

Chapter 10

What causes equity smiles?

In Chapter 8 we have characterized stochastic volatility smiles and how they are related to the model's specification: where do skew and curvature come from and how we can compute them approximately?

Then in Chapter 9 we have focused on dynamical aspects of smiles in stochastic volatility models: how do ATM volatilities move in these models? From the bare knowledge of the smile generated by a stochastic volatility model can we say anything about the joint dynamics of spot and implied volatilities in the model?

“But what about actual equity smiles?” is the restive reader bound to ask. What is it that is responsible for their skew and curvature? Large historical drawdowns in equity indexes are often purported to be responsible for the strong negative skew of implied index smiles. Is this correct? Are vanilla smiles in any way related to statistical properties of historical returns? Which payoffs do the latter impact? These are the questions we address.

The (un)related topic of jump-diffusion and Lévy models is dealt with in Appendix A.

10.1 The distribution of equity returns

Figure 10.1 shows the cumulative distribution of normalized negative and positive returns of the Dow Jones index. We have used daily closing quotes from January 1, 1900 to July 20, 2014.¹

We compute daily returns $r_i = \frac{S_{i+1}}{S_i} - 1$. The square root of the second moment of non-centered negative returns is 1.15% and that of non-centered positive returns is 1.08%. We separately normalize negative and positive returns by the square roots of second-order moments so that the non-centered second moment of normalized positive returns \bar{r}_i equals 1, as does that of negative returns. The lowest (normalized) negative return in our sample is -19.6 and the highest positive return is 14.3 .²

¹We gratefully acknowledge the website <http://stooq.com> for making these data available.

²With the second moments of non-centered positive and negative \bar{r}_i equal to 1, the standard deviation of the \bar{r}_i is almost exactly equal to 1.

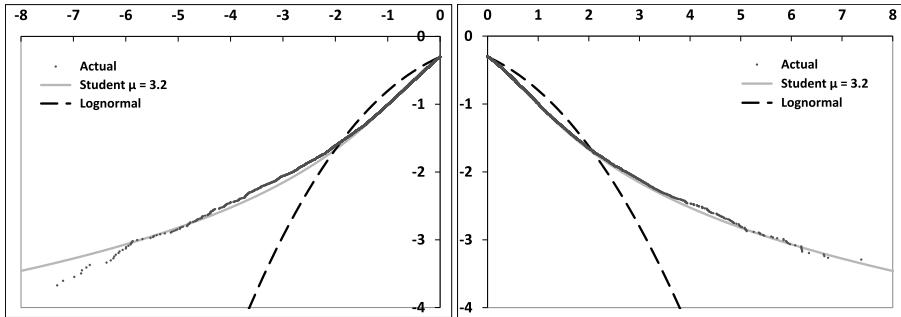


Figure 10.1: Logarithm (\log_{10}) of the empirical distribution function of normalized negative (left-hand graph) and positive (right-hand graph) returns of the Dow Jones index, together with the lognormal and Student distribution functions.

We then rank negative returns and define the empirical distribution function of normalized negative returns \bar{r} as:

$$P[\bar{r} \leq \bar{r}_i] = \frac{1}{2} \frac{i}{N^-}$$

where N^- is the number of negative returns in our sample ($N^- = 14103$) and the normalization factor $\frac{1}{2}$ ensures that $P[\bar{r} \leq 0] = \frac{1}{2}$. We do the same for positive returns ($N^+ = 15657$). In Figure 10.1, $\log_{10}(P[\bar{r} \leq r])$ (left-hand graph) and $\log_{10}(P[\bar{r} \geq r])$ (right-hand graph) are graphed as a function of r .

Together with the empirical distribution, we show two distributions: (a) the lognormal distribution function, (b) the Student distribution function with $\mu = 3.2$.

The Student distribution is well-suited to the modeling of fat-tailed random variables; its density is defined by:

$$\rho_\mu(x) = \frac{\Gamma(\frac{1+\mu}{2})}{\sqrt{\mu\pi}\Gamma(\frac{\mu}{2})} \frac{1}{\left(1 + \frac{x^2}{\mu}\right)^{\frac{1+\mu}{2}}} \quad (10.1)$$

where μ is customarily called “number of degrees of freedom”.

The smaller μ the thicker the tails. For large values of x the density scales like $\frac{1}{x^{1+\mu}}$, thus the one-sided cumulative distribution function scales like $\frac{1}{x^\mu}$. Only moments of order smaller than μ exist.

The variance of a Student random variable is $\frac{\mu}{\mu-2}$ and its kurtosis is $\frac{6}{\mu-4}$. For $\mu \leq 4$ the fourth moment diverges and so does the second moment for $\mu \leq 2$; for $\mu \rightarrow \infty$ the Student density converges to the Gaussian density.

The value $\mu = 3.2$ is obtained by least-square minimization of the difference of the logarithms of the empirical and Student distributions. When comparing empirical and Student distributions in Figure 10.1, one should observe that while the empirical

curve for negative returns consists of more than 14000 points, only a hundred of them correspond to values of $\bar{r} \leq -4$.

Student distribution functions for different values of μ are shown in Figure 10.2. Any value in the interval [3, 4] provides an acceptable fit to empirical data.³

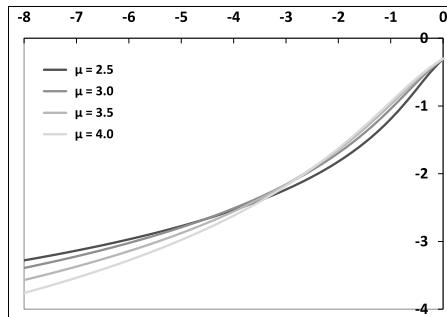


Figure 10.2: Logarithm (\log_{10}) of the left tail of the distribution function of a Student random variable, normalized so that its variance is 1, for different values of μ .

Thus our preliminary conclusions are:

- The empirical distribution of realized equity index returns is well approximated by a Student distribution with μ typically in the interval [3, 4], with tail probabilities much larger than those of a lognormal distribution. Figure 10.1 shows that the lognormal density underestimates the probability of a move larger than 4 standard deviations by a factor of 100.
- Negative and positive returns have very similar second-order moments and tail parameters. In contrast with what is frequently heard, there is no evidence that negative returns have thicker tails than positive ones.

These conclusions are by no means specific to the Dow Jones index, they apply broadly to all equity indexes. Figure 10.3 shows another example; a Hong Kong-based index – the Hang Seng China Enterprises Index (HSCEI), which is roughly twice as volatile as the Dow Jones index, using data from 1993 to 2014.

³For $\mu < 4$, the kurtosis κ of Student-distributed returns diverges. Obviously, the kurtosis estimator applied to any sample yields a finite number. However, the divergence of κ manifests itself in the fact that occasional large returns generate large jumps in the estimator for κ . Tails of the Student distribution for $\mu < 4$ are thick enough that these events occur sufficiently frequently that defining an average finite value for κ reliably becomes practically impossible. With respect to our Dow Jones data, even though our best fit yields values of $\mu < 4$, this does not mean that $\kappa = \infty$, as the very rare and large events that cause the divergence of κ for $\mu < 4$ may not be present in our sample.

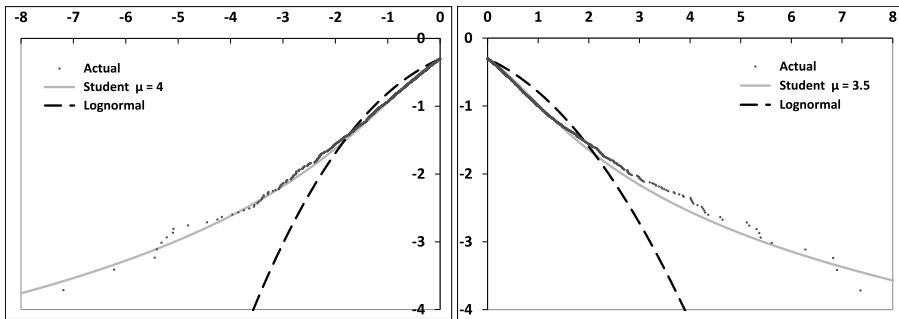


Figure 10.3: Logarithm (\log_{10}) of the empirical distribution function of normalized negative (left-hand graph) and positive (right-hand graph) returns of the HSCEI index, together with the lognormal and Student distribution functions.

10.1.1 The conditional distribution

Even though empirical curves in Figure 10.1 are remarkably smooth, one could question the pertinence of fitting a single distribution to one century worth of data, during which very dissimilar volatility regimes have existed.

In the analysis of the P&L of a delta-hedged option position in Chapter 1, daily returns r_i are modeled as:

$$r_i = \sigma_i \sqrt{\delta t} z_i \quad (10.2)$$

where σ_i is the instantaneous volatility and z_i are iid random variables with unit variance.

The rationale for this ansatz is that the variability of the probability distribution of r_i is condensed in that of the scale factor σ_i – the z_i are identically distributed.

When σ_i is stochastic the resulting distribution of the r_i is non-Gaussian even though the z_i may be. How much of the thickness of the tails of daily returns is generated by (a) the randomness of σ_i , (b) the distribution of z_i ?

(10.2) is a natural ansatz but accessing σ_i is difficult in practice. Figure 10.4 shows the same graphs as in Figure 10.1, except each return is normalized by the historical volatility calculated using the 200 previous returns, rather than that evaluated over the whole historical sample.

As expected, normalizing returns by an (heuristic) estimate of σ_i reduces the tail thickness, which is manifested in larger values for μ . A least-squares fit produces $\mu = 3.8$ for negative returns – a modest increase with respect to $\mu = 3.2$ – while we get $\mu = 6$ for positive returns.⁴

⁴The fact that the conditional value of μ for negative returns is little changed with respect to its unconditional value may be due to the fact that large negative returns are more likely to occur irrespective of current volatility levels: normalization by the latter does not shrink their tails substantially.

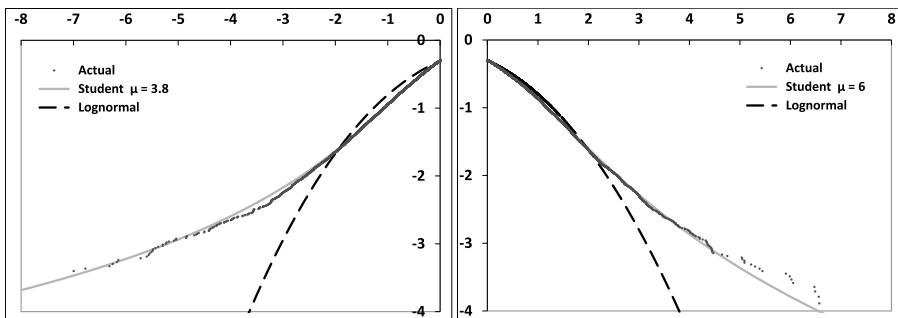


Figure 10.4: Logarithm (\log_{10}) of the empirical distribution function of conditional normalized negative (left-hand graph) and positive (right-hand graph) returns of the Dow Jones index, together with the lognormal and Student distribution functions.

10.2 Impact of the distribution of daily returns on derivative prices

Empirical distributions of returns – even conditional ones – have substantially fatter tails than the lognormal distribution. This has prompted some to argue that using lognormal models – or more generally diffusive stochastic volatility models – for pricing derivatives is inappropriate. Others have claimed that it is precisely the fat-tailed/non-lognormal nature of returns that accounts for the volatility smile. Is either of these statements correct?

Remember we have argued in Chapters 8 and 9 – see in particular the discussion in Section 9.10.5, page 376 – that the ATMF skew of the volatility smile is generated by the covariance of spot and *implied* volatilities. Do fat-tailed returns alter this picture?

We now investigate the following questions:

- Is the fat-tailed nature of returns manifested in any way in equity smiles? Stated differently, is there any trace of the one-day smile in smiles of standard maturities?
- Which payoffs are specifically sensitive to tails of daily returns/the one-day smile?

We answer these questions by using a model that enables us to separate the effects of (a) the distribution of daily returns, (b) the effect of stochastic (implied) volatilities.

10.2.1 A stochastic volatility model with fat-tailed returns

We use the two-factor model introduced in Section 7.4, page 226. While the model for forward variances ξ_t^τ is continuous, the spot returns are discrete and are simulated over one-day time intervals.

We cannot use a Student distribution for $\ln(\frac{S_{t+\Delta}}{S_t})$, as $E_t [S_{t+\Delta}]$ would be infinite, thus we use the Student distribution for $\frac{S_{t+\Delta}}{S_t} - 1$ and set:

$$S_{t+\Delta} = S_t [1 + (r - q) \Delta + \sigma_t \delta Z] \quad (10.3)$$

where σ_t is given by:

$$\sigma_t = \sqrt{\frac{1}{\Delta} \int_t^{t+\Delta} \xi_t^\tau d\tau}$$

and the dynamics of forward variances ξ_t^τ in the two-factor model is given by equation (7.28), page 226:⁵

$$d\xi_t^\tau = 2\nu \xi_t^\tau \alpha_\theta \left((1 - \theta) e^{-k_1(\tau-t)} dW_t^1 + \theta e^{-k_2(\tau-t)} dW_t^2 \right)$$

whose solution is given in equation (7.33), page 227, in terms of two Ornstein-Uhlenbeck processes X^1, X^2 , whose increments over $[t, t + \Delta]$ are given by:

$$\delta X^i = \int_t^{t+\Delta} e^{-k_i(t+\Delta-u)} dW_u^i$$

In (10.3) δZ is a random variable with mean 0 and variance Δ . In the standard version of the two-factor model δZ is simply replaced by: $\delta Z \equiv \delta W^S \equiv \int_t^{t+\Delta} dW_u^S$ and the covariances of the Gaussian random variables δW^S and δX^i are given by:

$$E [\delta X^i \delta X^j] = \rho_{ij} \frac{1 - e^{-(k_i+k_j)\Delta}}{k_i + k_j} \quad E [\delta X^i \delta W^S] = \rho_{iS} \frac{1 - e^{-k_i\Delta}}{k_i}$$

Here we need to draw δZ from a two-sided Student distribution with different values of μ for the left and right tails.

While in empirical distributions, probabilities for negative and positive returns are almost exactly equal to $\frac{1}{2}$, we would like to be able to set these probabilities, or in other words the price of the one-day at-the-money digital option: $E [1_{\delta Z \geq 0}]$, that is, have a handle on the one-day ATM skew – see the discussion in Section 10.2.2 below.

Let us call p_+, p_- the probabilities of positive and negative returns, and μ_+, μ_- the parameters of the corresponding Student distributions. We define δZ as:

$$\begin{cases} \delta Z = \sigma_+ \sqrt{\Delta} |X_{\mu_+}| & \text{with probability } p_+ \\ \delta Z = -\sigma_- \sqrt{\Delta} |X_{\mu_-}| & \text{with probability } p_- \end{cases}$$

⁵No need for volatility-of-volatility smile here.

where X_μ denotes a Student random variable with μ degrees of freedom. The density of δZ is given by:

$$\begin{cases} \rho(\delta Z) = \frac{2p_+}{\sigma_+\sqrt{\Delta}} \rho_{\mu+}\left(\frac{\delta Z}{\sigma_+\sqrt{\Delta}}\right) & \text{with probability } p_+ \\ \rho(\delta Z) = \frac{2p_-}{\sigma_-\sqrt{\Delta}} \rho_{\mu-}\left(\frac{\delta Z}{\sigma_-\sqrt{\Delta}}\right) & \text{with probability } p_- \end{cases}$$

The values of σ_+, σ_- must be such that:

$$E[\delta Z] = 0 \quad E[(\delta Z)^2] = \Delta \quad (10.4)$$

Since the standard deviation of a Student random variable is $\sqrt{\frac{\mu}{\mu-2}}$ it is more natural to rewrite σ_+, σ_- as:

$$\sigma_+ = \sqrt{\frac{\mu_+ - 2}{\mu_+}} \zeta_+ \quad \sigma_- = \sqrt{\frac{\mu_- - 2}{\mu_-}} \zeta_-$$

The conditions in (10.4) read:

$$\begin{aligned} p_+ \zeta_+ \alpha_+ - p_- \zeta_- \alpha_- &= 0 \\ p_+ \zeta_+^2 + p_- \zeta_-^2 &= 1 \end{aligned}$$

where $\alpha_+ = \frac{2}{\sqrt{\pi}} \frac{\sqrt{\mu_+-2}}{\mu_+-1} \frac{\Gamma(\frac{1+\mu_+}{2})}{\Gamma(\frac{\mu_+}{2})}$ and likewise for α_- . The solution is:

$$\zeta_+ = \frac{p_- \alpha_-}{\sqrt{p_+ (p_- \alpha_-)^2 + p_- (p_+ \alpha_+)^2}} \quad \zeta_- = \frac{p_+ \alpha_+}{\sqrt{p_+ (p_- \alpha_-)^2 + p_- (p_+ \alpha_+)^2}}$$

and δZ is given by:

$$\begin{cases} \delta Z = \zeta_+ \sqrt{\frac{\mu_+ - 2}{\mu_+}} \sqrt{\Delta} |X_{\mu+}| & \text{with probability } p_+ \\ \delta Z = -\zeta_- \sqrt{\frac{\mu_- - 2}{\mu_-}} \sqrt{\Delta} |X_{\mu-}| & \text{with probability } p_- \end{cases}$$

Student random variables can be generated in a number of ways. In our setting, we need (a) to correlate δZ with $\delta X^1, \delta X^2$ and (b) to be able to degenerate δZ into a Gaussian random variable to recover the standard form of the two-factor model.

It is then simpler to start with the Brownian increments δW^S supplied by the Monte Carlo engine of our standard two-factor model and map them into δZ according to:

$$\delta Z = \sqrt{\Delta} f \left(\frac{\delta W^S}{\sqrt{\Delta}} \right) \quad (10.5)$$

with f given by:

$$\begin{cases} x \leq \mathcal{N}_G^{-1}(p_-) & f(x) = \zeta_- \sqrt{\frac{\mu_- - 2}{\mu_-}} \mathcal{N}_{\mu_-}^{-1}\left(\frac{\mathcal{N}_G(x)}{2p_-}\right) \\ x \geq \mathcal{N}_G^{-1}(p_-) & f(x) = \zeta_+ \sqrt{\frac{\mu_+ - 2}{\mu_+}} \mathcal{N}_{\mu_+}^{-1}\left(\frac{1}{2} + \frac{\mathcal{N}_G(x) - p_-}{2p_+}\right) \end{cases} \quad (10.6)$$

where \mathcal{N}_G is the cumulative distribution function of the standard normal variable, \mathcal{N}_G^{-1} its inverse, and \mathcal{N}_{μ}^{-1} is the inverse cumulative distribution function of a Student random variable with μ degrees of freedom.⁶

Note that if $p_+ = p_- = \frac{1}{2}$, $\lim_{\mu_+, \mu_- \rightarrow \infty} \delta Z = \delta W$, as expected.

Rescaling spot/volatility correlations

We can now simulate the joint dynamics of S_t and forward variances using the Monte Carlo engine of our two-factor model and are one step away from vanilla smiles.

As we vary μ_+, μ_- to assess the influence of the one-day smile we must ensure that other features of the model are unchanged. As we replace δW^S with δZ in (10.3) the instantaneous volatilities of VS volatilities are unaffected, as is the instantaneous volatility of S , but spot/volatility covariances are altered.

We thus need to define new rescaled correlations ρ_{1S}^*, ρ_{2S}^* so that the instantaneous correlations – or covariances – of δZ and δX^1 and of δZ and δX^2 using ρ_{1S}^*, ρ_{2S}^* are identical to the covariances in the standard two-factor model parametrized with ρ_{1S}, ρ_{2S} , that is $E[\delta W^S \delta X^1]$ and $E[\delta W^S \delta X^2]$:

$$E_* [\delta Z \delta X^i] = E [\delta W^S \delta X^i]$$

where E_* denotes the expectation evaluated using the new correlations. The formula for rescaling correlations is easily obtained. Spot/volatility correlations are rescaled uniformly, according to:

$$\frac{\rho_{iS}^*}{\rho_{iS}} = \frac{\Delta}{E[\delta Z \delta W^S]} = \frac{1}{\int_{-\infty}^{+\infty} \phi(x) x f(x) dx} \quad (10.7)$$

where f is the mapping function in (10.6) and ϕ is the probability density of the standard normal variable.

The integral in the denominator of (10.7) is evaluated numerically once. We have $\rho_{iS}^* \geq \rho_{iS}$.⁷

⁶ $\mathcal{N}_G, \mathcal{N}_G^{-1}, \mathcal{N}_{\mu}^{-1}$ are readily available in standard numerical libraries. $f(x)$ should be cached – simulating δZ is then only marginally more expensive than simulating δW^S .

⁷ We have $\int \phi(x) x^2 dx = 1$ and $\int \phi(x) f(x)^2 dx = 1$, where f is the mapping function defined in (10.6). From Cauchy-Schwarz we get: $\int \phi(x) x f(x) dx \leq 1$. Unless we go to very low values of μ , ρ_{iS}^*/ρ_{iS} is not too large. For example, taking $\mu_+ = \mu_- = \mu$ and $p_+ = p_- = 1/2$ we have $\rho_{iS}^*/\rho_{iS} = 1.01$ for $\mu = 6$, 1.03 for $\mu = 4$, 1.09 for $\mu = 3$, 1.2 for $\mu = 2.5$.

10.2.2 Vanilla smiles

We now use the model described above to assess the impact of the one-day smile on the vanilla smile.

Smiles are obtained by straight pricing of vanilla payoffs. None of the techniques mentioned in Appendix A of Chapter 8 can be used here – be they the mixing solution, gamma/theta or timer techniques – as they rely on the assumption of a diffusion for S_t that is on the fact that the expansion of the P&L at order 2 in δS is exact for short time intervals – see the discussion below.

We use throughout a flat variance curve: $\xi_0^\tau = \xi_0 = 20\%^2$ and parameters in Table 10.1.⁸

ν	θ	k_1	k_2	ρ	ρ_{SX^1}	ρ_{SX^2}
257%	0.151	8.96	0.46	40%	-74.6%	-13.7%

Table 10.1: Parameters of the two-factor model

We first start with identical probabilities for positive and negative returns ($p_+ = p_- = \frac{1}{2}$) and use identical values for tail parameters: $\mu_+ = \mu_- = \mu$.

Figure 10.5 displays three-month and one-year smiles in the standard and fat-tailed version of the two-factor model.

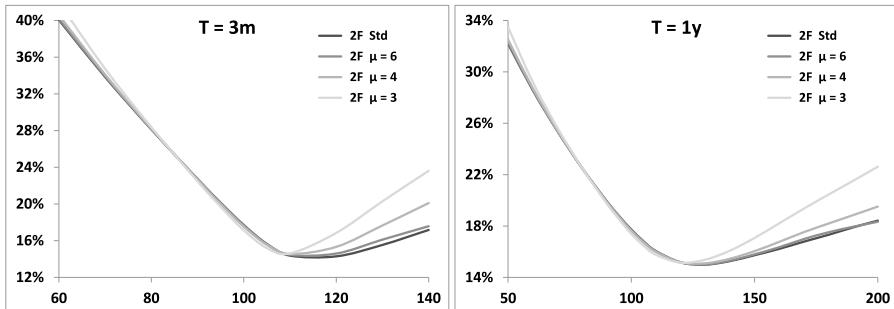


Figure 10.5: Smiles in the two-factor model with Student-distributed returns for different values of μ , $p_+ = p_- = 1/2$ and parameters in Table 10.1 along with the smile in the standard version of the model ($\mu = \infty$).

The near-ATMF smile is hardly affected by tails of daily returns, which mostly impact far out-of-the-money calls. The effect of μ is better appreciated by turning off stochastic volatility: the smile for maturity T is then generated by the 1-day smile only. The latter, together with the 1-year smile, is shown in Figure 10.6.

⁸These were typical of the Euro Stoxx 50 index as of July 2014.

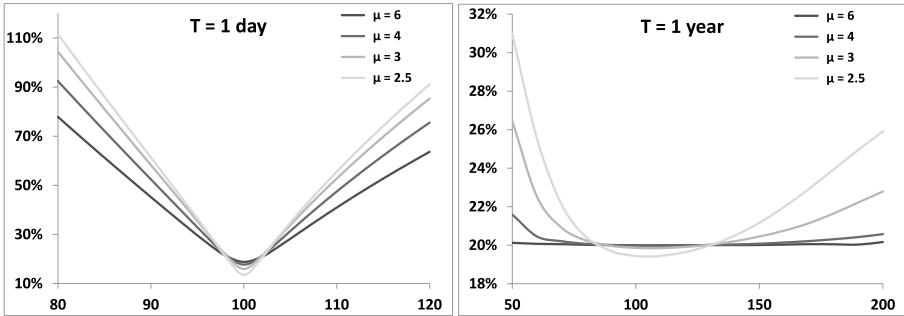


Figure 10.6: 1-day (left) and 1-year (right) smile for different values of $\mu = \mu_+ = \mu_-$ with $p_+ = p_- = 1/2$ and $\sigma = 20\%$ – no stochastic volatility ($\nu = 0$).

We now keep identical values for μ_+ and μ_- but use different values for p_+, p_- so that the ATM skew of the 1-day smile is non-vanishing. The resulting 1-year smile appears in Figure 10.7 for $p_+ = 30\%, 40\%, 50\%, 60\%, 70\%$.

Again let us turn off stochastic volatility; the 1-day and 1-year smiles are shown in Figure 10.8.

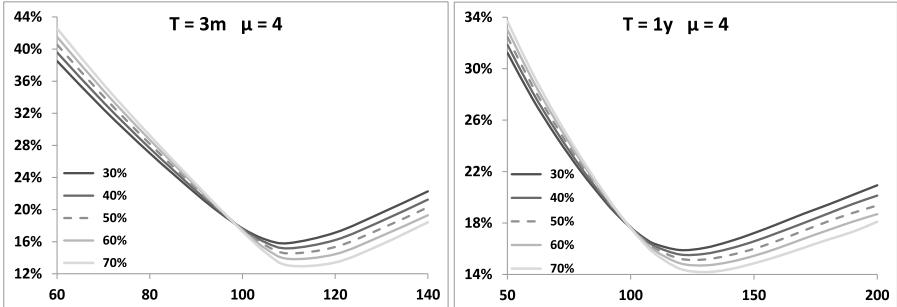


Figure 10.7: Smile in the two-factor model with Student-distributed returns for different values of p_+ , with $\mu_+ = \mu_- = 4$ and parameters in Table 10.1.

It is apparent that the difference $p_+ - p_-$ drives the ATM skew of the 1-day smile, which almost vanishes for $p_+ = p_- = \frac{1}{2}$. The fact that the one-day ATM skew steepens – and implied volatilities for strikes larger than 100 become lower – for larger values of p_+ is understood by noting that p_+ is the (undiscounted) price of a one-day digital that pays 1 if tomorrow's spot value is larger than today's.

Denoting by \mathcal{C}_K the undiscounted price of a call option of strike K , we have $\mathcal{C}_K = \mathcal{C}_K^{BS}(\widehat{\sigma}_K)$ where \mathcal{C}_K^{BS} is the Black-Scholes price and $\widehat{\sigma}_K$ the implied volatility

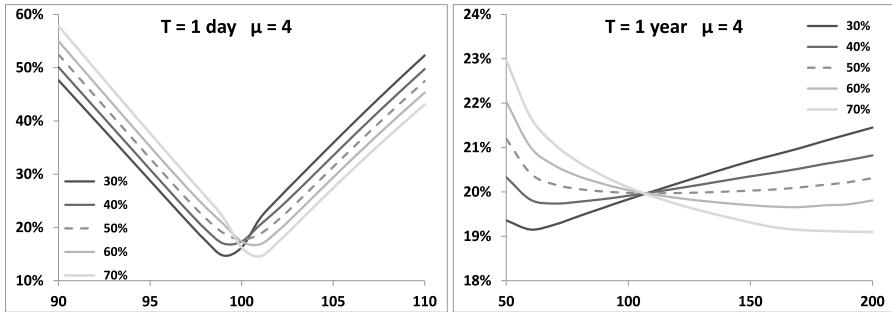


Figure 10.8: 1-day (left) and 1-year (right) smile for different values of p_+ , with $\mu_+ = \mu_- = 4$ and $\sigma = 20\%$ – no stochastic volatility ($\nu = 0$).

of strike K . Thus:

$$\begin{aligned} p_+ &= -\frac{dC_K}{dK} \\ &= -\left. \frac{dC_K^{BS}}{dK} \right|_{\hat{\sigma}_K} - \left. \frac{dC_K^{BS}}{d\hat{\sigma}} \right|_{\hat{\sigma}_K} \frac{d\hat{\sigma}_K}{dK} \end{aligned}$$

According to this formula, increasing p_+ with $\xi_0^T = \xi_0$ unchanged – thus leaving the ATM volatility almost unchanged – has the effect that $\frac{d\hat{\sigma}_K}{dK}$ decreases ($\frac{dC_K^{BS}}{d\hat{\sigma}}$ is positive), which is indeed what occurs in the left-hand graph of Figure 10.8. Note we have deliberately used an unrealistically wide range for p_+ .⁹

How does the contribution of the 1-day ATMF skew to the ATMF skew of maturity T scale with T ? Figure 10.9 shows the 95/105 skew as a function of T , together with a $\frac{1}{T}$ fit, with $\nu = 0$: the only source of smile is the 1-day smile.

The observant reader will remember coming across this scaling before, in the context of discrete forward variance models, in Section 7.8 of Chapter 7 – consider expression (7.119b), page 294, of the ATMF skew for maturity $N\Delta$, where Δ is the time scale of the discrete model.

These models generate the ATMF skew through two mechanisms: (a) the ATMF skew for time scale Δ , (b) the covariance of spot and forward variances – not unlike our fat-tailed stochastic volatility model.

See also Appendix A for more on this scaling.

10.2.3 Discussion

Let us assume no additional degrees of freedom besides S , and vanishing interest rate and repo without loss of generality. The P&L of a short delta-hedged option

⁹See the discussion of the impact of the skew on digital options in Section 1.3.1, page 19.

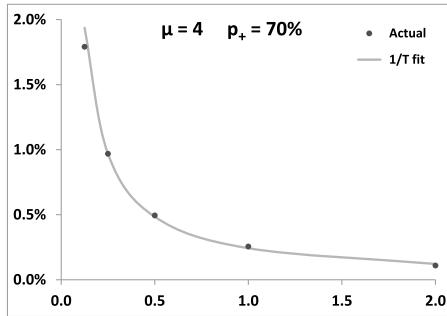


Figure 10.9: Difference of the 95% and 105% implied volatilities, as a function of maturity – in years – (dots) along with a $1/T$ fit (line), in the two-factor model with $\nu = 0$ (no stochastic volatility) with $\mu_+ = \mu_- = 4$ and $p_+ = 70\%$.

position during $\delta t = \Delta$ is:

$$P\&L = -(P(t + \Delta, S + \delta S) - P(t, S)) + \frac{dP}{dS} \delta S \quad (10.8)$$

where $P(t, S)$ is our pricing function. In the models used so far P can be expressed as an expectation, thus :

$$P(t, S) = E[P(t + \Delta, S_{t+\Delta}) | S_t = S]$$

Δ is assumed sufficiently small that the normal/lognormal distinction is not practically relevant. We then have:

$$P(t, S) = E[P(t + \Delta, S(1 + \sigma \delta Z))] \quad (10.9)$$

where δZ is a random variable with vanishing mean and variance Δ , and σ is the instantaneous volatility. Inserting expression (10.9) into (10.8) and expanding in powers of δZ yields:

$$\begin{aligned} P\&L &= -(P(t + \Delta, S + \delta S) - E_{t,S}[P(t + \Delta, S(1 + \sigma \delta Z))]) + \frac{dP}{dS} \delta S \\ &= -\frac{S^2}{2} \frac{d^2 P}{dS^2} \left(\frac{\delta S^2}{S^2} - \sigma^2 \Delta \right) - \sum_{k>2} \frac{S^k}{k!} \frac{d^k P}{dS^k} \left(\left(\frac{\delta S}{S} \right)^k - \sigma^k E[\delta Z^k] \right) \end{aligned} \quad (10.10)$$

In (10.10) we single out the order-two contribution in δS , as when we take the limit of short time intervals between successive delta rehedges, only this term survives in the cumulative P&L. Indeed consider an option of maturity T ; the cumulative P&L over the option's lifetime is the sum of $\frac{T}{\Delta}$ P&Ls of the form (10.10). In the

continuous-time diffusive models considered so far $\frac{\delta S}{S} \propto \sqrt{\Delta}$, thus the order-two contribution in (10.10) is of order Δ , hence generates a cumulative P&L that is finite. The k th contribution in (10.10), however, is of order $\Delta^{\frac{k}{2}}$, generating a cumulative P&L that scales like $\frac{T}{\Delta} \Delta^{\frac{k}{2}}$, hence tends to zero as $\Delta \rightarrow 0$.

- In previous chapters we have implicitly assumed that Δ was sufficiently small that the $\Delta \rightarrow 0$ limit was relevant, hence typical P&L expressions so far only included the first term in (10.10).¹⁰
- When $\Delta > 0$ this is no longer the case, as terms of order $k > 2$ contribute. These involve the quantity $E[\delta Z^k]$ which depends on the particular distribution of δZ . The dependence of vanilla option's prices to the latter is illustrated in figures 10.6 and 10.8 without stochastic volatility, that is with a constant σ , and in figures 10.5 and 10.7 with stochastic volatility.¹¹

The conclusion from figures 10.5 and 10.7 is that, except for exaggerate values of tail parameters, and possibly very short maturities, the impact of the conditional one-day smile on smiles for standard maturities is small. In particular the ATMF skew is predominantly generated by the covariance of spot and forward variances. Only a minute portion of it can be traced to the one-day smile or, equivalently, the conditional distribution of daily returns.

Moreover, high-strike – rather than low-strike – implied volatilities seem to be impacted the most.

The steep ATMF skews observed for equity indexes for typical maturities have thus nothing to do with the fact that historical distributions of daily equity returns may or may not exhibit large drawdowns. Rather, they supply an estimate of the implied level of spot/volatility covariances, just like VS implied volatilities supply an estimate of the implied level of spot volatility.

Recall expression (8.24), page 315, relating the ATMF skew to the integrated covariance of spot and VS volatility for the residual maturity, at order one in volatility of volatility:

$$\mathcal{S}_T = \frac{1}{2\hat{\sigma}_T^3 T} \int_0^T \frac{T-\tau}{T} \frac{\langle d\ln S_\tau \, d\hat{\sigma}_T^2(\tau) \rangle}{d\tau} d\tau$$

What about very short maturities? Smiles in figures 10.5 and 10.7 imply that vanilla smiles for maturities of the order of a few days will exhibit a large sensitivity

¹⁰Note that in jump-diffusion models $\frac{dS}{S}$ does not scale like $\sqrt{\Delta}$ anymore and all orders of δS^k contribute to the P&L, even for arbitrarily small values of Δ – see Section 10.3 below.

¹¹When using the fat-tailed stochastic volatility model of Section 10.2.1, the total P&L comprises contributions from $\delta\xi^\tau$ as well. The order-two contributions read $-\frac{S^2}{2} \frac{d^2 P}{dS^2} \left(\frac{\delta S^2}{S^2} - \sigma^2 \Delta \right) - \int_t^T S \frac{d^2 P}{dS \delta \xi^u} \left(\frac{\delta S}{S} \delta \xi^u - \mu(t, u, \xi) \Delta \right) du - \frac{1}{2} \iint_t^T \frac{d^2 P}{\delta \xi^u \delta \xi^{u'}} \left(\delta \xi^u \delta \xi^{u'} - \nu(t, u, u', \xi) \Delta \right) dudu'$ where μ and ν are defined in equation (7.2), page 218. The rescaling of spot/volatility correlations in (10.7) guarantees that, just as in the constant volatility case, the values of all instantaneous covariances do not depend on the distribution of δZ . Only higher-order contributions to the P&L will depend on the distribution of δZ .

to the one-day smile. Beside very short-dated vanilla options, are there other, longer-dated, options that are sensitive to the one-day smile?

Path-dependent options that involve daily returns are the natural candidates. One such payoff is the variance swap – see Chapter 5 for an introduction.

10.2.4 Variance swaps

In Section 5.3.4, page 162, we have compared the P&L of a delta-hedged log contract with the payoff of the VS and have shown that the difference is generated by terms δS^k with $k > 2$.

In the limit of short returns, in diffusive models, the implied volatilities $\hat{\sigma}_{VS,T}$ of the VS and $\hat{\sigma}_T$ of the log contract coincide. For daily returns, in a diffusive model, the difference $\hat{\sigma}_{VS,T} - \hat{\sigma}_T$ should still be negligible, while it is expected to be sizeable in fat-tailed models. Formula (5.38), page 162, for example, expresses the difference between $\hat{\sigma}_{VS,T}$ and $\hat{\sigma}_T$ generated by terms of order 3 in δS as a function of the (unconditional) skewness of daily returns.

Table 10.2 lists the values of $\hat{\sigma}_{VS,T} - \hat{\sigma}_T$ in our fat-tailed two-factor model, for different parameter configurations, for a 1-year VS.

μ	∞	6	4	3
$\nu = 0$	0%	0%	0.02%	0.16%
$\nu = 257\%$	0.02%	0.04%	0.10%	0.29%

p_+	30%	40%	50%	60%	70%
$\mu = 4, \nu = 257\%$	-0.11%	0%	0.10%	0.23%	0.40%

Table 10.2: Top: $(\hat{\sigma}_{VS,T} - \hat{\sigma}_T)$ for $T = 1$ year as a function of μ for $p_+ = p_- = 1/2$ with $\nu = 0$ (no stochastic volatility) and $\nu = 257\%$ – corresponds to smiles in Figure 10.5. Bottom: $(\hat{\sigma}_{VS,T} - \hat{\sigma}_T)$ as a function of p_+ for $\mu = 4, \nu = 257\%$ – these parameters correspond to smiles in Figure 10.7.

We can see in the top section of Table 10.2 that for $\mu = \infty$ – which corresponds to the standard version of the two-factor model – even for a large volatility of volatility, VS and log-contract volatilities are essentially identical. They become appreciably different as we widen the tails of returns (we decrease μ), and as we vary p^+ , thus increasing the ATMF skew (bottom section of table).¹²

The *relative* difference between log-contract and VS volatilities produced by our model for values of μ and p^+ that are consistent with historical distributions of daily returns ($\mu \simeq 4, p^+ = 50\%$) is $0.1\%/20\% = 0.5\%$.

¹²Because of expression (10.3) for the dynamics of S_t in the fat-tailed version of the two-factor model, the implied VS volatilities we input in the model are those of payoff $\Sigma_i (S_{i+1}/S_i - 1)^2$ rather than of payoff $\Sigma_i \ln^2 (S_{i+1}/S_i)$. This is not a problem; forward variances for this alternative definition of the VS are still driftless.

This is roughly the order of magnitude of the realized value of $\widehat{\sigma}_{VS,T} - \widehat{\sigma}_T$ in the backtest of Chapter 5 – see Figure 5.1, page 166.

10.2.5 Daily cliques

Daily cliques are cliques written on daily returns. An example is a one-year option on the Euro Stoxx 50 index that pays a daily coupon equal to the put payoff $(k - \frac{S_{i+1}}{S_i})^+$ where S_i, S_{i+1} are two consecutive daily closing values of the index.

Typical strikes for these daily puts are in the range of 75% to 90%. Daily cliques are also typically of the knock-out type – the option expires once one coupon has paid off and the premium is paid on a quarterly basis – making them very similar to CDS contracts.

Some popular variants involve put spreads – daily cliques whose coupons are capped at some specified level.

Table 10.3 shows prices for a 1-year clique of daily puts struck at 80% (no knock-out feature) for various values of μ_- with no stochastic volatility ($\nu = 0$). We have taken $\mu_+ = 4$ and $p_+ = p_- = \frac{1}{2}$ as well as $\sigma = 20\%$. The resulting one-day smiles appear in Figure 10.10.

μ_-	∞	6	4	3	2.5	2.2
	0.00%	0.00%	0.02%	0.15%	0.43%	0.62%

Table 10.3: Prices for a 1-year daily clique struck at 80% as a function of μ_- with $\mu_+ = 4$, $p_+ = p_- = \frac{1}{2}$, $\sigma = 20\%$ and no volatility of volatility.

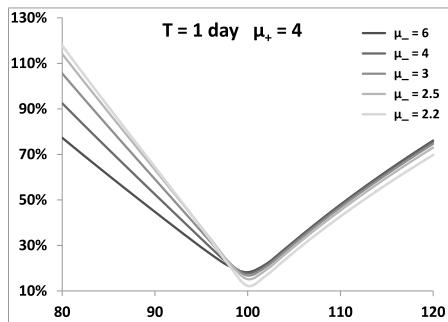


Figure 10.10: 1-day smile for different values of μ_- for $\mu_+ = 4$, $p_+ = p_- = 1/2$, $\sigma = 20\%$.

Obviously, daily cliques are worthless unless very large daily returns can be generated in the model. We have included in Table 10.3 the value $\mu_- = 2.2$ as values of μ_- this low are typically needed to match market prices.

This implied value of μ_- is thus much lower than its historical counterpart, which lies in the range [3, 4]. However, there is good reason for pricing daily cliques very conservatively. Daily cliques can be vega-hedged but cannot be replicated by delta and gamma-hedging as each coupon's maturity is one day. This leaves the seller saddled with the risk of a potentially very large negative unhedgeable daily P&L. Daily cliques are in fact insurance-type products that enable trading desks to exchange and mitigate stress-test risk.

Prices in Table 10.3 are calculated with $\nu = 0$. Once stochastic volatility is switched on prices of daily cliques increase, but are still mostly dependent on μ_- . For example, setting $\nu = 257\%$ increases the rightmost price in Table 10.3 by about 5 bps.¹³

10.3 Conclusion

- While historical (unconditional) distribution densities of equity returns exhibit fat tails, much thicker than can be achieved with typical diffusive models, the skew of vanilla smiles is overwhelmingly generated by the covariance of the spot with the implied volatility of the residual maturity, rather than the one-day smile.
- Unlike vanilla options, path-dependent options that involve daily returns do exhibit some sensitivity to the one-day smile, modest in the case of VSs, strong for daily cliques.
- These results have been obtained using a fat-tailed version of the two-factor stochastic volatility model. This is the two-factor forward variance model of Chapter 7, minimally adjusted so that:
 - daily returns are drawn with a two-sided Student distribution, rather than a Gaussian distribution. μ_+ , μ_- along with p_+ afford good control of the one-day conditional smile.
 - spot/volatility covariances remain unchanged. This is achieved by uniformly rescaling the native spot/volatility correlations.

¹³Stochastic volatility has the effect that the scales of successive returns are correlated, thus making non-knock-out daily cliques worth more than their knock-out counterparts. In the absence of stochastic volatility, because the probability of a knock-out event is so low, prices of both versions are practically identical.

Appendix A – jump-diffusion/Lévy models

This section is devoted to jump-diffusion/Lévy models – when are they called for and what do they do that stochastic volatility doesn't?

A.1 A stress-test reserve/remuneration policy

In the P&L of a delta-hedged option position, the order-one contribution in δS is cancelled by the delta, so the P&L starts with a δS^2 term. The first step in model building focuses on terms of order δS^2 – the gamma P&L. The essence of the Black-Scholes pricing equation lies in the provision of a deterministic theta term to offset the gamma P&L, given a break-even volatility σ :

$$P\&L = -\frac{1}{2}S^2 \frac{d^2P}{dS^2} \left(\frac{\delta S^2}{S^2} - \sigma^2 \delta t \right) \quad (10.11)$$

Mathematically, the solution P of the (parabolic) pricing equation has a probabilistic interpretation as the expectation of the payoff under a dynamics for S_t which is a diffusion. When other assets beside S_t – for example variances – are modeled as diffusive processes, other contributions to the P&L arise, of the same form as (10.11).¹⁴

As discussed in Section 1.2 of Chapter 1, the risk associated with gamma P&L is still sizeable and needs to be hedged away. This is done by dynamically trading vanilla options. With regard to δS , the residual risk is thus contributed by terms of order 3 and above.

Section 10.2.1 above has been devoted to the construction of a model that allows us to price these terms by separating the effect of the scale of δS from that of its distribution. We first choose a time scale – in our case one day – and write the return as:

$$\frac{\delta S}{S} = \sigma \delta Z$$

The instantaneous volatility σ – the scale of the return – is modeled with a two-factor forward variance model, and the distribution of δZ – or equivalently the conditional distribution of δS – is taken care of by the one-day smile parameters of the model.

Imagine we do not require this much sophistication and are looking for a way of pricing the effect of large returns using very simple assumptions. For example, we wish to assess the overall impact of typical adverse scenarios, or stress tests, on our hedged option position – that is the exotic option together with its vanilla hedge – and adjust the exotic option's price accordingly or set aside a reserve.

¹⁴The P&L of a delta-hedged, vega-hedged, position reads as in (7.3a), page 218, in the case of a forward variance model, or as in (2.105), page 69, in the case of the local volatility model.

This reserve is either (a) intended to offset on average the impact of the adverse scenarios and is released gradually to the trading desk or (b) charged by the bank to the trading desk on an ongoing basis to pay for the cost of allocating capital against adverse scenarios.

We already considered such a simple adjustment for variance swaps in Chapter 5 – see equation (5.42), page 166, for a VS together with its vanilla hedge.

Consider then a shock of fixed relative magnitude J on our underlying:

$$S \rightarrow S(1+J)$$

The P&L it generates on a short delta-hedged option position is:

$$P\&L = -(P(t, S(1+J)) - P(t, S)) + JS \frac{dP}{dS} \quad (10.12)$$

where $P(t, S)$ is our pricing function, and the second piece in (10.12) the contribution of our delta hedge. Expanding in powers of J :

$$P\&L = -\frac{1}{2}S^2 \frac{d^2P}{dS^2} J^2 - \frac{1}{6}S^3 \frac{d^3P}{dS^3} J^3 + \dots$$

Note that this P&L includes a δS^2 portion, which acts as an extra contribution to the volatility of S . Let us now assign an annualized frequency λ to our stress test. There occur on average $\lambda \delta t$ shocks during the time interval $[t, t + \delta t]$, whose P&L impact will be offset, on average, by a reserve given by:

$$\lambda \delta t \left[(P(t, S(1+J)) - P(t, S)) - JS \frac{dP}{dS} \right] \quad (10.13)$$

(10.13) can be interpreted in three ways:

- either as a theta contribution to offset on average the P&L generated by the stress-test scenario, which is assumed to occur with intensity λ .
- or as a tax levied on the trading desk by the bank to pay for the cost of capital allocated to stress-test risk. Assuming this capital is proportional to the stress-test P&L, with a proportionality coefficient β , and that the rate of return required by the bank on its capital is μ , the amount charged to the desk is then of the form (10.13) with $\lambda = \beta \mu$; λ is no longer interpreted as an intensity. Depending on the sign of its contribution to the overall stress-test P&L, a trading desk would thus be either taxed or rewarded.
- or as a minimal return we require on our consumption of stress-test limit. Trading desks are usually not charged directly for stress-test P&L, but are assigned stress-test P&L limits. This limit can be managed at the desk level by requiring that, for a given consumption of stress-test budget, commensurate revenue be generated. This amounts to demanding that our carry P&L comprises, on top of the usual gamma/theta P&L, a piece given by (10.13) with λ the (annualized) rate of remuneration of stress-test-budget usage.

Consider an option of maturity T . To calculate a reserve policy we use the initial estimate at time t , spot value S , of the P&L impact of a shock to compute an adjustment ΔP to the option price. ΔP is given by:

$$\Delta P(t, S) = \lambda(T - t) \left[(P(t, S(1 + J)) - P(t, S)) - JS \frac{dP}{dS} \right] \quad (10.14)$$

The price we quote at time t for our derivative is:

$$P + \Delta P$$

ΔP is proportional to $T - t$: at time passes the reserve is released and converges to 0 at $t = T$.

$\Delta P(t, S)$ is a reserve policy evaluated at t with the initial spot value S . Consider a move δS . While we started out with an amount $\Delta P(t, S)$, the new reserve we should now be holding is $\Delta P(t, S + \delta S)$. To generate this extra cash, we need to delta-hedge ΔP .

ΔP however includes no provision for the financing cost of its own delta hedge, and no theta to offset its own gamma.

These limitations are typical when one uses an ad-hoc reserve policy rather than a full-blown model. We now derive a pricing equation that takes care of these issues.

A.2 Pricing equation

We would like our P&L during two delta rehedges to read as in (10.11), but with an additional theta contribution given by (10.13):

$$P\&L = -\frac{1}{2}S^2 \frac{d^2P}{dS^2} \left(\frac{\delta S^2}{S^2} - \sigma^2 \delta t \right) + \lambda \left[(P(t, S(1 + J)) - P(t, S)) - JS \frac{dP}{dS} \right] \delta t \quad (10.15)$$

Proceeding as in Section 1.1, page 2, we can write down the corresponding pricing equation at once:

$$\frac{dP}{dt} + (r - q)S \frac{dP}{dS} + \frac{\sigma^2}{2} S^2 \frac{d^2P}{dS^2} + \lambda \left[(P(t, S(1 + J)) - P(t, S)) - JS \frac{dP}{dS} \right] = rP \quad (10.16)$$

With respect to the Black-Scholes equation, the last piece in the left-hand side provides for an additional theta to offset, on average, the P&L impact of our stress-test scenario.

Mathematically, the solution of (10.16) can be expressed as the expectation of the option's payoff under a dynamics for S_t that consists of a diffusion together with Poisson jumps:

$$dS_t = (r - q)S_t dt + \sigma S_t dW_t + JS_t (dN_t - \lambda dt) \quad (10.17)$$

where N_t is a counting process. Unlike the pricing equations we have encountered so far, (10.16) involves non-local terms: $\frac{dP}{dt}$ is a function not only of derivatives of P

with respect to S , which characterize P in the vicinity of S , but also on the value of P for $S(1+J)$.

The amplitude J of the stress-test scenario can be made a random variable. Let us call $\rho(J)$ its distribution. Our pricing equation becomes:

$$\begin{aligned} \frac{dP}{dt} + (r-q)S\frac{dP}{dS} + \frac{\sigma^2}{2}S^2\frac{d^2P}{dS^2} \\ + \lambda \int_{-1}^{\infty} \rho(J) \left[(P(t, S(1+J)) - P(t, S)) - JS\frac{dP}{dS} \right] dJ = rP \end{aligned} \quad (10.18)$$

Solving it

For a constant volatility σ and vanilla option payoffs (10.18) can be solved easily since it is homogeneous in $\ln S$. The procedure is similar to that in Appendix A of Chapter 8, page 347.

Define $\tau = T - t$ and:

$$x = \ln \frac{S}{K} + (r-q)\tau$$

where K is the strike of the vanilla option, set $P(t, S) = Se^{-q\tau}f(\tau, x)$ and introduce the Laplace transform $F(\tau, p)$ of f :

$$F(\tau, p) = \int_{-\infty}^{+\infty} e^{-px} f(\tau, x) dx$$

Replacing P by $Se^{-q\tau}f(\tau, x)$ in (10.18) yields:

$$\begin{aligned} -\frac{df}{d\tau} + \frac{\sigma^2}{2} \left(\frac{df}{dx} + \frac{d^2f}{dx^2} \right) \\ + \lambda \int_{-\infty}^{+\infty} \rho^*(u) \left(e^u f(x+u, \tau) - f - (e^u - 1) \left(f + \frac{df}{dx} \right) \right) du = 0 \end{aligned}$$

where $\rho^*(u)$ is the density of

$$u = \ln(1+J)$$

Taking now the Laplace transform of both sides leads to the following ODE for $F(\tau, p)$:

$$-\frac{dF}{d\tau} + \left[\frac{\sigma^2}{2}p(1+p) + \lambda(\psi(p) - (1+p)\psi(0)) \right] F = 0 \quad (10.19)$$

where $\phi(p)$ is defined as:

$$\psi(p) = \int_{-\infty}^{+\infty} \rho^*(u) (e^{(1+p)u} - (1+p)u - 1) du$$

The initial condition for F for $\tau = 0$, that is $t = T$, is provided by the option's payoff. For a call option $f(0, x) = (1 - e^{-x})^+$ thus:

$$F(0, p) = \int_{-\infty}^{+\infty} e^{-px} (1 - e^{-x})^+ dx = \frac{1}{p(1+p)}$$

$F(0, p)$ is defined for $\text{Re}(p) > 0$. For a put option $F(0, p)$ is identical except the condition is $\text{Re}(p) < -1$.

Integrating (10.19) we then get:

$$\begin{aligned} F(\tau, p) &= \frac{1}{p(1+p)} e^{\tau H(p)} \\ H(p) &= \frac{\sigma^2}{2} p(1+p) + \lambda (\psi(p) - (1+p)\psi(0)) \end{aligned} \quad (10.20)$$

Inverting $F(\tau, p)$ yields option prices.

Using as pricing function the solution P of (10.18) rather than $P^0 + \Delta P$ ensures that:

- the additional theta the model pays us in between two delta rehedges is exactly (10.13)
- $\frac{dP}{dS}$ incorporates the delta hedge of the reserve policy
- the P&L generated by jumps on the reserve policy itself is accounted for

One can back Black-Scholes implied volatilities out of $P(t, S)$ – what does the resulting smile look like? We now briefly analyze its ATMF skew.

A.3 ATMF skew

In order to use the perturbative result of Appendix B of Chapter 5 we need the density ρ_z of z_T , defined by:

$$z_T = \ln \frac{S_T}{S} - (r - q)(T - t)$$

The density of S_T is given by: $\rho(S_T) = e^{r(T-t)} \left. \frac{d^2 P}{dK^2} \right|_{K=S_T}$. Using that $P(t, S) = S e^{-q(T-t)} f(t, x)$ and the definition of z we get:

$$\rho_z(z) = (e^x (\partial_x + \partial_x^2) f)_{x=-z}$$

The cumulant-generating function of ρ_Z , $L(\tau, q)$, which is the logarithm of the characteristic function of ρ_z , is given by:

$$\begin{aligned} e^{L(\tau, q)} &= \int_{-\infty}^{+\infty} e^{-qz} \rho_z(z) dz \\ &= \int_{-\infty}^{+\infty} e^{-qz} (e^x (\partial_x + \partial_x^2) f(\tau, x))_{x=-z} dz \\ &= q(1+q) F(\tau, -(1+q)) \end{aligned} \quad (10.21)$$

In what follows we will be sitting at $t = 0$; we thus set $\tau \equiv T$. From expression (10.20) we thus get $L(T, q) = TH(p = -(1+q))$.

$$\begin{aligned} L(T, q) &= T \left(\frac{\sigma^2}{2} q(1+q) + \lambda(\psi(-(1+q)) + q\psi(0)) \right) \\ &= T(\phi(q) + q\phi(-1)) \end{aligned} \quad (10.22)$$

where $\phi(q)$ is given by:

$$\begin{aligned} \phi(q) &= \frac{\sigma^2}{2} q^2 + \psi(-(1+q)) \\ &= \frac{\sigma^2}{2} q^2 + \int_{-\infty}^{+\infty} \lambda \rho^*(u) (e^{-qu} + qu - 1) du \end{aligned} \quad (10.23)$$

Using J rather than u , $L(T, q)$ equivalently reads:

$$L(T, q) = T \left(\frac{\sigma^2}{2} q(1+q) + \int_{-1}^{\infty} \lambda \rho(J) ((1+J)^{-q} - 1 + qJ) dJ \right) \quad (10.24)$$

Note that $L(T, q)$ scales linearly with T : this is true of all processes for $\ln S$ with independent stationary increments, which is obviously the case for process (10.17).

As already mentioned in Appendix C of Chapter 8, and using equation (5.82), page 191, $L(T, q)$ has the following properties:

- the condition that the density of z_T integrate to 1 is $L(T, 0) = 0$
- the condition that the forward of S_T be $Se^{(r-q)T}$ is $L(T, -1) = 0$
- the log-contract implied volatility is given by: $\hat{\sigma}_T^2 = \frac{2}{T} \left. \frac{dL}{dq} \right|_{q=0}$

By construction, as can be checked using expression (10.22) for L , the two conditions are satisfied. When using an approximate form for L we need to make sure they still hold. Using (10.22) we get:

$$\hat{\sigma}_T^2 = \sigma^2 + 2\lambda \overline{e^u - u - 1} = \sigma^2 + 2\lambda \overline{J - \ln(1+J)}$$

where \overline{X} stands for the mean of random variable X .

The VS implied volatility, $\widehat{\sigma}_{VS,T}$, on the other hand, is simply given by the quadratic variation of $\ln S$:¹⁵

$$\widehat{\sigma}_{VS,T}^2 = \sigma^2 + \lambda \overline{\ln(1+J)^2}$$

For $J = 0$, the jump-diffusion model reduces to the Black-Scholes model and $L(q)$ is a polynomial of order 2. To analyze the ATMF skew of jump-diffusion models, let us assume that J is small and let us expand L in powers of J , stopping at the lowest non-trivial order. We get from (10.24):

$$L(T, q) = T \left(\frac{\sigma^2 + \lambda \overline{J^2}}{2} q(1+q) - \frac{\lambda \overline{J^3}}{6} q(1+q)(2+q) \right) \quad (10.25)$$

The contribution of order 2 in J merely shifts the volatility level: $\sigma^2 \rightarrow \sigma^2 + \lambda \overline{J^2}$, but the model remains of the Black-Scholes type – we need to go to order 3 in J . $L(T, q)$ in (10.25) complies with the two conditions above. The log-contract implied volatility, at this order, is:

$$\widehat{\sigma}_T^2 = \frac{2}{T} \left. \frac{dL}{dq} \right|_{q=0} = \sigma^2 + \lambda \overline{J^2}$$

For small values of J the ATMF skew S_T , at order one in s , is given by (5.93), page 194:

$$S_T = \frac{s}{6\sqrt{T}}$$

s is the skewness of $\ln S_T$, that is $s = \frac{\kappa_3}{(\widehat{\sigma}_T^2 T)^{3/2}}$ where κ_3 is the cumulant of order 3. Cumulants are defined through the expansion of $L(q)$ in powers of q :

$$L(T, q) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \kappa_n(T) q^n$$

From (10.25), $\kappa_3 = \lambda \overline{J^3} T$, thus $s = \frac{\lambda \overline{J^3} T}{(\sigma^2 + \lambda \overline{J^2})^{3/2} T^{3/2}}$. We thus get the following approximate formula for S_T :

$$S_T = \frac{\lambda \overline{J^3}}{6 \left(\sigma^2 + \lambda \overline{J^2} \right)^{3/2} T} = \frac{\lambda \overline{J^3}}{6 \widehat{\sigma}_T^3 T} \quad (10.26)$$

Thus, for small jump sizes, the ATMF skew of jump-diffusion models decays like $\frac{1}{T}$.

This scaling is illustrated in Figure 10.9, page 402. While data in Figure 10.9 are obtained with the fat-tailed two-factor model of Section 10.2.1, which moreover is a discrete model, because stochastic volatility is turned off, it generates independent stationary (discrete) increments for $\ln S$, hence the $\frac{1}{T}$ scaling of the ATMF skew.¹⁶

¹⁵We recover the expressions for $\widehat{\sigma}_T$ and $\widehat{\sigma}_{VS,T}$ in Section 5.3.2.

¹⁶For $T \rightarrow 0$, approximation (10.26) for S_T diverges, which is not the case of the actual ATMF skew. Why does the approximation break down in the short- T limit? Expanding $L(T, q)$ at order one in κ_3

A.4 Jump scenarios in calibrated models

By using equation (10.16) to risk-manage the exotic option at hand, we obtain a price and a delta-hedging strategy so that the P&L of our short option position together with its delta hedge in between two delta rehedges reads as in (10.15). In addition to the usual second-order gamma/theta P&L, we have an extra theta contribution, which corresponds to a regular release of the stress-test reserve policy where the stress test is specified by the frequency and distribution of the jumps in our model.

This would be fine if we were just delta-hedging our option: we quote as initial price $P(t = 0, S_0)$, the solution of (10.16), run our delta hedge and we are done.

In reality we use other derivatives – typically vanilla options – as hedge instruments, thus the price charged to the client needs to incorporate a stress-test reserve policy that offsets on average the stress-test P&L

$$P\&L = -(P(t, S(1 + J)) - P(t, S)) + JS \frac{dP}{dS}$$

of the *global* position, rather than that of the *naked* exotic option: P is the value of the *hedged* position. Consequently the hedging vanilla options need to be risk-managed using the same model.

Assume we are using a market model, for example the local volatility model, or one of the admissible local-stochastic volatility models of Chapter 12. We start with the pricing equation of one such model and insert in it the contribution from jumps – the last term in the left-hand sides of equation (10.16) or (10.18).

Is the resulting carry P&L still of the standard gamma/theta form typical of diffusive market models, with an additional theta corresponding to our stress-test/jump scenario? Let us assume here that we are using the local volatility model.

The pricing equation for $P^{LV}(t, S, \sigma)$ is (10.16) with σ replaced with the local volatility $\sigma(t, S)$, chosen so that market prices of vanilla options are recovered.¹⁷

With respect to the original pricing equation (2.102), page 68, of the local volatility model, we have an extra contribution to $\frac{dP^{LV}}{dt}$ coming from jumps.

amounts to perturbing ρ_z around a Gaussian density with Hermite polynomials – see expression (5.83), page 191. The coefficient of H_n is $\frac{\delta\kappa_n}{\Sigma^n \sqrt{n!}}$, which in our case is proportional to $T/T^{\frac{n}{2}} = T^{1-\frac{n}{2}}$, i.e. diverges for $n = 3$: this is no longer a small perturbation.

¹⁷Space prevents us from discussing here the calibration of such a model. In the presence of jumps, we can carry out the same derivation as in Section 2.2.1, page 27, to generate the following forward equation for vanilla option prices, which replaces (2.7):

$$\begin{aligned} \frac{dC}{dT} + (r - q - \lambda J) K \frac{dC}{dK} - \frac{\sigma^2(T, K)}{2} K^2 \frac{d^2C}{dK^2} - \lambda \left[(1 + J) C\left(\frac{K}{1 + J}, T\right) - C(K, T) \right] \\ = -(q + \lambda J) C \end{aligned}$$

From this we get the expression for $\sigma^2(T, K)$:

$$\sigma^2(T, K) = \sigma_{loc}^2(K, T) - 2\lambda \frac{[(1 + J) C_{K/(1+J), T} - C_{K, T}] - JC + \lambda JK \frac{dC}{dK}}{K^2 \frac{d^2C}{dK^2}} \quad (10.27)$$

In Section 2.7 of Chapter 2 we have analyzed the carry P&L of a delta-hedged/vega-hedged position and have shown that it has the usual gamma/theta form.

In a local volatility model with jumps, implied volatilities are still a function of t, S and $\sigma(t, S)$: $\hat{\sigma}_{KT} = \Sigma_{KT}^{\text{LV}}(t, S, \sigma)$, except Σ_{KT}^{LV} is a different function than in Section 2.7 as jumps are taken into account.

Going back to page 67, the reader can check that the derivation in the jump case is identical, except we have the following additional term in the right-hand side of equation (2.105), page 69:

$$+ \lambda \left[(P^{\text{LV}}(t, S(1 + J), \sigma) - P^{\text{LV}}(t, S, \sigma)) - JS \frac{P^{\text{LV}}}{dS} \right] \delta t$$

Expressing everything in terms of $P(t, S, \hat{\sigma}_{KT}) = P^{\text{LV}}(t, S, \sigma[t, S, \hat{\sigma}_{KT}])$, using (2.99), this P&L can be rewritten as:

$$\begin{aligned} &+ \lambda \left[\left(P(t, S(1 + J), \hat{\sigma}_{KT} + \Delta \Sigma_{KT}^{\text{LV}}(t, S)) - P(t, S, \hat{\sigma}_{KT}) \right) \right. \\ &\quad \left. - JS \left(\frac{dP}{dS} + \frac{dP}{d\hat{\sigma}_{KT}} \bullet \frac{d\Sigma_{KT}^{\text{LV}}}{dS} \right) \right] \delta t \end{aligned} \quad (10.28)$$

where $\Sigma_{KT}^{\text{LV}}(t, S)$ is the (additive) jump of implied volatility $\hat{\sigma}_{KT}$ generated at time t , spot S , by a relative jump of S of magnitude J , keeping the local volatility function – equal to $\sigma[t, s, \hat{\sigma}_{KT}]$ – unchanged:

$$\Delta \hat{\sigma}_{KT}(t, S) = \Sigma_{KT}^{\text{LV}}(t, S(1 + J), \sigma[t, s, \hat{\sigma}_{KT}]) - \hat{\sigma}_{KT}$$

The additional theta (10.28) is thus proportional to the P&L generated by *joint jumps* in spot and implied volatilities, minus the contribution from the delta-hedge.

The conclusion of this section is that, depending on the model we start with, adding jumps on S in the pricing equation results in general in a stress-test scenario that involves jumps not only on S but also on implied volatilities. In the local volatility model, the jump in implied volatilities is dictated by the jump scenario for S and the smile used for calibration.¹⁸

where $\sigma_{\text{loc}}(K, T)$ is the local volatility in the case with no jumps, given by the Dupire formula (2.3), page 26. We can check that, expanding the right-hand side of (10.27) at order two in J , we get:

$$\sigma^2(T, K) = \sigma_{\text{loc}}^2(K, T) - \lambda J^2$$

For numerical aspects related to (a) the calibration of the local volatility model with jumps, and (b) pricing in jump/diffusion models, we refer the reader to [2] and [82].

¹⁸How do we gain leverage on the jump scenario for implied volatilities? This is a hard question. We would need to define a set of local volatilities $\sigma_n(t, S)$ where n is the number of jumps before t . Denote by $P_n(t, S)$ the option price. P_n solves equation (10.16) or (10.18) with $\sigma(t, S)$ replaced by $\sigma_n(t, S)$.

A.5 Lévy processes

Imagine we add up different independent Poisson processes, each with its own intensity λ_i and jump density ρ_i^* . In expression (10.23) for $\phi(q)$, $\lambda\rho^*(u)$ is replaced with:

$$\lambda\rho^*(u) \rightarrow \sum_i \lambda_i \rho_i^*(u)$$

There are two situations:

- uninteresting: if $\sum_i \lambda_i$ is finite, this boils down to an effective Poisson process with intensity $\lambda = \sum_i \lambda_i$ and jump density $\rho^* = \frac{\sum_i \lambda_i \rho_i^*}{\sum_i \lambda_i}$.
- interesting: $\sum_i \lambda_i$ is infinite – jumps occur infinitely frequently. The expression of $\phi(q)$ is:

$$\phi(q) = \frac{\sigma^2}{2} q^2 + \int_{-\infty}^{+\infty} (e^{-qu} + qu - 1) k(u) du \quad (10.29)$$

The process for z_t is a non-trivial Lévy process and (10.29) is a particular form of the Lévy-Khintchine representation, with $k(u) = \sum_i \lambda_i \rho_i^*(u)$.

Mathematically, a Lévy process is a process with stationary independent increments. As such, the class of Lévy processes trivially contains Brownian motion and Poisson processes, but includes many other processes. We refer the interested reader to the many textbooks on this topic – see [33] for applications to derivatives pricing.

A.6 Conclusion

Jump-diffusion – and more generally Lévy – processes are useful tools for embedding in derivatives' prices a stress-test reserve policy corresponding to the scenarios that these processes express. As tools for modeling the actual dynamics of securities or for pricing derivatives, they are, in the author's view, not adequate, for mostly two reasons:

- the assumption of independent increments is violated in historical returns, and in models as well, in the latter because instantaneous volatility – which sets the scale of returns – is stochastic and needs to be so. As a representation of actual returns, jump/Lévy processes are a fairly unrealistic construct.

and the contribution from jumps is replaced with:

$$\lambda \left[(P_{n+1}(t, S(1+J)) - P_n(t, S)) - JS \frac{dP_n}{dS} \right]$$

We have a set of nested equations for the P_n – we obviously need to make additional assumptions in order to be able to calibrate $\sigma_n(t, S)$ to the market.

The case with two local-volatility functions, for the pre-default and post-default states, is relevant for stocks.

- jump processes are much more difficult to correlate than diffusions, except for those Lévy processes that have a representation in terms of time-changed Brownian motion.

Jump-diffusion models should then not be considered a bona fide characterization of the actual or implied dynamics of real securities but viewed as a means of calculating – and releasing – a reserve policy based on ad-hoc stress-test scenarios.

Using a diffusive process for pricing does not mean we actually assume that securities behave as diffusions. A diffusive pricing equation is merely a technical device for embedding in a derivative's price the time value needed to offset second order gamma P&Ls with given break-even levels.

Practically, properly offsetting a trading book's second-order sensitivities is already quite a challenge. Hedging P&Ls generated by higher-order moments is in practice hardly possible. The best we can hope for is either:

- (a) price them in a model that realistically models the tail behavior of actual returns
- or (b) estimate them using ad-hoc stress-test scenarios.

Case (b) is taken care of by jump/Lévy models.

As for case (a), we have explained in Section 10.2.1 above how to economically adjust an existing stochastic volatility model, with the benefit that the richness of the dynamics of volatilities is preserved – while achieving a realistic modeling of tail behavior.

Chapter's digest

10.1 The distribution of equity returns

► Unconditional distributions of daily returns of equity indexes exhibit fat tails that are well-approximated with a Student distribution, with similar parameters for left and right tail distributions in the range [3, 4].

► Conditional distributions, whereby returns are rescaled by an estimate of realized volatility, still exhibit fat-tailed distributions, with somewhat thinner tails – still much thicker than in the lognormal distribution.



10.2 Impact of the distribution of daily returns on derivative prices

► To study the effect of fat tails of daily returns – or equivalently the effect of the one-day smile – we develop a fat-tailed version of the two-factor model. The conditional distribution of daily returns is a two-sided Student distribution. The one-day smile is parametrized by (a) the parameters of the left and right tails, (b) the probability that returns are positive, which sets the one-day ATM skew.

► Forward variances are simulated in the two-factor model, as usual, and so are increments δW^S for the Brownian motion that drives S_t . S_t is simulated on a daily schedule according to:

$$S_{t+\Delta} = S_t [1 + (r - q) \Delta + \sigma_t \delta Z]$$

where $\delta Z = \sqrt{\Delta} f\left(\frac{\delta W^S}{\sqrt{\Delta}}\right)$. The mapping function is given by:

$$\begin{cases} x \leq \mathcal{N}_G^{-1}(p_-) & f(x) = \zeta_- \sqrt{\frac{\mu_+ - 2}{\mu_-}} \mathcal{N}_{\mu_-}^{-1}\left(\frac{\mathcal{N}_G(x)}{2p_-}\right) \\ x \geq \mathcal{N}_G^{-1}(p_-) & f(x) = \zeta_+ \sqrt{\frac{\mu_+ - 2}{\mu_+}} \mathcal{N}_{\mu_+}^{-1}\left(\frac{1}{2} + \frac{\mathcal{N}_G(x) - p_-}{2p_+}\right) \end{cases}$$

► Spot/volatility correlations need to be rescaled so that covariances of spot and implied volatilities remain identical in both standard and fat-tailed versions of the two-factor model. This rescaling is uniform with the ratio of correlations in the fat-tailed model to those in the standard model given by:

$$\frac{\rho_{iS}^*}{\rho_{iS}} = \frac{1}{\int_{-\infty}^{+\infty} \phi(x) x f(x) dx}$$

where $\phi(x)$ is the probability density of the standard normal variable.

► Vanilla smiles exhibit little sensitivity to the fat tails of daily returns, except for short maturities and out-of-the-money volatilities – mostly for high strikes. ATM skew are hardly affected.

- The one-day ATM skew does impact vanilla smiles, but the size of its contribution decays as $\frac{1}{T}$.
- The conclusion is that, for standard maturities, the fat-tailed nature of the conditional distribution of daily returns hardly impacts the ATMF skew, which is predominantly the product of the covariance of spot and implied volatilities.
- Derivatives whose payoffs involve daily returns are sensitive to the effect of the one-day smile – typical examples include variance swaps and daily cliques.



Appendix A – jump-diffusion/Lévy models

► Pricing equations based on jump-diffusion processes naturally arise when one needs to factor in the price of a derivative (a) the average impact of specific stress-test scenarios on the portfolio of the derivative together with its delta (and possibly its vega) hedge, (b) the cost of capital set aside to cover for the P&L generated by stress-test scenarios, (c) a minimal rate of return on the consumption of stress-test limits.

► Jumps increase the volatility in the model but also generate a smile. At the lowest non-trivial order in jump size – order 3 – the ATMF skew is given by:

$$\mathcal{S}_T = \frac{\lambda \overline{J^3}}{6\hat{\sigma}_T^3 T}$$

thus decays as $\frac{1}{T}$.

► Jump-diffusion and Lévy models are natural tools for embedding in the price of a derivative – and releasing it as time elapses – a reserve policy reflecting the average cost of specific stress-test scenarios.

► They are ill-suited to the modeling of the one-day smile as (a) their scale is fixed, (b) they cannot be easily correlated with other processes.

This page intentionally left blank