribution. It has told us nothing about how the underlying travels between expiries, and infinitely many processes are consistent with any given set of distributions. The question Dupire asked is whether at least *one* of them can be found in a particularly simple class.

Definition 9.4 (Local volatility model). A local volatility model is one in which the diffusion coefficient is a deterministic function of time and of the current level of the underlying:

$$dS_t = (r - q)S_t dt + \sigma_{\text{loc}}(t, S_t) S_t dW_t.$$

There is no new randomness here. Volatility is not a risk factor; it is a lookup table indexed by where the underlying happens to be and what time it is. That makes the model complete, in the sense of chapter 4 — the underlying and the money market still replicate everything.

Theorem 9.5 (Dupire). *If call prices $C(T, K)$ for all expiries and strikes come from a local volatility model, then*

$$\sigma_{\text{loc}}^2(T, K) = \frac{2 \left(\frac{\partial C}{\partial T} + (r - q)K \frac{\partial C}{\partial K} + q C \right)}{K^2 \frac{\partial^2 C}{\partial K^2}}. \quad (9.4)$$

Proof. Let $p(T, S)$ be the risk neutral density of S_T , which moves according to (9.2). Write the call price as $C(T, K) = e^{-rT} \int_K^\infty (S - K)p(T, S) dS$ and differentiate in T :

$$\frac{\partial C}{\partial T} = -rC + e^{-rT} \int_K^\infty (S - K) \frac{\partial p}{\partial T} dS.$$

Substitute the forward equation and take the two resulting integrals in turn. For the drift term, one integration by parts gives

$$\begin{aligned} \int_K^\infty (S - K) \left(-\frac{\partial}{\partial S} [(r - q)Sp] \right) dS &= \left[-(S - K)(r - q)Sp \right]_K^\infty + \int_K^\infty (r - q)Sp dS \\ &= (r - q) \int_K^\infty Sp dS, \end{aligned}$$

the boundary term vanishing at $S = K$ because of the factor $(S - K)$, and at infinity because the density decays. Splitting $S = (S - K) + K$ and using Theorem 9.3,

$$\int_K^\infty Sp dS = \int_K^\infty (S - K)p dS + K \int_K^\infty p dS = e^{rT}C - Ke^{rT} \frac{\partial C}{\partial K}.$$

For the diffusion term, integrate by parts twice. Writing $\phi(S) = \sigma_{\text{loc}}^2(T, S)S^2p(T, S)$ for brevity,

$$\begin{aligned} \frac{1}{2} \int_K^\infty (S - K) \frac{\partial^2 \phi}{\partial S^2} dS &= \frac{1}{2} \left(\left[(S - K) \frac{\partial \phi}{\partial S} \right]_K^\infty - \int_K^\infty \frac{\partial \phi}{\partial S} dS \right) \\ &= \frac{1}{2} \left(0 - [\phi]_K^\infty \right) = \frac{1}{2} \phi(K) \\ &= \frac{1}{2} \sigma_{\text{loc}}^2(T, K) K^2 p(T, K) = \frac{1}{2} \sigma_{\text{loc}}^2(T, K) K^2 e^{rT} \frac{\partial^2 C}{\partial K^2}. \end{aligned}$$

Collecting the pieces and cancelling e^{-rT} against e^{rT} ,

$$\begin{aligned} \frac{\partial C}{\partial T} &= -rC + (r - q) \left(C - K \frac{\partial C}{\partial K} \right) + \frac{1}{2} \sigma_{\text{loc}}^2(T, K) K^2 \frac{\partial^2 C}{\partial K^2} \\ &= -qC - (r - q)K \frac{\partial C}{\partial K} + \frac{1}{2} \sigma_{\text{loc}}^2(T, K) K^2 \frac{\partial^2 C}{\partial K^2}, \end{aligned}$$

and rearranging for σ_{loc}^2 gives (9.4). □

What has just happened deserves restating slowly, because the formula is usually met as something to be memorised.

We started with a question about a function of two variables — what is $\sigma_{\text{loc}}(t, S)$? — and we have answered it by rearranging an equation. There was no search, no fitting, and no approximation. Given a surface of call prices smooth enough to differentiate, formula (9.4) *tells* you the local volatility, at every point, in closed form. And since the surface was arbitrary, this says something remarkable:

Remark (Any smile can be fitted exactly). For any arbitrage-free surface of European option prices, there is a local volatility model reproducing every one of them exactly.

The simplest reading of the formula is dimensional. The denominator, $K^2 \partial^2 C / \partial K^2$, is the density at K weighted by K^2 , and the numerator is essentially the rate at which option value accrues as expiry is pushed out. So local volatility is the ratio of how much value time buys you at that strike to how much probability is sitting there. Where the market is confident the underlying will not go, the density is small, the denominator is small, and the local volatility that has to be assigned there is correspondingly large.

Before trusting it, we should check it on a case where we know the answer.

Example 9.2 (Dupire on a flat smile). Suppose the surface came from Black-Scholes with a constant σ , and take $r = q = 0$ so that (9.4) reduces to $\sigma_{\text{loc}}^2 = 2C_T / (K^2 C_{KK})$. From chapter 5,

$$\frac{\partial C}{\partial T} = \frac{S_0 n(d_1) \sigma}{2\sqrt{T}}, \quad \frac{\partial^2 C}{\partial K^2} = \frac{n(d_2)}{K \sigma \sqrt{T}},$$

so

$$\sigma_{\text{loc}}^2 = \frac{2}{K^2} \cdot \frac{S_0 n(d_1) \sigma}{2\sqrt{T}} \cdot \frac{K \sigma \sqrt{T}}{n(d_2)} = \sigma^2 \frac{S_0 n(d_1)}{K n(d_2)}.$$

It remains to show the last fraction is 1. Since $d_1 - d_2 = \sigma\sqrt{T}$ and $d_1 + d_2 = 2 \ln(S_0/K) / (\sigma\sqrt{T})$,

$$\frac{n(d_1)}{n(d_2)} = e^{-\frac{1}{2}(d_1^2 - d_2^2)} = e^{-\frac{1}{2}(d_1 - d_2)(d_1 + d_2)} = e^{-\ln(S_0/K)} = \frac{K}{S_0},$$

so $S_0 n(d_1) = K n(d_2)$ and $\sigma_{\text{loc}} = \sigma$. A flat smile implies a constant local volatility, as it must.

Remark (Why the formula is harder to use than to derive). Formula (9.4) asks for one derivative in T and two in K of a function known only at the strikes and expiries someone chose to quote. Differentiating twice amplifies noise ferociously, and the second derivative sits in the denominator, where a small error becomes a large one and a negative value becomes a catastrophe. In practice one never differences raw quotes: the surface is first fitted with something smooth and arbitrage-free by construction, and Dupire is applied to that. The formula is exact; the input never is.

Two things are commonly used for that fit, and they are not the same kind of object. SABR is a model, pressed into service as an interpolator; its implied volatility is an asymptotic expansion, and chapter 10 locates precisely where that expansion stops being arbitrage-free — at low strikes it can imply a negative density, which is the very quantity this formula divides by. The stochastic volatility inspired parameterisation, SVI, is not a model at all: it writes the total implied variance directly as a function of log-moneyness with five parameters to a slice, has no dynamics, and so cannot price a barrier or say how the smile moves. Chapter 20 sets it out, and says how far the shape is dictated by what a smile has to do and how far it is simply the simplest choice that does it. What it can do is be constrained to admit neither butterfly nor calendar arbitrage.

Which of the two is the better tool here follows from the remark above that any smile can be fitted exactly. Dupire does not need a model, because the model is already whatever the surface says it is. It needs a surface convex in the strike and increasing in total variance with maturity, and nothing besides. A parameterisation guaranteeing exactly those is doing the job that was asked; a stochastic volatility model in this role is doing more than was asked everywhere, and in the wings less than is needed.

9.6 Dupire in the Coordinates We Actually Have

Formula (9.4) is written in call prices, and nobody quotes call prices. What is quoted is implied volatility, so it pays to have Dupire in those coordinates too — and the translation turns out to reveal something the price version hides completely.

Use the two natural coordinates. Log-moneyness measures how far a strike is from the forward,

$$y = \ln \frac{K}{F},$$

and total implied variance measures how much uncertainty an option contains,

$$w(y, T) = \sigma_{\text{imp}}^2(y, T) T.$$

Total variance rather than volatility, because variance is what accumulates: two independent periods contribute variances that add, and it is w , not σ_{imp} , that is the natural quantity in time.

Substituting $C = \text{Black}(F, K, \sqrt{w/T}, T)$ into (9.4) and grinding through the chain rule — the algebra is mechanical and long, and adds nothing once you have seen where it goes — gives

$$\sigma_{\text{loc}}^2(y, T) = \frac{\frac{\partial w}{\partial T}}{1 - \frac{y}{w} \frac{\partial w}{\partial y} + \frac{1}{4} \left(-\frac{1}{4} - \frac{1}{w} + \frac{y^2}{w^2} \right) \left(\frac{\partial w}{\partial y} \right)^2 + \frac{1}{2} \frac{\partial^2 w}{\partial y^2}}. \quad (9.5)$$

Do not memorise it. Read what each piece is for. The numerator says local variance is fed by implied variance accumulating with maturity: if a surface's total variance were flat in T , no local volatility could be extracted at all. The denominator is a set of corrections for the surface having shape in the strike direction — the first for its slope, the second for the slope's interaction with the level, the third for its curvature. And since it is a denominator, it must stay positive, which is the arbitrage condition of the previous section reappearing in volatility coordinates: a surface whose skew is too steep for its level makes the denominator vanish and the implied local variance blow up, and that is the calendar or butterfly arbitrage announcing itself.

Now the case to do by hand, because it is the one chapter 10 is built on. Suppose that near the money the smile is a straight line,

$$\sigma_{\text{imp}}(y) = \sigma_0 + s y,$$

with a small slope s . Then to first order in s ,

$$\begin{aligned} w &= (\sigma_0 + s y)^2 T = \sigma_0^2 T + 2\sigma_0 s y T + O(s^2), \\ \frac{\partial w}{\partial y} &= 2\sigma_0 s T + O(s^2), \quad \frac{\partial^2 w}{\partial y^2} = O(s^2), \quad \frac{\partial w}{\partial T} = \sigma_0^2 + 2\sigma_0 s y + O(s^2). \end{aligned}$$

Every term in the denominator of (9.5) carrying $(\partial w / \partial y)^2$ or $\partial^2 w / \partial y^2$ is second order and drops, leaving

$$1 - \frac{y}{w} \frac{\partial w}{\partial y} = 1 - \frac{y}{\sigma_0^2 T} \cdot 2\sigma_0 s T = 1 - \frac{2s y}{\sigma_0}.$$

So

$$\sigma_{\text{loc}}^2 = \frac{\sigma_0^2 + 2\sigma_0 s y}{1 - 2s y / \sigma_0} \approx (\sigma_0^2 + 2\sigma_0 s y) \left(1 + \frac{2s y}{\sigma_0} \right) \approx \sigma_0^2 + 4\sigma_0 s y,$$

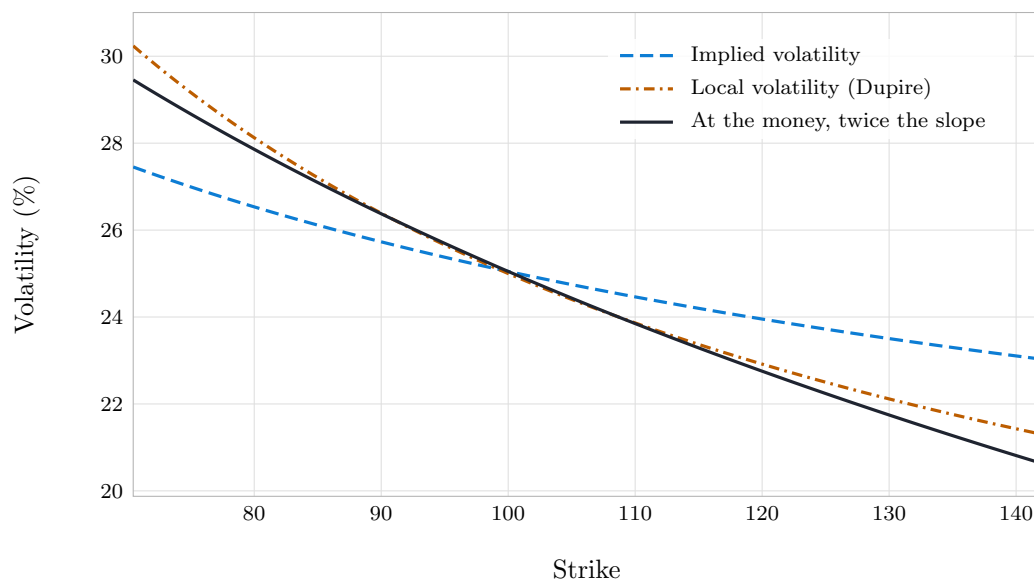


Figure 9.3: A displaced diffusion’s implied volatility, the local volatility Dupire’s formula extracts from it, and the straight line predicted by (9.6). The local volatility curve is twice as steep, and the prediction is a line drawn through the at-the-money point at twice the observed slope — not a fit. It tracks the extracted curve closely near the money and drifts away in the wings, which is what a first-order result should do.

Drawn by `dupire_local_vol`.

and taking the square root,

$$\sigma_{\text{loc}}(y) \approx \sigma_0 + 2sy. \quad (9.6)$$

Remark (The rule of two). The local volatility curve has *twice* the slope of the implied volatility curve that produced it. If the market's smile falls by half a volatility point per ten percent of moneyness, the local volatility falls by a full point.

An option struck at K does not care about volatility at K alone; the underlying travels from the forward to wherever it ends, and the option's implied volatility is a kind of average of the local volatility over that journey. Averaging spot and strike halves a slope. So to produce an observed implied slope s by averaging, the underlying local slope must have been $2s$ — the same factor of two, read backwards.

This apparently innocuous factor is the whole of chapter 10. It says that a local volatility model, having been fitted to today's skew, contains a volatility function twice as steep as that skew. That steepness is not a free parameter: it was forced on us by the fit. And it determines how the model thinks the smile moves when the underlying moves, which turns out to be badly.

9.7 Gyongyi's Theorem

The remark above — that *any* smile can be fitted — should be unsettling. We fitted every European option price exactly, using a model with no volatility risk in it at all — a model in which volatility is a deterministic table. Either European options contain no information about the randomness of volatility, or something has been quietly assumed. It is the former, and Gyongyi's theorem is the statement of exactly how much information they do contain.

Theorem 9.6 (Gyongyi). *Let X be an Itô process*

$$dX_t = \beta_t dt + \delta_t dW_t$$

with β and δ bounded adapted processes, not necessarily Markov and not necessarily functions of X . Define

$$a(t, x) = \mathbb{E}[\beta_t \mid X_t = x], \quad b^2(t, x) = \mathbb{E}[\delta_t^2 \mid X_t = x].$$

Then the Markov diffusion

$$dY_t = a(t, Y_t) dt + b(t, Y_t) dW_t, \quad Y_0 = X_0,$$

has the same distribution as X_t at every single time t .

Proof. Two processes have the same distribution at each time if the expectations of every smooth test function agree at each time, so fix such an f and track $\mathbb{E}[f(X_t)]$. By Itô's lemma,

$$df(X_t) = f'(X_t)\beta_t dt + \frac{1}{2}f''(X_t)\delta_t^2 dt + f'(X_t)\delta_t dW_t,$$

and the stochastic integral has zero expectation under the boundedness assumed, so

$$\frac{d}{dt}\mathbb{E}[f(X_t)] = \mathbb{E}\left[f'(X_t)\beta_t + \frac{1}{2}f''(X_t)\delta_t^2\right].$$

Now the one step that matters. Condition on X_t and use the tower property. Because $f'(X_t)$ and $f''(X_t)$ are functions of X_t alone, they pass through the inner conditional expectation as constants:

$$\mathbb{E}[f'(X_t)\beta_t] = \mathbb{E}[\mathbb{E}[f'(X_t)\beta_t \mid X_t]] = \mathbb{E}[f'(X_t)\mathbb{E}[\beta_t \mid X_t]] = \mathbb{E}[f'(X_t)a(t, X_t)],$$

and identically $\mathbb{E}[f''(X_t)\delta_t^2] = \mathbb{E}[f''(X_t)b^2(t, X_t)]$. Therefore

$$\frac{d}{dt}\mathbb{E}[f(X_t)] = \mathbb{E}\left[f'(X_t)a(t, X_t) + \frac{1}{2}f''(X_t)b^2(t, X_t)\right].$$

Running the same calculation for Y , whose coefficients are a and b by construction, produces the identical equation. So $\mathbb{E}[f(X_t)]$ and $\mathbb{E}[f(Y_t)]$ solve the same evolution equation from the same starting value, and where that equation has a unique solution they agree for all t and all f . $\square$

Remark (What the theorem says, without symbols). Take any process at all — let its volatility be random, let it depend on the whole history, let it be driven by other factors we cannot see. Now ask what it looks like at a single fixed date, ignoring how it got there. Gyongi says that this snapshot is always reproduced by some Markov diffusion, and tells you which one: at each point, use the *average* of the true volatility over all the ways the world could have arrived at that point.

The averaging is where the information goes. Two worlds, one in which volatility at $S = 100$ is always 20% and another in which it is 10% half the time and about 27% the other half, have the same average squared volatility there, and therefore the same distribution of S_T , and therefore the same European option prices. Nothing you can construct from options at a single expiry can tell them apart — because the option price is an expectation over that expiry's distribution, and the distributions are the same.

They are, of course, completely different worlds to be in if you have to hedge.

Applying the theorem to a general stochastic volatility model, in which $dS_t/S_t = \dots + \sigma_t dW_t$ with σ_t random, and comparing with Dupire gives the identity that the next two chapters revolve around.

Theorem 9.7 (Local volatility is a conditional expectation). *If a model with random instantaneous volatility σ_t reproduces the market's European option prices, then the Dupire local volatility of that same surface satisfies*

$$\sigma_{loc}^2(T, K) = \mathbb{E}[\sigma_T^2 \mid S_T = K]. \quad (9.7)$$

Proof. Both sides of (9.7) are read off the same object — the price surface — so the proof consists of computing that surface twice and comparing.

Write $C(T, K) = \mathbb{E}[(S_T - K)^+]$ for the undiscounted call surface of the stochastic volatility model, and differentiate it in maturity. By Itô's lemma applied to the payoff, taking the second derivative in the distributional sense since the payoff has the kink of chapter 3,

$$d(S_T - K)^+ = \mathbf{1}_{\{S_T > K\}} dS_T + \frac{1}{2}\delta(S_T - K)\sigma_T^2 S_T^2 dT,$$

and the first term has zero expectation because S is a martingale. So

$$\frac{\partial C}{\partial T} = \frac{1}{2}\mathbb{E}[\delta(S_T - K)\sigma_T^2 S_T^2] = \frac{1}{2}K^2 \mathbb{E}[\sigma_T^2 \mid S_T = K] p(T, K), \quad (9.8)$$

where the last step is the definition of a conditional expectation: integrating against the delta picks out the density $p(T, K)$ of S_T at K and, with it, the average of σ_T^2 over exactly the paths that arrive there. This is the step that produces the conditioning, and it produces it from nothing more than the fact that the second derivative of a call payoff is supported on a single level.

Now the same surface in the local volatility model. There the volatility is a function, so the identical computation gives

$$\frac{\partial C}{\partial T} = \frac{1}{2} K^2 \sigma_{\text{loc}}^2(T, K) p(T, K),$$

with no expectation to take. And the density is the same density: $\partial^2 C / \partial K^2 = p(T, K)$ for any model, which is the Breeden-Litzenberger identity, so two models agreeing on the surface agree on the marginal law of S_T .

Equating the two expressions for $\partial C / \partial T$ and cancelling $\frac{1}{2} K^2 p(T, K)$, which is non-zero wherever the density is, gives (9.7). $\square$

Remark (What the proof shows about the scope). Three features of the argument stand out, because they say exactly how far the theorem reaches.

It used the martingale property of S and nothing else about the drift, so it is indifferent to how many factors the volatility has, whether they are correlated with S , and whether σ is Markovian. Any of that would appear only inside the conditional expectation.

It matched one derivative of one surface. That is why the conclusion is about one-dimensional marginals and about European prices, and why nothing follows about the joint law across maturities. Chapter 10 is where that gap does damage: two models with the same σ_{loc} can disagree completely about a forward-starting payoff, and (9.7) does not notice.

It divided by $p(T, K)$. Where the density is small — far wings, short maturities — the local volatility is a ratio of two small numbers, which is the analytic reason for the numerical fragility of (9.4) rather than a defect of any particular scheme for evaluating it.

There is one case where the conditional expectation is available in closed form: a forward whose volatility is drawn once, at time zero, from two values. That is a genuine stochastic volatility — random and unknown — and its mimicking local volatility can be written down and simulated. The two models agree on every European price to within a few hundredths of a volatility point, which is a strong statement given that they are nothing like each other path by path: one has a volatility that never moves and is a function of the level, the other a volatility that is constant along each path and unknown at the start.¹

Constructively, it is the tool of chapter 11: it says that if you have a stochastic volatility model whose dynamics you like but whose fit you do not, you can multiply its volatility by whatever function makes (9.7) hold, and recover an exact fit without disturbing the dynamics. That is what a local-stochastic volatility model is.

Destructively, it says that local volatility is an average of something, and an average is a summary. The local volatility surface is not the market's volatility; it is the market's volatility with all of its own randomness integrated away. Every model matching the smile shares this one average, and they differ in everything the average discarded — which is to say, in everything that depends on more than one date at a time.

¹checked, with a guard against the check being vacuous — if the mixture's smile were flat a constant volatility would reproduce it and nothing would have been shown.

That is the subject of the next chapter. Local volatility fits every European option in the market exactly, and it is nevertheless the wrong model for almost everything else, in a way we can make quantitative rather than rhetorical.

9.8 Scheduled Events on the Surface

The curve chapter built the front end as a step function because the overnight rate moves only at meetings. The surface has the same structure and the same remedy, in the coordinate that matters here: variance rather than rate.

Between events, variance accumulates diffusively — linearly in time, which is what makes total variance the natural interpolation coordinate rather than volatility. At an event it jumps, because a policy decision or a labour report delivers a move in an instant. So the surface is

$$\sigma_{\text{imp}}^2(T) T = \int_0^T \sigma_{\text{diff}}^2(u) du + \sum_{\text{events } i \leq T} v_i, \quad (9.9)$$

with v_i the variance contributed by the i th event.

The construction that follows is the same manoeuvre as the meeting-date curve. Strip the event lumps out of the quoted total variances, interpolate what is left — which is now smooth, because the jumps have been removed — and add the lumps back at the right dates. Interpolating the raw surface instead produces the familiar pathology: two options expiring three days apart across a payroll date have implied volatilities that differ by several points, a smooth interpolation reads that as a volatility term structure, and every expiry in between is marked wrong.

Remark (Calibrating the lumps). The v_i are not free parameters to be fitted alongside everything else. They are pinned by the pairs of expiries that straddle each event: the difference in total variance between an option expiring just before a release and one expiring just after is v_i , up to the diffusive variance of the days between. Where such a pair is quoted, the event variance is a read-off rather than a fit.

Where it is not quoted — and for most events it is not — the lump is marked from history: the average squared move of the rate on past occurrences of that release. That is a genuinely different kind of number from the rest of the surface, being a historical estimate sitting inside a risk-neutral object, and it should be understood as a marking convention rather than as a calibration. It is also where a desk's view lives: a trader who thinks this meeting is more consequential than the last four marks the lump up, and that is a position rather than a parameter.

A jump or a lump of volatility?

Chapter 4 established that a policy decision has a finite and small outcome space — twenty-five basis point multiples, three or four possibilities — so it can be represented literally, as a discrete distribution over moves. The alternative is to represent it as extra Gaussian variance with the same second moment, which is what (9.9) does. The two are not the same model and the question is when the difference is paid for.

Calculation 9.8 (Where the two representations part). A meeting with outcomes $\{0, +25, +50\}$ basis points at probabilities $\{0.35, 0.55, 0.10\}$, a forward of 4%, and a diffusive normal volatility

of sixty basis points a year. Compare pricing the event as a three-point mixture against a single Gaussian with the same total variance.²

At the money they agree, to within a fraction of a per cent, because a straddle is priced off the variance and the variance was matched by construction. This is why the choice can go unnoticed indefinitely on an at-the-money book.

Across strikes they do not. The mixture puts mass at three places and less between them, so its implied volatility is not flat: at a one month expiry it runs from about 70 basis points at a strike a hundred below the forward to 82 at the money, an eleven basis point smile manufactured entirely by the event. A single Gaussian has no smile at all.

And the size depends on how much diffusion sits between the event and the expiry, which is the finding that decides everything. A digital struck between two outcomes prices at 0.194 under the mixture against 0.230 under the Gaussian three days after the meeting — a sixth of the price. At one month the two agree to four decimal places, and at a year they are indistinguishable. Seventeen basis points of diffusion against a twenty-five basis point spacing is enough to smear three atoms into something very nearly Gaussian.

Structure (The instrument decides, and so does the calendar). Calculation 9.8 gives a rule with two conditions rather than a preference.

What is the payoff reading? An instrument priced off the second moment — a straddle, a variance swap, anything approximately quadratic — cannot tell the two representations apart, because they were constructed to agree there. An instrument priced off the shape of the density — a digital, a tight butterfly, a range accrual, a barrier sitting between two policy outcomes — reads exactly what differs. So the choice is decided by the book, and a desk trading at-the-money volatility genuinely does not need the discrete representation.

And how much diffusion follows the event? The discrete structure survives only while the diffusion is small compared with the spacing between outcomes. For an option expiring days after a meeting that is the whole picture; a month later it is nearly gone; a year later the event is a lump of variance and nothing else. This is why the discrete representation is a front-end tool: it earns its complexity on the short-dated options that straddle a meeting, and is wasted effort further out.

Both conditions have to hold for the extra machinery to be worth it, which is a narrower criterion than the usual advice to model events properly. The general lesson is the one chapter 18 makes about copulas and chapter 10 about the forward smile: matching a moment is not matching a distribution, and whether the difference matters is a question about the payoff rather than about the model.

References

- Breeden, D. T., & Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4), 621–651.
- Dupire, B. (1994). Pricing with a smile. *Risk*, 7(1), 18–20.
- Gyongi, I. (1986). Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability Theory and Related Fields*, 71(4), 501–516.

²`events::PolicyEvent`, priced in closed form as a probability weighted sum of Bachelier prices rather than simulated.

- Derman, E., & Kani, I. (1994). Riding on a smile. *Risk*, 7(2), 32–39.