

# Chapter 6

---

## *An example of one-factor dynamics: the Heston model*

In the next chapter we cover models built on a specification of the dynamics of forward VS variances that can be calibrated to a term structure of VS volatilities, or a term-structure of vanilla implied volatilities of an arbitrary moneyness.

We now briefly pause to consider the Heston model, which is not a forward variance model, since in its native form it can be calibrated to the VS volatility of one single maturity.

It is instructive to assess the Heston model and its capabilities in the framework of forward variances. This exercise will help us introduce suitability criteria which we then apply to the forward variance models covered in Chapter 7. Moreover, the Heston model is an archetypal example among first-generation stochastic volatility models, that is models written in terms of the instantaneous variance  $V_t$ .

Unless stated otherwise, forward VS variances  $\xi_t^T$  will henceforth be called simply “forward variances” and we use  $\hat{\sigma}_T$  interchangeably for VS or log-contract implied volatilities.

---

### 6.1 The Heston model

The Heston model [60] is a first-generation model; it owes its popularity to the fact that, being an affine model, the Laplace transform of its moment-generating function for  $\ln S$  is analytically known. Numerical inversion of this transform then yields vanilla option prices – when there are no dividends in fixed cash amounts.

The analytics of the Heston model is abundantly covered in the literature – see for example [48], [69]. Rather, we concentrate on the joint spot/volatility dynamics that this model generates. We assume for simplicity zero interest rate and repo.

The Heston model is a diffusive one-factor model – instead of forward variances, the instantaneous variance is modeled, according to the following SDEs:

$$\begin{cases} dS_t = \sqrt{V_t} S_t dW_t \\ dV_t = -k(V_t - V^0)dt + \sigma \sqrt{V_t} dZ_t \end{cases} \quad (6.1)$$

$V_t$  is the instantaneous variance – with respect to our previous notation:

$$V_t = \xi_t^t = \bar{\sigma}_t^2$$

$\sigma$  is commonly called “volatility of volatility”, though it is not a lognormal volatility as it has dimension time<sup>-1</sup>. The Brownian motions  $W_t, Z_t$  are correlated, with correlation  $\rho$ .  $V^0$  is a constant.

---

## 6.2 Forward variances in the Heston model

Forward variances  $\xi_t^T$  are defined by:

$$\xi_t^T = E_t[V_T]$$

Taking the expectation of the second equation in (6.1) and using the compact notation  $\bar{V}_u = E_t[V_u]$  yields:

$$d\bar{V}_T = -k(\bar{V}_T - V^0)dT \quad (6.2)$$

whose solution is

$$\bar{V}_T = V^0 + e^{-k(T-t)}(V_t - V^0)$$

which, using our notation reads:

$$\xi_t^T = V^0 + e^{-k(T-t)}(\xi_t^t - V^0) \quad (6.3)$$

VS volatilities  $\hat{\sigma}_t^T$  are given by:

$$\hat{\sigma}_T^2(t) = \frac{1}{T-t} \int_t^T \xi_t^\tau d\tau = V^0 + \frac{1 - e^{-k(T-t)}}{k(T-t)}(V_t - V^0) \quad (6.4)$$

Differentiating (6.3) gives:

$$d\xi_t^T = \sigma e^{-k(T-t)} \sqrt{\xi_t^t} dZ_t$$

$\xi_t^T$  is driftless – as it should. In the framework of forward variances the Heston dynamics (6.1) reads:

$$\begin{cases} dS_t = \sqrt{\xi_t^t} S_t dW_t \\ d\xi_t^T = \sigma e^{-k(T-t)} \sqrt{\xi_t^t} dZ_t \end{cases} \quad (6.5)$$

The Heston model is thus a one-factor model for forward variances where the instantaneous volatility of all forward variances  $\xi_t^T$  is proportional to the instantaneous volatility  $\bar{\sigma}_t = \sqrt{\xi_t^t}$ . It is a Markov-functional model for forward variances, as  $\xi_t^T$  is a function of  $\xi_t^t$ , given by (6.3).

Still, the Heston model should not be considered as a particular version of the forward variance models of Chapter 7; a one-dimensional Markov representation exists *only* if the initial values  $\xi_{t=0}^T$  of forward variances satisfy condition (6.2):

$$\frac{d\xi_0^T}{dT} = -k(\xi_0^T - V^0) dT$$

It is not able to accommodate general term structure of VS volatilities.

Before analyzing further the dynamics of the Heston model, let us discuss the issue of the drift of  $V_t$  in first-generation models.

---

### 6.3 Drift of $V_t$ in first-generation stochastic volatility models

The traditional approach to these models typically found in papers and textbooks can be summarized as follows:

- Start with historical dynamics of the instantaneous variance:

$$dV_t = \mu(t, S, V, p) dt + \alpha dZ_t$$

where  $p$  are model parameters – such as  $k, V^0$  in the Heston model.

- In risk-neutral dynamics, drift of  $V_t$  is altered by “market price of risk”  $\lambda$ , which is an arbitrary function of  $t, S, V$ :

$$dV_t = (\mu(t, S, V, p) + \lambda(t, S, V)) dt + \alpha dZ_t$$

- A few lines down the road, jettison “market price of risk” and conveniently decide that risk-neutral drift has same functional form as historical drift, except parameters now have stars:

$$dV_t = \mu(t, S, V, p^*) dt + \alpha dZ_t$$

- Calibrate (starred) parameters on vanilla smile.

Discussions surrounding the “market price of risk” and the uneasiness generated by its *a priori* arbitrary form and hasty disposal are pointless – the “market price of risk” is a nonentity.

$V_t$  itself is an artificial object: for different  $t$ ,  $V_t$  represents a different forward variance. Its drift is then only a reflection of the term-structure of forward variances. Differentiating the identity  $V_t = \xi_t^t$  and using the following dynamics for  $\xi_t^T$ :

$$d\xi_t^T = \lambda_t^T dZ_t^T$$

yields

$$dV_t = \frac{d\xi_t^T}{dT} \Big|_{T=t} dt + \lambda_t^t dZ_t^t$$

The drift of the instantaneous variance is thus simply the slope at time  $t$  of the short end of the variance curve. One can check that taking the derivative at  $\xi_t^T$  with respect to  $T$  in (6.3) indeed yields the drift of  $V_t$  in the Heston model.

---

## 6.4 Term structure of volatilities of volatilities in the Heston model

Using (6.4) we get the following SDE for  $\hat{\sigma}_T$ :

$$d\hat{\sigma}_T = \bullet dt + \frac{\sigma}{2} \frac{1 - e^{-k(T-t)}}{k(T-t)} \frac{\hat{\sigma}_t}{\hat{\sigma}_T} dZ_t \quad (6.6)$$

where  $\hat{\sigma}_t = \sqrt{V_t}$ . Let us examine the term structure of the volatility of VS volatilities, that is the dependence of the volatility of  $\hat{\sigma}_T$  on  $T$ . We have the two following limiting regimes:

$$T - t \ll \frac{1}{k} \quad d\hat{\sigma}_T \simeq \bullet dt + \frac{\sigma}{2} \left( 1 - \frac{k(T-t)}{2} \right) \frac{\hat{\sigma}_t}{\hat{\sigma}_T} dZ_t \quad (6.7)$$

$$T - t \gg \frac{1}{k} \quad d\hat{\sigma}_T \simeq \bullet dt + \frac{\sigma}{2} \frac{1}{k(T-t)} \frac{\hat{\sigma}_t}{\hat{\sigma}_T} dZ_t \quad (6.8)$$

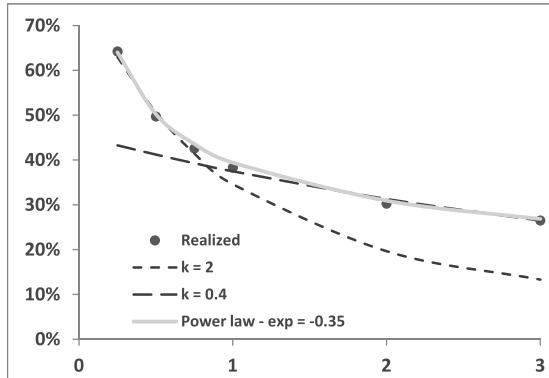
Thus, for long maturities, the instantaneous volatility of  $\hat{\sigma}_T$  decays like  $\frac{1}{T-t}$ . For a flat term structure of VS volatilities, (6.6) implies that the instantaneous lognormal volatility of  $\hat{\sigma}_T$  is:

$$\text{vol}(\hat{\sigma}_T) \propto \frac{1 - e^{-k(T-t)}}{k(T-t)} \quad (6.9)$$

Figure 6.1 shows the lognormal volatilities of VS implied volatilities of the Euro Stoxx 50 index, compared to expression (6.9), suitably rescaled, along with a power-law fit.

As Figure 6.1 shows, volatilities of VS volatilities are typically larger than volatilities of the underlying itself. This is not due to the fact that we are using VS volatilities – volatilities of ATM volatilities are similar.

As is apparent, while realized levels of volatilities of VS volatilities are consistent with a power law dependence on maturity, they cannot be captured by the dynamics of the Heston model over a wide range of maturities. The two values of  $k$  used in Figure 6.1, 2 and 0.4, have been chosen so as to best match the short and long end of the term structure of the volatilities of VS volatilities in the graph.



**Figure 6.1:** Volatility of VS volatilities of the Euro Stoxx 50 index as a function of maturity (years), evaluated on the period [2005, 2010] (dots), along with (a) volatilities of VS volatilities in the Heston model given by (6.9) for two different values of  $k$  (dotted lines), (b) a power-law fit  $\propto T^{-0.35}$ .

## 6.5 Smile of volatility of volatility

In diffusive models, for short maturities, the ATMF implied volatility is approximately equal to the instantaneous volatility, which is equal to  $\sqrt{V}$ . From (6.1) we get:

$$d\sqrt{V} = \bullet dt + \frac{\sigma}{2} dZ_t$$

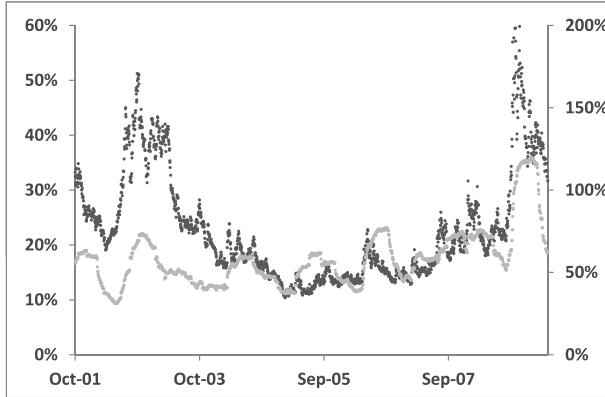
Thus, in the Heston model, short ATMF implied volatilities are approximately normal, rather than lognormal, with a normal volatility equal to  $\frac{\sigma}{2}$ . Figure 6.2 shows the 3-month ATMF implied volatility as well as its realized (lognormal) volatility computed over a 6-month sliding window.

Inspection of the scales of the left-hand and right-hand axes again confirms that volatilities of short-dated volatilities are larger than volatilities themselves. Also, Figure 6.2 shows that high levels of volatility tend to coincide with high levels of volatility of volatility. In this respect, implied volatilities are in fact more than lognormal: their dynamics seems to be of the type:

$$d\hat{\sigma}_{\text{ATM}} = \bullet dt + \hat{\sigma}_{\text{ATM}}^\gamma dZ_t$$

with  $\gamma > 1$ . This should be compared with the value  $\gamma = 0$  that the Heston model generates for short maturities.

What about implied volatilities for longer maturities? Equation (6.8) shows that the instantaneous volatility of long-dated VS volatilities decays like  $\frac{1}{T-t}$ . Equation



**Figure 6.2:** The 3-month ATMF implied volatility of the Euro Stoxx 50 index (darker dots, left-hand axis) together with its (lognormal) volatility, evaluated with a six-month sliding window (lighter dots, right-hand axis) from October 2001 to May 2009.

(6.4) implies that, since  $V_t$  is positive,  $\hat{\sigma}_T(t)$  has a floor:

$$\hat{\sigma}_T(t) \geq \hat{\sigma}_T^{\min}(t) = \sqrt{V^0} \sqrt{1 - \frac{1 - e^{-k(T-t)}}{k(T-t)}}$$

Figure 6.3 shows  $\frac{\hat{\sigma}_T^{\min}(t)}{\sqrt{V^0}}$  as a function of  $T - t$ ; note that, from (6.4),  $\sqrt{V^0}$  is the level of VS volatility for long-dated maturities, calibrated at  $t = 0$ . Forward VS volatilities are thus floored at a fraction of the initial long-run VS volatility level: The consequence for the smile of volatility of volatility is that volatilities of VS volatilities vanish as VS volatilities come near the floor. As is apparent in Figure 6.3, this floor on  $\hat{\sigma}_T(t)$  is not a minor effect.

## 6.6 ATM skew in the Heston model

We now turn our attention to the dependence of the ATMF skew on maturity and volatility level. We obtain an approximation of the ATMF skew at order one in the volatility of volatility  $\sigma$ .

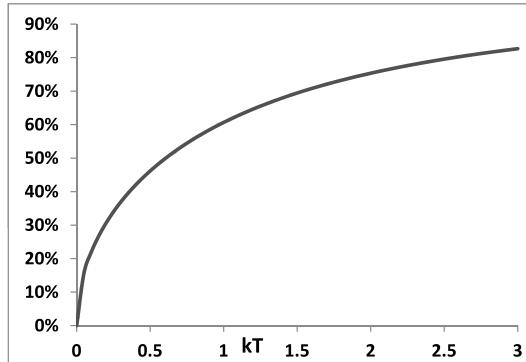


Figure 6.3:  $\frac{\hat{\sigma}_T^{\min}(t)}{\sqrt{V^0}}$  as a function of  $k(T - t)$ .

### 6.6.1 The smile at order one in volatility of volatility

While we provide in Chapter 8 a general expression of the ATMF skew at order one in the volatility of volatility for general stochastic volatility models, we now carry out a derivation for the specific case of the Heston model.

The Heston model is homogeneous, thus vanilla implied volatilities are not a separate function of spot and strike, but of the ratio of the strike to the forward of the option's maturity. To reduce bookkeeping we now set interest rate and repo to zero and reinstate them once we have our final formulas.

From SDEs (6.1) we get the following equation for the price  $P(t, S, V)$  of a European option:

$$\frac{dP}{dt} - k(V - V^0) \frac{dP}{dV} + \frac{V}{2} S^2 \frac{d^2 P}{dS^2} + \frac{\sigma^2}{2} V \frac{d^2 P}{dV^2} + \rho \sigma V S \frac{d^2 P}{dS dV} = 0 \quad (6.10)$$

with the terminal condition at maturity  $P(T, S, V) = f(S)$ , where  $f$  is the option's payoff. Denote by  $P^0(t, S, V)$  the solution of (6.10) with  $\sigma = 0$ :

$$\frac{dP^0}{dt} - k(V - V^0) \frac{dP^0}{dV} + \frac{V}{2} S^2 \frac{d^2 P^0}{dS^2} = 0 \quad (6.11)$$

with terminal condition  $P^0(T, S, V) = f(S)$ , where  $f$  is the option's payoff.

To gain some intuition on the solution of (6.11) let us go back to the stochastic representation (6.1). For  $\sigma = 0$ ,  $V$  is deterministic: the SDE for  $V_t$  in (6.1) becomes an ODE, identical to (6.2). Given the value  $V_t$  at time  $t$ , the value of  $V_\tau$  at a later time  $\tau$  is given by

$$V_\tau(V_t) = V^0 + (V_t - V^0)e^{-k(\tau-t)} \quad (6.12)$$

For  $\sigma = 0$ , the Heston model becomes a lognormal model with deterministic time-dependent volatility  $\sigma(\tau)$  given by:

$$\sigma^2(\tau) = V_\tau(V_t)$$

The implied volatility at time  $t$  for maturity  $T$  is given by:

$$\hat{\sigma}_T^2(t, V) = \frac{1}{T-t} \int_t^T V_\tau(V) d\tau = V^0 + (V - V^0) \frac{1 - e^{-k(T-t)}}{k(T-t)} \quad (6.13)$$

where  $V$  is the instantaneous variance at time  $t$ .  $P^0$  is thus simply given by the Black-Scholes formula evaluated with implied volatility  $\hat{\sigma}_T(t, V)$ :

$$P^0(t, S, V) = P_{BS}(t, S, \hat{\sigma}_T(t, V))$$

Let us expand  $P$ :

$$P = P^0 + \delta P$$

where  $\delta P$  is of order one in  $\sigma$  and let us insert this expression in equation (6.10), keeping terms up to order one in  $\sigma$ . Using the fact that  $P^0$  obeys (6.11) we are left with the following equation for  $\delta P$ :

$$\frac{d\delta P}{dt} - k(V - V^0) \frac{d\delta P}{dV} + \frac{V}{2} S^2 \frac{d^2 \delta P}{dS^2} = -\rho\sigma V S \frac{d^2 P^0}{dS dV} \quad (6.14)$$

with the terminal condition  $\delta P(T, S, V) = 0$ .  $\delta P$  is generated by the source term in the right-hand side. The second derivative of  $P_0$  with respect to  $V$  does not appear, as it is multiplied by  $\sigma^2$  – at order 1 in  $\sigma$  only the mixed derivative of  $P_0$  with respect to  $S, V$  contributes.

The solution of (6.14) at time 0 is given by:

$$\delta P(t, S, V) = \begin{aligned} & E_t^0 \left[ \int_t^T \rho\sigma V_\tau S_\tau \frac{d^2 P^0}{dS dV} \Big|_{\tau, S_\tau, V_\tau} d\tau \right] \\ & S_t = S \\ & V_t = V \end{aligned}$$

where the subscript 0 indicates that the expectation is taken with respect to the dynamics (6.1) with  $\sigma = 0$ , that is in a Black-Scholes model with deterministic time-dependent volatility  $\sigma(\tau)$ .

In the Black-Scholes model the vega of a European option is related to its dollar gamma – see (5.66), page 181:

$$\frac{dP_{BS}}{d\hat{\sigma}} = S^2 \frac{d^2 P_{BS}}{dS^2} \hat{\sigma}(T - \tau)$$

We then have:

$$\begin{aligned} \frac{dP_0}{dV} &= S^2 \frac{d^2 P^0}{dS^2} \hat{\sigma}_T(\tau, V)(T - \tau) \frac{d\hat{\sigma}_T(\tau, V)}{dV} \\ &= \frac{1 - e^{-k(T-\tau)}}{2k} S^2 \frac{d^2 P^0}{dS^2} \end{aligned}$$

where we have used expression (6.13) for  $\hat{\sigma}_T(\tau, V)$ .  $\delta P$  now reads:

$$\delta P(t, S, V) = \begin{aligned} & E_t^0 \left[ \int_t^T \frac{\rho\sigma}{2} V_\tau(V_t) \frac{1 - e^{-k(T-\tau)}}{k} S_\tau \frac{d}{dS} S^2 \frac{d^2 P^0}{dS^2} \Big|_{\tau, S_\tau, V_\tau} d\tau \right] \\ & S_t = S \\ & V_t = V \end{aligned}$$

where  $V_\tau$  is given by expression (6.12) as a function of  $V_t$ , the instantaneous variance at time 0.

We now take  $t = 0$  and simply denote by  $V$  the instantaneous variance at  $t = 0$ . We have shown in Appendix A of Chapter 5 that, in the Black-Scholes model with deterministic time-dependent volatility,  $e^{-r(\tau-t)} \frac{d^n P^0}{d \ln S^n}(\tau, S_\tau)$  is a martingale. Using:

$$S \frac{d}{dS} S^2 \frac{d^2}{dS^2} = \frac{d^3}{d \ln S^3} - \frac{d^2}{d \ln S^2}$$

we get:

$$\delta P = \frac{\rho\sigma}{2} \left[ \int_0^T V_\tau(V) \frac{1 - e^{-k(T-\tau)}}{k} d\tau \right] \left( \frac{d^3 P^0}{d \ln S^3} - \frac{d^2 P^0}{d \ln S^2} \right)_{t=0, S, V} \quad (6.15)$$

where  $V_\tau(V)$  is given by:

$$V_\tau(V) = V^0 + (V - V^0)e^{-k\tau}$$

Compare (6.15) with expression (5.88), page 193:  $\delta P$  has the same expression as a function of the second and third order derivatives of  $P^0$  with respect to  $\ln S$ : perturbing at order one in the volatility of volatility amounts to perturbing at order one in the third-order cumulant with fixed forward variances.<sup>1</sup>

This is not surprising. Indeed, the ODE for  $E[V_t]$  in the Heston model – see equation (6.2) – does not involve  $\sigma$ . Consequently, the perturbation in powers of  $\sigma$  leaves forward variances unchanged at all orders.

The interpretation of the integral over  $\tau$  in the prefactor is not straightforward at this stage. It will become clear when we carry out the derivation for general stochastic volatility models – see Section 8.6.

The expansion of the implied volatility  $\widehat{\sigma}_{KT}$  at order one in  $\sigma$  is:  $\widehat{\sigma}_{KT} = \widehat{\sigma}_T(0, V) + \delta\widehat{\sigma}_{KT}$  where  $\delta\widehat{\sigma}_{KT}$  is given by:

$$\begin{aligned} \delta\widehat{\sigma}_{KT} &= \left( \frac{dP_{BS}}{d\widehat{\sigma}} \right)^{-1} \delta P \\ &= \frac{1}{\widehat{\sigma}_T(0, V) T} \left[ \int_0^T \frac{\rho\sigma}{2} V_\tau(V) \frac{1 - e^{-k(T-\tau)}}{k} d\tau \right] \frac{\frac{d^3 P^0}{d \ln S^3} - \frac{d^2 P^0}{d \ln S^2}}{\frac{d^2 P^0}{d \ln S^2} - \frac{d P^0}{d \ln S}} \\ &= \frac{1}{\widehat{\sigma}_T(0, V) T} \left[ \int_0^T \frac{\rho\sigma}{2} V_\tau(V) \frac{1 - e^{-k(T-\tau)}}{k} d\tau \right] \frac{d}{d \ln S} \ln \left( \frac{d^2 P^0}{d \ln S^2} - \frac{d P^0}{d \ln S} \right) \end{aligned}$$

---

<sup>1</sup>The mistrustful reader is encouraged to compute  $\kappa_3$  at order one in  $\sigma$  – an easy task as the characteristic function of  $\ln S$  is analytically known in the Heston model – to verify that one indeed recovers (6.15) from (5.88).

where we have expressed vega in terms of gamma using relationship (5.66). Using the formula for the Black-Scholes dollar gamma, we get:

$$\delta\hat{\sigma}_{KT} = \frac{1}{\hat{\sigma}_T^3 T^2} \left[ \frac{\rho\sigma}{2} \int_0^T V_\tau(V) \frac{1 - e^{-k(T-\tau)}}{k} d\tau \right] \left( \frac{\hat{\sigma}_T^2 T}{2} + \ln \frac{K}{S} \right)$$

We now reinstate interest rate and repo. As mentioned above, we only need to replace  $\frac{K}{S}$  with  $\frac{K}{F_T}$ , where  $F_T$  is the forward for maturity  $T$ :

$$\delta\hat{\sigma}_{KT} = \frac{1}{\hat{\sigma}_T^3 T^2} \left[ \frac{\rho\sigma}{2} \int_0^T V_\tau(V) \frac{1 - e^{-k(T-\tau)}}{k} d\tau \right] \left( \frac{\hat{\sigma}_T^2 T}{2} + \ln \frac{K}{F_T} \right) \quad (6.16)$$

We have used the shorthand notation  $\hat{\sigma}_T = \hat{\sigma}_T(0, V) = \sqrt{V^0 + (V - V^0) \frac{1 - e^{-kT}}{kT}}$ . The notation  $\hat{\sigma}_T$  is appropriate, as  $\hat{\sigma}_T$  is the VS volatility both in the order-zero and order-one expansion in  $\sigma$ . Denoting the ATM skew by  $S_T$  and the ATM volatility by  $\hat{\sigma}_{F_T T}$ , (6.16) gives:

$$\hat{\sigma}_{F_T T} = \hat{\sigma}_T \left( 1 + \frac{\hat{\sigma}_T T}{2} S_T \right) \quad (6.17a)$$

$$S_T = \left. \frac{d\hat{\sigma}_{KT}}{d \ln K} \right|_{F_T} = \frac{1}{\hat{\sigma}_T^3 T^2} \frac{\rho\sigma}{2} \int_0^T V_\tau \frac{1 - e^{-k(T-\tau)}}{k} d\tau \quad (6.17b)$$

### Short maturities

Let us take the limit  $kT \ll 1$ . Formula (6.16) translates at lowest order in  $kT$  in the following expressions for  $\hat{\sigma}_{F_T T}$  and  $S_T$ :

$$\hat{\sigma}_{F_T T} = \sqrt{V} \left( 1 + \frac{\rho\sigma T}{8} \right) \quad (6.18a)$$

$$S_T = \frac{\rho\sigma}{4\sqrt{V}} = \frac{\rho\sigma}{4\hat{\sigma}_{F_T T}} \quad (6.18b)$$

where the second equality in (6.18b) is correct at order 1 in  $\sigma$ .

### Long maturities

We now take the limit  $kT \gg 1$ . Keeping only terms of order 1 in  $1/kT$  we get:

$$\hat{\sigma}_{F_T T} = \sqrt{V^0} \left( 1 + \frac{\rho\sigma}{4k} \right) + \frac{1}{2kT} \left( \frac{V - V^0}{\sqrt{V^0}} + \frac{\rho\sigma}{4k} \frac{V - 3V^0}{\sqrt{V^0}} \right) \quad (6.19a)$$

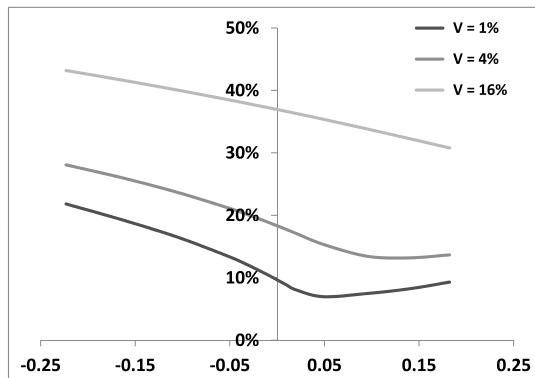
$$S_T = \frac{\rho\sigma}{2\sqrt{V^0}} \frac{1}{kT} \quad (6.19b)$$

### 6.6.2 Example

Consider the example of a smile for a three-month maturity generated with the following parameters, whose values are typical of index smiles:  $V^0 = 0.04$ ,  $k = 1$ ,  $\sigma = 0.6$ ,  $\rho = -80\%$ .

Figure 6.4 shows  $\hat{\sigma}_{KT}$  as a function of  $\ln(\frac{K}{F_T})$  for three different values of  $V$ : 1%, 4%, 16%, corresponding approximately to ATMF volatilities around 10%, 20%, and 40%, respectively.

The reason why we choose to vary  $V$  while keeping other parameters constant is that  $V$  is the only state variable of the Heston model. Figure 6.4 provides an illustration of future smiles that the Heston model generates at time  $T_1$  for maturity  $T_2 = T_1 + 3$  months – as a function of  $V_{T_1}$ .



**Figure 6.4:** Implied volatilities  $\hat{\sigma}_{KT}$  as a function of  $\ln(\frac{K}{F_T})$  for a maturity  $T = 3$  months, in the Heston model, generated with the following parameters:  $V^0 = 0.04$ ,  $k = 1$ ,  $\sigma = 0.6$ ,  $\rho = -80\%$ , for three different values of  $V$ .

The quality of approximations (6.17a, 6.17b) for these parameter values is assessed in Tables 6.1 and 6.2.

$V$	1%	4%	16%
$S_T^{\text{real}}$	-8.1	-6.0	-3.1
$S_T^{\text{approx}}$	-8.8	-5.5	-3.0

**Table 6.1:** ATM skew in the Heston model for a three-month maturity, for different values of  $V$  with  $V^0 = 0.04$ ,  $k = 1$ ,  $\sigma = 0.6$ ,  $\rho = -80\%$ , computed either exactly:  $S_T^{\text{real}}$  or using approximation (6.17b):  $S_T^{\text{approx}}$ . Results have been multiplied by 10 to represent approximately the difference in volatility points of implied volatilities for strikes  $0.95F_T$  and  $1.05F_T$ .

$V$	1%	4%	16%
$\widehat{\sigma}_T$	11.6	20	38.2
$(\widehat{\sigma}_{F_T T} - \widehat{\sigma}_T)_{\text{real}}$	-1.9	-1.7	-1.2
$(\widehat{\sigma}_{F_T T} - \widehat{\sigma}_T)_{\text{approx}}$	-0.1	-0.3	-0.5

**Table 6.2:**  $\widehat{\sigma}_T$  and the difference  $\widehat{\sigma}_{F_T T} - \widehat{\sigma}_T$ , both in volatility points, in the Heston model for a three-month maturity, for different values of  $V$  with  $V^0 = 0.04$ ,  $k = 1$ ,  $\sigma = 0.6$ ,  $\rho = -80\%$ , computed either exactly (real) or using approximation (6.17a) (approx).

The ATMF skew is acceptably captured by an approximation at order one in  $\sigma$ : the maximum relative error is about 10%.

In contrast, the difference  $\widehat{\sigma}_{F_T T} - \widehat{\sigma}_T$ , is poorly estimated by (6.17a). Though both  $\mathcal{S}_T$  and  $\widehat{\sigma}_{F_T T} - \widehat{\sigma}_T$  are of order one in  $\sigma$ , the approximation for  $\mathcal{S}_T$  is more robust.

This is generally observed in stochastic volatility models for equities: one typically needs to carry out the expansion of the ATMF volatility at order two in volatility of volatility to reach acceptable accuracy – more on this in Section 8.2.

### 6.6.3 Term structure of the ATMF skew

Equations (6.18b), (6.19b) show that while the ATMF skew tends to a constant for  $T \rightarrow 0$ , it decays like  $\frac{1}{T}$  for long maturities, at order one in  $\sigma$ . The  $\frac{1}{T}$  decay for long maturities is expected: because  $V$  is mean-reverting, for maturities  $T$  such that  $T \gg \frac{1}{k}$ , the distribution of returns of  $\ln S$  over long periods becomes independent on the initial value of  $V$ , hence returns of  $\ln S$  over long time scales become independent. The third-order cumulant  $\kappa_3$  of  $\ln S$  then scales like  $T$ , the skewness  $s$  of  $\ln S$  scales like  $\frac{1}{\sqrt{T}}$  and (5.92) then implies that  $\mathcal{S}_T \propto \frac{1}{T}$ .

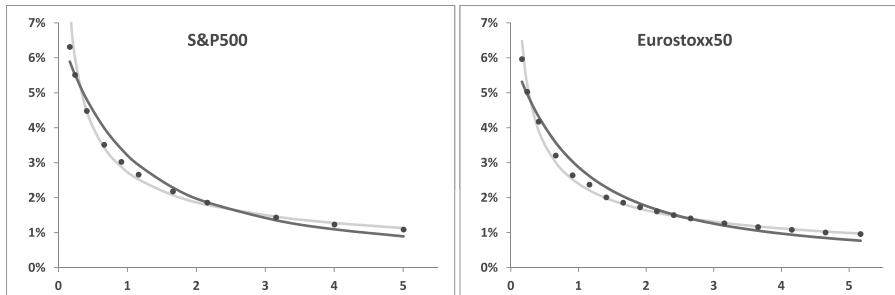
In the special case when  $V = V^0$ , the term structure of VS volatilities is flat,  $V_t$  is constant and (6.16) takes the following simple form:

$$\left. \frac{d\widehat{\sigma}_{KT}}{d \ln K} \right|_{F_T} = \frac{\rho\sigma}{2\sqrt{V^0}} \frac{kT + e^{-kT} - 1}{(kT)^2} \quad (6.20)$$

Figure 6.5 shows an example of the term structure of the ATMF skew for the Euro Stoxx 50 and S&P 500 indexes, together with a power-law fit and a best fit using formula (6.20). The maturity dependence of the market skew indeed exhibits a power-law-like behavior, which cannot be captured by the Heston model for both short and long maturities: the Heston model is a one-factor model, with an embedded time scale  $1/k$ .

The issue here is not only about whether we are or aren't able to fit the vanilla smile. Rather, when risk-managing cliques we may need to carry a naked forward smile position: it is then necessary to assess whether the model is able to gener-

ate *future* smiles – that is vanilla smiles at future dates – that are comparable to historically observed vanilla smiles – see the discussion below.



**Figure 6.5:** The ATMF skew as a function of maturity for the Euro Stoxx 50 and S&P 500 indexes (dots), observed on October 22, 2010, expressed as the difference in volatility points of the implied volatilities of strikes  $0.95F_T$  and  $1.05F_T$ . A best fit using a power law with exponent 0.55 (lighter line) as well as a best fit using formula (6.20) ( $k = 2.9$ ) (darker line) are shown as well.

#### 6.6.4 Relationship between ATMF volatility and skew

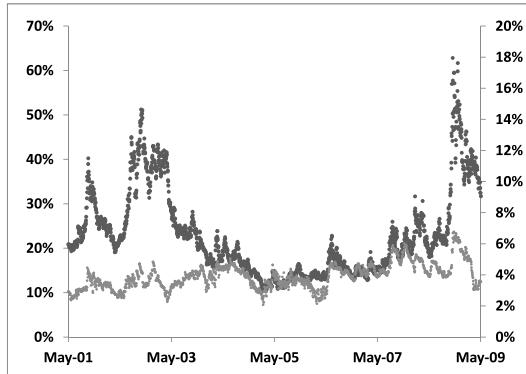
For short maturities, equation (6.18b) shows that, in the Heston model the ATMF skew is inversely proportional to the ATMF implied volatility – this is also evidenced in Figure 6.6. Is this dependence observed in reality?

Figure 6.6 displays the ATM volatility together with the ATMF skew for a 3-month maturity, for the Euro Stoxx 50 index.<sup>2</sup> We can see that while skew and volatility seem to behave fairly independently, they are, if anything, positively correlated rather than negatively. It seems unreasonable to hard-wire an inverse dependence of the ATMF skew to the ATMF volatility in our model.

## 6.7 Discussion

The above analysis has highlighted some discrepancies between the spot/volatility dynamics generated by the Heston model on one hand, and observed in reality on the other hand – see also [8]. What makes the Heston model unsuitable for handling exotics however, rather than its inability to reproduce *exactly* the observed historical dynamics, is the lack of flexibility it affords.

<sup>2</sup>We have used the ATM volatility and skew for simplicity – using ATMF data would have yielded a similar graph.



**Figure 6.6:** The ATM skew (lighter dots, right-hand axis) as the difference of the implied volatilities of the 95% and 105% strikes, and the ATM volatility (darker dots, left-hand axis) of the Euro Stoxx 50 index, for a 3-month maturity.

Indeed, from a trading perspective, one may choose to use parameter levels different than their historical averages and structural dependencies other than what is historically observed, even for parameters that have no implied counterpart. For example, even though the historical realized correlation between two quantities may be negligibly small, we will use bid/offer levels that are different than zero, depending on the size and sign of the option's sensitivity to correlation.

Likewise, imagine that in reality the short ATMF skew was indeed approximately inversely proportional to the short ATMF volatility, thus in line with the behavior generated in the Heston model. Still, when selling an exotic option that has positive forward ATMF volatility/skew cross-gamma:  $\frac{d^2 P}{d\sigma_{F_T T} dS_T} > 0$ , we may choose to conservatively price with a model that generates positive or vanishing covariance between ATMF volatility and skew – rather than negative.

Making the level of the short ATMF skew independent on the level of short ATMF volatility is easy: we only need to replace the SDE for  $V_t$  in (6.1) with:

$$dV_t = \bullet dt + \nu V_t dZ_t$$

This would also have the advantage of making the short ATMF volatility lognormal rather than normal.

Other deficiencies of the Heston model are structural. For example, the scaling with maturity of the volatility of volatility in (6.6) and of the ATMF skew in (6.20) is intimately related to the fact that the Heston model is a one-factor model. Some have advocated making the parameters of the Heston model time-dependent so as to alter the dependence of  $S_T$  on  $T$  and achieve accurate calibration of the vanilla smile.

Making  $V^0$  time-dependent so as to best match the term structure of VS volatilities is appropriate, as forward variances can be hedged by trading variance swaps.<sup>3</sup> Making  $\sigma$  and  $\rho$  time-dependent is more questionable.

Indeed, consider the example of a call-spread cliquet paying at time  $T_2$  the payoff  $(\frac{S_{T_2}}{S_{T_1}} - 95\%)^+ - (\frac{S_{T_2}}{S_{T_1}} - 105\%)^+$ , where  $T_2 - T_1 \ll \frac{1}{k}$ . This call spread has negligible sensitivity to volatility but is very sensitive to the ATMF skew observed at  $T_1$  for maturity  $T_2$ . The order-one expansion in  $\sigma$  in (6.18b) yields the following expression for the ATMF skew at  $T_1$  for maturity  $T_2$ :

$$\mathcal{S}_{T_2}(T_1) = \frac{\rho(T_1)\sigma}{4\sqrt{V_{T_1}}}$$

where we have allowed  $\rho$  to be time-dependent so as to match the term structure of the market *vanilla* ATMF skew. By pricing the cliquet with the Heston model thus calibrated we are betting on a given level of *forward* skew  $\mathcal{S}_{T_2}(T_1)$ , derived from calibration to today's term structure of the ATMF skew of vanilla options. Worse, the model will actually generate hedge ratios for our forward-start call spread on vanilla options of maturities  $T_1$  and  $T_2$ .

This is not justified, as it is not possible to hedge forward-skew risk by trading vanilla options – we refer the reader to the illuminating experiment in Section 3.1.7. In case we are unable to take an offsetting position in a different exotic that has comparable forward-smile risk and have to keep the cliquet and its forward skew exposure on our book, it is more reasonable to *select* a conservative level of forward skew we are comfortable with, rather than having it dictated by calibration to market prices of instruments that are incapable of hedging it. The same goes for volatility-of-volatility risk.

Only if we are able to trade cliquets of varying maturities is it licit to make  $\sigma$  or  $\rho$  time-dependent, so as to match the term structure of the implied forward skew.

Fundamentally, the problem with the Heston model lies not so much with the model itself, its lack of flexibility, or even its peculiar idiosyncrasies, but with its usage (and its users): which practical pricing or hedging issue naturally calls for SDE (6.5)?

---

<sup>3</sup>Exact calibration is not guaranteed as  $V_0(t)$  has to remain positive.

## Chapter's digest

- In terms of forward variances, the SDE of the Heston model reads:

$$\begin{cases} dS_t = \sqrt{\xi_t^t} S_t dW_t \\ d\xi_t^T = \sigma e^{-k(T-t)} \sqrt{\xi_t^t} dZ_t \end{cases}$$

with the following constraint on the initial values of forward variances:

$$\frac{d\xi_0^T}{dT} = -k(\xi_0^T - V^0) dT$$

VS volatilities are given by:

$$\hat{\sigma}_T^2(t) = V^0 + \frac{1 - e^{-k(T-t)}}{k(T-t)}(V_t - V^0)$$

thus are floored.

- The drift of the instantaneous variance  $V_t$  in stochastic volatility models has nothing to do with the “market price of risk”. It is related to the initial slope of the variance curve:

$$dV_t = \left. \frac{d\xi_t^T}{dT} \right|_{T=t} dt + \lambda_t^t dZ_t^t$$

- For a flat term-structure of VS volatilities, at order one in volatility of volatility, the ATMF skew of the Heston model is given by:

$$\mathcal{S}_T = \frac{\rho\sigma}{2\sqrt{V^0}} \frac{kT + e^{-kT} - 1}{(kT)^2}$$

and the (lognormal) volatility of the VS volatility of maturity  $T$ ,  $\hat{\sigma}_T$ , has the form:

$$\text{vol}(\hat{\sigma}_T) \propto \frac{1 - e^{-kT}}{kT}$$

- At order one in volatility of volatility, the ATMF skew of the Heston model for short maturities is given by:

$$\mathcal{S}_T = \frac{\rho\sigma}{4\sqrt{V}}$$

and for long maturities:

$$\mathcal{S}_T = \frac{\rho\sigma}{2\sqrt{V^0}} \frac{1}{kT}$$

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

# **Chapter 7**

---

## **Forward variance models**

We catch up to where we left off, at the end of Chapter 4. We examine diffusive stochastic volatility models built on the dynamics of continuous forward VS variances – they are exactly calibrated to an initial term structure of VS volatilities, by construction. They can alternatively be calibrated to a term structure of implied volatilities of other payoffs, for example ATMF vanilla options, or power payoffs.

We concentrate on the control of the term structure of volatility of volatility, the term structure of the ATMF skew, the smile of volatility of volatility and cover options on realized variance and VIX instruments.

The last section deals with discrete forward variance models, a type of stochastic volatility model that is particularly suited to the analysis of exotic option risks in terms of volatility of volatility, spot/volatility covariance and future skew.

---

### **7.1 Pricing equation**

The Heston model – studied in the preceding chapter – is an elementary attempt at designing a model such that implied volatilities are not frozen anymore and have their own dynamics. This is done by specifying an SDE for the instantaneous variance  $V_t$ , a non-physical object. It has then been our task to extract the dynamics of implied volatilities that this SDE gives rise to.

In this chapter we model implied volatilities directly. From the discussion in Section 4.3.6, page 148, and Section 5.5, page 168, it is clear that the easiest objects to model are VS forward variances  $\xi^T$ .

The  $\xi^T$  will be our state variables, in addition to  $S$ , and we will design models so that we have a direct handle on the volatilities of VS volatilities – instantaneous or discrete, forward or spot-starting. The price of an option in such a model is given by:

$$P(t, S, \xi)$$

where  $\xi$  is the variance curve.

Consider a short position in an option of maturity  $T$  – first unhedged. Our P&L during  $\delta t$  is

$$-\left[ P(t + \delta t, S + \delta S, \xi + \delta \xi) - (1 + r\delta t) P(t, S, \xi) \right]$$

The delta hedge consists of  $\frac{dP}{dS}$  shares and  $\frac{\delta P}{\delta \xi^u}$  forward VS contracts of maturity  $u$ , for all  $u$  in  $[t, T]$ , where  $\frac{\delta P}{\delta \xi^u}$  is a functional derivative, since  $\xi^u$  is a function of  $u$ . Our total P&L during  $\delta t$ , including now our delta- and vega-hedge, is:

$$\begin{aligned} P\&L = & - [P(t + \delta t, S + \delta S, \xi + \delta \xi) - (1 + r\delta t) P(t, S, \xi)] \\ & + \frac{dP}{dS} (\delta S - (r - q)S\delta t) + \int_t^T \frac{\delta P}{\delta \xi^u} \delta \xi^u \end{aligned}$$

We remind the reader of the fact that (forward) VSs – which provide exact delta hedges for the  $\xi^u$  – can be entered into at zero cost. The  $\xi^u$  have no financing cost – thus zero risk-neutral drift – hence the simple form of the contribution of  $\delta \xi^u$  to the P&L.

Expanding at order one in  $\delta t$  and two in  $\delta S$  and  $\delta \xi^u$ :<sup>1</sup>

$$\begin{aligned} P\&L = & - \frac{dP}{dt} \delta t + rP\delta t - (r - q)S \frac{dP}{dS} \delta t \\ & - \frac{S^2}{2} \frac{d^2 P}{dS^2} \frac{\delta S^2}{S^2} - \frac{1}{2} \int_t^T du \int_t^T du' \frac{d^2 P}{\delta \xi^u \delta \xi^{u'}} \delta \xi^u \delta \xi^{u'} - \int_t^T du S \frac{d^2 P}{dS \delta \xi^u} \frac{\delta S}{S} \delta \xi^u \end{aligned} \quad (7.1)$$

Using the same criteria that led us to the Black-Scholes equation in Section 1.1, we specify break-even levels for the random second-order terms in the P&L. Denote by  $\sigma(t)$  the instantaneous break-even volatility of  $S_t$  and by  $\nu(t, u, u')$  and  $\mu(t, u)$  the instantaneous break-even covariances for, respectively,  $\delta \xi^u, \delta \xi^{u'}$  and  $\delta \xi^u, \frac{\delta S}{S}$ :

$$\mu(t, u) \delta t = \left\langle \frac{\delta S}{S} \delta \xi^u \right\rangle_t \quad (7.2a)$$

$$\nu(t, u, u') \delta t = \left\langle \delta \xi^u \delta \xi^{u'} \right\rangle_t \quad (7.2b)$$

Obviously  $\mu$  is only defined for  $u \geq t$  and  $\nu$  for  $u \geq t, u' \geq t$ .

Break-even variances and covariances can be at most a function of our state variables  $S$  and  $\xi$ , unless we introduce additional degrees of freedom in the model. We thus write:  $\sigma(t, S, \xi)$ ,  $\nu(t, u, u', S, \xi)$ ,  $\mu(t, u, S, \xi)$ . In the models we will work with in the sequel,  $\nu$  and  $\mu$  do not depend on  $S$ , so let us write our covariance functions more simply as:  $\nu(t, u, u', \xi)$ ,  $\mu(t, u, \xi)$ .

We would like our carry P&L at order one in  $\delta t$ , two in  $\delta S, \delta \xi$  to read:

$$P\&L = - \frac{S^2}{2} \frac{d^2 P}{dS^2} \left( \frac{\delta S^2}{S^2} - \sigma(t, S, \xi)^2 \delta t \right) \quad (7.3a)$$

$$- \frac{1}{2} \int_t^T du \int_t^T du' \frac{d^2 P}{\delta \xi^u \delta \xi^{u'}} \left( \delta \xi^u \delta \xi^{u'} - \nu(t, u, u', \xi) \delta t \right) \quad (7.3b)$$

$$- \int_t^T du S \frac{d^2 P}{dS \delta \xi^u} \left( \frac{\delta S}{S} \delta \xi^u - \mu(t, u, \xi) \delta t \right) \quad (7.3c)$$

---

<sup>1</sup>We use economical notations for what should really read:  $\frac{\delta^2 P}{\delta \xi^u \delta \xi^{u'}}$  and  $\frac{d\delta P}{dS \delta \xi^u}$ .

The gamma P&L in (7.3a) can be offset by trading a short-maturity VS. So that no free theta is generated, we must impose that the break-even volatility at time  $t$  is equal to the instantaneous VS volatility:  $\sigma(t, S, \xi)^2 = (\hat{\sigma}_t^t)^2 = \xi^t$ .

Identifying the  $\delta t$  terms in (7.1) and (7.3) supplies us with the pricing equation in our model:

$$\begin{aligned} \frac{dP}{dt} + (r - q)S \frac{dP}{dS} + \frac{\xi^t}{2} S^2 \frac{d^2P}{dS^2} \\ + \frac{1}{2} \int_t^T du \int_t^T du' \nu(t, u, u', \xi) \frac{d^2P}{d\xi^u d\xi^{u'}} + \int_t^T du \mu(t, u, \xi) S \frac{d^2P}{dS d\xi^u} = rP \end{aligned} \quad (7.4)$$

with the terminal condition  $P(t, S, \xi, t = T) = g(S)$  where  $g$  is the option's payoff, in the case of a European option.

Generally,  $g$  can depend on the full path of  $S_t$  for a path-dependent option, or the full path of  $S_t$  and of the  $\xi_t^T$  if the payoff involves observations of VS volatilities.<sup>2</sup>

The probabilistic interpretation of (7.4) is that  $P$  is given – in the case of a European payoff – by:

$$P = E[g(S_T) | S_t = S, \xi_t^u = \xi^u]$$

with the following SDEs for  $S_t$  and  $\xi_t^u$ :

$$\begin{cases} dS_t &= (r - q)S_t dt + \sqrt{\xi_t^t} S_t dW_t^S \\ d\xi_t^u &= \lambda_t^u dW_t^u \end{cases}$$

with  $\lambda_t^u$  and correlations between  $W_t^S$  and  $W_t^u$  such that:

$$\lim_{dt \rightarrow 0} \frac{1}{dt} E_t[d\ln S_t d\xi_t^u] = \sqrt{\xi_t^t} \lambda_t^u \frac{1}{dt} E_t[dW_t^S dW_t^u] = \mu(t, u, \xi) \quad (7.5)$$

$$\lim_{dt \rightarrow 0} \frac{1}{dt} E_t[d\xi_t^u d\xi_t^{u'}] = \lambda_t^u \lambda_t^{u'} \frac{1}{dt} E_t[dW_t^u dW_t^{u'}] = \nu(t, u, u', \xi) \quad (7.6)$$

## 7.2 A Markov representation

While equation (7.4) is general, it is an infinite-dimensional equation that is not solvable unless the  $\xi^u$  possess a Markov-functional representation. Failing that, in a Monte Carlo simulation the (infinitely many)  $\xi^u$  need to be evolved individually. This is not possible, unless one resorts to an approximation. If instead a Markov-functional representation exists, the  $\xi^u$  can be expressed as a function of a small set of state variables.

<sup>2</sup>Consider for example VS swaptions.

In addition to this technical condition, we also require that the dynamics of forward variances be financially motivated.<sup>3</sup>

Consider a forward variance  $\xi^T$ , where  $T$  is an arbitrary date, and let us start with a lognormal dynamics for  $\xi^T$  – Figure 6.2 in Section 6.5 suggests that this assumption is a reasonable starting point. How should the volatility of  $\xi_t^T$  depend on  $t$  and  $T$ ?

If there existed a market of options on  $\xi^T$  with maturities ranging from  $t$  to  $T$ , the volatility risk of  $\xi^T$  could be hedged away and the volatility of  $\xi^T$  would be derived from market implied volatilities. However, volatility of volatility is only traded in very special forms, for example through options on realized variance, through VIX futures and options, or cliques. We have already considered cliques beforehand and will analyze in detail other instruments further on, and characterize the type of volatility-of-volatility risk they are sensitive to.

In general we will have no choice but to carry a position on the realized volatility of  $\xi^T$  and thus will need to make assumptions that we will depend on. It is then reasonable to make the assumption of time-homogeneity: the volatility of  $\xi^T$  only depends on  $T-t$ , with a dependence that is adjustable so that, for example, volatilities of spot-starting VS volatilities  $\hat{\sigma}_T$  in the model can be made to match their historical counterparts. Let us then write:

$$d\xi_t^T = \omega(T-t) \xi_t^T dW_t^T \quad (7.7)$$

The solution of this SDE is

$$\ln(\xi_t^T) = \ln(\xi_0^T) - \frac{1}{2} \int_0^t \omega^2(T-\tau) d\tau + \int_0^t \omega(T-\tau) dW_\tau^T \quad (7.8)$$

Imagine the same Brownian motion  $W_t$  drives the dynamics of all  $\xi^T$ . Equation (7.8) expresses  $\ln \xi^T$  as a weighted average of increments of  $W_t$ , with a weight  $\omega(T-t)$  that depends on  $T$ , hence is specific to forward variance  $\xi^T$ . Even though a single Brownian motion drives our model, simulation of the forward variance curve at time  $t$  requires knowledge of the full path of  $W_t$ , as weights for different forward variances are different – all of the  $\xi^T$  have to be simulated individually.

However, if  $\omega$  is of the form:

$$\omega(u) = \omega e^{-ku} \quad (7.9)$$

$$\int_0^t \omega(T-\tau) dW_\tau = e^{-kT} \int_0^t e^{k\tau} dW_\tau$$

The dependence on  $T$  factors out and knowledge of one quantity –  $\int_0^t e^{k\tau} dW_\tau$  – allows the construction of the full variance curve at time  $t$ : a Markov-functional representation exists.<sup>4</sup>

---

<sup>3</sup>See [15] for a characterization of the conditions on the volatility structure of a futures curve such that the resulting dynamics admits a finite-dimensional Markov representation. The  $\xi^T$  are indeed akin to a futures curve, as they are driftless.

<sup>4</sup>In [20] Hans Buehler studies Markov representations of the variance curve of the type:  $\xi_t^T = G(\mathbf{X}_t, T-t)$  where  $\mathbf{X}_t$  is a vector diffusive process, and provides a few examples of  $(\mathbf{X}_t, G)$  couples

Choosing an exponentially decaying volatility function is equivalent to driving the dynamics of forward variances with one Ornstein–Uhlenbeck (OU) process  $X_t$ :

$$dX_t = -kX_t dt + dW_t, \quad X_0 = 0$$

$X_t$  and its variance are given by:

$$X_t = \int_0^t e^{-k(t-\tau)} dW_\tau \quad E[X_t^2] = \frac{1 - e^{-2kt}}{2k}$$

With  $\omega(u)$  of the form (7.9), the solution of SDE (7.7) reads:

$$\xi_t^T = \xi_0^T \exp \left( \omega e^{-k(T-t)} X_t - \frac{\omega^2}{2} e^{-2k(T-t)} E[X_t^2] \right) \quad (7.10)$$

$\omega$  is the lognormal volatility of  $\xi_t^{T=t}$ , a forward variance with vanishing maturity.

---

### 7.3 $N$ -factor models

Let us use  $N$  Brownian motions and write the SDE of  $\xi_t^T$  as:

$$d\xi_t^T = \omega \alpha_w \xi_t^T \sum_i w_i e^{-k_i(T-t)} dW_t^i \quad (7.11)$$

where  $\alpha_w$  is a normalizing factor such that the instantaneous lognormal volatility of  $\xi_t^{T=t}$  is  $\omega$ . Volatilities of volatilities are more natural objects than volatilities of variances. We thus introduce the lognormal volatility  $\nu$  of a VS volatility of vanishing maturity, which is the square root of  $\xi_t^T$ . Its instantaneous volatility is half that of  $\xi_t^T$ . We have:

$$\omega = 2\nu \quad (7.12a)$$

$$\alpha_w = \frac{1}{\sqrt{\sum_{ij} w_i w_j \rho_{ij}}} \quad (7.12b)$$

The solution of (7.11) is given by:

$$\xi_t^T = \xi_0^T \exp \left( \omega \sum_i w_i e^{-k_i(T-t)} X_t^i - \frac{\omega^2}{2} \sum_{ij} w_i w_j e^{-(k_i+k_j)(T-t)} E[X_t^i X_t^j] \right) \quad (7.13)$$

where the  $N$  OU processes  $X^i$  are defined by:

$$dX_t^i = -k_i X_t^i dt + dW_t^i, \quad X_{t=0}^i = 0 \quad (7.14)$$

---

that ensure that  $\xi_t^T$  are martingales. This amounts to enforcing a parametric representation of the variance curve – arbitrary VS term structures cannot be accommodated. He later relaxes this constraint by setting  $\xi_t^T = \xi_0^T G(\mathbf{X}_t, T-t)$ .

### 7.3.1 Simulating the $N$ -factor model

Because it is driven by Ornstein–Uhlenbeck processes the  $N$ -factor model is easily and exactly simulable. We start at  $t = 0$  with:

$$X_0^i = 0, \quad E[X_0^i X_0^j] = 0$$

Imagine we have  $X_t^i$ , and  $E[X_t^i X_t^j]$  at time  $t = \tau_n$  and we need to generate them at time  $\tau_{n+1} = \tau_n + \delta\tau$ . The solution of (7.14) at time  $t$  is given by:

$$X_t^i = e^{-k_i t} X_0^i + \int_0^t e^{-k_i(t-u)} dW_u^i$$

$X_{\tau_{n+1}}^i$  thus reads:

$$\begin{aligned} X_{\tau_{n+1}}^i &= e^{-k_i \tau_{n+1}} X_0^i + \int_0^{\tau_{n+1}} e^{-k_i(\tau_{n+1}-u)} dW_u^i \\ &= e^{-k_i \delta\tau} X_{\tau_n}^i + \int_{\tau_n}^{\tau_{n+1}} e^{-k_i(\tau_{n+1}-u)} dW_u^i \end{aligned}$$

Introducing the Gaussian random variable  $\delta X^i = \int_{\tau_n}^{\tau_{n+1}} e^{-k_i(\tau_{n+1}-u)} dW_u^i$ , with zero mean,  $X_{\tau_{n+1}}^i$  is generated from  $X_{\tau_n}^i$  through:

$$X_{\tau_{n+1}}^i = e^{-k_i \delta\tau} X_{\tau_n}^i + \delta X^i \quad (7.15)$$

Using this expression for  $X_{\tau_{n+1}}^i$  and taking expectations:

$$E[X_{\tau_{n+1}}^i X_{\tau_{n+1}}^j] = e^{-(k_i+k_j)\delta\tau} E[X_{\tau_n}^i X_{\tau_n}^j] + E[\delta X^i \delta X^j] \quad (7.16)$$

To generate the Gaussian random variables  $\delta X^i$  we only need their covariance matrix, which is given by:

$$E[\delta X^i \delta X^j] = \rho_{ij} \frac{1 - e^{-(k_i+k_j)\delta\tau}}{k_i + k_j} \quad (7.17)$$

where  $\rho_{ij}$  is the correlation of Brownian motions  $W_t^i$  and  $W_t^j$ .

Thus, in case we do not need to simulate the spot process – for example if we are dealing with payoffs on realized or implied variance – no time stepping is required: the  $X_t^i$  are generated exactly for times  $t$  at which instantaneous or VS variances are needed, as mandated by the derivative's term sheet.

The stochastic volatility degrees of freedom of the  $N$ -factor lognormal model are easily and exactly simulated; this is a very attractive feature, especially when compared with the Heston model.<sup>5</sup>

---

<sup>5</sup>It is a well-known fact that mere simulation of the (one-factor) process  $V_t$  in the Heston model is excessively arduous, especially for large volatilities of volatilities. This one issue has contributed its fair share of papers to the mathematical finance literature.

In addition, as shown in Section 7.7 further below, we can easily relax the lognormality of forward variances while preserving the Markov-functional property of the model and retaining the capability of exactly simulating the dynamics of the variance curve.

Note that, unlike the  $X_t^i$ , the ‘‘convexity terms’’  $E[X_t^i X_t^j]$  are non-random, hence do not need to be simulated by the time-stepping process in (7.16); they can simply be computed in advance for times  $t$  of interest.

### Simulating the spot process

Over the interval  $[\tau_n, \tau_{n+1}]$  the process for  $\ln S$  is discretized as:

$$\delta \ln S = \left( r - q - \frac{\xi_t^t}{2} \right) \delta t + \sqrt{\xi_t^t} \delta W^S$$

where  $\delta W^S$  is a Gaussian random variable of variance  $\delta t$ . The covariance of  $\delta W^S$  and  $\delta X^i$  can be computed at once using their expressions:

$$\delta X^i = \int_{\tau_n}^{\tau_{n+1}} e^{-k_i(\tau_{n+1}-u)} dW_u^i \quad \delta W^S = \int_{\tau_n}^{\tau_{n+1}} dW_u^S$$

We get:

$$E [\delta W^S \delta X^i] = \rho_{iS} \frac{1 - e^{-k_i \delta \tau}}{k_i} \quad (7.18)$$

where  $\rho_{iS}$  is the correlation between  $W^i$  and  $W^S$ .

Using expressions (7.17) and (7.18) for the various covariances, Gaussian random variables  $\delta W^S$  and  $\delta X^i$  are easily generated.

In case we are only interested in obtaining the vanilla smile generated by our model, there are more efficient techniques than simulating  $S_t$  and evaluating vanilla payoffs – we refer the reader to Appendix A of Chapter 8, page 336.

### 7.3.2 Volatilities and correlations of variances

The instantaneous volatility of  $\xi_t^T$  is, from SDE (7.11):

$$\omega(T-t) = (2\nu)\alpha_w \sqrt{\sum_{ij} w_i w_j \rho_{ij} e^{-(k_i+k_j)(T-t)}} \quad (7.19)$$

and the instantaneous correlation of two forward variances  $\xi_t^T, \xi_t^{T'}$  is given by:

$$\rho_t(\xi_t^T, \xi_t^{T'}) = \frac{\sum_{ij} w_i w_j \rho_{ij} e^{-(k_i(T-t)+k_j(T'-t))}}{\sqrt{\sum_{ij} w_i w_j \rho_{ij} e^{-(k_i+k_j)(T-t)}} \sqrt{\sum_{ij} w_i w_j \rho_{ij} e^{-(k_i+k_j)(T'-t)}}} \quad (7.20)$$

Consider the VS volatility for maturity  $T$ ,  $\hat{\sigma}_T(t)$ :  $\hat{\sigma}_T^2(t) = \frac{1}{T-t} \int_t^T \xi_t^\tau d\tau$ . The dynamics of  $\hat{\sigma}_T(t)$  is given by:

$$d\hat{\sigma}_T = \nu \alpha_w \frac{1}{\hat{\sigma}_T} \sum_i w_i \left( \frac{1}{T-t} \int_t^T \xi_t^\tau e^{-k_i(\tau-t)} d\tau \right) dW_t^i + \bullet dt \quad (7.21)$$

We now introduce the notation  $\nu_T(t)$  for the instantaneous lognormal volatility of  $\hat{\sigma}_T$  at time  $t$ .  $\nu_T(t)$  is given by:

$$\begin{cases} \nu_T(t) = \nu \alpha_w \sqrt{\sum_{ij} w_i w_j \rho_{ij} f_i(t, T) f_j(t, T)} \\ f_i(t, T) = \frac{\int_t^T \xi_t^\tau e^{-k_i(\tau-t)} d\tau}{\int_t^T \xi_t^\tau d\tau} \end{cases} \quad (7.22)$$

The instantaneous volatility of a very short VS volatility is  $\nu$ :

$$\nu_t(t) = \nu$$

As is clear from (7.22)  $\nu$  is a global scale factor for volatilities of volatilities.

What about volatilities of forward VS volatilities? Consider two dates  $T_1, T_2$  with  $t \leq T_1 \leq T_2$  and define the forward VS volatility  $\hat{\sigma}_{T_1 T_2}$  as:

$$\hat{\sigma}_{T_1 T_2}(t) = \sqrt{\frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \xi_t^\tau d\tau}$$

The instantaneous volatility  $\nu_{T_1 T_2}(t)$  of  $\hat{\sigma}_{T_1 T_2}$  is given by:

$$\begin{cases} \nu_{T_1 T_2}(t) = \nu \alpha_w \sqrt{\sum_{ij} w_i w_j \rho_{ij} f_i(t, T_1, T_2) f_j(t, T_1, T_2)} \\ f_i(t, T_1, T_2) = \frac{\int_{T_1}^{T_2} \xi_t^\tau e^{-k_i(\tau-t)} d\tau}{\int_{T_1}^{T_2} \xi_t^\tau d\tau} \end{cases} \quad (7.23)$$

### Flat term structure of VS volatilities

In the case of a flat term structure of VS volatilities,  $\xi_t^\tau$  does not depend on  $\tau$ . The integral in  $f_i(t, T)$  in (7.22) can be evaluated analytically and we get the following simple formula for the instantaneous volatility of  $\hat{\sigma}_T$  at time  $t$ :

$$\nu_T(t) = \nu \alpha_w \sqrt{\sum_{ij} w_i w_j \rho_{ij} I(k_i(T-t)) I(k_j(T-t))} \quad (7.24)$$

where

$$I(x) = \frac{1 - e^{-x}}{x} \quad (7.25)$$

Likewise, the instantaneous volatility  $\nu_{T_1 T_2}(t)$  of the forward VS volatility  $\hat{\sigma}_{T_1 T_2}$  is given by:

$$\nu_{T_1 T_2}(t) = \nu \alpha_w \sqrt{\sum_{ij} w_i w_j \rho_{ij} I(k_i(T_2-T_1)) I(k_j(T_2-T_1)) e^{-(k_i+k_j)(T_1-t)}} \quad (7.26)$$

As is clear from (7.24), whenever the VS term structure is flat at time  $t$ ,  $\nu_T(t)$  and  $\nu_{T_1 T_2}(t)$  are respectively functions of  $T - t$  and  $T_1 - t, T_2 - t$  only: the model is time-homogeneous.

In what follows, for the sake of setting model parameters, we will frequently use this situation as a reference case.

### 7.3.3 Vega-hedging in finite-dimensional models

Imagine we use  $N$  OU processes. In a lognormal model for forward variances we then have a Markov-functional representation: all forward variances can be written as *functions* of the  $N$  OU processes  $X^i$  and time. One might argue that we should delta-hedge – in our case vega-hedge – forward variance risk using  $N$  variance swaps of different maturities only, so as to neutralize sensitivities with respect to the  $N$  factors  $X^i$ . In the case of a one-factor model we could pick a particular maturity and our delta hedge would consist of one variance swap only.

However the function of a delta – in our case vega – strategy is to immunize our position at order one against all deformations  $\delta\xi^T$  of the variance curve – not only those allowed by the covariance structure of the model. Only if the deltas  $\frac{dP}{d\xi^T}$  are traded are we then able to materialize during  $\delta t$  the usual gamma/theta P&L with break-even levels specified by the covariance functions  $\mu$  and  $\nu$  in the pricing equation (7.4).

From SDE (7.11) for  $\xi_t^T$  and SDE:

$$dS_t = (r - q)S_t dt + \sqrt{\xi_t^T} S_t dW_t^S$$

for  $S_t$ , we get the spot/variance and variance/variance covariance functions in the  $N$ -factor model:

$$\mu(t, u, \xi) = \omega \alpha_w \sqrt{\xi_t^T \xi_t^u} \sum_i \rho_{SX^i} w_i e^{-k_i(u-t)} \quad (7.27a)$$

$$\nu(t, u, u', \xi) = \omega^2 \alpha_w^2 \xi_t^u \xi_t^{u'} \sum_{ij} \rho_{ij} w_i w_j e^{-k_i(u-t)} e^{-k_j(u'-t)} \quad (7.27b)$$

where  $\rho_{ij}$  is the correlation of  $W^i$  and  $W^j$  and  $\rho_{SX^i}$  the correlation of  $W^i$  and  $W^S$ .

Thus, with regard to deltas, the deformation modes of the variance curve generated by the  $N$  processes have no special significance. Model factors simply set the structure and rank of the break-even covariance matrix of the gamma/theta P&L of a hedged position. We refer the reader to the discussion of a similar issue – the delta in the local volatility model – in Section 2.7.8, page 77.

It is important to stress that calculation of deltas is not connected in any way to the covariance structure of the hedging instruments in the model at hand.

## 7.4 A two-factor model

How many OU processes should we use? How should we select their time scales  $1/k_i$ ? Let us start with a one-factor model:

$$d\xi_t^T = (2\nu) e^{-k(T-t)} \xi_t^T dW_t$$

From (7.24) the instantaneous volatility of  $\hat{\sigma}_T$  in the case of a flat term structure of VS volatilities at time  $t$  is:

$$\nu_T(t) = \nu I(k(T-t)) = \nu \frac{1 - e^{-k(T-t)}}{k(T-t)}$$

Observe that this expression is identical to formula (6.9) in the Heston model – we know from our study in Chapter 6 that one factor does not offer sufficient flexibility with regard to the dynamics of forward variances.

We now try with two OU processes  $X^1$  and  $X^2$ . Denote by  $k_1, k_2$  their mean-reversion constants, and by  $\rho_{12}$  the correlation between the Brownian motions driving  $X^1$  and  $X^2$ . We introduce the mixing parameter  $\theta \in [0, 1]$  and denote by  $\alpha_\theta$  the normalization constant previously noted  $\alpha_w$  in (7.12) – recall that the instantaneous lognormal volatility of the instantaneous variance  $\xi_t^t$  is equal to  $2\nu$ :

$$d\xi_t^T = (2\nu)\xi_t^T \alpha_\theta \left( (1-\theta) e^{-k_1(T-t)} dW_t^1 + \theta e^{-k_2(T-t)} dW_t^2 \right) \quad (7.28)$$

$$\alpha_\theta = 1/\sqrt{(1-\theta)^2 + \theta^2 + 2\rho_{12}\theta(1-\theta)} \quad (7.29)$$

We introduce processes  $x_t^T$  defined by:

$$x_t^T = \alpha_\theta \left[ (1-\theta) e^{-k_1(T-t)} X_t^1 + \theta e^{-k_2(T-t)} X_t^2 \right] \quad (7.30)$$

where  $X_t^1, X_t^2$  are OU processes:

$$\begin{cases} dX_t^1 = -k_1 X_t^1 dt + dW_t^1, & X_0^1 = 0 \\ dX_t^2 = -k_2 X_t^2 dt + dW_t^2, & X_0^2 = 0 \end{cases}$$

$x_t^T$  is a driftless Gaussian process:

$$dx_t^T = \alpha_\theta \left[ (1-\theta) e^{-k_1(T-t)} dW_t^1 + \theta e^{-k_2(T-t)} dW_t^2 \right] \quad (7.31a)$$

whose quadratic variation is given by:

$$\langle (dx_t^T)^2 \rangle = \eta^2 (T-t) dt \quad (7.31a)$$

$$\eta(u) = \alpha_\theta \sqrt{(1-\theta)^2 e^{-2k_1 u} + \theta^2 e^{-2k_2 u} + 2\rho_{12}\theta(1-\theta)e^{-(k_1+k_2)u}} \quad (7.31b)$$

By definition of  $\alpha_\theta$ ,  $\eta(0) = 1$ .

SDE (7.28) now simply reads:

$$d\xi_t^T = (2\nu)\xi_t^T dx_t^T \quad (7.32)$$

Its solution is:

$$\xi_t^T = \xi_0^T f^T(t, x_t^T) \quad (7.33)$$

$$f^T(t, x) = e^{\omega x - \frac{\omega^2}{2}\chi(t, T)} \quad (7.34)$$

where  $\omega = 2\nu$  and  $\chi(t, T)$  is given by:

$$\begin{aligned} \chi(t, T) &= \int_{T-t}^T \eta^2(u) du \\ &= \alpha_\theta^2 \left[ (1-\theta)^2 e^{-2k_1(T-t)} \frac{1-e^{-2k_1 t}}{2k_1} + \theta^2 e^{-2k_2(T-t)} \frac{1-e^{-2k_2 t}}{2k_2} \right. \\ &\quad \left. + 2\theta(1-\theta)\rho_{12} e^{-(k_1+k_2)(T-t)} \frac{1-e^{-(k_1+k_2)t}}{k_1+k_2} \right] \end{aligned} \quad (7.35)$$

(7.33) expresses the property that  $\xi_t^T$  has a Markov representation as a function of  $x_t^T$  – a Gaussian process. We have a Markov-functional model for  $\xi_t^T$ . The reason for introducing  $x_t^T$  will become clear further below when we consider VIX futures.

Presently, the mapping function  $f$  is just an exponential, thus forward variances are lognormally distributed, but we will use other forms for  $f$  in Section 7.7.1.

We take  $k_1 > k_2$  without loss of generality and call  $X^1$  the short factor and  $X^2$  the long factor. From (7.21):

$$\begin{aligned} \frac{d\hat{\sigma}_T}{\hat{\sigma}_T} &= \nu\alpha_\theta \left( (1-\theta) \frac{\int_t^T \xi_t^\tau e^{-k_1(\tau-t)} d\tau}{\int_t^T \xi_t^\tau d\tau} dW_t^1 + \theta \frac{\int_t^T \xi_t^\tau e^{-k_2(\tau-t)} d\tau}{\int_t^T \xi_t^\tau d\tau} dW_t^2 \right) + \bullet dt \\ &= \nu\alpha_\theta \left( (1-\theta)A_1 dW_t^1 + \theta A_2 dW_t^2 \right) + \bullet dt \end{aligned} \quad (7.36)$$

with  $A_i$  given by:

$$A_i = \frac{\int_t^T \xi_t^\tau e^{-k_i(\tau-t)} d\tau}{\int_t^T \xi_t^\tau d\tau} \quad (7.38)$$

The instantaneous volatility of a VS volatility  $\nu_T(t)$  is given by:

$$\nu_T(t) = \nu\alpha_\theta \sqrt{(1-\theta)^2 A_1^2 + \theta^2 A_2^2 + 2\rho_{12}\theta(1-\theta)A_1 A_2} \quad (7.39)$$

For a flat term-structure of VS volatilities:

$$A_i = I(k_i(T-t)) = \frac{1 - e^{-k_i(T-t)}}{k_i(T-t)}$$

### 7.4.1 Term structure of volatilities of volatilities

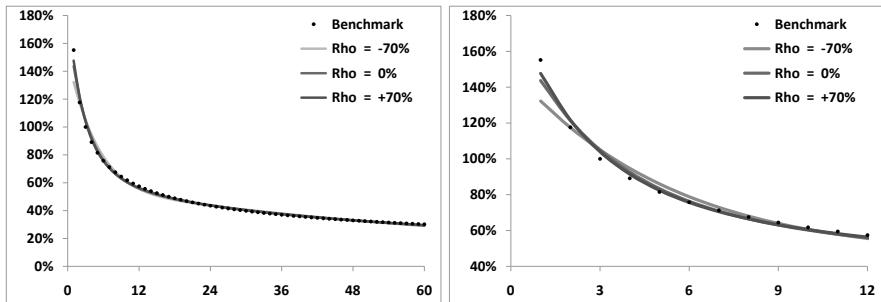
How flexible is a two-factor model? As illustrated in Figure 6.1, for equity indexes, volatilities of VS volatilities usually display a power-law dependence on maturity, with an exponent that typically lies between 0.3 and 0.6.

In the sequel we will make frequent use of the following time-homogeneous benchmark form for  $\nu_T(t)$ :

$$\nu_T^B(t) = \sigma_0 \left( \frac{\tau_0}{T-t} \right)^\alpha \quad (7.40)$$

where  $\tau_0$  is a reference maturity and  $\sigma_0$  is the volatility of  $\widehat{\sigma}_{t+\tau_0}(t)$ . Typically we will take  $\alpha = 0.4$ ,  $\tau_0 = 3$  months and  $\sigma_0 = 100\%$ . Figure 6.1 shows that the realized volatility of a 3-month VS volatility is around 60% for the Euro Stoxx 50 index. Implied levels for  $\sigma_0$  derived from prices of options on realized variance are about twice as large – hence our choice for  $\sigma_0$ .

Figure 7.1 shows  $\nu_T^B(t)$ , as well as expression (7.39) for  $\nu_T(t)$  generated by a two-factor model for a flat term structure of VS volatilities, at  $t = 0$ . We have chosen three different sets of parameters, differentiated by the value of the correlation between processes  $X^1$  and  $X^2$ . We have used  $\rho_{12} = -70\%$ ,  $0\%$ ,  $70\%$  and have selected the remaining parameters  $\nu, \theta, k_1, k_2$  so as to best match our benchmark (7.40) for maturities from one month to 5 years.



**Figure 7.1:** The left-hand graph displays the term structure of instantaneous volatilities at  $t = 0$  of VS volatilities  $\nu_T(t)$  ( $y$  axis) as a function of  $T$  ( $x$  axis, in months) generated by the benchmark form (7.40) as well as the two-factor model, with the different sets of parameters listed in Table 7.1. The right-hand graph focuses on maturities less than 1 year.

As is clear from Figure 7.1, the two-factor model is able to capture a power-law dependence for volatilities of volatilities over a wide range of maturities – similarly good agreement is achieved for other values of  $\alpha$ . Moreover, for a given  $\alpha$ , many different sets of parameters exist that provide an equally acceptable fit to our benchmark  $\nu_T^B(t)$ . Table 7.1 displays the parameters used.

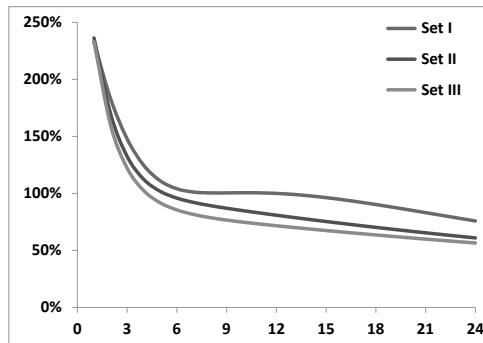
	$\nu$	$\theta$	$k_1$	$k_2$	$\rho_{12}$
Set I	150%	0.312	2.63	0.42	-70%
Set II	174%	0.245	5.35	0.28	0%
Set III	186%	0.230	7.54	0.24	70%

**Table 7.1:** Three sets of parameters matching  $\nu_T^B(t)$  in (7.40) with  $\sigma_0 = 100\%$ ,  $\tau_0 = 0.25$ ,  $\alpha = 0.4$ , for maturities up to 5 years. The resulting term structures of volatility of volatility are shown in Figure 7.1.  $\nu$  is the instantaneous (lognormal) volatility of a VS volatility of vanishing maturity.

Notice how the time scales of the OU processes  $1/k_1$ ,  $1/k_2$  are clearly separated, thus generating a volatility-of-volatility term structure that cannot be captured in a one-factor model. Figure 7.1 demonstrates that very similar term structures of instantaneous volatilities of spot-starting VS volatilities are obtained in sets I, II, III, employing very different time scales.

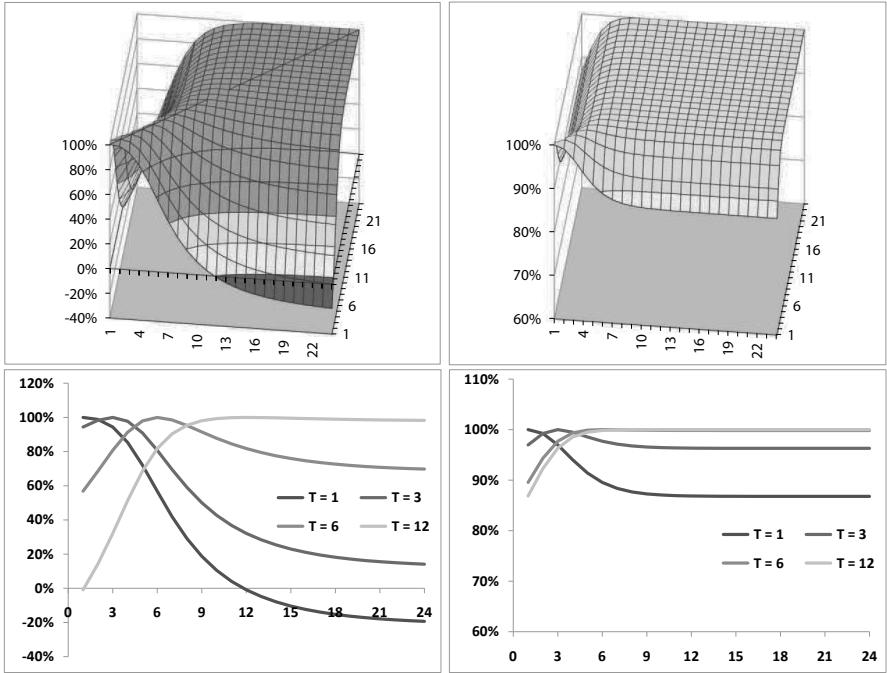
#### 7.4.2 Volatilities and correlations of forward variances

What distinguishes parameter sets corresponding to the same value of  $\alpha$ ? VS variances are equally weighted baskets of forward variances:  $\hat{\sigma}_T^2 = \frac{1}{T} \int_0^T \xi^\tau d\tau$ . It is instructive to look at the volatilities and correlations of the forward variances themselves. Figure 7.2 shows the instantaneous volatilities at  $t = 0$  of forward variances (not volatilities) using the three sets in Table 7.1, while Figure 7.3 displays  $\rho(\xi_t^T, \xi_{t'}^{T'})$ . We have used expressions (7.19) and (7.20), specialized to the case of the two-factor model.



**Figure 7.2:** Instantaneous volatilities of forward variances  $\xi_t^T$  at  $t = 0$  as a function of  $T$  (in months), using parameter sets in Table 7.1.

While volatilities of forward variances are higher in Set I than in Set III, the opposite is true of correlations. This is natural since the volatilities of VS variances –



**Figure 7.3:** Top:  $\rho(\xi_t^T, \xi_t^{T'})$  at  $t = 0$  as a function of  $T$ ,  $T'$  (in months) in Set I (left) and Set III (right). Bottom: slices of  $\rho(\xi_t^T, \xi_t^{T'})$  for  $T' = 1$  month, 3 months, 6 months, 12 months, in Set I (left) and Set III (right).

which are baskets of forward variances – are almost identical in both sets as they have been calibrated to the same benchmark: the higher the volatilities of the basket components, the lower their correlations. Observe at the top of Figure 7.3 how  $\rho(\xi_t^T, \xi_t^{T'})$  becomes almost constant, equal to one, for  $T$ ,  $T'$  larger than a given threshold, especially for Set III.

Inspection of expression (7.20) for  $\rho(\xi_t^T, \xi_t^{T'})$  shows that correlations are unchanged if all  $k_i$  are shifted by the same constant: the relevant time scales for correlations are not the  $\frac{1}{k_i}$ , but the quantities  $\frac{1}{k_i - k_j}$ . In a two-factor model  $\rho(\xi_t^T, \xi_t^{T'})$  is thus only a function of  $k_1 - k_2$ : the correlation structure has a single time scale  $\frac{1}{k_1 - k_2}$ .

In Set III, the values of  $k_1$ ,  $k_2$  are, respectively 7.54 and 0.24. For  $T - t \gg \frac{1}{k_1 - k_2} = 1.64$  months, the contribution of the short factor is negligible and the  $\xi_t^T$  behave as in a one-factor model, with 100% correlations among themselves. Their correlations with variances  $\xi_t^{T'}$  with  $T' - t \gg \frac{1}{k_1 - k_2}$  do not depend on  $T$  anymore – this is clearly seen in the slices of  $\rho(\xi_t^T, \xi_t^{T'})$  for  $T - t = 6$  months and  $T - t = 12$

months in the right-hand graph at the bottom of Figure 7.3. Also note that, while long-dated variances are driven by process  $X^2$ , short-dated variances are driven by the linear combination  $\theta X^1 + (1 - \theta) X^2$ : even with  $\rho_{12} = 0$ , there is a fair amount of correlation between short- and long-dated variances; in this respect Set II is more akin to Set III than to Set I.

### 7.4.3 Smile of VS volatilities

We have chosen to model instantaneous forward variances  $\xi^T$  as lognormal processes, based on historical evidence that VS volatilities are lognormal rather than normal – see Figure 6.2. VS variances  $\hat{\sigma}_T^2$ , which are baskets of the  $\xi^T$ , will not be exactly lognormal and neither will VS volatilities  $\hat{\sigma}_T$ . Their non-lognormality can be assessed by pricing variance swaptions, that is options to enter at  $T_1$  into a long position in a VS of maturity  $T_2$  with a strike  $K$ . The payoff of such a VS at  $T_2$  is  $(\sigma_r^2 - K)$ , where  $\sigma_r$  is the realized volatility over  $[T_1, T_2]$ .

The option is exercised at  $T_1$  only if the forward VS volatility  $\hat{\sigma}_{T_1 T_2}$  observed at  $T_1$  is larger than  $\sqrt{K}$ : we exercise the swaption and sell a VS struck at the market implied VS volatility  $\hat{\sigma}_{T_1 T_2}(T_1)$ . The payout of this strategy at  $T_2$  is:<sup>6</sup>

$$(\sigma_r^2 - K) - (\sigma_r^2 - \hat{\sigma}_{T_1 T_2}^2(T_1)) = \hat{\sigma}_{T_1 T_2}^2(T_1) - K$$

The underlying of the VS swaption is thus the *forward* VS volatility  $\hat{\sigma}_{T_1 T_2}$  and the VS swaption is a call option of maturity  $T_1$  on its square.<sup>7</sup> Expressing  $\hat{\sigma}_{T_1 T_2}^2(T_1)$  as a function of forward variances observed at  $T_1$ , the swaption payoff reads:

$$\left( \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \xi_{T_1}^u du - K \right)^+$$

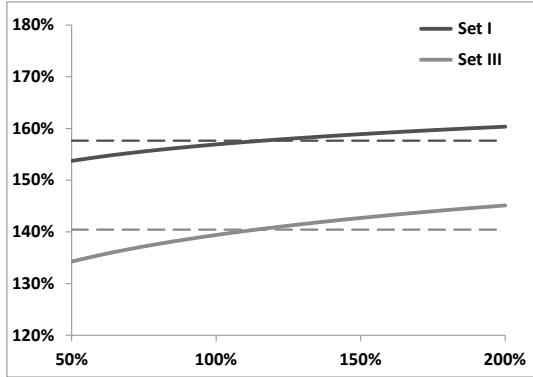
Figure 7.4 shows the smile of variance swaptions with  $T_1 = 3$  months and  $T_2 = 6$  months, in Set I and Set III, in the case of a flat term structure of VS volatilities. Implied volatilities for  $\hat{\sigma}_{T_1 T_2}^2(T_1)$  have been computed by simply inverting the Black-Scholes formula as  $\hat{\sigma}_{T_1 T_2}^2$  is driftless. In the two-factor model  $\hat{\sigma}_{T_1 T_2}(T_1)$  is a function of two Gaussian variables  $X_{T_1}^1, X_{T_1}^2$ : variance swaptions are simply priced by two-dimensional quadrature.<sup>8</sup>

While the term structure of VS volatilities is flat at  $t = 0$ , it is not at future dates. For the sake of computing the instantaneous volatility of  $\hat{\sigma}_{T_1 T_2}$  at time  $t$ , let us make the approximation that the VS term structure at time  $t$  is flat. The instantaneous volatility of  $\hat{\sigma}_{T_1 T_2}$  at  $t$  is then given by  $\nu_{T_1 T_2}(t)$  in (7.26). The instantaneous volatility of  $\hat{\sigma}_{T_1 T_2}^2(t)$  is twice as large. We then get the following strike-independent approximation of the implied volatility  $2\hat{\nu}_{T_1 T_2}(T_1)$  by integrating the square of  $\nu_{T_1 T_2}(t)$  in

<sup>6</sup>See the footnote on page 151 for the normalization of VS payoffs in actual VS term sheets.

<sup>7</sup>Note the similarity with cliquits – see Section 1.3.2.

<sup>8</sup>Since call and put payoffs are not smooth functions, one should employ for best performance a Gaussian quadrature with abscissas and weights determined for the one-sided Gaussian density.



**Figure 7.4:** Implied volatilities of variance swaptions – that is of  $\widehat{\sigma}_{T_1 T_2}^2(T_1)$  – as a function of *volatility moneyness*:  $\sqrt{K}/\widehat{\sigma}_{T_1 T_2}(t=0)$  with  $T_1 = 3$  months,  $T_2 = 6$  months in sets I and III. Dotted lines correspond to the strike-independent level  $2\widehat{\nu}_{T_1 T_2}(T_1)$ , where the integral in (7.41) has been computed numerically.

(7.26) over  $[0, T_1]$ .

$$2\widehat{\nu}_{T_1 T_2}(T_1) = 2\sqrt{\frac{1}{T_1} \int_0^{T_1} \nu_{T_1 T_2}^2(t) dt} \quad (7.41)$$

$2\widehat{\nu}_{T_1 T_2}(T_1)$  appears in Figure 7.4 as a dashed line.

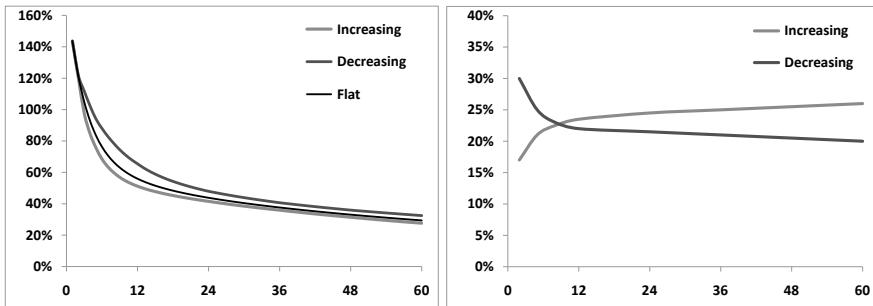
Figure 7.4 displays the weak positively sloping smile that is typical of baskets of lognormal underlyings: while not exactly lognormal, VS volatilities are close to lognormal and approximation (7.41) is fairly accurate.

We will introduce further on an extension of the model that allows for full control of the smile of forward variances.

#### 7.4.4 Non-constant term structure of VS volatilities

We now consider the effect of the shape of the term structure of VS volatilities on their instantaneous volatilities. Figure 7.5 shows the instantaneous volatilities at  $t=0$  of VS volatilities  $\widehat{\sigma}_T$  as a function of  $T$  for a positively sloping, a negatively sloping, and a flat term structure of VS volatilities.

Consider formula (7.22) for  $\nu_T(t)$ . Volatilities of VS volatilities are larger for negatively sloping  $\widehat{\sigma}_T$ : this can be understood by noting that in our model short-dated instantaneous variances  $\xi^T$  have larger volatilities than longer-dated ones. A negatively sloping term structure of VS volatilities implies that the initial values of these shorter-dated variances are larger than those of longer-dated variances. This increases their relative weight in the expression of  $\widehat{\sigma}_T^2$ , thus increasing  $\nu_T$  – see expression (7.22) for  $\nu_T(t)$ .



**Figure 7.5:** Left: instantaneous volatilities at  $t = 0$  of VS volatilities  $\nu_T$  as a function of  $T$  (months) computed in Set II for three different term structures of VS volatilities. Right: increasing and decreasing term structures of VS volatilities used in left-hand graph.

Note that simply multiplying all  $\hat{\sigma}_T$  by the same constant leaves volatilities of volatilities unchanged, as the  $\xi^T$  are lognormal.

The smiles of VS volatilities shown in Figure 7.4, computed for a flat VS curve then have an additional dependence on the slope of the term structure of VS volatilities.

#### 7.4.5 Conclusion

A two-factor model provides sufficient control on volatilities of forward variances so that the benchmark form  $\nu_T^B(t)$  in (7.40) can be matched over a wide range of maturities. Furthermore, very similar term structures of volatilities of spot-starting VS volatilities  $\hat{\sigma}_T$  can be obtained using different sets of parameters, which allows for separation of volatilities of (a) *spot-starting* VS volatilities and (b) *forward* VS volatilities.

While instantaneous volatilities of *spot-starting* VS volatilities are identical in these different sets, instantaneous volatilities and correlations of forward variances are different, hence volatilities of *forward* VS volatilities differ. This is clearly seen in the case of the 3-months in 3-months swaption implied volatilities in Figure 7.4: we get values around 160% using Set I and 140% using Set III – this disconnection of volatilities of *spot-starting* and *forward-starting* VS volatilities cannot be achieved within a one-factor model.

The correlation structure of forward variances in the two-factor model is rather poor as it is determined by a single time scale  $\frac{1}{k_1 - k_2}$ ; making it richer would be the primary motivation for introducing a third factor. Finally, while not exactly lognormal, VS volatilities are almost lognormal – Figure 7.5 highlights the fact that their volatilities will depend somewhat on the term structure of VS volatilities.

The question of which parameter set to use can only be settled on a case-by-case basis by analyzing the nature of the volatility-of-volatility risk of the payoff at hand.

In case the underlyings of our option are *spot-starting* volatilities, all sets corresponding to the same value of  $\alpha$  will yield very similar prices. If instead *forward* volatilities are the real underlyings, we will need to discriminate among sets generating equal levels of volatilities of *spot-starting* volatilities, yet different levels of volatilities and correlations of *forward* volatilities.

---

## 7.5 Calibration – the vanilla smile

What do vanilla smiles look like in the two-factor model? This question should be asked jointly with another question: which financial observables should it be calibrated to, and which instruments used as hedges?

The natural building blocks of forward variance models are (a) the spot and (b) forward variances – or VS volatilities.

Calibrating a model amounts to *deciding* which (vanilla) instruments our exotic option price is a function of, along with the spot. The consequence is these instruments are our hedge instruments.

Calibrating our forward variance model to VSs implies these are the hedge instruments we use.

Alternatively, we can calibrate the  $\xi_0^t$  so that the term structure of ATMF or ATM volatilities – or implied volatilities for an arbitrary moneyness – is recovered.<sup>9</sup> Using the corresponding vanilla options as hedges – together with the spot – leads to a well-defined gamma/theta carry P&L for a hedged position that involves covariances of implied volatilities of the calibrated instruments with each other and with the spot.<sup>10</sup>

In practice, lognormal volatilities of ATMF volatilities in the two-factor model are not much different than those of VS volatilities – thus formula (7.39) can be used to set model parameters so that desired levels of volatilities of ATMF volatilities are obtained.

Forward variance models can thus equivalently be viewed as genuine market models for the spot and the term structure of implied volatilities for a given moneyness, with the capability – as illustrated by the example of the two-factor model – of accommodating an exogenously specified dynamics for this one-dimensional set of instruments.

Once we have calibrated a term-structure of implied volatilities, say VS or ATMF, we can select the parameters of the model so as to achieve a best-fit of the whole

---

<sup>9</sup>Efficient techniques for generating vanilla smiles are surveyed in Appendix A of Chapter 8.

<sup>10</sup>Derivation of this practically relevant result is very similar to how we derive the carry P&L in the local volatility model – see Section 2.7 of Chapter 2, page 66 – or in local-stochastic volatility models – see Section 12.3 of Chapter 12, page 463.

smile. This calibration is of a very different nature than calibration to the VS/ATMF volatility term structure. In the latter case the “calibration” process amounts to inputting the values of the model’s underliers, that is ATMF or VS volatilities – this can hardly be called a calibration.

Inferring model *parameters* from market prices of vanilla options is a quite different matter. These parameters are then used to price more exotic structures. The resulting hedge ratios may be meaningless as they are likely to reflect structural relationships that are model-specific rather than express genuine matching of risks of a congruent nature.<sup>11</sup>

Characterizing the smiles generated by forward variance models and their determinants is then an important issue which is dealt with in detail in Chapter 8.

Smiles generated by the two-factor model are discussed in Section 8.7, page 326.

Efficient techniques for generating vanilla smiles in stochastic volatility models are surveyed in Appendix A of Chapter 8 – see page 336.

The relationship of the vanilla smile to the dynamics of implied volatilities is another important subject that is covered in Chapter 9 – the special case of the two-factor model is examined in Section 9.7, page 363.

## 7.6 Options on realized variance

Typically, options on realized variance comprise call and put payoffs on the realized variance of an equity underlying – usually an index – with the same convention for realized volatility as that of variance swaps. The payoff of a call on realized variance is:

$$\frac{1}{2\widehat{\sigma}_{\text{ref}}} (\sigma_r^2(T) - \widehat{\sigma}^2)^+ \quad \text{with} \quad \sigma_r^2(T) = \frac{252}{N} \sum_{i=0}^{N-1} \ln^2 \left( \frac{S_{i+1}}{S_i} \right)$$

where  $\widehat{\sigma}$  is the (volatility) strike,  $\sigma_r(T)$  is the realized volatility over the option’s maturity, and  $\widehat{\sigma}_{\text{ref}}$  is usually chosen equal to the VS volatility  $\widehat{\sigma}_T$  for the option’s maturity. The reason for this normalization is that, for an at-the-money option,  $\widehat{\sigma} = \widehat{\sigma}_T$  and at order one in  $\sigma_r(T) - \widehat{\sigma}_T$ , the payoff of the ATM call on variance is simply  $(\sigma_r(T) - \widehat{\sigma}_T)^+$ .

Volatility swaps, whose payoff is:

$$\sigma_r(T) - \widehat{\sigma}$$

trade as well; calls and puts on realized volatility trade occasionally.

---

<sup>11</sup>Think for example of volatility of volatility in the Heston model. This parameter drives both the curvature of the vanilla smile near the money and the volatility of VS volatilities, which affects prices of options on realized variance.

A call option on realized variance can be promptly priced in any stochastic volatility model: this produces a number that acquires the status of a price once we have identified the main risks of our option, the instruments that can be used to hedge them, and appropriate carry levels for the residual risks. To this end we now develop a simple model (SM) that is sufficiently robust that it can be used practically. The path we follow is similar to how one derives a well-known approximation for Asian options – see footnote 14 below.<sup>12</sup>

### 7.6.1 A simple model (SM)

We drop for now the  $\frac{1}{2\sigma_{\text{ref}}}$  factor without loss of generality and focus on the payoff  $(\sigma_r^2 - \hat{\sigma}^2)^+$  – we will also take vanishing interest rate and repo for simplicity. Let us use the notation  $Q_t$  to denote the quadratic variation defined as the sum of daily squared log-returns over the interval  $[0, t]$ :

$$Q_t = \sum_0^t \ln^2 \left( \frac{S_{i+1}}{S_i} \right) = t\sigma_r^2(t)$$

where  $\sigma_r(t)$  is the realized volatility over  $[0, t]$ . The option's payoff is  $(\frac{1}{T}Q_T - \hat{\sigma}^2)^+$ . This suggests that the underlying of our option is  $Q_t$  – can it be replicated?

The most natural candidate for hedging  $Q$  is a VS: consider entering a VS contract of maturity  $T$  at time  $t$ , struck at an implied volatility  $\hat{\sigma}_T(t)$ , whose payoff at  $T$  is:

$$\frac{1}{T-t} \sum r_i^2 - \hat{\sigma}_T^2(t), \quad r_i = \ln \left( \frac{S_{i+1}}{S_i} \right)$$

where the sum runs over all returns over the interval  $[t, T]$ .

Imagine entering this VS contract at time  $t$  – at no cost – and consider its value at some later time  $t'$ . While at  $t$  the market value of the realized quadratic variation over  $[t, T]$  was  $(T-t)\hat{\sigma}_T^2(t)$ , at  $t'$ , the quadratic variation over  $[t, t']$  has already been realized and is equal to  $Q_{t'} - Q_t$ , while the market value of the quadratic variation over  $[t', T]$  is  $(T-t')\hat{\sigma}_T^2(t')$ . The P&L over  $[t, t']$  of our VS position is thus:

$$\begin{aligned} P\&L &= \frac{1}{T-t} [(Q_{t'} - Q_t) + (T-t')\hat{\sigma}_T^2(t')] - \hat{\sigma}_T^2(t) \\ &= \frac{1}{T-t} (Q_{t'} + (T-t')\hat{\sigma}_T^2(t')) - \frac{1}{T-t} (Q_t + (T-t)\hat{\sigma}_T^2(t)) \\ &= \frac{T}{T-t} \left[ \frac{Q_{t'} + (T-t')\hat{\sigma}_T^2(t')}{T} - \frac{Q_t + (T-t)\hat{\sigma}_T^2(t)}{T} \right] \end{aligned} \quad (7.42)$$

The above equation expresses the property that taking at time  $t$  a position in  $\frac{T-t}{T}$  variance swaps of maturity  $T$  perfectly hedges the “underlying”  $U$  defined by:

$$U_t = \frac{Q_t + (T-t)\hat{\sigma}_T^2(t)}{T} \quad (7.43)$$

---

<sup>12</sup> A presentation based on this same idea was given by Zhenyu Duanmu in 2004 – see [38].

over any finite interval  $[t, t']$  as, from (7.42), the P&L of such a position is simply  $P\&L = (U_{t'} - U_t)$ .

The conclusion is that, while  $Q_t$  itself is not hedgeable, the package consisting of  $Q_t + (T-t) \hat{\sigma}_T^2(t)$  can be exactly replicated:  $U_t$  is a legitimate underlying. Moreover the pricing drift of  $U$  is zero, since taking a VS position involves no cash outlay. This is consistent with the expression of  $U$  in terms of forward variances:

$$U_t = \frac{1}{T} \left( \int_0^t \xi_\tau^\tau d\tau + \int_t^T \xi_t^\tau d\tau \right) \quad (7.44)$$

where the first piece in the right-hand side of (7.44) corresponds to past observations. As the  $\xi^\tau$  are driftless, so is  $U$ . From definition (7.43), the values of  $U$  at  $t = 0$  and  $t = T$  are, respectively:

$$U_0 = \hat{\sigma}_T^2(0) \quad U_T = \frac{1}{T} Q_T$$

The payoff of the call on realized variance is then simply:

$$(U_T - \hat{\sigma}^2)^+$$

and the price  $P$  of this option can be expressed as the expectation of the payoff  $(U_T - \hat{\sigma}^2)^+$  under a dynamics of  $U_t$  that is driftless, with the initial condition  $U_0 = \hat{\sigma}_T^2(0)$ .

### A dynamics for $U_t$

From (7.43) we have:

$$dU_t = \frac{T-t}{T} \underbrace{d(\hat{\sigma}_T^2(t))}_{\text{Diffusive portion}} \quad (7.45)$$

where we are only keeping the diffusive portion of  $d(\hat{\sigma}_T^2(t))$  as  $U_t$  is driftless. Let us assume that  $\hat{\sigma}_T(t)$  is lognormal with deterministic volatility  $\nu_T(t)$ :  $\hat{\sigma}_T^2(t)$  is lognormal with volatility  $2\nu_T(t)$ . We get the following SDE for  $U_t$ :

$$dU_t = 2 \frac{T-t}{T} \nu_T(t) \hat{\sigma}_T^2(t) dW_t$$

which we rewrite as:

$$\frac{dU_t}{U_t} = 2R_t \frac{T-t}{T} \nu_T(t) dW_t \quad (7.46)$$

with  $R_t$  given by:

$$R_t = \frac{\hat{\sigma}_T^2(t)}{U_t} = \frac{T\hat{\sigma}_T^2(t)}{Q_t + (T-t)\hat{\sigma}_T^2(t)} = \frac{T\hat{\sigma}_T^2(t)}{t\sigma_r^2(t) + (T-t)\hat{\sigma}_T^2(t)} \quad (7.47)$$

SDE (7.46) is problematic as the dynamics of  $U_t$  is not autonomous – it involves  $\hat{\sigma}_T^2(t)$  through  $R_t$ . But for this prefactor, the dynamics of  $U_t$  would be lognormal. The quadratic variation of  $\ln U$  over  $[t, T]$  is:

$$\int_t^T 4R_\tau^2 \left( \frac{T-\tau}{T} \right)^2 \nu_T^2(\tau) d\tau \quad (7.48)$$

$R_\tau$  is the ratio of two integrated variances, thus is a number of order one.

We now make an approximation: let us replace  $R_\tau$  with a constant – we approximate  $R_\tau$  by its value at time  $t$ , the pricing date:

$$R_\tau \simeq R_t$$

SDE (7.46) now becomes:

$$\frac{dU_\tau}{U_\tau} = 2R_t \frac{T-\tau}{T} \nu_T(\tau) dW_\tau$$

where  $t$  is now the pricing date.  $U$  is then lognormal and the option price  $P(t, U)$  is given by a Black-Scholes formula evaluated with an effective volatility  $\sigma_{\text{eff}}$ :

$$P(t, U) = P_{\text{BS}}(t, U, \sigma_{\text{eff}}, T) \quad (7.49a)$$

$$\sigma_{\text{eff}}^2 = \frac{1}{T-t} \int_t^T 4R_t^2 \left( \frac{T-\tau}{T} \right)^2 \nu_T^2(\tau) d\tau \quad (7.49b)$$

The number of VS contracts of maturity  $T$  needed at time  $t$  to hedge our option is:<sup>13</sup>

$$\frac{T-t}{T} \frac{dP}{dU}$$

where the ratio  $\frac{T-t}{T}$  expresses the fact that it takes  $\frac{T-t}{T}$  VS contracts of maturity  $T$  at time  $t$  to replicate  $U$ .

At inception  $U_{t=0} = \hat{\sigma}_T^2(0)$ , so that  $R_{t=0} = 1$ . The option premium is then given by:<sup>14</sup>

---

<sup>13</sup> Actual VS term-sheets specify the VS payoff as  $\frac{1}{2\hat{\sigma}_T} (\sigma_r^2 - \hat{\sigma}_T^2)$ . We are not using the  $1/(2\hat{\sigma}_T)$  prefactor for now.

<sup>14</sup> A similar route can be followed to derive a well-known approximation for the price of an Asian option in the Black-Scholes model, that is an option that pays at time  $T$  a European payoff on  $\frac{1}{T} \int_0^T S_t dt$ . Assume zero interest rate/repo for simplicity and call  $M_t$  the running average at time  $t$ :  $M_t = \frac{1}{t} \int_0^t S_\tau d\tau$ . While  $M_t$  cannot be exactly hedged by taking a position on  $S$ , the package  $U_t$  defined by  $U_t = \frac{tM_t + (T-t)S_t}{T}$  is exactly replicated over  $[t, t+dt]$  by trading  $\frac{T-t}{T}$  units of  $S$ :  $U_t$  is driftless. The initial and final values of  $U$  are:  $U_0 = S_0$ ,  $U_T = M_T$ . The SDE for  $U_t$  is:  $\frac{dU_t}{U_t} = R_t \frac{T-t}{T} \sigma dW_t$  with  $R_t = \frac{S_t}{U_t}$ , where  $\sigma$  is the volatility of  $S$ . Let us make the approximation  $\frac{S_t}{U_t} = 1$ .  $U_t$  is then lognormal with an effective volatility over  $[0, T]$  given by:  $\sigma_{\text{eff}}^2 = \frac{1}{T} \int_0^T \left( \frac{T-t}{T} \right)^2 \sigma^2 dt = \frac{\sigma^2}{3}$ . The option price is then simply given by the Black-Scholes formula  $P_{\text{BS}}(t=0, S_0, \sigma_{\text{eff}}, T)$ , with  $\sigma_{\text{eff}} = \sqrt{\frac{\sigma^2}{3}}$ .

$$P(t=0) = P_{\text{BS}}(t=0, \hat{\sigma}_T^2(0), \sigma_{\text{eff}}, T) \quad (7.50a)$$

$$\sigma_{\text{eff}}^2 = \frac{1}{T} \int_0^T 4 \left( \frac{T-\tau}{T} \right)^2 \nu_T^2(\tau) d\tau \quad (7.50b)$$

These are the basic equations of the SM.

### 7.6.2 Preliminary conclusion

So far, our main results are, assuming the approximation  $R_\tau \equiv R_t$  is satisfactory – this will be checked further on:

- Options on realized variance are simply hedged by dynamically trading variance swaps of the option's residual maturity.
- Their value does not depend on the detailed dynamics of instantaneous forward variances – it only depends on the instantaneous volatility  $\nu_T(t)$  of  $\hat{\sigma}_T(t)$ , for all times  $t \in [0, T]$ : the only pricing ingredient is the curve  $\nu_T(t)$ , such as those in Figure 7.1.
- The SM is a legitimate, arbitrage-free, model. The only approximation we are making is in the volatility of  $U_t$  – hence in the value of  $\sigma_{\text{eff}}$ . We have replaced  $R_\tau$  with its value at the pricing date,  $R_t$  so that  $U_t$  is lognormal.

In the following section we will assess the accuracy of the SM using various forms for  $\nu_T(t)$ .

### 7.6.3 Examples

Our conclusion suggests that options on realized variance are in fact simple instruments as their only volatility-of-volatility risk is the exposure to the realized volatility over  $[0, T]$  of the VS volatility of maturity  $T$ .

We now check this prediction by pricing an at-the-money call on realized variance of maturity 6 months and 1 year in the two-factor model and in the SM.

$\nu_T(\tau)$  in formula (7.50b) for  $\sigma_{\text{eff}}$  is the instantaneous volatility of the VS volatility for maturity  $T$  observed at  $t$ ,  $\hat{\sigma}_T(t)$ . In our benchmark (7.40),  $\nu_T(\tau)$  is specified directly:

$$\nu_T^B(t) = \sigma_0 \left( \frac{\tau_0}{T-t} \right)^\alpha \quad (7.51)$$

In the two-factor model, instead, the model generates its own dynamics for  $\hat{\sigma}_T(t)$ . Set II parameters are such that, for a flat VS term structure, instantaneous volatilities of VS volatilities at  $t = 0$  – given by (7.39) – best match the benchmark form (7.51). What about instantaneous volatilities of VS volatilities at *future* times  $t$  – which enter expression (7.50b) for  $\sigma_{\text{eff}}$ ?

In expression (7.50b) for  $\sigma_{\text{eff}}$  we use the following expression for  $\nu_T(t)$ , which we denote  $\nu_T^0(t)$ :

$$\nu_T^0(t) = \nu \alpha_\theta \sqrt{(1 - \theta)^2 f_1^0(t, T)^2 + \theta^2 f_2^0(t, T)^2 + 2\rho_{12}\theta(1 - \theta) f_1^0(t, T) f_2^0(t, T)} \quad (7.52)$$

with  $f_i^0(t, T)$  given by:

$$f_i^0(t, T) = \frac{\int_t^T \xi_0^\tau e^{-k_i(\tau-t)} d\tau}{\int_t^T \xi_0^\tau d\tau}$$

The reader can check – see expression (7.22) – that the actual expression of the instantaneous volatility of  $\hat{\sigma}_T(t)$  at time  $t$ ,  $\nu_T(t)$ , specialized to the two-factor model, is identical to (7.52), except,  $f_i(t, T)$  is used in place of  $f_i^0(t, T)$ :

$$f_i(t, T) = \frac{\int_t^T \xi_t^\tau e^{-k_i(\tau-t)} d\tau}{\int_t^T \xi_t^\tau d\tau}$$

Using the time-deterministic form (7.52) instead of  $\nu_T(t)$  – a process – amounts to making the assumption that at time  $t$  forward variances  $\xi_t^\tau$  are equal to their initial values  $\xi_0^\tau$ . For a flat term structure of VS volatilities,  $\nu_T^0(t) = \nu_T(t)$  as given by expression (7.39), page 227.

Even though the VS term structure is flat at  $t = 0$ , its shape at future times in the two-factor model will not be flat: the instantaneous volatility of  $\hat{\sigma}_T(t)$  will depend on the VS term structure prevailing at time  $t$  and may differ from  $\nu_T^0(t)$ .

How good is the SM with  $\sigma_{\text{eff}}$  in (7.50) computed with  $\nu_T^0(t)$ , compared with prices produced by a Monte Carlo simulation of the two-factor model?

Option prices appear in Table 7.2. We have used a flat initial term structure of VS volatilities at 20% and have followed the standard market practice for the normalization of the call payoff:

$$\frac{1}{2\hat{\sigma}_T} (\sigma_r^2 - \hat{\sigma}^2)^+$$

Table 7.2 also lists prices produced by (7.50), with  $\nu_T(t) = \nu_T^B(t)$ . In the latter case  $\sigma_{\text{eff}}$  is given by:

$$\sigma_{\text{eff}} = \frac{2\sigma_0}{\sqrt{3 - 2\alpha}} \left( \frac{\tau_0}{T} \right)^\alpha \quad (7.53)$$

Observe how close prices generated by the SM with  $\nu_T(t) = \nu_T^0(t)$  are to exact prices. Also note how prices generated with parameter sets I, II, III almost match, and how they almost match prices produced by the SM with  $\nu_T(t) \simeq \nu_T^B(t)$ . This confirms that, indeed, the primary risk of options on realized volatility is the exposure to the realized volatility of the VS volatility for the residual maturity.

		Set I	Set II	Set III	Benchmark
6 months	exact	2.97%	2.96%	2.94%	
	SM	2.93%	2.88%	2.86%	2.82%
1 year	exact	3.13%	3.08%	3.06%	
	SM	3.02%	2.99%	2.98%	3.01%

**Table 7.2:** Prices of an ATM call option on realized variance computed in the two-factor model with parameter sets in Table 7.1, page 229, either in a Monte Carlo simulation (exact), or using the SM with  $\nu_T^0(t)$  given by (7.52) (SM). Prices in the last column are computed using the SM with  $\nu_T(t)$  given by benchmark (7.51) with  $\sigma_0 = 100\%$ ,  $\tau_0 = 0.25$ ,  $\alpha = 0.4$ . The term structure of VS volatilities is flat at 20%.

Finally, notice how prices very weakly depend on maturity. This is due to the fact that the value of  $\alpha$  in our benchmark, 0.4, is close to 0.5. For  $\alpha = 0.5$ ,  $\sigma_{\text{eff}}$  in (7.53) is inversely proportional to  $\sqrt{T}$ ,  $\sigma_{\text{eff}}\sqrt{T}$  is constant and the price given by (7.50a) does not depend on  $T$ .

#### 7.6.4 Accounting for the term structure of VS volatilities

So far, in the derivation of the SM, we have made the assumption  $R_\tau \equiv R_t$ . At inception  $R_0 = 1$ . However, (7.47) shows that  $R_\tau$  will be different than 1 whenever the realized volatility over  $[0, \tau]$ ,  $\sigma_r(\tau)$ , is different than  $\hat{\sigma}_T(\tau)$ .

Prices in Table 7.2 have been computed with flat VS volatilities equal to 20%. Within the two-factor model, while  $\sigma_r(\tau)$  and  $\hat{\sigma}_T(\tau)$  are random, the expectation of their squares is equal to 20%<sup>2</sup>, thus the expectations of the numerator and denominator of  $R_\tau$  within the model are equal, and approximation  $R_\tau \equiv R_0 = 1$  is appropriate. The good agreement of exact and SM prices in Table 7.2 shows that the fluctuation of  $R_\tau$  around 1 has little impact on the option's premium.

Obviously, from a pricing point of view, the assumption  $R_\tau \equiv 1$  will need to be corrected for non-constant term structures of VS volatilities. The dependence on the term structure of VS volatilities that the option price then acquires will need to be hedged, however.

In fact, regardless of the shape of the term structure of VS volatilities – and even for flat ones – the realized volatility, hence  $Q_t$ , will in practice be whatever it wants to be and realized values of  $R_\tau$  will be substantially different than 1.

For example, imagine that, in reality, realized volatilities are systematically lower than implied VS volatilities – which is usually the case for indexes:  $Q_t/t < \hat{\sigma}_T^2(t)$ .  $R_t$  will be systematically larger than 1 and (7.46) indicates that the realized volatility of  $U$  will be larger than the level we have priced. If we have sold a call on realized variance, our daily gamma/theta P&L on  $U$  will be negative, thus we will lose money steadily even though the instantaneous realized volatility at time  $\tau$  of  $\hat{\sigma}_T$  matches

our pricing level  $\nu_T(\tau)$ . Can we hedge this exposure of  $R_\tau$  to the realized volatility of  $S_t$ ?

As we will see shortly, the issues of (a) accounting for the term structure of VS volatilities in the option price, and (b) hedging the exposure of  $R_\tau$  to the realized volatility of the underlying are connected.

Let us first amend (7.49b) to take into account the term structure of VS volatilities. For general term structures the expectations at time  $t$  of the numerator and denominator of  $R_\tau$  in (7.47) are:

$$\begin{aligned} E_t[U_\tau] &= U_t \equiv \frac{Q_t + (T-t)\hat{\sigma}_T^2(t)}{T} \\ E_t[\hat{\sigma}_T^2(\tau)] &= \frac{1}{T-\tau} \int_\tau^T \xi_t^u u = \hat{\sigma}_{\tau T}^2(t) \end{aligned}$$

where  $\hat{\sigma}_{\tau T}(t)$  is the forward VS volatility at time  $t$  for the interval  $[\tau, T]$ . We now make the approximation:

$$R_\tau \simeq \frac{E_t[\hat{\sigma}_T^2(\tau)]}{E_t[U_\tau]} = \frac{T\hat{\sigma}_{\tau T}^2(t)}{Q_t + (T-t)\hat{\sigma}_T^2(t)} \quad (7.54)$$

Formula (7.49b) for  $\sigma_{\text{eff}}$  now becomes:

$$\sigma_{\text{eff}}^2 = \frac{4}{T-t} \int_t^T \left( \frac{T-\tau}{T} \right)^2 \left( \frac{T\hat{\sigma}_{\tau T}^2(t)}{Q_t + (T-t)\hat{\sigma}_T^2(t)} \right)^2 \nu_T^2(\tau) d\tau \quad (7.55)$$

At time  $t = 0$ ,  $Q_0 = 0$ ,  $\sigma_{\text{eff}}$  and the option's price are given by:

$$\sigma_{\text{eff}}^2 = \frac{4}{T} \int_0^T \left( \frac{T-\tau}{T} \right)^2 \left( \frac{\hat{\sigma}_{\tau T}^2(0)}{\hat{\sigma}_T^2(0)} \right)^2 \nu_T^2(\tau) d\tau \quad (7.56a)$$

$$P(t=0) = P_{\text{BS}}(t=0, \hat{\sigma}_T^2(0), \sigma_{\text{eff}}, T) \quad (7.56b)$$

In what follows we use this new expression for  $\sigma_{\text{eff}}$  instead of (7.50b).

$\sigma_{\text{eff}}$  now depends on the full term structure of VS volatilities up to  $T$ , with the consequence that the option price acquires an exposure to intermediate VS volatilities which it is necessary to hedge.

### 7.6.5 Vega and gamma hedges

The continuous density  $\lambda(\tau)$  of intermediate VSs is given by the functional derivative of  $P$  with respect to  $\hat{\sigma}_\tau^2$ . Writing  $\hat{\sigma}_{\tau T}^2(t)$  as:

$$\hat{\sigma}_{\tau T}^2(t) = \frac{(T-t)\hat{\sigma}_T^2(t) - (\tau-t)\hat{\sigma}_\tau^2(t)}{T-\tau}$$

(7.55) can be rewritten as:

$$\sigma_{\text{eff}}^2 = \frac{4}{T-t} \int_t^T \left( \frac{(T-t)\hat{\sigma}_T^2(t) - (\tau-t)\hat{\sigma}_\tau^2(t)}{Q_t + (T-t)\hat{\sigma}_T^2(t)} \right)^2 \nu_T^2(\tau) d\tau \quad (7.57)$$

$\lambda(\tau)$  is given by:

$$\begin{aligned}\lambda(\tau) &= \left. \frac{\delta P}{\delta \hat{\sigma}_\tau^2} \right|_t \\ &= -8 \frac{dP}{d\sigma_{\text{eff}}^2} \frac{(\tau-t)(T-\tau)}{(T-t)T} \frac{T \hat{\sigma}_{\tau T}^2(t)}{(Q_t + (T-t) \hat{\sigma}_T^2(t))^2} \nu_T^2(\tau)\end{aligned}\quad (7.58)$$

Differentiating  $\sigma_{\text{eff}}^2$  with respect to  $\hat{\sigma}_T^2$  also produces a discrete quantity of variance swaps of maturity  $T$  which we denote by  $\mu_T$ :

$$\mu_T = \frac{dP}{d\sigma_{\text{eff}}^2} \frac{d\sigma_{\text{eff}}^2}{d\hat{\sigma}_T^2}$$

These come in addition to the VS of maturity  $T$  that offsets  $\frac{dP}{dU}$ .

While we trade these intermediate VSs as well as an additional VS of maturity  $T$  to hedge the mark-to-market P&L generated by the dependence of  $R_\tau$  – hence of  $\sigma_{\text{eff}}$  – on intermediate VS volatilities, these VSs will also generate spurious gamma/theta P&Ls.

These P&Ls are in fact offsetting similar P&Ls from the short option position. To see that this is indeed the case, rewrite expression (7.57) of  $\sigma_{\text{eff}}^2$  as:

$$\sigma_{\text{eff}}^2 = \frac{4}{T-t} \int_t^T \left( \frac{(Q_t + (T-t) \hat{\sigma}_T^2(t)) - (Q_t + (\tau-t) \hat{\sigma}_\tau^2(t))}{Q_t + (T-t) \hat{\sigma}_T^2(t)} \right)^2 \nu_T^2(\tau) d\tau \quad (7.59)$$

Recall that a position in a VS of maturity  $T$  replicates the package  $Q_t + (T-t) \hat{\sigma}_T^2(t)$  exactly – see equation (7.42). As is manifest in equation (7.59)  $\hat{\sigma}_\tau^2(t)$  does not appear alone, but associated with  $Q_t$  in such a way that the VS position that hedges the exposure to  $\hat{\sigma}_\tau^2(t)$  actually hedges  $Q_t + (\tau-t) \hat{\sigma}_\tau^2(t)$  – the same goes for the exposure to  $\hat{\sigma}_T^2(t)$ . In other words, the extra VS position that hedges the sensitivity of  $\sigma_{\text{eff}}^2$  to the VS term structure also hedges the sensitivity of  $R_\tau$  to  $Q_t$ .

The conclusion is that in the SM, the gamma/theta P&L generated by the VS position that hedges the exposure of  $\sigma_{\text{eff}}^2$  to the term structure of VS volatilities exactly offsets the gamma/theta P&L generated by the dependence of  $\sigma_{\text{eff}}^2$  on the quadratic variation  $Q_t$  and VS volatilities  $\hat{\sigma}_\tau$  for  $\tau \in [0, T]$ .

Thus our vega hedge also functions as a gamma/theta hedge. We will need to check whether this property, obtained in the SM with the help of approximation (7.54), holds more generally.

The continuous density  $\lambda(\tau)$  and discrete quantity  $\mu_T$  of hedging VSs are related by a simple equation. Expression (7.56a) shows that  $\sigma_{\text{eff}}$  is unchanged if all VS volatilities are rescaled uniformly – thus for a small change  $\delta \hat{\sigma}_T^2 = \varepsilon \hat{\sigma}_T^2$ , at order one in  $\varepsilon$ ,  $\delta \sigma_{\text{eff}}^2 = 0$ :

$$\int_0^T \lambda(\tau) \delta \hat{\sigma}_\tau^2 d\tau + \mu_T \delta \hat{\sigma}_T^2 = 0$$

which implies that

$$\int_0^T \lambda(\tau) \hat{\sigma}_\tau^2 d\tau + \mu_T \hat{\sigma}_T^2 = 0$$

For the case of a flat term structure of VS volatilities, this simplifies to:

$$\int_0^T \lambda(\tau) d\tau + \mu_T = 0 \quad (7.60)$$

The aggregate vega of the VS position that hedges  $\sigma_{\text{eff}}$  thus vanishes: the total vega of the VS hedge reduces to the vega of the VS position of maturity  $T$  that hedges  $U$ .

### 7.6.6 Examples

We now use expression (7.56a) for  $\sigma_{\text{eff}}^2$  to study the nature of the VS hedge, concentrating first on the case of a flat term structure of VS volatilities. The pricing date is January 1, 2010 and the option's maturity is January 1, 2011. We have used market conventions: the payoff at  $T$  of the ATM call on realized variance and the payoff at  $\tau$  of a VS of maturity  $\tau \in [0, T]$  are, respectively:

$$\frac{1}{2\hat{\sigma}_T} \left( \frac{Q_T}{T} - \hat{\sigma}^2 \right)^+, \quad \frac{1}{2\hat{\sigma}_\tau} \left( \frac{Q_\tau}{\tau} - \hat{\sigma}_\tau^2 \right) \quad (7.61)$$

The vega of a VS contract is thus 1.

#### With the benchmark

We first use the SM with our benchmark form for  $\nu_T^B(t)$ . We use a flat term structure of VS volatilities: while the *price* computed using either expression (7.56a) or expression (7.50b) for  $\sigma_{\text{eff}}$  is identical, the VS *hedge* is different. The VS hedge, together with the dollar gammas generated by the hedging VSs are reported in Table 7.3.

In practice, VSs do not trade for all maturities, but for maturities corresponding to monthly or quarterly expiries of listed options: we have chosen monthly expiries, including an unrealistic one-day maturity. For the sake of computing hedge ratios for the intermediate maturities  $\tau_i$  in Table 7.3 we have used a simple piecewise affine interpolation of  $\tau \hat{\sigma}_\tau^2$  over each interval  $[\tau_i, \tau_{i+1}]$ . This ensures that as  $\hat{\sigma}_{\tau_i}$  is shifted, only forward variances  $\xi_u$  for  $u \in [\tau_{i-1}, \tau_{i+1}]$  vary – in particular our variance option will have no vega on VS volatilities of maturities longer than the option's maturity.<sup>15</sup>

The rightmost column in Table 7.3 lists the dollar gammas generated by each intermediate VS. As expressed by equation (5.9) the dollar gamma of a VS – without the  $\frac{1}{2\hat{\sigma}_\tau}$  prefactor – is  $2e^{-r(T-t)}$ . Using the market convention in (7.61) and zero interest rate, the dollar gamma of a VS of maturity  $\tau$  is given by:

$$S^2 \frac{d^2 \text{VS}}{dS^2} = \frac{1}{\hat{\sigma}_\tau \tau} \quad (7.62)$$

---

<sup>15</sup>This is not guaranteed if less rustic interpolation schemes, such as splines, are used.

Maturities	SM - benchmark	
	VS hedge	VS dollar gamma
02-Jan-10	-0.01%	-13%
31-Jan-10	-0.4%	-25%
02-Mar-10	-0.8%	-25%
02-Apr-10	-1.2%	-25%
02-May-10	-1.6%	-25%
02-Jun-10	-2.0%	-24%
02-Jul-10	-2.2%	-22%
01-Aug-10	-2.6%	-22%
01-Sep-10	-2.8%	-21%
01-Oct-10	-3.0%	-20%
01-Nov-10	-3.1%	-18%
01-Dec-10	-2.9%	-16%
01-Jan-11	87.7%	438%
<b>Aggregate</b>	<b>65.1%</b>	<b>182%</b>

Single maturity		
01-Jan-11	65.1%	325%

**Table 7.3:** VS vega hedges and the corresponding dollar gammas for an ATM call on realized variance of maturity 1 year starting on January 1, 2010, computed in the SM with  $\nu_T(t)$  given by the benchmark form (7.51) with  $\sigma_0 = 100\%$ ,  $\tau_0 = 0.25$ ,  $\alpha = 0.4$ , as well as aggregate vega and dollar gamma. The term structure of VS volatilities is flat at 20%. The bottom line shows the same results when only the VS volatility for the option's maturity is used for pricing and hedging.

Table 7.3 shows that a call option on realized variance has negative sensitivities to VS volatilities for intermediate maturities. These sensitivities are identical for a call and a put struck at the same strike, as they are proportional to  $\frac{dP}{d\sigma_{\text{eff}}}$  – see (7.58).<sup>16</sup>

Recall that the VS hedge for the option's maturity consists of two pieces: one piece hedges  $U$  and acts as a delta; the other hedges  $\sigma_{\text{eff}}$  and acts as a vega: as is apparent from the bottom of Table 7.3, the former is the largest contributor to the vega of the aggregate VS hedge.

The bottom of Table 7.3 also shows the vega and dollar gamma of the VS hedge obtained if we discard intermediate maturities and price our option using only the VS volatility for the option's maturity, in which case  $\sigma_{\text{eff}}$  is given by (7.50b). Notice how the aggregate vegas are identical in both situations – which is expected from (7.60), which expresses that for a flat VS term structure the aggregate vega of the VSs that hedge  $\sigma_{\text{eff}}$  vanishes.

<sup>16</sup>Besides, a long position in a call combined with a short position in a put with matching strikes is a forward, which is perfectly hedged with a VS of maturity  $T$  and has zero sensitivity to intermediate VS volatilities.

Notice though how the aggregate gammas are very different, depending upon whether we use the full term structure of VS volatilities or not. Indeed, short-maturity VSs, despite their small vega, do contribute large dollar gammas<sup>17</sup> – this is clear from formula (7.62).

In practice VSs for such short maturities do not trade: a *long* position on a call or put on realized variance generates a *short* gamma position that must be offset with vanilla options.

### Checking the gamma in the SM

What about the gamma of the option on realized variance? We have proved in Section 7.6.4 that in the SM the vega hedge also functions as a perfect gamma hedge. In the SM, the option's dollar gamma can then be computed by summing the dollar gammas of the hedging VSs. From Table 7.3 the dollar gamma thus computed is 182%.

This is only applicable to the SM. There exists however a model-independent expression of the dollar gamma for models such that the dynamics of forward variances is independent on the spot level. This is the case for the forward variance models considered so far.

Imagine that  $S$  is shifted by  $\delta S$ : the quadratic variation is shifted from  $Q$  to  $Q + (\frac{\delta S}{S})^2$ . The dollar gamma of our option is then given by:

$$S^2 \frac{d^2 P}{dS^2} = 2 \frac{dP}{dQ}$$

We have:

$$\frac{dP}{dQ} = \frac{P(Q_0 + \delta Q) - P(Q_0)}{\delta Q}$$

for  $\delta Q$  small where  $Q_0 = 0$  is the quadratic variation at  $t = 0$ , and  $\delta Q$  is a small increment of  $Q$ . Quadratic variation is additive thus, all things being equal, a small perturbation of  $Q_0$  generated by a change of the initial spot value results in the same perturbation in  $Q_T$ .<sup>18</sup> Using convention (7.61) for the variance option payoff and denoting by  $\hat{\sigma}$  its (volatility) strike, at first order in  $\delta Q$ :

$$\begin{aligned} P(Q_0 + \delta Q, \hat{\sigma}) &= \frac{1}{2\hat{\sigma}_T} E \left[ \left( \frac{Q_T + \delta Q}{T} - \hat{\sigma}^2 \right)^+ \right] \\ &= \frac{1}{2\hat{\sigma}_T} E \left[ \left( \frac{Q_T}{T} - \left( \hat{\sigma}^2 - \frac{\delta Q}{T} \right) \right)^+ \right] = P\left(Q_0, \hat{\sigma} - \frac{\delta Q}{2\hat{\sigma}T}\right) \end{aligned}$$

The dollar gamma of the call option is thus related to the sensitivity to its (volatility) strike through:

$$S^2 \frac{d^2 P}{dS^2} = -\frac{1}{\hat{\sigma}T} \frac{dP}{d\hat{\sigma}} \quad (7.63)$$

---

<sup>17</sup>This is apparent in the contribution of the one-day maturity VS to the aggregate dollar gamma – this was the reason for including this unrealistic VS maturity in the hedge portfolio.

<sup>18</sup>This is not the case in local volatility or mixed local/stochastic volatility models: a change in the initial spot value affects both the values and volatilities of forward variances:  $Q_T$  is not simply shifted by  $\delta Q$ .

Let us stress that while (7.63) holds for the forward variance models we are studying, it does not hold in local volatility or mixed local-stochastic volatility models – more generally models such that covariances of forward variances depend on the spot level.

Using (7.63) with  $P$  given by the SM yields a dollar gamma of 175% for our 1-year ATM call. The fact that the value of 182% obtained – in the framework of the SM – by summing the dollar gammas of the VS hedge is close to this is another check of the validity of approximation (7.54) for  $R_\tau$  in the SM.

### With the two-factor model

We now use the two-factor model with Set II parameters, and compare exact VS hedges with those given by the SM with  $\nu_T(t) = \nu_T^0(t)$ . Results appear in Table 7.4. Hedge ratios computed either in a Monte Carlo simulation of the two-factor model or by using the SM are very similar.

Maturities	VS hedge		VS dollar gamma	
	Exact	SM	Exact	SM
02-Jan-10	-0.02%	-0.01%	-31%	-27%
31-Jan-10	-0.9%	-0.8%	-55%	-46%
02-Mar-10	-1.2%	-1.0%	-37%	-30%
02-Apr-10	-1.2%	-1.5%	-23%	-30%
02-May-10	-1.0%	-1.2%	-14%	-18%
02-Jun-10	-0.8%	-1.1%	-9%	-13%
02-Jul-10	-0.7%	-1.0%	-7%	-10%
01-Aug-10	-0.5%	-0.7%	-5%	-6%
01-Sep-10	-0.4%	-0.3%	-3%	-2%
01-Oct-10	-0.1%	-0.1%	0%	-1%
01-Nov-10	0.5%	1.3%	3%	8%
01-Dec-10	1.9%	2.4%	10%	13%
01-Jan-11	68.2%	68.8%	341%	344%
Aggregate	63.8%	64.9%	169%	183%

Single maturity				
01-Jan-11	63.8%	64.9%	319%	325%

**Table 7.4:** VS vega hedge for an ATM call on realized variance of maturity 1 year starting on January 1, 2010, computed with (a) a Monte Carlo simulation of the two-factor model with Set II parameters (Exact); (b) the SM with  $\nu_T^0(t)$  given by (7.52) – as well as the corresponding dollar gammas. The term structure of VS volatilities is flat at 20%. The bottom line shows the same results when only the VS volatility for the option’s maturity is used for pricing and hedging.

Comparison of Table 7.3 and Table 7.4 shows, however, that VS hedge ratios in the two-factor model or in the benchmark are different, even though the aggregate vegas are almost exactly equal: the VS hedge is distributed differently. Why is this?

Both the second column of Table 7.3 and the third column of Table 7.4 are computed using the SM except in the first case expression (7.56a) of  $\sigma_{\text{eff}}^2$  makes use of  $\nu_T^B(t)$ , while in the latter case  $\nu_T^0(t)$  in (7.52) is used.

Set II parameters are such that, with the initial flat term structure of VS volatilities,  $\nu_T^0(t) \simeq \nu_T^B(t)$ .  $\nu_T^B$  is by construction independent on the term structure of VS volatilities.

Unlike  $\nu_T^B$ ,  $\nu_T^0(t)$  depends on this term structure – see expression (7.52). In the two-factor model, the instantaneous volatility of  $\hat{\sigma}_T(t)$  depends on the term structure of VS volatilities – this is illustrated in Figure 7.5, page 233.

This sensitivity of the instantaneous volatility of  $\hat{\sigma}_T(t)$  to the VS term structure is reflected in an additional sensitivity of the option on realized variance to the VS term structure. This accounts for the difference of the VS hedges in Tables 7.3 and 7.4.

The reason why aggregate vegas in (a) the second column of Table 7.3, (b) the third column of Table 7.4 almost exactly match is that we are using a flat term structure of VS volatilities. Indeed, in a uniform rescaling of VS volatilities  $\nu_T^0(t)$  is unchanged, as can be checked on expression (7.52). With a flat term structure, a rescaling of VS volatilities is akin to a uniform shift. Thus, in a uniform shift of VS volatilities,  $\nu_T^0(t)$  is unchanged, which implies that the sum of the additional VS hedges contributed by the sensitivity of  $\nu_T^0(t)$  to the VS term structure must vanish.

Turning now to the dollar gamma, from Table 7.4 the aggregate dollar gamma of the VS hedge is 183%, when the exact dollar gamma, computed through the sensitivity to the option's strike is 166%. The SM property that the vega hedge of options on realized variance functions as a gamma hedge still holds approximately.

## Conclusion

We have pinpointed an additional element of model-dependence of options on realized variance: the dependence of volatilities of volatilities on the slope – not the level – of the VS term structure.

Depending on which assumption is made, one obtains different VS hedges – column 2 in Table 7.3 or column 3 in Table 7.4 – even though the initial price is identical. Both hedges are acceptable as they are generated by legitimate models – indeed the benchmark can be mimicked in the two-factor model, at the expense of time-homogeneity.<sup>19</sup> Making volatilities of VS volatilities independent on the term structure is a simpler assumption to work with and a book of options on realized variance can be risk-managed using the SM.

On the other hand the dependence generated by the two-factor model is not implausible – see Figure 7.5, page 233. In the two-factor model, volatilities of volatilities are larger for decreasing term structures of VS volatilities, which is what one would indeed expect. Also, the two-factor model can be used to risk-manage options

---

<sup>19</sup>One only needs to: (a) make it a one-factor model ( $\theta = 0$ ) and set  $k_1 = 0$ , (b) make  $\nu$  in (7.28) time-dependent:  $\nu \rightarrow \nu(t) = \sigma_0 \left( \frac{\tau_0}{T-t} \right)^\alpha$ . Instantaneous volatilities of  $\hat{\sigma}_T(t)$  are then exactly equal to  $\nu_T^B(t)$  and are independent on the term structure of VS volatilities. Because the maturity  $T$  is hard-wired in the model, the latter is not time-homogeneous anymore.

on both spot-starting and forward-starting variance, as well as other payoffs that have volatility-of-volatility risk.

It is then more reasonable to pick *one* parameter set for the two-factor model that matches the benchmark with given  $\sigma_0, \tau_0, \alpha$ , and use it to risk-manage all options in a book.

### 7.6.7 Non-flat VS volatilities

Here we focus on the case of sloping VS term structures. In the two-factor model, for a given parametrization, instantaneous volatilities of VS volatilities depend on the VS term structure.

To be able to measure the “intrinsic” sensitivity of the option on realized variance to the VS term structure, we now recalibrate parameters  $\nu, \theta, k_1, k_2$  of the two-factor model so that, for each term structure at hand,  $\nu_T^0(t) \simeq \nu_T^B(t)$ , where  $\nu_T^0(t)$  is given by (7.52). For  $\nu_T^B(t)$ , we take  $\sigma_0 = 100\%$ ,  $\tau_0 = 0.25$ ,  $\alpha = 0.4$ .

Table 7.5 lists prices of an ATM call on realized variance of maturities 6 months and 1 year in the two-factor model, for two term structures of VS volatilities – respectively decreasing and increasing – shown in Figure 7.6.

	1 year		6 months		Parameters				
	exact	approx	exact	approx	$\nu$	$\theta$	$k_1$	$k_2$	$\rho_{12}$
Decreasing	3.46%	3.32%	4.28%	4.11%	2.27	0.30	12.8	1.06	0%
Increasing	4.46%	4.36%	3.03%	2.95%	2.22	0.29	12.0	0.96	0%

**Table 7.5:** Prices of an ATM call on realized variance computed in the two-factor model for the case of a decreasing (resp. increasing) term structure of VS volatilities, either in a Monte Carlo simulation (exact) or in the approximate model with  $\nu_T^0(t)$  given by (7.52) (approx). Parameter values appear in right-hand side of table.

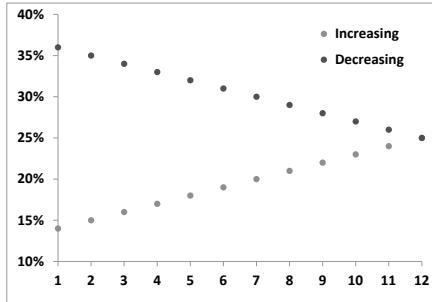
For the 1-year maturity case, the option’s strike  $\hat{\sigma}_T$  is 25% whereas in the 6-month case  $\hat{\sigma}_T = 31\%$  (resp.  $\hat{\sigma}_T = 19\%$ ) for the decreasing (resp. increasing) term structure of VS volatilities.

As Table 7.5 shows, exact and approximate prices are very close, considering the unrealistically steep term structures we have used. This again confirms that the volatility of  $U_t$  is the main contributor to the value of the option on realized variance.

For the 1-year maturity case, the option’s strike is  $\hat{\sigma}_T = 25\%$ , for both increasing and decreasing term structures. Observe the sizeable impact of the term structure of VS volatilities on the option’s price – of the order of 1%.

### 7.6.8 Accounting for the discrete nature of returns

In the model developed so far, the time value of an option on realized variance is generated by the volatility of  $\hat{\sigma}_T(t)$ , which is the only source of randomness.



**Figure 7.6:** Term structures of VS volatilities at  $t = 0$  used for generating prices in Table 7.5. Maturities are expressed in months.

Setting  $\nu_T(t) \equiv 0$  in expression (7.56a) leads to  $\sigma_{\text{eff}} = 0$ : the price of an ATM call on realized variance vanishes.

However, the volatility of  $U_t$  in (7.46) vanishes when  $\nu_T(t) \equiv 0$  only because we have chosen a continuous-time framework: the quadratic variation  $Q_t$  accrues continuously in our model. In realized variance payoffs, however,  $Q_t$  is expressed as the sum of squared daily log-returns up to time  $t$ .

In the Black-Scholes model with constant volatility, for example, using the standard definition for quadratic variation yields:

$$Q_t = \lim_{\Delta t \rightarrow 0} \sum_0^{t-1} \ln^2 \left( \frac{S_{i+1}}{S_i} \right) = \sigma^2 t$$

which implies that  $Q_t$  is deterministic and the price of an ATM call on realized variance vanishes. Consider instead the following definition for  $Q_t$ :

$$Q_t = \sum_0^{t-1} \ln^2 \left( \frac{S_{i+1}}{S_i} \right)$$

$Q_t$  is the standard estimator of realized variance over  $[0, t]$ . The log-return  $r = \ln \left( \frac{S_{i+1}}{S_i} \right)$  is given by:

$$r = \sigma \sqrt{\Delta t} Z - \frac{\sigma^2 \Delta t}{2} \quad (7.64a)$$

$$\simeq \sigma \sqrt{\Delta t} Z \quad (7.64b)$$

where  $Z$  is a gaussian random variable with unit variance and  $\Delta t$  is the interval between two observations of  $S$ , i.e. 1 day.

Volatility levels of typical equity underlyings –  $\sigma \sqrt{\Delta t}$  are of the order of 1% to 2% – are such that  $\sigma \sqrt{\Delta t} \ll 1$ , justifying the approximation in (7.64b). While the expectation of  $Q_t$  is still (approximately)  $\sigma^2 t$ ,  $Q_t$  is a finite sum of squared

random variables and has non-vanishing variance. Even in a Black-Scholes model with constant volatility, an ATM call option on variance acquires a non-zero value. This value is generated by the intrinsic variance of the volatility estimator we are using.

Let us then switch to discrete time and consider the evolution of  $U_t$  over the time interval  $[t, t + \Delta t]$ . Let  $r$  be the log-return of  $S_t$  over  $[t, t + \Delta t]$ . The variation of  $U_t$  during  $\Delta t$  reads:

$$\begin{aligned} U_{t+\Delta t} - U_t &= \frac{(Q_t + r^2) + (T - t - \Delta t) \hat{\sigma}_T^2(t + \Delta t)}{T} - \frac{Q_t + (T - t) \hat{\sigma}_T^2(t)}{T} \\ &= \frac{T - t - \Delta t}{T} \left( \hat{\sigma}_T^2(t + \Delta t) - \hat{\sigma}_T^2(t) \right) + \frac{1}{T} \left( r^2 - \hat{\sigma}_T^2(t) \Delta t \right) \end{aligned} \quad (7.65)$$

The first piece in (7.65) is the discrete-time counterpart of the  $\frac{T-t}{T} d(\hat{\sigma}_T^2)$  term in (7.45), which is the only contribution to  $dU_t$  in a continuous-time setting.

Let us assume that the volatility of  $\hat{\sigma}_T^2$  vanishes so that this contribution vanishes and let us concentrate on the contribution of the second piece to the variance of  $U_{t+\Delta t} - U_t$ .

Write the daily return over  $\Delta t$  as:

$$r = \sigma_t \sqrt{\Delta t} Z \quad (7.66)$$

where  $Z$  is a (possibly non-gaussian) random variable with unit variance, and  $\sigma_t$  is the instantaneous (discrete) volatility over  $[t, t + \Delta t]$ . Let us make the approximation that the term structure of the VS curve at  $t$  is flat:  $\sigma_t = \hat{\sigma}_T(t)$ . The second piece in (7.65) then reads:

$$\frac{\hat{\sigma}_T^2(t) \Delta t}{T} (Z^2 - 1)$$

where, just as in (7.64b), we have discarded terms of higher order in  $\sigma \sqrt{\Delta t}$ . Computing now the variance of  $\frac{U_{t+\Delta t}}{U_t}$  contributed by the second term in (7.65) we get:

$$\begin{aligned} E \left[ \left( \frac{U_{t+\Delta t}}{U_t} - 1 \right)^2 \right] &= \frac{1}{U_t^2} \left( \frac{\hat{\sigma}_T^2 \Delta t}{T} \right)^2 E \left[ (Z^2 - 1)^2 \right] \\ &= \left( \frac{\hat{\sigma}_T^2}{U_t} \right)^2 \left( \frac{\Delta t}{T} \right)^2 (2 + \kappa) \end{aligned} \quad (7.67)$$

where  $\kappa$  is the (excess) kurtosis of  $Z$ :  $\kappa = E[Z^4] - 3$ . Let us assume that the term structure of VS volatilities is flat, so that  $E[U_t] = \hat{\sigma}_T^2$  and let us replace  $U_t$  with its expectation. The prefactor in (7.67) is then simply equal to 1. This yields:

$$E \left[ \left( \frac{U_{t+\Delta t}}{U_t} - 1 \right)^2 \right] \simeq \left( \frac{\Delta t}{T} \right)^2 (2 + \kappa) \quad (7.68)$$

We now revert to the continuous-time framework. The quadratic variation of  $U_t$  over  $[0, T]$  now acquires an extra contribution and reads :

$$\int_0^T 4R_\tau^2 \left( \frac{T-\tau}{T} \right)^2 \nu_T^2(\tau) d\tau + N \left( \frac{\Delta t}{T} \right)^2 (2 + \kappa) \quad (7.69)$$

where  $N = \frac{T}{\Delta t}$  is the number of returns in the interval  $[0, T]$ . Dividing now (7.69) by  $T$  provides the following amended expression for  $\sigma_{\text{eff}}^2$  which supersedes (7.56a):

$$\sigma_{\text{eff}}^2 = \frac{1}{T} \int_0^T 4 \left( \frac{T-\tau}{T} \right)^2 \left( \frac{\hat{\sigma}_{\tau T}^2(0)}{\hat{\sigma}_T^2(0)} \right)^2 \nu_T^2(\tau) d\tau + \frac{2 + \kappa}{NT} \quad (7.70)$$

where  $N$  is the number of returns over  $[0, T]$ :  $N = \frac{T}{\Delta t}$ . The second piece in (7.70) is generated by the intrinsic variance of the variance estimator itself: as expected, its relative contribution to  $\sigma_{\text{eff}}^2$  is largest for short maturities.

Which value should we pick for  $\kappa$ ?  $\kappa$  is the *conditional* kurtosis of daily log-returns, that is the kurtosis of  $Z$  in (7.66). It is the portion of the *unconditional* kurtosis – the kurtosis of  $r$  – that is not generated by fluctuations of  $\sigma_t$  – i.e. the scale – of daily returns. In practice, as already mentioned in Section 1.2.2, measuring the *unconditional* kurtosis is already tricky; 5 is a typical level for equity underlyings.

Sorting out which portion of  $\kappa$  is attributable to fluctuations of  $\sigma_t$  – a quantity that is not directly observable – or to the intrinsic kurtosis of scaled log-returns ( $Z$ ) is even more challenging. In numerical examples below we have used  $\kappa = 2$ . The case of conditional lognormal daily returns corresponds to  $\kappa = 0$ .

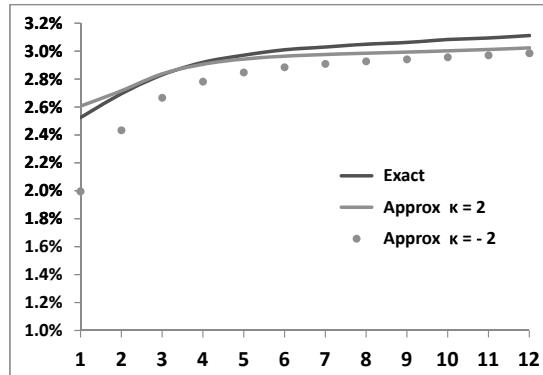
Setting  $\kappa = -2$  in formula (7.70) suppresses the contribution of the kurtosis of daily returns to  $\sigma_{\text{eff}}$  altogether.

Figure 7.7 shows prices of an ATM call option on realized variance for maturities 1 month to 1 year. We have used Set II parameters, flat VS volatilities equal to 20% and have used a ratio of 21 returns per month. We compare approximate prices obtained with expression (7.70) for  $\sigma_{\text{eff}}$  with  $\kappa = 2$  or  $\kappa = -2$  with (exact) prices computed in a Monte Carlo simulation of the two-factor model. While the dynamics of forward variances is generated in standard fashion, we draw daily returns with a Student distribution with  $\nu = 7$  degrees of freedom, so that their conditional kurtosis is equal to 2.<sup>20</sup>

Figure 7.7 highlights the fact that the discrete nature of returns is mostly apparent for short-maturity options – say 3 months or less. Observe how prices computed in our approximate model with  $\sigma_{\text{eff}}$  given by (7.70) are in good agreement with prices computed in a Monte Carlo simulation of the full-blown forward variance model –

---

<sup>20</sup>The density of the Student distribution is given in (10.1), page 392. Recall that the kurtosis of the Student distribution is  $\frac{6}{\nu-4}$ , and that only moments of order less than  $\nu$  exist. Using a Student distribution for the log-return has the consequence that the expectation of the return itself (the forward, equal to  $\exp(r)$ ) is infinite. Here we are only concerned with the estimation of moments of  $r$  up to order 4. We refer the reader to Chapter 10 for an example of stochastic volatility model with Student-distributed returns.



**Figure 7.7:** Price of an ATM call on realized variance as a function of maturity, from 1 month to 1 year, computed either in a Monte Carlo simulation of the forward variance model (exact), or in the approximate model with  $\sigma_{\text{eff}}$  given by (7.70) with  $\kappa = 2$  and  $\kappa = -2$ .

with Student-like daily returns for  $S_t$ . The slight negative bias of approximate prices with respect to exact prices is also seen in Table 7.5.

### 7.6.9 Conclusion

- The primary risk of options on realized variance is the exposure to the realized volatility of  $\widehat{\sigma}_T(t)$ , the implied volatility for the option's maturity. This risk is well captured with an approximate model for the dynamics of  $\widehat{\sigma}_T(t)$ : an option on realized variance is delta-hedged with VSs of the option's maturity. Forward variance models generating different volatilities and correlations of forward variances but the same volatility for  $\widehat{\sigma}_T(t)$  yield very similar prices.
- Additional VSs for intermediate maturities are needed to hedge the residual exposure to the VS term structure and to properly gamma-hedge the option on realized variance against the realized volatility of the underlying. The approximate model does not price the gamma costs arising from dynamical trading of these intermediate VSs. Comparison of approximate and exact prices suggests however that these costs are small.
- Quadratic variation does not accrue continuously but is expressed as a discrete sum of squared daily log-returns: this results in a specific contribution of the intrinsic variance of the volatility estimator to the option's price.

### 7.6.10 What about the vanilla smile? Lower and upper bounds

Nowhere in our analysis have we calibrated our model to the vanilla smile – except for determining VS volatilities. This is natural as we are only using VSs as hedges. Can the information in the vanilla smile be somehow used? In what measure do market prices of vanilla options restrict prices of options on realized variance?

We know that the density of  $S_T$  can be extracted from the vanilla smile of maturity  $T$  – see equation (2.8), page 29. If we now assume that the process for  $S_t$  is a diffusion – which in practitioner terms means that the expansion of the daily P&L of a delta-hedged position at order two in  $\frac{\delta S}{S}$  is adequate – much more information can be extracted from the vanilla smile.

Indeed, by delta-hedging European options we are able to create payoffs that involve the realized variance weighted by the dollar gamma, a function of  $t$  and  $S$ . The VS is an example of such a payoff – with a constant weight – but more complex payoffs involving  $S_t$  and increments of the quadratic variation  $Q_t$  can be synthesized.

What about call and put payoffs on  $Q_T$ ? These are not replicable by a combination of a static position in a European payoff and a delta strategy on  $S$ . Given a particular market smile, can we nevertheless derive model-independent bounds for the prices of calls and puts on realized variance?

Model-independent lower and upper bounds for prices of options on realized variance are most naturally determined by solving the stochastic control problem touched upon in Appendix A of Chapter 2, in the context of the Lagrangian Uncertain Volatility Model. The higher bound  $\bar{\mathcal{P}}(F)$  is obtained as the highest price generated by processes for the instantaneous volatility  $\sigma_t$  such that vanilla option prices are recovered:

$$\bar{\mathcal{P}}(F) = \max_{\substack{\sigma_t \in [\sigma_{\min}, \sigma_{\max}] \\ E_{\sigma}[O_i] = \mathcal{P}_{\text{Market}}(O_i)}} E_{\sigma}[F]$$

where we set  $\sigma_{\min} = 0$ ,  $\sigma_{\max} = +\infty$ .  $F$  is the payoff of the option on realized variance, and the  $O_i$  are payoffs of vanilla options.

Likewise, the model-independent lower bound  $\underline{\mathcal{P}}(F)$  is obtained by minimizing  $E_{\sigma}[F]$ . In the literature, this problem has been tackled differently, making use of particular solutions of the Skorokhod embedding problem – see [42] as well as [62], [28], [34].

The  $\underline{\mathcal{P}}(F)$ ,  $\bar{\mathcal{P}}(F)$  bounds thus derived are sharp: in case they are violated, an arbitrage strategy consisting in a static position in an option on realized variance plus a European payoff together with a dynamic delta strategy nets a positive P&L. Both the European payoff and the delta strategy have to be determined numerically. The delta strategy consists in delta-hedging the European payoff, either in timer-option fashion ( $\sigma = \sigma_{\max} = \infty$ ), or at zero volatility ( $\sigma = \sigma_{\min} = 0$ ).<sup>21</sup>

---

<sup>21</sup>For options that involve the quadratic variation  $Q$  in addition to  $S$ , equation (2.130), page 88, is replaced – with zero rates and repo – with:

To give the reader an intuition for how these delta-hedging regimes come into play, we now go through the derivation of a heuristic lower bound for a call on realized variance proposed by Bruno Dupire<sup>22</sup> and a heuristic upper bound, proposed by Peter Carr and Roger Lee and published in [28].<sup>23</sup>

Both are non-optimal bounds in the sense that the lower/upper bound lie lower/higher than  $\underline{\mathcal{P}}(F)/\overline{\mathcal{P}}(F)$ .

### A lower bound

Assume zero interest rate and repo and consider a short position in a European option of maturity  $T$  with convex payoff  $f(S)$ . Let us delta-hedge this option in timer option manner with a quadratic variation budget given by  $\mathcal{Q} = \widehat{\sigma}^2 T$ , where  $\widehat{\sigma}$  is the (volatility) strike of the option on realized variance.<sup>24</sup>. The option's value is given by:  $\mathcal{P}_{BS}^f(S_t, Q_t; \mathcal{Q})$ , where  $\mathcal{P}_{BS}^f$  is given by expression (5.68), page 183, and  $Q_t$  is the realized quadratic variation since  $t = 0$ .

If  $Q_T < \widehat{\sigma}^2 T$ , there remains at maturity some residual quadratic variation budget and we make a positive P&L equal to:

$$P\&L_{\text{Final}}^{Q_T < \mathcal{Q}} = \mathcal{P}_{BS}^f(S_T, Q_T; \mathcal{Q}) - \mathcal{P}_{BS}^f(S_T, Q_T; Q_T) \quad (7.71)$$

If instead we exhaust our budget at time  $\tau < T$ :  $Q_\tau = \mathcal{Q}$ , we delta-hedge our option at zero implied volatility on  $[\tau, T]$  and our final P&L is:

$$P\&L_{\text{Final}}^{Q_T > \mathcal{Q}} = -\frac{1}{2} \int_\tau^T S_t^2 \left. \frac{d^2 f}{dS^2} \right|_{S_t} \bar{\sigma}_t^2 dt \quad (7.72)$$

where  $\bar{\sigma}_t$  is the instantaneous realized volatility.

Let us now assume that the convex profile  $f$  is such that  $\frac{S^2}{2} \frac{d^2 f}{dS^2}$  either vanishes or is equal to a positive constant,  $\Gamma$ :  $\frac{S^2}{2} \frac{d^2 f}{dS^2} = \theta(S)\Gamma$  and  $\Gamma \geq 0$ , with  $\theta(S) = 0$  or 1. Then:

$$P\&L_{\text{Final}}^{Q_T > \mathcal{Q}} \geq -\Gamma \int_\tau^T \bar{\sigma}_t^2 dt \quad (7.73)$$

$$\geq -\Gamma(Q_T - Q_\tau) = -\Gamma(Q_T - \mathcal{Q}) \quad (7.74)$$

$$\frac{dP}{dt} + \max_{\sigma=\sigma_{\min}, \sigma_{\max}} \sigma^2 \left( \frac{S^2}{2} \frac{d^2 P}{dS^2} + \frac{dP}{dQ} \right) = 0$$

with  $\sigma_{\min} = 0$ ,  $\sigma_{\max} = +\infty$ . In this equation  $P$  is the price of payoff  $F - \sum \lambda_i O_i$  where  $F$  is the exotic option's payoff – in our case a call on realized variance – and the  $O_i$  are vanilla options. Because  $\sigma_{\min} = 0$ ,  $\sigma_{\max} = +\infty$ , the solution of this PDE is then such that (a)  $\frac{dP}{dt} = 0$ ; (b) in regions of  $S$  where the max is obtained with  $\sigma = \sigma_{\max} = +\infty$ , we have  $\frac{S^2}{2} \frac{d^2 P}{dS^2} + \frac{dP}{dQ} = 0$ . These are exactly conditions (5.70), page 184, that characterize timer options. I thank Pierre Henry-Labordère for pointing this out to me.

<sup>22</sup>Presented at the 2005 Global Derivatives conference.

<sup>23</sup>In what follows all  $P\&L$ s are computed at order one in  $\delta t$  and order two in  $\delta S$ ; provided this is adequate, the sub- and super-replicating strategies derived below indeed work. Mathematically, they hold for processes of  $S_t$  that are continuous semimartingales. Practically, they hold as long as contributions of order three and higher in  $\delta S$  to our daily  $P\&L$  can be ignored.

<sup>24</sup>See Section 5.9 on timer options.

Consider a call option on realized variance with strike  $\widehat{\sigma}$ , whose payoff is

$$\frac{1}{2\widehat{\sigma}_T} \left( \frac{Q_T}{T} - \widehat{\sigma}^2 \right)^+ = \frac{1}{2\widehat{\sigma}_T T} (Q_T - \mathcal{Q})^+$$

where  $\widehat{\sigma}_T$  is the VS volatility for maturity  $T$ . Denote by  $C_{\text{Mkt}}^{\widehat{\sigma}}$  its market price and let us buy  $2\widehat{\sigma}_T T$  units of it.

In case  $Q_T > \mathcal{Q}$ , this option's payout more than offsets the negative P&L  $P\&L_{\text{Final}}^{Q_T > \mathcal{Q}}$ : our total final P&L is positive.

If instead  $Q_T < \mathcal{Q}$ , the option on realized variance expires worthless, and our final total P&L, given by  $P\&L_{\text{Final}}^{Q_T < \mathcal{Q}}$  is positive as well.

Thus our short position in the European payoff  $f(S)$  combined with a long position in  $2\widehat{\sigma}_T T$  options on realized variance of strike  $\widehat{\sigma}$  nets a profit in all cases. It must then be that the cost for entering this position is positive. Setting up this position requires a cash amount equal to the market price of the realized variance call minus the difference between market and model prices of the European payoff at  $t = 0$ , which is equal to:

$$P_{\text{Mkt}}^f - \mathcal{P}_{BS}^f(S_0, 0; \mathcal{Q})$$

The condition that the cash amount needed at  $t = 0$  be positive then reads:

$$(2\widehat{\sigma}_T T) C_{\text{Mkt}}^{\widehat{\sigma}} - \left( P_{\text{Mkt}}^f - \mathcal{P}_{BS}^f(S_0, 0; \mathcal{Q}) \right) \geq 0$$

This yields the following lower bound for  $C_{\text{Mkt}}^{\widehat{\sigma}}$ :

$$C_{\text{Mkt}}^{\widehat{\sigma}} \geq \frac{1}{2\widehat{\sigma}_T T} \frac{1}{\Gamma} \left( P_{\text{Mkt}}^f - \mathcal{P}_{BS}^f(S_0, 0; \widehat{\sigma}^2 T) \right) \quad (7.75)$$

where we have replaced  $\mathcal{Q}$  with  $\widehat{\sigma}^2 T$ . This condition holds for any European payoff  $f$  such that  $\frac{S^2}{2} \frac{d^2 f}{dS^2}$  either vanishes or is equal to  $\Gamma$ . Which payoff should we pick, so that the lower bound for  $C_{\text{Mkt}}^{\widehat{\sigma}}$  provided by the right-hand side of (7.75) is highest?

$f$  can be replicated with cash, forwards and vanilla options of all strikes, of which only the latter contribute to the right-hand side of (7.75). From formula (3.6) for the replication of European payoffs, the density of vanilla options of strike  $K$  in the replicating portfolio is equal to the second derivative of the payoff, thus is equal to  $2\theta(K) \frac{\Gamma}{K^2}$ . The right-hand side of (7.75) can thus be written as:

$$\frac{1}{2\widehat{\sigma}_T T} \frac{1}{\Gamma} \int_0^\infty 2\theta(K) \frac{\Gamma}{K^2} \left( P_{\text{Mkt}}^K - \mathcal{P}_{BS}^K(S_0, 0; \widehat{\sigma}^2 T) \right) dK \quad (7.76)$$

where  $P_{\text{Mkt}}^K$  is the market price for a vanilla option of strike  $K$  and  $\mathcal{P}_{BS}^K$  is its Black-Scholes price. There is no need to distinguish between calls and puts, since by call-put parity  $P_{\text{Mkt}}^K - \mathcal{P}_{BS}^K$  is identical for a call or a put struck at the same strike.

Let us introduce the implied volatility for strike  $K$ ,  $\widehat{\sigma}_K$ . By definition of  $\widehat{\sigma}_K$ :  $P_{\text{Mkt}}^K = \mathcal{P}_{BS}^K(S_0, 0; \widehat{\sigma}_K^2 T)$ . (7.76) is thus equal to:

$$\frac{1}{2\widehat{\sigma}_T T} \int_0^\infty \frac{2}{K^2} \theta(K) \left( \mathcal{P}_{BS}^K(S_0, 0; \widehat{\sigma}_K^2 T) - \mathcal{P}_{BS}^K(S_0, 0; \widehat{\sigma}^2 T) \right) dK$$

The highest value for this expression is obtained by setting  $\theta(K) = 1$  for strikes such that  $\widehat{\sigma}_K > \widehat{\sigma}$  and  $\theta(K) = 0$  otherwise. We thus get our final expression for the (sub-optimal) lower bound :

$$C_{\text{Mkt}}^{\widehat{\sigma}} \geq \frac{1}{2\widehat{\sigma}_T T} \int_0^\infty \mathbf{1}_{\widehat{\sigma}_K > \widehat{\sigma}} \frac{2}{K^2} \left( \mathcal{P}_{BS}^K(S_0, 0; \widehat{\sigma}_K^2 T) - \mathcal{P}_{BS}^K(S_0, 0; \widehat{\sigma}^2 T) \right) dK \quad (7.77)$$

This idea can be extended to the case of a call on forward-starting variance – see [28].

Imagine that implied volatilities all lie above the (volatility) strike of the call on realized variance:  $\mathbf{1}_{\widehat{\sigma}_K > \widehat{\sigma}} = 1 \forall K$ . This implies also that  $\widehat{\sigma}_T \geq \widehat{\sigma}$ : the call on realized variance is in the money. We then recognize in the right-hand side of (7.77) the difference between the market price for the VS for maturity  $T$  – see equation (5.16), page 154 – and its price for a flat smile equal to  $\widehat{\sigma}$ . This yields:

$$C_{\text{Mkt}}^{\widehat{\sigma}} \geq \frac{1}{2\widehat{\sigma}_T} (\widehat{\sigma}_T^2 - \widehat{\sigma}^2)$$

which expresses that, with zero rates, the price of a call option is larger than its intrinsic value.

Likewise, if  $\mathbf{1}_{\widehat{\sigma}_K > \widehat{\sigma}} = 0 \forall K$ ,  $\widehat{\sigma}_T \leq \widehat{\sigma}$ : the call on realized variance is out of the money, and (7.77) again expresses that  $C_{\text{Mkt}}^{\widehat{\sigma}}$  lies above the intrinsic value, in this case zero.

### An upper bound

We present here one version of Peter Carr and Roger Lee's super-replicating strategy – the upper bound in [28] is sharper.

We work with zero interest rate and repo and temporarily omit the  $\frac{1}{2\widehat{\sigma}_T T}$  prefactor in the payoff of the call on realized variance:

$$(Q_T - \Sigma)^+$$

with  $\Sigma = \widehat{\sigma}^2 T$  where  $\widehat{\sigma}$  is the strike of the option on realized variance. Pick two barriers  $L$  and  $H$  such that  $L \leq S_0 \leq H$  where  $S_0$  is the initial spot level. It is possible to generate model-independently the payoff  $(Q_\tau - \Sigma)^+$  where  $\tau$  is defined as the time when  $S_\tau$  first hits  $L$  or  $H$  – see Section 5.9 on timer options.

This payoff is the timer equivalent of a perpetual barrier option that pays a rebate  $R(t) = (t - \Sigma)^+$  when  $S$  first hits  $L$  or  $H$ .

Denote by  $\mathcal{B}_{LH}(S_0, Q; \Sigma)$  the price of this option – as a timer option price it does not depend on  $t$ .  $\mathcal{B}_{LH}$  solves the following PDE:

$$\frac{S^2}{2} \frac{d^2 \mathcal{B}_{LH}}{dS^2} + \frac{d\mathcal{B}_{LH}}{dQ} = 0 \quad (7.78)$$

with boundary conditions:  $\mathcal{B}_{LH}(L, Q; \Sigma) = \mathcal{B}_{LH}(H, Q; \Sigma) = (Q - \Sigma)^+$

- If  $\tau > T$ , at  $t = T$  we have:

$$\begin{aligned} \mathcal{B}_{LH}(S_T, Q_T; \Sigma) &= E_T[(Q_\tau - \Sigma)^+] = E_T[(Q_\tau - Q_T + Q_T - \Sigma)^+] \\ &\geq (Q_T - \Sigma)^+ \end{aligned} \quad (7.79)$$

where  $E_T$  is a shorthand notation for  $E_{T, S_T, Q_T}$ , an expectation taken with respect to PDE (7.78), and we have used the property that the quadratic variation increases with time:  $Q_\tau - Q_T \geq 0$ . Thus, if  $\tau > T$ , a long position in the timer barrier option that pays  $(Q_\tau - \Sigma)^+$  super-replicates the payoff of the option on realized variance of maturity  $T$ .

- What if  $\tau \leq T$ ? If  $S$  hits either  $L$  or  $H$  at time  $\tau < T$  our timer option pays us  $(Q_\tau - \Sigma)^+$ . Since our aim is to super-replicate payoff  $(Q_T - \Sigma)^+$ , we need to trade an additional instrument of maturity  $T$  that generates  $(Q_T - Q_\tau)$ . At second order in  $\delta S$ , delta-hedging the profile  $-2 \ln S$  at zero implied volatility generates the realized quadratic variation:

$$d(-2 \ln S_t) + \frac{2}{S_t} dS_t = dQ_t$$

The P&L from the costless delta strategy  $\frac{2}{S_t} dS_t$ , combined with a long position in an option that pays  $-2 \ln S_T$  at  $T$ , generates the quadratic variation up to  $T$ . Since we only need to generate the quadratic variation starting at  $\tau$ , when  $S$  hits either  $L$  or  $H$ , let us adjust  $-2 \ln S$  by an affine function of  $S$  that preserves the convexity of  $-2 \ln S$  but ensures that the resulting payoff profile vanishes for  $S = L$  and  $S = H$ . The resulting payoff  $f_{LH}(S)$  is:

$$f_{LH}(S) = -2 \ln S + 2 \left( \frac{\ln H - \ln L}{H - L} (S - L) + \ln L \right)$$

$f_{LH}(S) \leq 0$  for  $S \in [L, H]$  and is positive otherwise. We have:

$$Q_T - Q_\tau = (f_{LH}(S_T) - f_{LH}(S_\tau)) - \int_\tau^T \frac{df_{LH}}{dS} \Big|_{S_t} dS_t \quad (7.80a)$$

$$= f_{LH}(S_T) - \int_\tau^T \frac{df_{LH}}{dS} \Big|_{S_t} dS_t \quad (7.80b)$$

since by construction  $f_{LH}(S_\tau) = 0$ . By buying the timer option discussed above as well as the European option that pays  $f_{LH}(S_T)$  we generate at  $T$

the amount:  $(Q_\tau - \Sigma)^+ + (Q_T - Q_\tau)$ . It is easy to check that this again super-replicates the call on variance:

$$(Q_\tau - \Sigma)^+ + (Q_T - Q_\tau) \geq (Q_T - \Sigma)^+$$

Our super-replicating portfolio thus comprises, in addition to the barrier timer option, a European option of maturity  $T$  that pays  $f_{LH}(S_T)$ . In the case  $\tau > T$ , however, the value of our portfolio at  $T$  is  $\mathcal{B}_{LH}(S_T, Q_T; \Sigma) + f_{LH}(S_T)$ . Since  $f_{LH}(S_T) \leq 0$  for  $S \in [L, H]$ , to ensure that the super-replication in (7.79) still works, we just need to replace  $f_{LH}(S_T)$  with  $g_{LH}(S_T)$  given by:

$$g_{LH}(S) = \max(f_{LH}(S), 0)$$

Because  $g_{LH} \geq f_{LH}$ , the super-replication still holds for  $\tau \leq T$ , provided we trade the delta  $\frac{df_{LH}}{dS}$  over  $[\tau, T]$ .

In conclusion, reverting to the usual normalization for the payoff of calls on realized variance:  $\frac{1}{2\hat{\sigma}_T T}(Q_T - \hat{\sigma}^2 T)$ , whose price is  $C_{\text{Mkt}}^{\hat{\sigma}}$ , we have:

$$C_{\text{Mkt}}^{\hat{\sigma}} \leq \frac{1}{2\hat{\sigma}_T T} \left( \mathcal{B}_{LH}(S_0, 0; \hat{\sigma}^2 T) + g_{LH}(S_0, T) \right) \quad (7.81)$$

where  $\mathcal{G}_{LH}(S_0, T)$  is the market price of the option that pays  $g_{LH}(S_T)$  at  $T$ . As it is a European payoff, it can be replicated with vanilla options:  $\mathcal{G}_{LH}(S_0, T)$  only depends on the vanilla smile for maturity  $T$ .

$\mathcal{B}_{LH}(S_0, 0; \hat{\sigma}^2 T)$  instead, only depends on  $L$  and  $H$ . The best higher bound is thus obtained for the  $(L, H)$  couple that minimizes the right-hand side of (7.81).

The reasoning for the upper bound can be extended to calls on forward-starting variance – see [28].

As a sanity check, take  $L = H = S_0$ . Then  $\mathcal{B}_{LH}(S_0, 0; \hat{\sigma}^2 T) = 0$  and  $g_{LH}(S)$  is given by:

$$\begin{aligned} g_{LH}(S) &= f_{LH}(S) \\ &= -2 \ln\left(\frac{S}{S_0}\right) + \frac{2}{S_0}(S - S_0) \end{aligned}$$

We know from (5.12) that, with zero rate and repo, the market price of an option that pays  $-2 \ln(\frac{S_T}{S_0})$  is  $\sigma_T^2 T$ . Thus,  $\mathcal{G}_{LH}(S_0, T) = \sigma_T^2 T$  and (7.81) expresses the fact that the price of a call option is bounded above by the forward. This higher bound is not indecent; it is the limit of a call price when volatility is taken to infinity.

## Conclusion

Information in the vanilla smile can be used to bound prices of options on realized variance. Using these bounds as bid/offer levels leads to prices that are, in practice, too conservative.<sup>25</sup>

---

<sup>25</sup>These bounds can be interpreted as a measure of model risk – in the absence of jumps, that is, in practitioner terms, when only terms of order up to two in  $\delta S$  are considered in our daily P&L. When

### 7.6.11 Options on forward realized variance

Consider two dates  $T_1, T_2$ . An option on forward realized variance pays at  $T_2$  a vanilla payoff on the variance realized over the interval  $[T_1, T_2]$ . This type of option is attractive when the term structure of VS volatilities is such that forward VS implied volatilities appear to be much lower/higher than reasonable estimates of future realized volatility.

At  $t = T_1$  the option is simply a spot-starting option on realized variance, which we have just extensively analyzed. Its price at  $T_1$  is a function of VS volatilities  $\hat{\sigma}_T(T_1), T \in [T_1, T_2]$  given, for example in the SM, by expressions (7.56), but where we sit at  $t = T_1$  rather than at  $t = 0$ :

$$\begin{aligned}\sigma_{\text{eff}}^2 &= \frac{4}{T_2 - T_1} \int_{T_1}^{T_2} \left( \frac{T_2 - \tau}{T_2 - T_1} \right)^2 \left( \frac{\hat{\sigma}_{\tau T_2}^2(T_1)}{\hat{\sigma}_{T_2}^2(T_1)} \right)^2 \nu_{T_2}^2(\tau) d\tau \\ P(t = T_1) &= P_{\text{BS}}(t = T_1, \hat{\sigma}_{T_2}^2(T_1), \sigma_{\text{eff}}, T_2)\end{aligned}\quad (7.82)$$

The value of our option at  $T_1$  is then a particular form of a swaption payoff. The price at  $t = 0$  of an option on forward realized variance is thus the price of a particular variance swaption of maturity  $T_1$ . It depends on assumptions about the volatility of forward variances  $\xi_t^\tau, \tau \in [T_1, T_2]$ , over the interval  $t \in [0, T_1]$ .

Consider replacing  $\sigma_{\text{eff}}$  with 0 in (7.82); the value at  $T_1$  of our option is then simply the call – or put – payoff applied to  $\hat{\sigma}_{T_2}^2(T_1)$ . This is the payoff of a standard VS swaption, that is the option to enter at  $T_1$  a VS contract of maturity  $T_2$ .<sup>26</sup> An option on forward realized variance is thus more expensive than the corresponding VS swaption.<sup>27</sup>

Spot-starting options on realized variance can be economically priced and risk-managed in the SM as the only ingredient is the curve  $\nu_T(\tau), \tau \in [0, T]$ : the instantaneous volatility of the VS volatility for the residual maturity. Options on forward realized variance, on the other hand, require a specification of volatilities of *forward* VS volatilities. These cannot be backed out of the curve  $\nu_T(\tau), \tau \in [0, T]$ . An option on forward realized variance can then only be priced in a full-blown model.

In the two-factor model we can choose parameters such that, while volatilities of *spot-starting* VS volatilities are identical, volatilities of *forward* VS volatilities are different. Sets I, II, III in Table 7.1, page 229, have this property. They are chosen so as to match the benchmark  $\nu_T^B(t)$  in (7.40) with  $\sigma_0 = 100\%$ ,  $\tau_0 = 0.25$ ,  $\alpha = 0.4$ .

---

higher-order terms are taken into account – which is the case if jumps are allowed – even the bounds on a simple VS become sizeably spaced; see for example [63]. By construction, the processes that generate these bounds are worst-case scenarios and can hardly be considered realistic.

<sup>26</sup>Variance swaption smiles in the two-factor model are shown in Section 7.4.3.

<sup>27</sup>This is because  $\hat{\sigma}_{T_2}^2(T_1)$ , which is the underlying of the VS swaption, is the expectation at  $t = T_1$  of  $\sigma_r^2$ , where  $\sigma_r$  is the realized volatility over  $[T_1, T_2]$ . We then have, in model-independent fashion:  $E_{T_1}[(\sigma_r^2 - K)^+] \geq (E[\sigma_r^2] - K)^+ = (\hat{\sigma}_{T_2}^2(T_1) - K)^+$ .

Table 7.6 lists prices of a 6 month-in-6 month ATM call option on forward realized variance in Sets I, II, III, together with (a) prices of spot-starting options with the same maturity from Table 7.2, page 241, (b) prices of 6 month-in-6 month variance swaptions. We can see that (a) options on forward realized variance are indeed more expensive than both spot-starting options and swaptions, (b) the price in Set I is appreciably higher than that in Set III, even though prices of spot-starting options are almost identical.

The difference between the second and third line of Table 7.6 is a measure of the additional time value contributed by the volatility of forward volatilities  $\xi_t^\tau$  during the interval  $[T_1, T_2]$ .

	Set I	Set II	Set III
Spot-starting	2.97%	2.96%	2.94%
Foward-starting	4.25%	4.09%	3.94%
Swaption	3.12%	2.90%	2.69%

**Table 7.6:** Prices of an ATM call option on forward realized variance (top) and ATM call VS swaption (bottom) with  $T_1 = 6$  months,  $T_2 = 1$  year, computed in the two-factor model with parameter sets in Table 7.1, page 229.

Prices for a spot-starting option of the same maturity, from Table 7.2, are shown for reference. The term structure of VS volatilities is flat at 20%.

## 7.7 VIX futures and options

We discuss here the application of the two-factor model to VIX instruments. This will allow us to relax the assumption of lognormality of forward variances.

VIX futures trade on the CBOE and expire on the morning of the Wednesday that is exactly 30 days prior to the monthly expiration dates of listed S&P 500 options. The settlement value of the expiring VIX future is the 30-day log-contract implied volatility, computed using market prices of listed S&P 500 options.<sup>28</sup> Ignoring the effect of fixed cash-amount dividends, the log-contract implied volatility at time  $t$  for maturity  $T$  is given by expression (5.12):

$$\widehat{\sigma}_T^2(t) = \frac{e^{r(T-t)}}{T-t} (Q_{\text{mkt},t}^T - Q_{\widehat{\sigma}=0,t}^T)$$

where  $Q_{\text{mkt},t}^T$  is the market price at time  $t$  of the payoff  $-2 \ln S_T$  and  $Q_{\widehat{\sigma}=0,t}^T$  is the price of the same payoff using vanishing volatility. We now set  $t = T_i$ , the expiration

<sup>28</sup>The settlement value of the VIX future is in fact one hundred times this VS volatility – we ignore this multiplier in what follows.

date of the  $i$ th future, and  $T - t = \Delta = 30$  days, and use the notation  $\widehat{\sigma}_{\text{VIX},T_i}$  for the settlement value of the VIX future expiring at  $T_i$ . As indicated in Section 5.2, the payoff  $-2 \ln S_T$  is replicated by a continuous density  $\frac{2}{K^2}$  of vanilla options, which yields:

$$\widehat{\sigma}_{\text{VIX},T_i}^2(T_i) = \frac{2e^{r\Delta}}{\Delta} \left( \int_0^{F_{T_i}^{T_i+\Delta}} P_{\text{mkt},T_i}^{K,T_i+\Delta} \frac{dK}{K^2} + \int_{F_{T_i}^{T_i+\Delta}}^{\infty} C_{\text{mkt},T_i}^{K,T_i+\Delta} \frac{dK}{K^2} \right) \quad (7.83)$$

where  $P_{\text{mkt},t}^{KT}$  (resp.  $C_{\text{mkt},t}^{KT}$ ) is the (discounted) market price at time  $t$  of a put (resp. call) option of strike  $K$ , maturity  $T$  on the S&P 500 and  $F_{T_i}^{T_i+\Delta}$  is the forward of the S&P 500 index for maturity  $T_i + \Delta$  observed at  $T_i$ . In the formula actually used by the CBOE, the integrals above are discretized using the trapezoidal rule and the integration does not run from 0 to  $\infty$  but is cut off for small and large strikes wherever a zero bid price is encountered for two consecutive strikes.<sup>29</sup> The resulting approximation is quite good, as strikes of listed S&P 500 options are closely spaced and out-of-the-money strikes are reasonably liquid, typically from 50% to 125%.

Note that  $\widehat{\sigma}_{\text{VIX},T_i}$  does not share the same convention as VS payoffs for annualizing volatility: in (7.83)  $\frac{1}{\Delta}$  is used instead of  $\frac{252}{N}$ . Depending on the exact number of trading days for the 30-day period at hand, this difference in conventions may translate into a difference of the order of one point of volatility. Further below, we will use the same  $\frac{1}{\Delta}$  convention when comparing VIX and log-contract – or VS – quotes.

Exchange-traded options on VIX futures exist as well, with the same expiration dates as the underlying VIX futures. Figure 7.8 shows the values of the 5 VIX futures for expiries July through December, observed on January 14, 2011, as well as the smiles of the associated options – see Figure 7.9 below for the expiry dates of the VIX futures.<sup>30</sup>

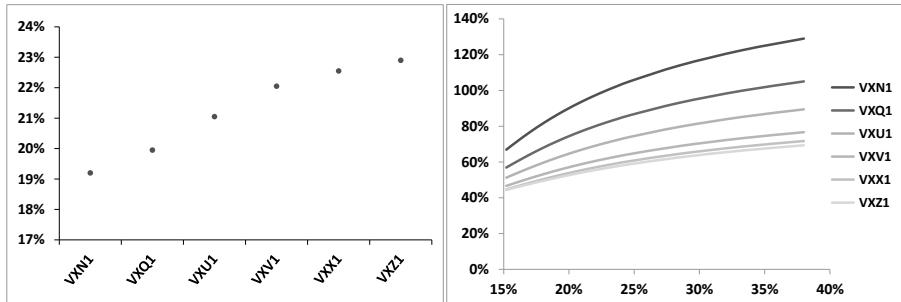
The settlement value of a VIX future is a log-contract volatility, not a VS volatility. However, for ease of exposition, since forward VS variances and forward log-contract variances are both driftless and can be modeled identically – see the discussion in Section 5.5, page 168 – we make no distinction in what follows between  $\widehat{\sigma}_T$  and  $\widehat{\sigma}_{\text{VS},T}$ .  $\widehat{\sigma}_T = \widehat{\sigma}_{\text{VS},T}$  and forward log-contract and VS variances are identical, equal to  $\xi_t^T$ .

The distinction between  $\widehat{\sigma}_T$  and  $\widehat{\sigma}_{\text{VS},T}$  becomes relevant again in the discussion of the arbitrage between S&P 500 VSs and VIX instruments, in Section 7.7.4.

---

<sup>29</sup>We refer the reader to [www.cboe.com/micro/vix/vixwhite.pdf](http://www.cboe.com/micro/vix/vixwhite.pdf) for the exact procedure for computing  $\widehat{\sigma}_{\text{VIX},T}$ .

<sup>30</sup>The VIX itself is a number published in real time:  $\widehat{\sigma}_{\text{VIX},t}$  is the 30-day VS volatility obtained through an interpolation of VS volatilities for two consecutive expiration dates of S&P 500 listed options, computed with the same methodology as for the settlement value of VIX futures. While the VIX index has become a popular indicator of real-time market temperature we do not consider it in what follows as – just like the temperature in New York City – it cannot be traded.



**Figure 7.8:** Left: VIX futures as of June 14, 2011 for expiries ranging from July (VXN1) to December (VXZ1). Right: Smiles of VIX futures.

### 7.7.1 Modeling VIX smiles in the two-factor model

Let  $F_t^i$  be the VIX future for expiry  $T_i$ , observed at time  $t$ . At  $t = T_i$ ,  $F_t^i$  is equal to the VIX index, that is the 30-day VS volatility observed at  $T_i$ :

$$F_{t=T_i}^i = \hat{\sigma}_{\text{VIX},T_i}(T_i) \quad (7.84)$$

where we denote by  $\hat{\sigma}_{\text{VIX},T_i}$  the forward VS volatility corresponding to the VIX future expiring at  $T_i$  – it is defined for  $t < T_i$ :

$$\hat{\sigma}_{\text{VIX},T_i}^2(t) = \frac{1}{\Delta} \int_{T_i}^{T_i+\Delta} \xi_t^T dT = E_t[(F_{T_i}^i)^2] \quad (7.85)$$

$F_t^i$  is the value of a future, hence it is driftless. For  $t < T_i$ ,  $F_t^i$  is thus given by:

$$F_t^i = E[F_{T_i}^i] \quad (7.86)$$

So far, we have modeled instantaneous forward variances as lognormal processes. While the resulting discrete forward variances are not exactly lognormal, they are close to lognormal – see for example swaption smiles in Figure 7.4. As is apparent in Figure 7.8, implied volatilities of VIX futures exhibit substantial smiles: we need to relax the lognormality of instantaneous forward variances, while preserving a Markov representation, if possible.

Let us start with the two-factor model specified in Section 7.4, page 226. Forward variances  $\xi_t^T$  are given by (7.33) and (7.34):

$$\begin{aligned} \xi_t^T &= \xi_0^T f^T(t, x_t^T) \\ f^T(t, x) &= e^{\omega x - \frac{\omega^2}{2} \chi(t, T)} \end{aligned}$$

In the basic version of the two-factor model, the mapping function is an exponential, hence  $\xi_t^T$  is lognormal.

### A Markov-functional model

More general forms for  $f(t, x)$  can be considered. The condition that  $\xi_t^T$  is driftless translates into the following PDE for  $f^T(t, x)$ :

$$\frac{df^T}{dt} + \frac{\eta^2(T-t)}{2} \frac{d^2 f^T}{dx^2} = 0 \quad (7.87)$$

where  $\eta^2(T-t)$  is the instantaneous variance of  $x_t^T$ .  $\eta$  is given in (7.31b):

$$\eta(u) = \alpha_\theta \sqrt{(1-\theta)^2 e^{-2k_1 u} + \theta^2 e^{-2k_2 u} + 2\rho_{12}\theta(1-\theta)e^{-(k_1+k_2)u}}$$

Once a terminal condition at time  $t = T$ ,  $f^T(T, x)$  is chosen, the mapping function  $f^T(t, x)$  is generated for dates  $t < T$  by solving (7.87). Any solution of (7.87) needs to be suitably normalized so that the initial value of the forward variance  $\xi_t^T$  is recovered at  $t = 0$ , for  $x = 0$  – which is another way of saying that  $E[\xi_t^T] = \xi_0^T$ ,  $\forall t$ :

$$f^T(0, 0) = 1$$

The dynamics of  $\xi^T$  reads:

$$\frac{d\xi_t^T}{\xi_t^T} = \frac{d \ln f^T}{dx}(t, x_t^T) dx_t^T$$

The instantaneous volatility of  $\xi^T$  is thus given by:

$$\text{vol}(\xi_t^T) = \eta(T-t) \left| \frac{d \ln f^T}{dx} \right|_{x=f^{T^{-1}}(\xi_t^T, t)} \quad (7.88)$$

$f^{T^{-1}}$  is well-defined only if  $f^T$  is monotonic; this is verified in the parametric model used in the section that follows. The right-hand side of (7.88) is a function of  $\xi_t^T$  and  $t$ , thus what we have is a local volatility model for  $\xi_t^T$ .

### A simple parametrization

Any linear combination of exponential solutions (7.34) solves (7.87) – using positive weights ensures that  $f^T(t, x) \geq 0$ , for all  $t, x$ . With such an ansatz for  $f^T(t, x)$  smiles for  $\xi^T$  are positively sloping, as for large values of  $x$  the linear combination is dominated by terms with large values of  $\omega$ . This is not a serious limitation as VIX smiles for liquid strikes are usually positively sloping. Let us then use just two exponentials – this ansatz was proposed in [10]. We introduce the volatility-of-volatility smile parameters  $\gamma_T$ ,  $\beta_T$ ,  $\omega_T$  and set:

$$f^T(t, x) = (1 - \gamma_T) e^{\omega_T x - \frac{\omega_T^2}{2} \chi(t, T)} + \gamma_T e^{\beta_T \omega_T x - \frac{(\beta_T \omega_T)^2}{2} \chi(t, T)} \quad (7.89)$$

$\gamma_T \in [0, 1]$  is a mixing parameter. For  $\gamma_T = 0$  or  $\gamma_T = 1$ ,  $\xi^T$  is lognormal.  $\beta_T$  is also chosen in  $[0, 1]$  – for  $\beta_T = 0$ ,  $\xi^T$  is simply a displaced lognormal. The instantaneous volatility of  $\xi_T$  is given by (7.88). Setting  $t = 0$ ,  $x = 0$  yields:

$$\text{vol}(\xi_t^T)|_{t=0} = \omega_T ((1 - \gamma_T) + \beta_T \gamma_T) \eta(T)$$

Let us introduce the dimensionless parameter  $\zeta^T$  and write  $\omega_T$  as:

$$\omega_T = \frac{2\nu}{(1 - \gamma_T) + \beta_T \gamma_T} \zeta_T \quad (7.90)$$

The instantaneous volatility of  $\xi_t^T$  at  $t = 0$  is then given by:

$$\text{vol}(\xi_t^T)|_{t=0} = 2\nu \zeta_T \eta(T)$$

$\zeta_T$  is thus simply a scale factor for the instantaneous volatility of  $\xi_t^T$  at  $t = 0$ . We now use parameters  $\gamma^T, \beta^T, \zeta^T$  to adjust the smile of VIX futures, for a given set of parameters  $\nu, \theta, k_1, k_2, \rho$ .

### Calibration of VIX futures and options

We first choose a set of parameters  $\nu, \theta, k_1, k_2, \rho$  to generate the underlying basic dynamics of our model – specifically one of the three sets in Table 7.1.

Then for each expiry  $T_i$  we determine forward variances  $\xi_0^T$  for  $T \in [T_i, T_i + \Delta]$  as well as parameters  $\gamma^T, \beta^T, \zeta^T$  so that market prices of (a) the VIX future  $F_t^i$  expiring at  $T_i$  and (b) VIX options maturing at  $T_i$  are matched.<sup>31</sup> Mathematically our model needs to ensure that:

$$\begin{aligned} E_t[\hat{\sigma}_{\text{VIX}, T_i}(T_i)] &= F_t^i \\ E_t[(\hat{\sigma}_{\text{VIX}, T_i}(T_i) - K)^+] &= \mathcal{C}_t^{Ki}, \quad E_t[(K - \hat{\sigma}_{\text{VIX}, T_i}(T_i))^+] = \mathcal{P}_t^{Ki} \end{aligned}$$

where  $\mathcal{C}_t^{Ki}, \mathcal{P}_t^{Ki}$  are, respectively, *undiscounted* market prices of call and put options with strike  $K$ , maturity  $T_i$  on the VIX index, and  $\hat{\sigma}_{\text{VIX}, T_i}$  is given by (7.85) as a function of forward variances. We assume that within each interval  $[T_i, T_i + \Delta]$  forward variances at  $t = 0$  are flat and denote them by  $\xi_0^i$ . Similarly we use constant values for  $\gamma^T, \beta^T, \zeta^T$  over  $[T_i, T_i + \Delta]$  – which we denote by  $\gamma^i, \beta^i, \zeta^i$ .

The forward variance for interval  $[T_i, T_{i+1}]$  that underlies the VIX future expiring at  $T_i$  is thus a function of  $t, X_t^1, X_t^2$  given by:

$$\hat{\sigma}_{\text{VIX}, T_i}^2(t, X_t^1, X_t^2) = \xi_0^i \frac{1}{\Delta} \int_{T_i}^{T_{i+1}} f^\tau(t, x_t^\tau) d\tau \quad (7.91)$$

and both the VIX future and prices of options on  $\hat{\sigma}_{\text{VIX}, T_i}(T_i)$  are obtained through a two-dimensional Gaussian quadrature on  $(X_{T_i}^1, X_{T_i}^2)$ .

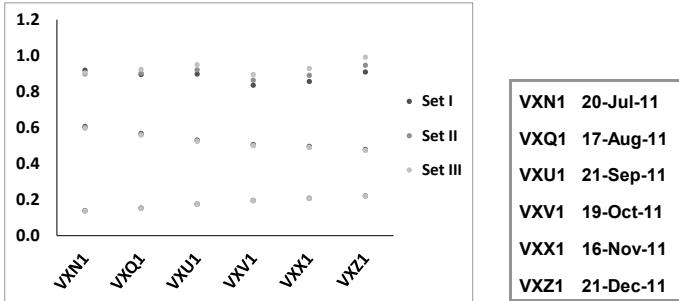
Figure 7.9 shows the values of parameters  $\gamma^i, \beta^i, \zeta^i$  calibrated to the market smiles of VIX futures as of June 14, 2011 for strikes in the interval [15%, 40%]. In this strike range the difference between market and calibrated implied volatilities lies well within the bid/offer spread which is about 3 points of volatility – the smiles

---

<sup>31</sup>Note that there is no guarantee that the variance curve thus determined matches SP500 log-contract implied volatilities – more on this in Section 7.7.4 below.

in Figure 7.8 are in fact those generated by our model with Set II parameters.<sup>32</sup> Calibration is performed by least-squares minimization with a sufficiently large weight on the future itself that it is exactly calibrated.<sup>33</sup>

It should be mentioned that for very large strikes – that are much less liquid – market implied volatilities fall off, a feature that our parametrization is unable to capture.



**Figure 7.9:** Left: values of  $\gamma^i, \beta^i, \zeta^i$  calibrated on market smiles of VIX futures on June 14, 2011, using parameter sets in Table 7.1 – bottom:  $\beta^i$ , middle:  $\gamma^i$ , top:  $\zeta^i$ . Right: expiries of VIX futures. The corresponding futures and their smiles appear in Figure 7.8.

Calibration has been performed using the three sets of parameters  $\nu, \theta, k_1, k_2, \rho$  listed in Table 7.1, page 229. Figure 7.9 also lists the expiry dates of the corresponding VIX futures.<sup>34</sup> Calibrated values of  $\gamma^i, \beta^i, \zeta^i$  for the six futures considered are similar; moreover they hardly depend on which set is used: only  $\zeta$  is appreciably larger in Set III than in Sets I and II. This is expected. Sets I, II, III generate almost identical term

<sup>32</sup>Contrary to what is claimed at times, simultaneous jumps in the underlying and its volatility are not needed to generate market-compatible VIX smiles. While in this section we focus exclusively on VIX futures, that is on the dynamics of forward variances, we provide in Section 8.7.2 an example of parametrization for the correlations of  $S_t$  with  $X_t^1, X_t^2$  that generates a term structure of the ATMF skew that agrees with market smiles of vanilla options on  $S_t$ . It is thus possible to approximately match both the S&P 500 smile and VIX smiles without resorting to simultaneous jumps – provided the mismatch between VS volatilities either derived from the S&P 500 VS market or derived from the VIX market is small – see the discussion in Section 7.7.4 below.

<sup>33</sup>The reader may wonder why, rather than calibrating  $\xi^i$  along with  $\gamma^i, \beta^i, \zeta^i$ , we do not determine  $\xi^i$  using equation (7.98) in Section 7.7.4 below and then calibrate  $\gamma^i, \beta^i, \zeta^i$  on  $F_t^i$  and the smile of VIX options maturing at  $T_i$ . Imagine that our model is able to perfectly calibrate VIX futures and options; then the (calibrated) value of  $\xi^i$  will obey equation (7.98). This will not be the case, however, if market and model smiles differ significantly outside the strike range used for calibration. In the latter case, using (7.98) will generate a value for  $\xi^i$  that may make it impossible to find values of  $\gamma^i, \beta^i, \zeta^i$  such that  $F_t^i$  and market prices of near-the-money VIX options are recovered.

<sup>34</sup>Because expiry dates of VIX futures are not spaced 30 days apart, intervals  $[T_i, T_i + \Delta]$  for consecutive futures may overlap – see the case of VXU1 and VXV1 for example.  $\gamma, \beta, \zeta$  are then assumed constant over each interval  $[T_i, \min(T_i + \Delta, T_{i+1})]$ .

structures for volatilities of spot-starting VS volatilities, however, as manifested in Figure 7.2, volatilities of forward volatilities are lowest in Set III – this is compensated for by an increase in  $\zeta$ .

Calibration is equally satisfactory for Sets I, II, III. Which set should one choose? Different sets for parameters  $\nu, \theta, k_1, k_2, \rho$  generate different distributions of the realized variance of  $F_t^i$ , as well as different correlation structures for the  $F_t^i$ . Define  $\sigma_t$  as the square root of the expected instantaneous variance of  $F_t^i$ :

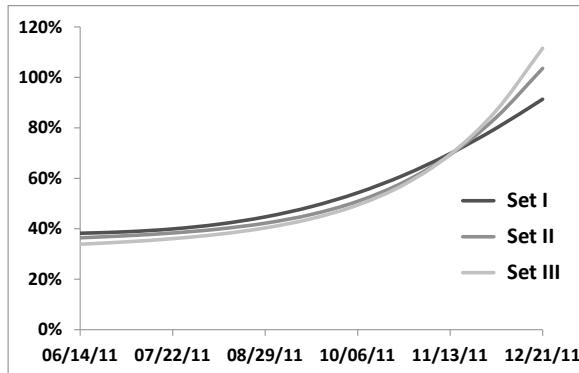
$$\sigma_t^2 = E_0 \left[ \frac{\langle (d \ln F_t^i)^2 \rangle}{dt} \right] \quad (7.92)$$

In our two-factor Markov-functional model,  $F_t^i$  is a function of  $X_t^1, X_t^2$ , given by numerical evaluation of the expectation in (7.86).  $\sigma_t$  is thus simply given by:

$$\sigma_t = \sqrt{E [\alpha_{X^1}^2 + \alpha_{X^2}^2 + 2\rho_{12}\alpha_{X^1}\alpha_{X^2}]} \quad (7.93)$$

where  $\alpha_{X^1}, \alpha_{X^2}$  are the sensitivities of  $\ln F_t^i$  with respect to  $X_t^1, X_t^2$  and the expectation is taken over  $X_t^1, X_t^2$  – this is efficiently evaluated through a two-dimensional Gaussian quadrature.

Figure 7.10 displays  $\sigma_t$  for the VIX future expiring on December 21, 2011, for dates  $t$  ranging from June 16, 2011 to the future's expiry.



**Figure 7.10:**  $\sigma_t$  – as defined in (7.92) – for VIX future VXZ1, for dates ranging from June 14 to its expiry date, in Sets I, II, III.

While the integrated value of  $\sigma_t^2$  is identical in Sets I, II, III, the distribution is different.<sup>35</sup>

<sup>35</sup>By definition,  $\sigma_t^2$  is the expected instantaneous realized variance of  $F_t^i$ . The integral of  $\sigma_t^2$  over  $[0, T_i]$  is thus the implied variance of the log contract payoff:  $-2 \ln(F_{T_i}^i)$ , which only depends on the smile of  $F_{T_i}^i$  – the VIX smile for maturity  $T_i$ .

Controlling the term structure of the instantaneous volatilities of VIX futures is then the criterion for choosing one particular set for  $\nu, \theta, k_1, k_2, \rho$ .

While curves in Figure 7.10 are different, they are not terribly different; the reader may think that the range of volatility distributions of forward volatilities spanned by the two-factor model is limited.

That this is not the case is illustrated in the right-hand graph of Figure 7.13 below, where we use a one-factor model for simplicity and vary  $k_1$  – still maintaining calibration to VIX smiles. Notice how different the volatility distributions are.

In Figure 7.10 parameters of Sets I, II, III have been used. They have the property that they generate approximately the same volatilities for spot-starting volatilities of all maturities. This additional constraint accounts for the narrower range of volatility distributions in Figure 7.10 than in Figure 7.13.

### Characterizing the dynamics in the model

What about the dynamics of VIX futures in our model? The variance curve observed at time  $t$  is a function of  $(X_t^1, X_t^2)$ , however each instantaneous forward variance  $\xi^T$  is a function of  $x_t^T$ . We have a one-factor Markov-functional model for each instantaneous forward variance  $\xi^T$  – i.e. a local volatility model<sup>36</sup> – whose local volatility function is given by (7.88).

Processes  $x_t^T$  for  $T \in [T_i, T_i + \Delta]$  are different linear combinations of  $X_t^1, X_t^2$ , thus, literally  $F_t^i$  is a function of  $(X_t^1, X_t^2)$ ; however since typically  $\Delta \ll \frac{1}{k_1}, \frac{1}{k_2}$ ,  $F_t^i$  can practically be considered a function of the single quantity  $x_t^{T_i}$ . We thus have essentially a multi-asset local volatility model for VIX futures.

### 7.7.2 Simulating VIX futures in the two-factor model

In the parametric model specified by (7.89), page 264, continuous forward variances  $\xi_t^T$  are modeled as a function of  $x_t^T$ . Forward variances  $\widehat{\sigma}_{\text{VIX}, T_i}^2(t)$  are explicitly known for all  $t$ .

Consider VIX future  $F^i$ . At  $t = T_i$ ,  $F_{T_i}^i = \sqrt{\widehat{\sigma}_{\text{VIX}, T_i}^2(T_i)}$ : the values of VIX futures *at their settlement dates* are readily available – this is sufficient for pricing VIX options.

Consider however a path-dependent payoff that depends on  $F_t^i$  for  $t < T_i$ , for example an option on a VIX ETF or ETN – see Section 7.7.3 below. Pricing such an option requires simulation of  $F_t^i$  at dates  $t < T_i$ . In the continuous forward variance model specified by (7.89),  $F_{t=T_i}^i$  is a function of  $X_{T_i}^1$  and  $X_{T_i}^2$  given by (7.84).  $F_t^i$  is given by:

$$F_t^i = E_t[F_{T_i}^i(X_{T_i}^1, X_{T_i}^2)] \quad (7.93)$$

---

<sup>36</sup>See Section 2.10.

While  $\hat{\sigma}_{VIX, T_i}^2(t)$  is readily available,  $F_t^i$  for  $t < T_i$  has to be computed by two-dimensional quadrature on  $X_{T_i}^1, X_{T_i}^2$ .<sup>37</sup>

In case  $F_t^i$  is needed for many dates – say on a daily basis – it is preferable to use a discrete forward variance model of the type discussed in Section 7.8.2 below. In these models VIX futures – rather than forward variances – are modeled directly.

### 7.7.3 Options on VIX ETFs/ETNs

VIX ETFs or ETNs typically maintain a rolling position in VIX futures.<sup>38</sup> Denoting by  $X_t$  the value of the ETF:

$$\frac{dX_t}{X_t} = rdt + \sum_{i, T_i > t} w_t^i \frac{dF_t^i}{F_t^i} \quad (7.94)$$

where the ETF's allocation strategy is expressed in the weight  $w_t^i$ . The VXX, one of the most popular ETNs, maintains a long position in the first and second nearby futures, so that the weighted duration of both futures is approximately 30 days. It would be most natural to set  $w_t^i \equiv w(T_i - t)$  with  $w(\tau)$  given by:

$$\begin{cases} w(\tau) = \frac{\tau}{\Delta} & \tau \in [0, \Delta] \\ w(\tau) = 2 - \frac{\tau}{\Delta} & \tau \in [\Delta, 2\Delta] \\ w(\tau) = 0 & \tau > 2\Delta \end{cases} \quad (7.95)$$

where  $\Delta$  is the interval between two VIX futures' expiries.  $w(\tau)$  appears in Figure 7.11. This allocation strategy results in the following dynamics for the VXX:

$$\frac{dX_t}{X_t} = rdt + \left[ w(T^{1st} - t) \frac{dF_t^{1st}}{F_t^{1st}} + (1 - w(T^{1st} - t)) \frac{dF_t^{2nd}}{F_t^{2nd}} \right]$$

where  $T^{1st}(t)$  is the expiry of the first nearby future  $F_t^{1st}$ , and  $T^{2nd}(t)$  that of the following future.

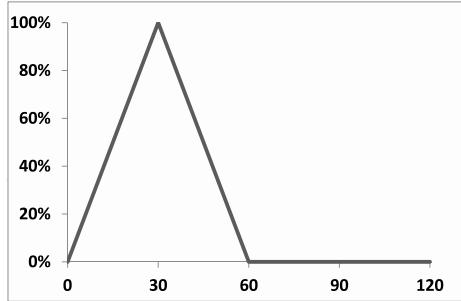
The VXX prospectus states that it is the *number* of futures that is proportional to  $\tau$  and  $\Delta - \tau$ , rather than the *notional* invested in both futures.<sup>39</sup>

$$\begin{aligned} w_{1st} &= \frac{(T^{1st} - t) F_t^{1st}}{(T^{1st} - t) F_t^{1st} + (\Delta - (T^{1st} - t)) F_t^{2nd}} \\ w_{2nd} &= 1 - w_{1st} \end{aligned}$$

<sup>37</sup>This is achieved efficiently by first finding the linear combination of  $X_{T_i}^1$  and  $X_{T_i}^2$  that accounts for the bulk of the variance of  $F_{T_i}^i$ . The resulting quadrature is then almost one-dimensional, especially since  $\Delta$  is small. The chosen algorithm should ensure that  $F_t^i$  is a martingale so that self-financing strategies that invest in VIX futures – such as VIX ETNs – have the correct forward.

<sup>38</sup>ETF stands for “exchange traded fund” – it is a fund whose shares trade much like stocks. ETN stands for “exchange traded note”. It is very similar to an ETF except there are no segregated assets backing the ETN: the holder of an ETN bears the credit risk of the note’s issuer. In theory (7.94) the drift of  $X_t$  should be supplemented with the credit spread of the issuer. Market appetite for borrowing the ETN – this is the case for the VXX – may however be such that the repo is large enough that it more than offsets the credit spread.

<sup>39</sup>See the prospectus of the VXX ETN at [www.ipathetn.com/static/pdf/vix-prospectus.pdf](http://www.ipathetn.com/static/pdf/vix-prospectus.pdf)



**Figure 7.11:**  $w(\tau)$  for the VXX ETN, as a function of  $\tau$  (days). We have made the simple assumption that VIX expiries are spaced 30 days apart.

If  $F_t^{1\text{st}} = F_t^{2\text{nd}}$  the weights become identical to those given by expression (7.95) for  $w(\tau)$ , which is the convention we use in what follows for the sake of simplicity.

The VXX smiles in Figure 7.13 below are computed in a Monte Carlo simulation that uses the proper convention – using either convention generates very similar prices, unless the term structure of VIX futures is unreasonably steep.

Consider now options on the VXX, which are listed. Can VXX options be priced off VIX smiles?

VIX implied volatilities quantify the realized volatility of VIX futures up to their expiry date. In contrast, SDE (7.94) for the VXX and the profile of  $w(\tau)$  in Figure 7.11 show that returns of the VXX are a weighted average of returns of two VIX futures that always have less than two months to expiry. Thus two natural questions arise:

- Assuming we are calibrated to VIX smiles, what is the impact of the distribution of the volatility of VIX futures?
- What is the impact of correlation between VIX futures?

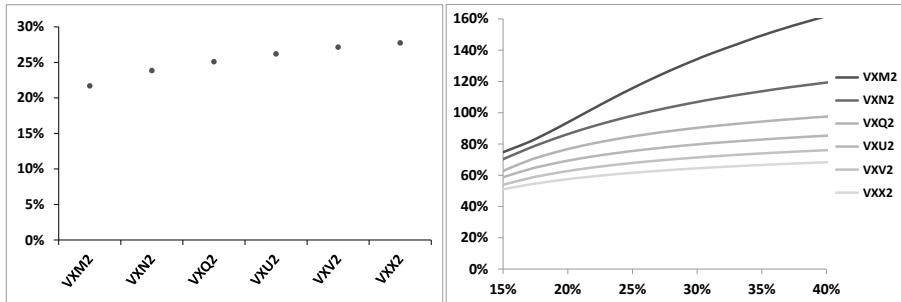
### Volatilities of VIX futures

In what follows we use the parametric model of Section 7.8.2 where VIX futures are modeled directly. We wish to assess the effect of different distributions of the realized volatility of each future throughout its life. To this end we use a one-factor model ( $\theta = 0$ ) so that the instantaneous correlation of VIX futures is constant, equal to 100%, and vary the value of  $k_1$  while remaining calibrated to VIX smiles.

We use VIX market data as of June 8, 2012; the calibrated smiles along with the levels of VIX futures appear in Figure 7.12.<sup>40</sup>

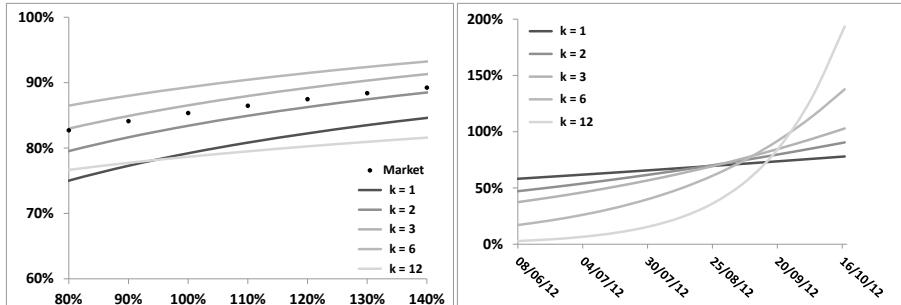
---

<sup>40</sup>Note that, as we vary  $k_1$ , calibration to VIX smiles remains identical. Indeed, as we vary  $k_1$ , the variance of  $x_{T_i}^{T_i}$  changes. However, calibration makes up for this through a change in  $\zeta_i$  so that the variance of  $\omega_i x_{T_i}^{T_i}$  is unchanged. From (7.104),  $F_{T_i}^i$  is a function of the Gaussian variable  $\omega_i x_{T_i}^{T_i}$ : if its variance is unchanged as we vary  $k_1$ , so is the density of  $F_{T_i}^i$ . As a result, calibrated values of  $\gamma_i, \beta_i$  do not depend on  $k_1$  and the VIX smiles generated by the model do not depend on  $k_1$  either.



**Figure 7.12:** Left: VIX futures as of June 8, 2012 for expiries ranging from June 20 (VXM2) to November 21 (VXX2). Right: Smiles of VIX futures.

Figure 7.13 shows the VXX smiles generated by different values of  $k_1$  for the listed maturity of December 21, 2012 – together with the market smile.<sup>41</sup> It also shows the instantaneous volatility of VIX future VXV2, which expires on October 17, 2012. This instantaneous volatility,  $\sigma_t$ , is obtained as the square root of the expectation of the instantaneous variance:  $\sigma_t^2 = E[\left(\frac{dF_t^i}{F_t^i}\right)^2]$ . This expectation is easily computed by Gaussian quadrature on  $x_t^{T_i}$ .



**Figure 7.13:** Left: VXX smiles as of June 8, 2012 for the December 21st maturity, for different values of  $k_1$ , compared to the market smile. Right: instantaneous volatility of the VIX future expiring on October 17, 2012 (VXV2) for different values of  $k_1$ .

Consider the case  $k_1 = 1$ . For this (low) value of  $k_1$  the instantaneous volatilities of VIX futures have little term structure and are distributed more or less evenly over their lives (see Figure 7.13 for the case of the VXV2 future, whose ATM im-

<sup>41</sup>Listed options on the VXX are American. However, because of the low level of interest rates, the fact that the VXX pays no dividends and the high volatility of the VXX, they can in practice be considered as European for a wide range of strikes around the money.

plied volatility is 70%) with no particular concentration on the final two months before expiry. The resulting VXX implied volatilities are lower than market implied volatilities, which is not surprising.

Conversely, consider the case  $k_1 = 12$ . As the graph in the right-hand side of Figure 7.13 shows, the volatility of VIX futures is now concentrated right before expiry. However, as is clear from Figure 7.11, this is where their weight in the VXX vanishes: again we expect low VXX implied volatilities, which the left-hand graph in Figure 7.13 confirms.

One can see that the highest implied volatilities are obtained for  $k_1 \simeq 6$ , with  $k_1 = 2, k_1 = 3$  generating VXX smiles that most closely approximate the market VXX smile.<sup>42</sup>

### Correlations of VIX futures

We have thus far used 100% correlation for simplicity. The realized correlation of the first two nearby futures typically ranges from 85% to 100%. How does correlation impact VXX implied volatilities?

Let us make the simple assumption that the volatilities of the two nearby futures  $F^i$  and  $F^{i+1}$  are identical and constant, equal to  $\sigma$ , and that their correlation is constant, equal to  $\rho$ . The instantaneous volatility of the VXX,  $\sigma_X(t)$ , is given for  $t \in [T_{i-1}, T_i]$  by:

$$\sigma_X^2(t) = \sigma^2 (w^2(T_i - t) + w^2(T_{i+1} - t) + 2\rho w(T_i - t)w(T_{i+1} - t))$$

where  $w(\tau)$  appears in Figure 7.11. If weights  $w(T_i - t)$  and  $w(T_{i+1} - t)$  were constant, equal to  $\frac{1}{2}$ , we would get the standard formula for the volatility of an equally weighted basket:  $\sigma_X^2 = \frac{1+\rho}{2}\sigma^2$ .

In our case the product  $w(T_i - t)w(T_{i+1} - t)$  vanishes both for  $t = T_{i-1}$  and for  $t = T_i$ : the effect of the cross term will be comparatively smaller. Integrating  $\sigma_X^2(t)$  over  $[T_{i-1}, T_i]$ :

$$\hat{\sigma}_X^2 = \frac{1}{\Delta} \int_{T_{i-1}}^{T_i} \sigma_X^2(t) dt = \frac{2+\rho}{3}\sigma^2$$

Consider the two cases  $\rho_{\min} = 80\%$ ,  $\rho_{\max} = 100\%$ . The ratio of implied volatilities  $\frac{\hat{\sigma}_X^{\max}}{\hat{\sigma}_X^{\min}}$  is then equal to  $\sqrt{\frac{2+\rho_{\max}}{2+\rho_{\min}}} = 103.5\%$ : a difference of correlation of 20 points gives rise to a variation of VXX implied volatilities of a few points only.

### Conclusion

In conclusion, VIX smiles do not provide enough information for confining prices of options on VIX ETF(N)s sufficiently. The example of the VXX demonstrates that, even though consistency with VIX smiles is enforced, implied volatilities of the VXX are very dependent on assumptions about how the volatility of each VIX future

---

<sup>42</sup>For  $k_1 = 3$  the volatility-of-volatility adjustment factors  $\zeta_i$  become (serendipitously) essentially identical for all VIX futures, thus making the model time-homogeneous.

is distributed throughout its life, and also depend moderately on the correlation structure of VIX futures.

Thus VXX options are not redundant instruments – they supply information on the volatility distribution of VIX futures.<sup>43</sup>

#### 7.7.4 Consistency of S&P 500 and VIX smiles

From market prices of VIX futures and options we can derive forward log-contract volatilities for the S&P 500 index. The square of the forward volatility for time interval  $[T_i, T_i + \Delta]$  is given by:

$$\hat{\sigma}_{T_i, T_{i+\Delta}}^2(t) = E_t [ (F_{T_i}^i)^2 ] \quad (7.96)$$

We know from Section 3.1.3 that any European payoff on  $F_{T_i}^i$  can be replicated by a static position consisting of cash, forwards (or futures) and vanilla options on  $F_{T_i}^i$ . The decomposition in (3.6) applied to the function  $f(S) = S^2$  reads:

$$S^2 = S_*^2 + 2S_*(S - S_*) + 2 \int_0^{S_*} (K - S)^+ dK + 2 \int_{S_*}^{\infty} (S - K)^+ dK \quad (7.97)$$

where  $S_*$  is arbitrary. We now apply this identity to  $S = F_{T_i}^i$  with  $S_* = F_t^i$  and translate this equality of payoffs in an equality of prices.

Adding up the (undiscounted) prices of the different components in the right-hand side of (7.97) yields the following consistency condition relating S&P 500 forward volatilities to market prices of VIX futures and options:

$$\hat{\sigma}_{T_i, T_{i+\Delta}}^2(t) = (F_t^i)^2 + 2 \int_0^{F_t^i} \mathcal{P}_t^{Ki} dK + 2 \int_{F_t^i}^{\infty} \mathcal{C}_t^{Ki} dK \quad (7.98)$$

where  $\mathcal{P}_t^{Ki}$  (resp.  $\mathcal{C}_t^{Ki}$ ) are undiscounted market prices of put (resp. call) options on the VIX of maturity  $T_i$ . We have used the fact that the price of a payoff linear in  $(F_T^i - F_t^i)$  – the second piece in (7.97) – vanishes. In contrast to the replication of the log contract, the densities of calls and puts on the VIX are constant, so that  $\hat{\sigma}_{T_i, T_{i+\Delta}}^2(t)$  has little dependence on the exact cutoff used in the integrals in the above expression.

#### Log-contract versus VS volatility

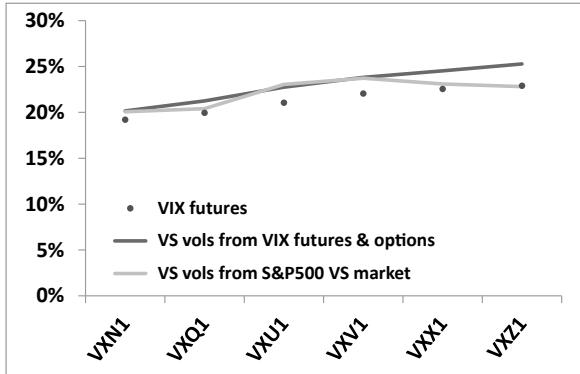
Because of the definition of the settlement value of VIX futures,  $\hat{\sigma}_{T_i, T_{i+\Delta}}$  as defined in (7.96) is a log-contract volatility. (7.98) thus expresses an identity between market prices of VIX instruments and of log-contracts, which, unlike VSs, do not trade.

---

<sup>43</sup>Note the similarity with the issue of pricing interest rate swaptions in a model calibrated on LIBOR caps/floors.

We will thus assume that the difference between  $\widehat{\sigma}_T$  and  $\widehat{\sigma}_{VS,T}$  is small, so that we can use VSs *in lieu* of log contracts.<sup>44</sup>

Figure 7.14 shows  $F_t^i$ ,  $\widehat{\sigma}_{T_i, T_{i+\Delta}}(t)$  as given by replication on the VIX market through (7.98) and  $\widehat{\sigma}_{T_i, T_{i+\Delta}}(t)$  as generated by interpolation of market quotes of S&P 500 VS implied volatilities observed on January 14, 2011.



**Figure 7.14:** VIX futures, VS volatilities as generated by (7.98) and VS volatilities as derived from the S&P 500 VS market, all in 365 convention.

The contribution from the time values of VIX calls and puts in (7.98) is of the order of one to two points of volatility. Notice how VS volatilities derived either from the S&P 500 VS market or from the VIX market do not coincide. This suggests an arbitrage strategy: imagine that a particular forward VS volatility as derived from the VIX market through (7.98) lies higher than its counterpart derived from the regular S&P 500 VS market. We sell VIX futures and options in the proportions expressed by (7.98) and buy a *forward*  $[T_i, T_i + \Delta]$  VS. At  $t = T_i$ , upon settlement of VIX futures and options we unwind the regular VS position – which by then is a spot-starting VS – at an implied volatility equal to the settlement value of the expiring VIX future: the P&L of this strategy is  $\widehat{\sigma}_{T_i, T_{i+\Delta}}^2(t)_{VIX\ mkt} - \widehat{\sigma}_{T_i, T_{i+\Delta}}^2(t)_{VS\ mkt}$ . Practically, however, setting up this strategy is not as straightforward:

- Short-maturity forward variance swaps on the S&P 500 index are not liquid – they are built by combining a long position in a VS of maturity  $T_i + \Delta$  with a short position in a VS of maturity  $T_i$ . Even though we may be charged a bid/offer spread on one leg only, the resulting spread for the  $[T_i, T_i + \Delta]$  forward VS volatility will be sizeable. Moreover, S&P 500 VS contracts only trade for maturities corresponding to the expiries of listed S&P 500 options –

<sup>44</sup>As discussed in Section 5.5, there are valid reasons why  $\widehat{\sigma}_T$  and  $\widehat{\sigma}_{VS,T}$  could be different. Typically, because sellers of VSs would lose on large drawdowns of the underlying index, we expect that  $\widehat{\sigma}_{VS,T} > \widehat{\sigma}_T$ . It so happens that usually – see below – equivalent log-contract volatilities derived from the VIX market lie higher than  $\widehat{\sigma}_{VS,T}$ . In case  $\widehat{\sigma}_{VS,T} > \widehat{\sigma}_T$ , the discrepancy with the S&P 500 market is even stronger.

the third Friday of each month. While  $T_i + \Delta$  is an S&P 500 expiry by definition of the VIX index,  $T_i$  is not as it falls 30 days before, on a Wednesday. Forward VS volatilities that can practically be traded on the VS market are either two days shorter or five days longer than the corresponding VS volatilities synthesized from the VIX market. Considering that a 30-day interval comprises approximately 20 returns, carrying an open gamma position on two – or three<sup>45</sup> – squared returns entails appreciable risk. It is then preferable to group together packages corresponding to several adjacent VIX futures – say three – so that bid/offer costs and risks are mitigated.

- VIX options are only available for discrete strikes. The continuous portfolio of VIX call and put options in (7.98) is in practice replaced by a discrete portfolio. Quantities of VIX futures and options are then determined so as to achieve the most expensive sub-replication of the  $(F_{T_i}^i)^2$  payoff – when the forward VS volatility derived from the VIX market is higher than that of the S&P 500 VS market – or the cheapest super-replication of  $(F_{T_i}^i)^2$  in the opposite case.<sup>46</sup>
- At  $t = T_i$  we need to unwind the forward VS position at an implied VS volatility that is exactly equal to  $\hat{\sigma}_{VIX, T_i}$ . This is achieved by selling vanilla options – which will subsequently be delta-hedged until  $T_i + \Delta$  – on the S&P 500 in exactly the same quantities and for the same prices used in the calculation of the settlement of  $F^i$  by the CBOE. This is possible as VIX futures settle at the open of the S&P 500 options' market: orders can be placed and are executed at the open at the same prices that are used for the calculation of  $F_{T_i}^i$ . Still we know from Section 5.3.7 that the package consisting of a VS together with its offsetting vanilla replication is not risk-free. It is thus best to unwind the arbitrage position before  $T_i$ , should an opportunity arise.

Violations of (7.98) can then only be arbitAGED by S&P 500 volatility market makers.

In the author's experience, arbitrage opportunities that used to arise involved, most often than not, VIX-synthesized VS volatilities lying higher than their S&P 500 VS counterparts. As of the time of writing, these opportunities seem to hardly occur anymore.

Are there other structural connections between S&P 500 and VIX smiles that could be practically arbitAGED when violated? For examples are implied volatilities of (a) S&P 500 put options, (b) VIX call options related?

In [36] Stefano de Marco and Pierre Henry-Labordère consider the sub- and super-replication of VIX options using VIX futures, the S&P 500 index and S&P 500 options. They derive optimality conditions, which can be solved numerically,

---

<sup>45</sup>There are three business days from Friday to Wednesday.

<sup>46</sup>The most expensive sub-replicating and cheapest super-replicating portfolios are determined with the simplex algorithm. In the latter case, because of the convexity of the parabola, super-replication only holds for a limited range of values of  $F_{T_i}^i$ .

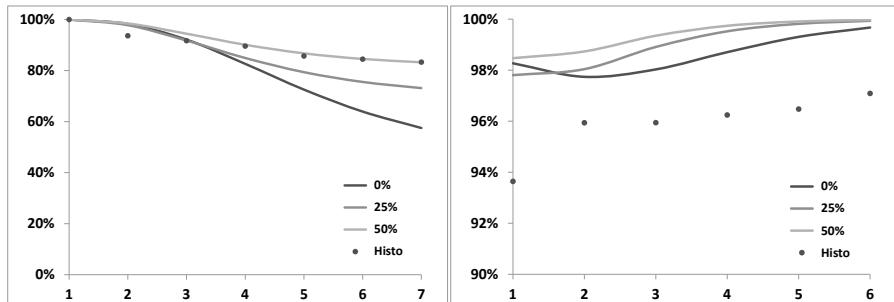
and also obtain analytical non-optimal upper and lower bounds; with the upper bound optimal under a condition involving the S&P 500 smiles for maturities  $T_i$  and  $T_{i+1}$ . These bounds are widely spaced, but more interestingly it does not seem that there is much more information to be extracted from the S&P 500 smile, other than log-contract implied volatilities.

See also reference [77], where Andrew Papanicolaou derives a bound on the moment-generating function of a squared VIX future from prices of moments of the S&P 500 index.

### 7.7.5 Correlation structure of VIX futures

How do correlations of VIX futures generated by the two-factor model compare with those observed in reality?

Consider the first 7 futures  $F_t^i, i = 1 \dots 7$ . Figure 7.15 shows instantaneous correlations  $\rho(F_t^1, F_t^i)$  (left-hand graph) and  $\rho(F_t^i, F_t^{i+1})$  (right-hand graph) in the two-factor model, for an observation date  $t$  that lies 15 days before the expiry of the first VIX future; this represents an “average” correlation level. The dots correspond to historical correlations evaluated from February 16, 2010 to February 15, 2012 – from which we have excluded roll dates.



**Figure 7.15:** Left: correlations  $\rho(F_t^1, F_t^i)$  of the first 7 VIX futures with the first future in the two-factor model for different values of the factor/factor correlation, and as observed in reality (dots), as a function of  $i$ . Right: correlations  $\rho(F_t^i, F_t^{i+1})$  of contiguous futures.

We have used three parameter sets for the two-factor model, all calibrated to the benchmark form (7.40) for  $\nu_T^B(t)$ , page 228, with  $\sigma_0 = 100\%$ ,  $\tau_0 = 3$  months,  $\alpha = 0.4$ , characterized by different levels of correlation  $\rho_{12}$  between the two factors:  $\rho_{12} = 0\%, 25\%, 50\%$ . The set with  $\rho = 0\%$  is Set II in Table 7.1.

It is apparent that correlations of the first future with other futures are acceptably captured with  $\rho_{12} = 50\%$ . The right-hand side graph highlights however that correlations of adjacent futures are then systematically higher in the two-factor model than in reality. In the two-factor model, VIX futures with long expiries have almost 100% correlation.

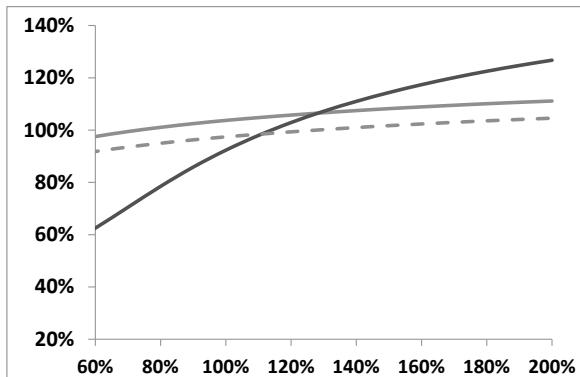
This is due to the fact, already pointed out in Section 7.4.2, page 229, that in the two-factor model correlations between forward variances involve one time scale only:  $\frac{1}{k_1 - k_2}$ .

We cannot realistically expect to achieve a good fit of historical correlations of VIX futures by employing one single time scale. Regaining some flexibility with respect to the correlation structure would be a valid motivation for introducing a third factor in the model.

### 7.7.6 Impact on smiles of options on realized variance

Consider an option on realized variance starting on June 14, 2011 and maturing in January 20, 2012, that is 30 days after the expiry of the VIXZ1 future, and let us price it using Set II parameters and the corresponding volatility-of-volatility smile parameters  $\gamma^i, \beta^i, \zeta^i$  in Figure 7.9.

Figure 7.16 shows the implied volatility of this option as a function of *volatility moneyness*, computed by inverting the Black-Scholes formula with the (driftless) underlying equal to the square of the VS volatility for the option's maturity. These results are generated by a Monte Carlo simulation of the two-factor model with conditionally Gaussian returns for  $\ln S$ . Two other sets of implied volatilities are plotted as well – see caption.



**Figure 7.16:** Implied volatilities of options on realized variance. The inception date is June 14, 2011 and the maturity is January 20, 2012. Implied volatilities of the realized variance are shown as a function of *volatility moneyness* using Set II parameters and three configurations for  $\gamma^i, \beta^i, \zeta^i$ : dark line: values in Figure 7.9; light line: same but setting  $\gamma^i = 0$ ; dotted line: setting  $\gamma^i = 0, \zeta^i = 1$ .

Comparison of the dark and light curves in Figure 7.16 shows that the impact of the smile of forward variances is by no means small. The continuous light curve shows the implied volatilities as generated by Set II parameters with  $\gamma^i = 0$ : it incorporates the volatility-of-volatility adjustment factors  $\zeta^i$  as derived from cali-

bration on the VIX smiles. The dotted line is obtained using Set II parameters only:  $\gamma^i = 0$ ,  $\zeta^i = 1$ . Comparison of the light continuous and dotted lines indeed confirms that  $\zeta$  is a simple adjustment factor for volatilities of forward variances: implied volatilities are altered approximately uniformly.

While Figure 7.16 is an indication of the effect of the smile of forward variances on the price of an option on realized variance, the size of this effect is highly model-dependent.

Calibration on the smiles of VIX options sets the value of the expected integrated variance of VIX futures,  $\int_t^{T_i} \sigma_u^2 du$ , but not its distribution. Furthermore, the correlation of VIX futures is left undetermined, yet it is instrumental in determining the volatility of VS volatilities for the option's maturity, which is the main determinant of its price.

### 7.7.7 Impact on the vanilla smile

The effect of VIX smiles on the vanilla smile of the underlying itself is covered in Section 8.10.

---

## 7.8 Discrete forward variance models

So far in this chapter we have modeled instantaneous forward variances  $\xi_t^T$ . We parametrize the two-factor model so that it generates:

- the desired term structure of volatilities of volatilities,
- the desired term structure of ATMF skew – this is discussed in Chapter 8 – or the desired level of future ATMF skew, if we have forward-start options in mind.

As we vary volatility-of-volatility parameter  $\nu$ , we alter volatilities of volatilities in the model, but also the ATMF skew of the vanilla smile, and also future ATMF skews.

Likewise, changing spot/volatility correlations alters both the ATMF skew of the vanilla smile, and the skew of forward-starting options. The fact that we do not have independent handles on:

- the vanilla smile,
- the smile of forward-start options,
- volatilities of volatilities,

is a typical shortcoming of continuous forward variance models, a particularly worrisome one when one needs to risk-manage complex path-dependent options that are subject to these three types of risk.

In what follows, we present models that allow for separation of these risks and permit an assessment of their individual impact on exotic option prices; they were first presented in [9].<sup>47</sup>

We refer the reader to the discussion of the risks of forward-start options in Section 3.1.6 of Chapter 3, page 111, if she/he has not read it yet.

Most exotic payoffs that have forward-smile risk involve returns for a set time scale, for example monthly or quarterly returns. One typical example is an accumulator: take 12 monthly returns of the S&P 500 index, cap them individually at, say 3%, then sum them up and pay this sum, floored at zero, as an annual coupon.

A discrete forward variance model is tied to a particular schedule. The latter is defined in the term sheet of the exotic option at hand, thus dates  $T_i$  in the schedule are typically uniformly spaced, say by a month or a quarter. Nonetheless, we assume an arbitrary schedule in what follows.

Our aim is to separately control the future smiles over each individual time interval  $[T_i, T_{i+1}]$  and also the vanilla smile, in addition to the term structure of volatilities of volatilities.

The model is built in two stages:

- first define a dynamics for discrete forward variances over intervals  $[T_i, T_{i+1}]$ ,
- then specify a dynamics for  $S_t$  over each interval.

A benefit of discrete models is that VIX futures can be modeled directly, rather than forward variances – see Section 7.8.2 below. Our first step can be replaced with:

- first define a dynamics for VIX futures.

### 7.8.1 Modeling discrete forward variances

Let  $\xi_t^i$  be the discrete forward VS variance for interval  $[T_i, T_{i+1}]$ . It is similar to the continuous forward variances employed so far, except it is defined as a finite difference rather than a derivative:

$$\xi_t^i \equiv \hat{\sigma}_{T_i, T_{i+1}}^2(t) = \frac{(T_{i+1} - t)\hat{\sigma}_{T_{i+1}}^2(t) - (T_i - t)\hat{\sigma}_{T_i}^2(t)}{T_{i+1} - T_i}$$

---

<sup>47</sup>There cannot be complete disconnection between the vanilla smile and future smiles. Consider vanilla smiles for maturities  $T_1$  and  $T_2 > T_1$  and the corresponding densities  $\rho_1(S_1)$ ,  $\rho(S_2)$ . Given  $\rho_1$  and  $\rho_2$ , the transition density  $\rho_{12}(S_2|S_1, \bullet)$ , which determines future smiles generated by the model – where  $\bullet$  stands for state variables other than  $S$  – cannot be chosen arbitrarily. It has to comply with the Chapman-Kolmogorov condition:

$$\rho_2(S_2) = \int E[\rho_{12}(S_2|S_1, \bullet) | S_1] \rho_1(S_1) dS_1, \quad \forall S_2$$

Still, the presence of the  $\bullet$  state variables ( $X_t^1, X_t^2$ , in the two-factor model) affords considerable freedom in selecting  $\rho_{12}$ . We refer the reader to Section 3.1.7 of Chapter 3, page 113, for examples of how loosely cliquet prices are constrained by the vanilla smile.

where  $\widehat{\sigma}_T(t)$  is the VS volatility for maturity  $T$ , and  $\widehat{\sigma}_{T_i, T_{i+1}}(t)$  the forward VS volatility for interval  $[T_i, T_{i+1}]$ , observed at  $t$ . In the diffusive models we work with, implied volatilities of VSs and log-contracts are identical, thus  $\widehat{\sigma}_{T_i, T_{i+1}}(t)$  is also the implied volatility at  $t$  of the payoff that pays  $\ln(\frac{S_{T_{i+1}}}{S_{T_i}})$  at  $T_{i+1}$ .

As with continuous variance models, just because we use forward variances as basic building blocks does not mean we necessarily use VSs as hedge instruments. Our model can be calibrated to a term structure of implied volatilities for a given moneyness for maturities  $T_i$  – for example ATMF volatilities.

The corresponding vanilla options are then our hedge instruments, along with the spot, and the carry P&L of a hedged position is of the genuine gamma/theta form. We refer the reader to the discussion in Section 7.5 in the context of continuous models, whose conclusions apply to discrete models as well.

Just as their continuous counterparts, the  $\xi_T^i$  are driftless.<sup>48</sup> We can thus recycle the two-factor model and, mimicking (7.28), write the SDE of  $\xi_t^i$  as:

$$\begin{aligned} d\xi_t^i &= (2\nu_i)\xi_t^i \alpha_{\theta_i} \left( (1 - \theta_i) e^{-k_1(T_i - t)} dW_t^1 + \theta_i e^{-k_2(T_i - t)} dW_t^2 \right) \quad (7.99) \\ \alpha_{\theta_i} &= 1/\sqrt{(1 - \theta_i)^2 + \theta_i^2 + 2\rho_{12}\theta_i(1 - \theta_i)} \end{aligned}$$

where index  $i$  for parameters  $\theta$  and  $\nu$  keeps track of the forward variance  $\xi^i$  they apply to.

While  $\theta$  and  $\nu$  depend on  $i$ , we use the same values for  $k_1, k_2, \rho_{12}$  for all intervals – otherwise we lose the two-dimensional Markov representation of the  $\xi_t^i$ . We also use the same values as in the continuous model, so that discrete and continuous versions of the model can be mapped onto another – see below.

The solution of (7.99) reads:

$$\begin{aligned} \xi_t^i &= \xi_0^i e^{\omega_i x_t^{T_i} - \frac{\omega_i^2}{2} \chi(t, T_i)} \quad (7.100) \\ x_t^{T_i} &= \alpha_{\theta_i} \left[ (1 - \theta_i) e^{-k_1(T_i - t)} X_t^1 + \theta_i e^{-k_2(T_i - t)} X_t^2 \right] \end{aligned}$$

with  $\omega_i = 2\nu_i$ . The driftless processes  $x_t^{T_i}$  are defined in (7.30) and  $\chi(t, T)$  is defined in (7.35), page 227.

### Mapping a continuous to a discrete model

The spacing between two successive dates  $T_i, T_{i+1}$  is specific to each exotic payoff – it is different for different payoffs. Still, risks of the same nature should be priced at the same level across the book, for example volatility-of-volatility risk. We thus need to parametrize our discrete model so that some model features remain unchanged with respect to its continuous counterpart.

---

<sup>48</sup>The exposure to  $\xi^i$  can be delta-hedged by going long  $T_{i+1}$  VSs of maturity  $T_{i+1}$  and short  $T_i$  VSs of maturity  $T_i$ , with no cash borrowing or lending involved.

Consider the forward volatility  $\widehat{\sigma}_{T_i, T_{i+1}}(t)$ . It is given, in the discrete and continuous model, respectively, by:

$$\begin{cases} \widehat{\sigma}_{T_i, T_{i+1}}(t) = \sqrt{\xi_t^i} \\ \widehat{\sigma}_{T_i, T_{i+1}}(t) = \sqrt{\frac{1}{T_{i+1}-T_i} \int_{T_i}^{T_{i+1}} \xi_t^\tau d\tau} \end{cases}$$

The SDE of forward variance  $\widehat{\sigma}_{T_i, T_{i+1}}^2(t)$  is given, in both models, by:

$$\frac{d\widehat{\sigma}_{T_i, T_{i+1}}^2}{\widehat{\sigma}_{T_i, T_{i+1}}^2} = 2\nu_i \alpha_{\theta_i} \left( (1 - \theta_i) e^{-k_1(T_i - t)} dW_t^1 + \theta_i e^{-k_2(T_i - t)} dW_t^2 \right) \quad (7.101a)$$

$$\frac{d\widehat{\sigma}_{T_i, T_{i+1}}^2}{\widehat{\sigma}_{T_i, T_{i+1}}^2} = 2\nu \alpha_\theta \left( (1 - \theta) A_i^1(t) e^{-k_1(T_i - t)} dW_t^1 + \theta A_i^2(t) e^{-k_2(T_i - t)} dW_t^2 \right) \quad (7.101b)$$

where  $A_i^n(t)$  reads:

$$A_i^n(t) = \frac{\int_{T_i}^{T_{i+1}} \xi_t^\tau e^{-k_n(\tau - T_i)} d\tau}{\int_{T_i}^{T_{i+1}} \xi_t^\tau d\tau}$$

(7.101a) stems from (7.99) directly while (7.101b) is adapted from the corresponding expression (7.36), page 227, for a spot-starting VS volatility.

The two SDEs in (7.101) cannot be identical in both models, for all  $t$ , for all configurations of the forward variance curve  $\xi_t^\tau$ , but let us demand that they coincide for all  $t$ , for forward variances  $\xi_t^\tau$  equal to their initial values  $\xi_0^\tau$ , that is with  $A_i^n$  equal to  $A_i^n(0)$ .

A quick glance at the right-hand sides of (7.101a) and (7.101b) shows this is possible only if  $k_1, k_2, \rho$  are identical in both models, where  $\rho$  is the correlation between  $W_t^1$  and  $W_t^2$ .

The conditions on  $\nu_i, \theta_i$  read:

$$\begin{aligned} \nu_i \alpha_{\theta_i} (1 - \theta_i) &= \nu \alpha_\theta (1 - \theta) A_i^1(0) \\ \nu_i \alpha_{\theta_i} \theta_i &= \nu \alpha_\theta \theta A_i^2(0) \end{aligned}$$

This yields:

$$\begin{cases} \theta_i = \frac{\theta A_i^2(0)}{\theta A_i^2(0) + (1 - \theta) A_i^1(0)} \\ \nu_i = \nu \frac{\alpha_\theta \theta}{\alpha_{\theta_i} \theta_i} A_i^2(0) \end{cases} \quad (7.102)$$

Provided  $\theta_i$  and  $\nu_i$  are given by (7.102), the dynamics of forward volatilities  $\widehat{\sigma}_{T_i, T_{i+1}}(t)$  is identical in the discrete and continuous versions of the model – for forward variances  $\xi_t^\tau$  equal to their initial values  $\xi_0^\tau$ . Consequently:

- instantaneous volatilities of VS volatilities of maturities  $T_i$

- instantaneous correlations of forward VS volatilities  $\hat{\sigma}_{T_i, T_{i+1}}$  and  $\hat{\sigma}_{T_j, T_{j+1}}$

are identical as well, in both models, for forward variances  $\xi_t^\tau$  equal to their initial values  $\xi_0^\tau$ .

Our criterion for mapping consists in requiring that the SDEs in (7.101) match  $\forall t$  for  $\xi_t^\tau = \xi_0^\tau$ . We now give an illustration of the fact that, indeed, the dynamics of the  $\xi^i$  in both discrete and continuous models is very similar.

### An example

Consider the case of a flat term structure of VS volatilities, equal to 20%. In this case

$$A_i^n(0) = \frac{1 - e^{-k_n(T_{i+1} - T_i)}}{k_n(T_{i+1} - T_i)}$$

$\theta_i$  and  $\nu_i$  then only depend on  $T_{i+1} - T_i$ . They are shown in Figure 7.17.

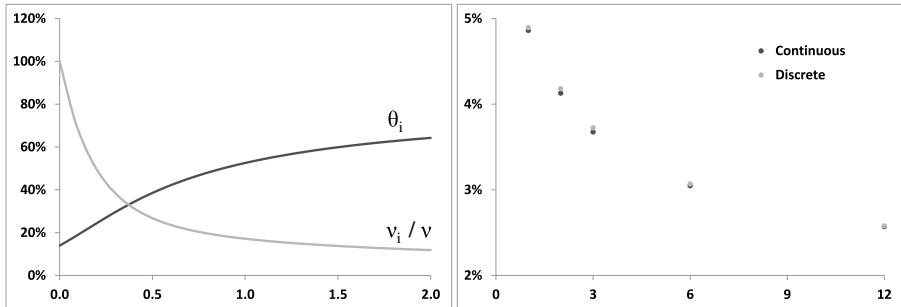
Figure 7.17 also shows prices of a VS ATM swaption, whose payoff is

$$\frac{1}{2\hat{\sigma}_{T, T+\Delta}(0)} \left( \hat{\sigma}_{T, T+\Delta}^2(T) - \hat{\sigma}_{T, T+\Delta}^2(0) \right)^+$$

for various values of  $\Delta$ .

The expiry  $T$  of the swaption is 1 year, and the discrete model is mapped for each value of  $\Delta$  using (7.102) with  $T_i = T$ ,  $T_{i+1} = T + \Delta$ .

Parameters of the continuous model appear in Table 7.7. They are chosen so as to generate a term-structure of VS volatilities that closely fits benchmark (7.51), page 239, with  $\alpha = 0.6$ ,  $\tau_0 = 3$  months,  $\sigma_0 = 125\%$ .



**Figure 7.17:** Left:  $\theta_i$  and  $\frac{\nu_i}{\nu}$  as a function of  $\Delta$  (years). Right: prices of a 1-year ATM VS swaption in both continuous and discrete models, as a function of the maturity of the underlying VS volatility (months).

Prices in both discrete and continuous models are very similar, thus illustrating that with mapping (7.102), both models generate very similar dynamics for discrete forward VS variances.

$\nu$	$\theta$	$k_1$	$k_2$	$\rho_{12}$
310%	0.139	8.59	0.47	0%

Table 7.7: Parameters of the continuous two-factor forward variance model.

### Smiling discrete forward variances

With the dynamics in (7.99), forward volatilities  $\widehat{\sigma}_{T_i, T_{i+1}}$  are lognormal, thus VS swaption smiles are flat.<sup>49</sup> For the sake of generating upward-sloping smiles, we can use the same ansatz as in (7.89), page 264, and replace (7.100) with:

$$\xi_t^i = \xi_0^i \left( (1 - \gamma_i) e^{\omega_i x_t^{T_i} - \frac{\omega_i^2}{2} \chi(t, T_i)} + \gamma_i e^{\beta_i \omega_i x_t^{T_i} - \frac{(\beta_i \omega_i)^2}{2} \chi(t, T_i)} \right)$$

with the normalization in (7.90):

$$\omega_i = \frac{2\nu_i}{(1 - \gamma_i) + \beta_i \gamma_i} \zeta_i$$

#### 7.8.2 Direct modeling of VIX futures

VIX instruments are introduced in Section 7.7, in the context of the two-factor (continuous) forward variance model.

VIX futures  $F_t^i$  at dates  $t < T_i$  are accessible in the continuous model, at the cost of a two-dimensional quadrature – this is explained in Section 7.7.2. For payoffs requiring frequent observations of VIX futures, using a model where VIX futures – rather than variances – are modeled directly is preferable.

We now discuss this particular breed of discrete forward variance models – they are Markov-functional models for VIX futures.

Denote by  $T_i$  the expiry of future  $F_t^i$ . At expiry, a VIX future is equal to the log-contract implied volatility for maturity  $T_i + \Delta$ , where  $\Delta = 30$  days. Borrowing the notations of Section 7.7.1:

$$F_{T_i}^i = \widehat{\sigma}_{\text{VIX}, T_i}(T_i) = \widehat{\sigma}_{T_i, T_i + \Delta}(T_i)$$

VIX futures expire on Wednesdays, 30 days before the expiry of listed S&P 500 options, thus two consecutive VIX expiries are spaced either (a) 28 days or (b) 35 days apart. In case (a),  $T_i + \Delta > T_{i+1}$ : the forward VS volatilities  $\widehat{\sigma}_{T_i, T_i + \Delta}$  and  $\widehat{\sigma}_{T_{i+1}, T_{i+1} + \Delta}$  that underlie, respectively, futures  $F^i$  and  $F^{i+1}$  overlap. In case (b),  $T_i + \Delta < T_{i+1}$ , and there is no overlap. We assume that  $T_{i+1} - T_i = \Delta$ .<sup>50</sup>

<sup>49</sup>In the continuous version of the model, instantaneous forward variances  $\xi_t^T$  are lognormal, but discrete forward variances, hence forward VS volatilities, are not. Figure 7.4, page 232, highlights the resulting slight positive skew of VS swaptions.

<sup>50</sup>Practically, in case (b), we set  $\widehat{\sigma}_{T_i, T_{i+1}}(T_i) \equiv \widehat{\sigma}_{T_i, T_i + \Delta}(T_i)$ , whereas in case (a), when both underlying forward variances overlap, we set:  $\widehat{\sigma}_{T_i + \Delta, T_{i+1} + \Delta}(T_{i+1}) \equiv \sqrt{\frac{\Delta \widehat{\sigma}_{T_{i+1}, T_{i+1} + \Delta}^2(T_{i+1}) - (T_i + \Delta - T_{i+1}) \widehat{\sigma}_{T_i, T_i + \Delta}^2(T_i)}{T_{i+1} - T_i}}$ .

### A parametric form

Since VIX futures are driftless – just like forward variances  $\xi^\tau$  – we can use the discrete two-factor model directly on VIX futures and, mirroring (7.99) write:

$$\begin{aligned} dF_t^i &= \nu_i F_t^i \alpha_{\theta_i} \left( (1 - \theta_i) e^{-k_1(T_i - t)} dW_t^1 + \theta_i e^{-k_2(T_i - t)} dW_t^2 \right) \quad (7.103) \\ \alpha_{\theta_i} &= 1 / \sqrt{(1 - \theta_i)^2 + \theta_i^2 + 2\rho_{12}\theta_i(1 - \theta_i)} \end{aligned}$$

where we use  $\nu_i$  instead of  $2\nu_i$  since  $F_t^i$  is a volatility rather than a variance.

The solution of (7.103) is given by:

$$\begin{aligned} F_t^i &= F_0^i e^{\omega_i x_t^{T_i} - \frac{\omega_i^2}{2} \chi(t, T_i)} \\ x_t^{T_i} &= \alpha_{\theta_i} \left[ (1 - \theta_i) e^{-k_1(T_i - t)} X_t^1 + \theta_i e^{-k_2(T_i - t)} X_t^2 \right] \end{aligned}$$

with  $\omega_i = \nu_i$ .

Upward-sloping smiles can be generated by using the same ansatz as in (7.89):

$$F_t^i = F_0^i \left( (1 - \gamma_i) e^{\omega_i x_t^{T_i} - \frac{\omega_i^2}{2} \chi(t, T_i)} + \gamma_i e^{\beta_i \omega_i x_t^{T_i} - \frac{(\beta_i \omega_i)^2}{2} \chi(t, T_i)} \right) \quad (7.104)$$

with:

$$\omega_i = \frac{\nu_i}{(1 - \gamma_i) + \beta_i \gamma_i} \zeta_i$$

VIX smiles generated with this parametrization are very similar to those generated by the equivalent parametrization for continuous forward variances in Section 7.7.1.

### A non-parametric form

While ansatz (7.104) is adequate for capturing VIX smiles, as a parametric form it only allows for certain types of smile shapes. We now build a model that can be calibrated to arbitrary VIX smiles, as long as they are non-arbitrageable; it is an example of the Markov-functional models discussed in Section 2.10.

Reasoning as in Section 7.7.1, let us write  $F_t^i$  as:

$$F_t^i = F_0^i f^i(t, x_t^{T_i}) \quad (7.105)$$

Note that (7.104) is but a particular form of (7.105), with  $f^i$  the sum of two exponentials.

The mapping function  $f^i(t, x)$  has to be such that (a) at  $t = T_i$ ,  $F_{t=T_i}^i$  is distributed so that the corresponding VIX smile observed at  $t = 0$  is recovered, (b)  $F_t^i$  is driftless.

Condition (b) implies that  $f^i$  obeys PDE (7.87):

$$\frac{df^i}{dt} + \frac{\eta_i^2 (T_i - t)}{2} \frac{d^2 f^i}{dx^2} = 0 \quad (7.106)$$

where  $\eta_i^2(T_i - t)$  is the instantaneous variance of  $x_t^{T_i}$ .  $\eta_i$  is given in (7.31b):

$$\eta_i(u) = \alpha_{\theta_i} \sqrt{(1 - \theta_i)^2 e^{-2k_1 u} + \theta_i^2 e^{-2k_2 u} + 2\rho_{12}\theta_i(1 - \theta_i)e^{-(k_1+k_2)u}}$$

Once the terminal profile  $f^i(T_i, x)$  is specified, solving (7.106) produces  $f^i(t, x)$  for all  $t \leq T_i$ .

The terminal condition for  $f^i, f^i(T_i, x)$  must be such that the mapping  $x_{T_i}^{T_i} \rightarrow f^i(T_i, x_{T_i}^{T_i})$  generates the VIX market smile for maturity  $T_i$ .

Consider a VIX level  $K$  and denote by  $\mathcal{D}^K$  the undiscounted market price of a digital option of strike  $K$ , maturity  $T_i$  that pays 1 if  $F_{T_i}^i < K$  and zero otherwise.  $\mathcal{D}^K$  is straightforwardly derived from the vanilla smile of future  $F^i$ , as a digital is essentially a narrow put spread:  $\mathcal{D}^K = \frac{d\mathcal{P}^K}{dK}$ . We have:

$$\begin{aligned} \mathcal{D}^K &= P(F_{T_i}^i < K) = P(f^i(T_i, x_{T_i}^{T_i}) < k) \\ &= \mathcal{N}_i(f^{i-1}(T_i, k)) \end{aligned}$$

where  $k$  is the moneyness:  $k = K/F_0$  and  $\mathcal{N}_i$  is the cumulative distribution function of the centered Gaussian random variable  $x_{T_i}^{T_i}$ , whose variance is known in closed form.

This yields  $f^{i-1}(T_i, k) = \mathcal{N}_i^{-1}(\mathcal{D}^K)$ . By choosing a large number of values of moneyness  $k$  we determine  $f^{i-1}$ . Since  $\mathcal{D}^K$  is an increasing function of  $K$ , so is  $f^{i-1}$ .  $f^{i-1}$  is monotonic thus  $f^i$  is well-defined and monotonic as well.<sup>51</sup>

### Characterizing the dynamics of VIX futures

What kind of dynamics does (7.105) generate for  $F_t^i$ ? Making use of (7.106):

$$\frac{dF_t^i}{F_t^i} = \frac{d\ln f^i}{dx}(t, x_t^{T_i}) dx_t^{T_i}$$

The instantaneous volatility of  $F^i$  is thus given by:

$$\text{vol}(F^i) = \eta(T_i - t) \left| \frac{d\ln f^i}{dx} \right|_{x=f^{i-1}(t, F_t^i)} \quad (7.107)$$

---

<sup>51</sup>Do we have an assurance that  $f^i(0, 0) = 1$  – so that we get indeed the right initial value for the VIX future? This is equivalent to  $E[f^i(T_i, x_{T_i}^{T_i})] = 1$ , that is the mapping function  $f^i$  ensures the forward of  $F_{T_i}^i$  is correctly priced. This is not guaranteed: even though the finite set of digitals  $\mathcal{D}^K$  we have used to build  $f^i$  is correctly priced, their integral – which is equal to the forward – depends on how we interpolate/extrapolate  $f^{-1}$  or  $f$ . We thus may need to uniformly rescale  $f^i$  to make sure  $f^i(0, 0) = 1$ .

$f^{i-1}$  is well-defined only if  $f^i$  is monotonic, which is the case.<sup>52</sup> The right-hand side of (7.107) is a function of  $F_t^i$  and  $t$ , thus what we have is really a local volatility model for  $F_t^i$ .

### What about forward VS volatilities?

The benefit of modeling VIX futures directly is that they are readily accessible in a simulation. Forward VS volatilities  $\widehat{\sigma}_{T_i, T_i + \Delta}$  are directly accessible at  $T_i$  since, by definition of the settlement value of VIX futures:

$$F_{T_i}^i = \widehat{\sigma}_{T_i, T_i + \Delta}(T_i)$$

What if we also need  $\widehat{\sigma}_{T_i, T_i + \Delta}(t)$  for  $t \leq T_i$ ?

VIX futures are given by:

$$F_{T_i}^i = F_0^i f^i(t, x_t^{T_i})$$

Moreover, forward variances  $\widehat{\sigma}_{T_i, T_i + \Delta}^2$  are driftless. We can thus represent  $\widehat{\sigma}_{T_i, T_i + \Delta}$  as:

$$\widehat{\sigma}_{T_i, T_i + \Delta}(t) = F_0^i \sqrt{g^i(t, x_t^{T_i})}$$

The terminal condition of  $g^i$  is:

$$g^i(T_i, x) = f^i(T_i, x)^2$$

and  $g^i(t, x)$  for  $t \leq T_i$  is obtained by solving PDE (7.106):

$$\frac{dg^i}{dt} + \frac{\eta_i^2 (T_i - t)}{2} \frac{d^2 g^i}{dx^2} = 0$$

In conclusion, we have a model calibrated to VIX smiles where all VIX futures and the corresponding forward VS volatilities are easily generated. We only need to simulate two Ornstein-Uhlenbeck processes:  $X_t^1$  and  $X_t^2$ .<sup>53</sup>

---

<sup>52</sup>The mapping function at  $T_i$ ,  $f^i(T_i, x)$  is monotonic, as it is derived from market prices of digital options of maturity  $T_i$  – no-arbitrage requires monotonicity of digital option prices with respect to their strike. Next, (7.106) implies that if  $f^i(T_i, x)$  is monotonic, so is  $f^i(t, x)$ . Indeed, take the derivative of (7.106) with respect to  $x$ :  $\frac{df^i}{dx}(t, x)$  solves the same PDE as  $f^i$ . It can thus be written as an expectation:  $\frac{df^i}{dx}(t, x) = E_t[\frac{df^i}{dx}(T_i, x_{T_i}^{T_i}) | x_t^{T_i} = x]$ .  $\frac{df^i}{dx}(T_i, x) \geq 0 \forall x$  then implies  $\frac{df^i}{dx}(t, x) \geq 0 \forall x, \forall t$ .

<sup>53</sup>The ease with which we build multi-asset, multi-factor Markov-functional models may look suspicious to readers with a fixed income background. In fixed income, multi-factor Markov-functional models are notoriously difficult to build, because determination of the final mapping function of an underlying – swap or LIBOR rate – involves an annuity ratio that depends on the mapping function of a different, contiguous, asset. In our context, it is as if prices of European options of maturity  $T_i$  on VIX future  $F^i$  no longer read  $E[h(f^i(T_i, x_{T_i}^{T_i}))]$ , but  $E[h(f^i(T_i, x_{T_i}^{T_i}))f^{i+1}(T_i, x_{T_i}^{T_{i+1}})]$ . In the multi-factor case  $x_t^{T_i}$  and  $x_t^{T_{i+1}}$  are different processes, hence the simple calibration procedure outlined in Section 7.8.2 would no longer work. Fortunately, unlike swap or LIBOR rates, forward variances or VIX futures are martingale under the same measure.

### 7.8.3 A dynamics for $S_t$

Having specified a dynamics for forward variances, we now define a dynamics for the underlying – how does the former constrain the latter?

As  $t$  reaches date  $T_i$  the VS volatility for interval  $[T_i, T_{i+1}]$ ,  $\widehat{\sigma}_{T_i, T_{i+1}}(t = T_i)$  is known. The dynamics of  $S_t$  for  $t \in [T_i, T_{i+1}]$  has to comply with this value of  $\widehat{\sigma}_{T_i, T_{i+1}}$ .

We now specify an SDE for  $S_t$ ,  $t \in [T_i, T_{i+1}]$  that meets the three following requirements:

- The VS, or log-contract, implied volatility at  $T_i$  for maturity  $T_{i+1}$  is equal to  $\widehat{\sigma}_{T_i, T_{i+1}}(T_i)$ .
- The probability density of  $\frac{S_{T_{i+1}}}{S_{T_i}}$  is independent on  $S_{T_i}$ . This is an essential condition for ensuring that future and spot-starting smiles are decoupled. With this provision, prices of cliques of the form:

$$\sum_i \omega_i f\left(\frac{S_{T_{i+1}}}{S_{T_i}}\right)$$

have no sensitivity to correlations between the Brownian motion driving  $S_t$  and those driving forward variances. While these correlations have zero impact on future smiles, they do impact spot-starting vanilla smiles – this is how we decouple spot and future smiles in the model.

- Scenarios of future smiles on  $[T_i, T_{i+1}]$ , as a function of  $\widehat{\sigma}_{T_i, T_{i+1}}(T_i)$ , should be set at will.

Consider a path-dependent local volatility dynamics for  $S_t$ ,  $t \in [T_i, T_{i+1}]$ , given by:

$$dS_t = (r - q)S_t dt + \sigma^i \left( \frac{S_t}{S_{T_i}} \right) S_t dW_t^S \quad (7.108)$$

where the correlations of  $W_t^S$  with  $W_t^1$  and  $W_t^2$  are denoted by  $\rho_{SX^1}$  and  $\rho_{SX^2}$ . Setting  $s_t = \frac{S_t}{S_{T_i}}$ , we have:

$$\begin{aligned} ds_t &= (r - q)s_t dt + \sigma^i(s_t)s_t dW_t^S \\ s_{T_i} &= 1 \end{aligned}$$

thus the density of  $s_{T_{i+1}}$  is indeed independent on  $S_{T_i}$ .

We choose the following expression for  $\sigma^i$ :

$$\sigma^i(s) = \sigma_0^i \frac{n^i}{n^i - 1} \frac{(n^i - 1) s^{\beta^i - 1}}{(n^i - 1) + s^{\beta^i - 1}} \quad (7.109)$$

which is parametrized by three numbers:  $\sigma_0^i$ ,  $\beta^i$  and  $n^i$ .

- $\sigma_0^i$  is the local volatility for  $S = S_{T_i}$ .
- $\beta^i$  controls the skew of maturity  $T_{i+1}$ .  $\frac{d\sigma^i}{d \ln S} \Big|_{S=S_{T_i}} = \sigma_0^i \frac{n^i - 1}{n^i} (\beta^i - 1)$ . For  $(\beta^i - 1)$  small, assuming zero interest rate and repo, the ATM skew of maturity  $T_{i+1}$  is given, at order one in  $(\beta^i - 1)$  by formula (2.50a), page 47:

$$\frac{d\widehat{\sigma}_{K,T_{i+1}}}{d \ln K} \Big|_{S_{T_i}} \simeq \frac{1}{2} \frac{d\sigma^i}{d \ln S} \Big|_{S_{T_i}} = \frac{\sigma_0^i}{2} \frac{n^i - 1}{n^i} (\beta^i - 1) \quad (7.110)$$

- $n^i$  prevents the divergence of  $\sigma$  for small values of  $S$ ; the maximum level of volatility is  $n^i \sigma_0^i$ . In practice,  $n^i$  can be used to control other features of the smile, for example the difference between the VS volatility and the ATMF volatility of the smile of maturity  $T_{i+1}$ , observed at  $T_i$ .

$\sigma_0^i, \beta^i, n^i$  can be set at will, as a function of  $\widehat{\sigma}_{T_i, T_{i+1}}(T_i)$ , provided the log-contract implied volatility at  $T_i$  for maturity  $T_{i+1}$  is indeed equal to  $\widehat{\sigma}_{T_i, T_{i+1}}(T_i)$ .

How do we control the future ATMF skew scenarios generated by our model? This is done by expressing how the ATMF skew  $\mathcal{S}_{T_{i+1}}(T_i)$  depends on the level of ATMF volatility. Two natural choices are:

- a fixed ATMF skew, irrespective of the level of ATMF volatility:  $\mathcal{S}_{T_{i+1}}(T_i) = \mathcal{S}_i$ , where we are free to choose the level of the ATMF skew  $\mathcal{S}_i$  for each interval  $[T_i, T_{i+1}]$ .
- a specific dependence of the ATMF skew to the ATMF volatility, for example parametrized by a power-law:

$$\mathcal{S}_{T_{i+1}}(T_i) = \left( \frac{\widehat{\sigma}_{\text{ATMF}, T_{i+1}}(T_i)}{\widehat{\sigma}_{\text{ATMF, ref}}^i} \right)^{\gamma^i} \mathcal{S}_{\text{ref}}^i \quad (7.111)$$

where  $\mathcal{S}_{\text{ref}}^i$  and  $\widehat{\sigma}_{\text{ATMF, ref}}^i$  are reference levels for ATMF skew and volatility.  $\gamma^i, \mathcal{S}_{\text{ref}}^i$  and  $\widehat{\sigma}_{\text{ATMF, ref}}^i$  can be chosen differently for each interval  $[T_i, T_{i+1}]$ .

Once the type of dependence of  $\mathcal{S}_{T_{i+1}}(T_i)$  on  $\widehat{\sigma}_{\text{ATMF}, T_{i+1}}(T_i)$  is chosen, we need to determine two functions  $\sigma_0^i()$ ,  $\beta^i()$  for each interval  $[T_i, T_{i+1}]$  such that setting  $\sigma_0^i = \sigma_0^i(\widehat{\sigma}_{T_i, T_{i+1}}(T_i))$  and  $\beta^i = \beta^i(\widehat{\sigma}_{T_i, T_{i+1}}(T_i))$  produces the desired behavior.

The ability to choose different fixed ATMF skew  $\mathcal{S}_i$  or different values of  $\gamma^i, \mathcal{S}_{\text{ref}}^i$  and  $\widehat{\sigma}_{\text{ATMF, ref}}^i$  for different intervals  $[T_i, T_{i+1}]$  is not superfluous.

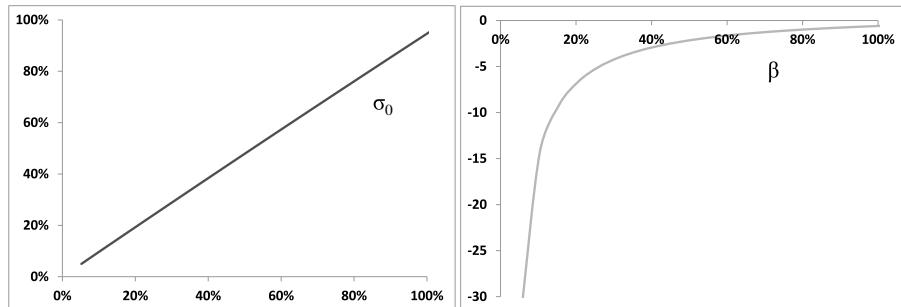
It does not make sense to offset the sensitivity to the  $[T_i, T_{i+1}]$  future skew with an opposite sensitivity to the future skew for a different interval. Thus, even in the unlikely case that the future skew  $\mathcal{S}_i$  implied from market prices of cliques happens to be constant, we still need to separately calculate and manage the sensitivities to parameters controlling the future skew of each interval  $[T_i, T_{i+1}]$  – hence  $\sigma_0^i$  and  $\beta^i$ .

## Two examples

Consider an interval  $[T_i, T_{i+1}]$ . Functions  $\sigma_0(\hat{\sigma}_{T_i, T_{i+1}}(T_i))$  and  $\beta(\hat{\sigma}_{T_i, T_{i+1}}(T_i))$  – we omit the  $i$  index to lighten notation – are obtained as follows:

- Select discrete values  $\sigma_0^k$  of  $\sigma_0$  spanning a sufficiently wide range.
- For each value of  $\sigma_0^k$  find the value  $\beta^k$  of  $\beta$  such that the ATM skew has the desired value, either constant or specified by (7.111). For each trial value of  $\beta$ , the vanilla smile at  $T_i$  for maturity  $T_{i+1}$  is obtained by numerically solving the forward equation of the local volatility model (2.7), page 29.
- Numerically solve PDE (5.49), page 173, to generate the VS volatility for maturity  $T_{i+1}$ :  $\hat{\sigma}_{\text{VS}}^k$ .
- Store the couples  $(\hat{\sigma}_{\text{VS}}^k, \sigma_0^k)$  and  $(\hat{\sigma}_{\text{VS}}^k, \beta^k)$  and proceed to the next value of  $\sigma_0$ .
- Finally, interpolate the discrete couples  $(\hat{\sigma}_{\text{VS}}^k, \sigma_0^k)$  to generate the function  $\sigma_0(\hat{\sigma}_{\text{VS}})$ , and likewise for  $\beta(\hat{\sigma}_{\text{VS}})$ .

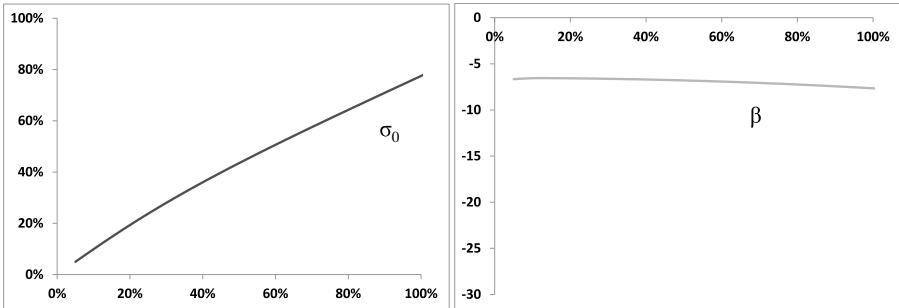
Figure 7.18 shows functions  $\sigma_0()$  and  $\beta()$  such that the ATM skew is fixed, equal to 5%. We use the difference between implied volatilities of the 95% and 105% strikes rather than  $\frac{d\hat{\sigma}_{K, T_{i+1}}}{d \ln K} \Big|_{S_{T_i}}$  as a measure of ATM skew. We have used zero interest rate and repo and a monthly schedule:  $T_{i+1} - T_i = 1$  month, and  $n$  is set to 3.



**Figure 7.18:**  $\sigma_0$  (left) and  $\beta$  (right) as a function of  $\hat{\sigma}_{T_i, T_{i+1}}(T_i)$ , such that the difference of implied volatilities of the 95% and 105% strikes is equal to 5%. We have taken  $n_i = 3$  and  $T_{i+1} - T_i = 1$  month.

Figure 7.19 shows functions  $\sigma_0()$  and  $\beta()$  such that the ATM skew is of the form in (7.111) with  $\gamma = 1$ ,  $S_{\text{ref}} = 5\%$ ,  $\hat{\sigma}_{\text{ATMF, ref}} = 20\%$ : the ATM skew is proportional to the ATM volatility.

The shapes of  $\beta(\hat{\sigma}_{\text{VS}})$  are consistent with approximation (7.110) which implies that  $\beta$  is constant for a skew that is proportional to the ATM volatility while  $(\beta - 1)$



**Figure 7.19:**  $\sigma_0$  (left) and  $\beta$  (right) as a function of  $\hat{\sigma}_{T_i, T_{i+1}}(T_i)$ , such that the ATMF skew is of the form in (7.111) with  $\gamma = 1$ ,  $S_{\text{ref}} = 5\%$ ,  $\hat{\sigma}_{\text{ATMF, ref}} = 20\%$ . We have taken  $n_i = 3$  and  $T_{i+1} - T_i = 1$  month.

should be inversely proportional to the VS volatility in order to generate a skew that is independent on the level of volatility.<sup>54</sup>

- In the first situation, our discrete forward variance model mimics the two-factor continuous forward variance model, which generates an ATMF skew that is approximately independent on the level of VS volatility. See formula (8.55), page 330, for the ATMF skew at order one in volatility of volatility and Figure 8.4, page 331, for an illustration of the (in)dependence of the ATMF skew on the level of VS volatility.
- In the second situation, with an ATMF skew proportional to the ATMF volatility, we mimic, for short maturities, the behavior of the  $\frac{3}{2}$  model. See the corresponding SDE in (8.45), with  $\gamma = \frac{3}{2}$ , and the short-maturity ATMF skew in (8.49), page 326.
- What if we make the ATMF skew *inversely* proportional to the ATMF volatility, by using (7.111) with  $\gamma^i = -1$ ? We would then be mimicking, for short maturities, the behavior of the Heston model – see expression (6.18a), page 210.

Discrete forward variance models thus afford a great deal of flexibility as to the dependence of the short-maturity future skew to the short future ATMF volatility, while still leaving us the freedom of choosing spot/volatility correlations, which impact spot-starting smiles.

<sup>54</sup>For short maturities, approximation (3.28), page 121, shows that the VS implied volatility is equal to  $\sigma_0$  at order zero in the slope of the local volatility function. For the sake of obtaining the ATMF skew at order one in the slope of the local volatility function, we can replace  $\sigma_0$  with the VS or ATMF volatility in (7.110).

Two more aspects are worth commenting, before we turn to the vanilla smile.

- What if there are cash-amount dividends? In this case functions  $\sigma_0(\xi_{T_i}^i)$  and  $\beta(\xi_{T_i}^i)$  should also depend on  $S_{T_i}$ , if they are to ensure (a) that the VS volatility at  $T_i$  for maturity  $T_{i+1}$  is indeed  $\xi_{T_i}^i$ , (b) that the ATMF skew at  $T_i$  for maturity  $T_{i+1}$  still agrees with our specification.

The case of cash-amount dividends is taken care of economically by (a) making  $\beta$  a function of  $\sigma_0$ , with the dependence obtained in the calibration of the  $\sigma_0$  and  $\beta$  functions with  $S_{T_i}$  set to the forward for maturity  $T_i$ , (b) generating the mapping  $\sigma_0(\xi^i)$  for a range of values of  $S_{T_i}$ , thus obtaining in effect a mapping  $\sigma_0(\xi^i, S_{T_i})$ . While not exact, step (a) ensures, in practice, that the target forward skew scenarios are obtained with good accuracy.<sup>55</sup>

- While the local volatility function in (7.109) is adequate for generating the desired ATMF skew scenarios, it is not able to generate very large spreads between VS and ATMF volatilities. Also, as  $t$  reaches  $T_i$ , forward-start options become in effect European options of maturity  $T_{i+1}$ ; our local volatility function should be such that it is able to match the market smile for maturity  $T_{i+1}$  observed at  $T_i$ . For these reasons it is a good idea to include an additional quadratic component  $\alpha s^2$ .

#### 7.8.4 The vanilla smile

Two mechanisms contribute to the smile of discrete forward models:

- the correlations between the Brownian motions driving  $S_t$  and the  $\xi_t^i$
- the local volatility functions  $\sigma^i(s)$ , which generate the future skews for intervals  $[T_i, T_{i+1}]$ .

Consider a discrete model with lognormal dynamics (7.99), page 280, for  $\xi^i$  and  $\sigma^i(s)$  given by (7.109).

Setting  $\nu_i = 0$  and  $\beta^i = 1, \forall i$  turns our model into a Black-Scholes model with deterministic volatility: on each interval  $[T_i, T_{i+1}]$ ,  $S_t$  is lognormal with constant volatility  $\sqrt{\xi^i}$ .

As  $\nu_i \neq 0$  and  $(\beta^i - 1) \neq 0$  volatility becomes stochastic. We now derive an expression of the ATMF skew at order one in  $\nu_i$  and in  $(\beta^i - 1)$ .

We assume a constant tenor  $\Delta$  so that our schedule is given by  $T_i = i\Delta$ . We also use the same values for  $\nu_i, \theta_i, \alpha$ . These numbers then only depend on tenor  $\Delta$ , through the mapping relationship (7.102), and we thus denote them by  $\nu_\Delta, \theta_\Delta, \alpha_{\theta_\Delta}$ . Likewise, we omit the  $i$  index in  $\sigma_0, \beta$ .

We use a constant term structure of forward variances:  $\xi_0^i = \widehat{\sigma}^2$ , and functions  $\sigma_0$  and  $\beta$  are chosen so that the ATMF skew of maturity  $\mathcal{S}_{T_{i+1}}(T_i)$  does not depend on  $\widehat{\sigma}_{T_i, T_{i+1}}(T_i)$  and is equal to  $\mathcal{S}_\Delta$ .

---

<sup>55</sup>I am indebted to Julien Tijou for developing this enhancement to the original model.

We start from expression (8.29) of vanilla option prices derived in Section 8.4 of Chapter 8 and the resulting expression of the ATMF skew at order one in volatility of volatility (8.32), page 319:

$$\mathcal{S}_T = \frac{1}{2\hat{\sigma}^3 T} \int_0^T \frac{T-t}{T} \frac{\langle d \ln S_t d\hat{\sigma}_T^2(t) \rangle_0}{dt} dt \quad (7.112)$$

where the 0 subscript means the covariation is evaluated in the unperturbed state, that is with forward variances equal to their values at  $t = 0$  and with the instantaneous volatility of  $S_t$  read off the initial term structure of VS volatilities. In our context, expanding at order one in volatility of volatility corresponds to expanding at order one both in  $\nu$  and  $(\beta - 1)$ .

We calculate  $\mathcal{S}_T$  for maturities  $T$  that are multiples of tenor  $\Delta$ :  $T = N\Delta$ . Using the definition of  $\hat{\sigma}_T(t)$ ,  $(T-t)\hat{\sigma}_T^2(t) = \int_t^T \xi_t^\tau d\tau$ , where  $\xi_t^\tau$  is the *instantaneous* forward variance for date  $\tau$ , (7.112) can be rewritten as:

$$\begin{aligned} \mathcal{S}_{N\Delta} &= \frac{1}{2\hat{\sigma}^3 T^2} \int_0^T \left\langle d \ln S_t d \left( \int_t^T \xi_t^\tau d\tau \right) \right\rangle_0 \\ &= \frac{1}{2\hat{\sigma}^3 T^2} \sum_{i=0}^{N-1} \int_{T_i}^{T_{i+1}} \left\langle d \ln S_t d \left( \int_t^{T_N} \xi_t^\tau d\tau \right) \right\rangle_0 \end{aligned}$$

For  $\tau \in [T_k, T_{k+1}]$ ,  $\xi_t^\tau = \xi_t^k$ . Thus:

$$\int_t^{T_N} \xi_t^\tau d\tau = \int_t^{T_{i+1}} \xi_t^\tau d\tau + \Delta \sum_{j=i+1}^{N-1} \xi_t^j = (T_{i+1} - t)\hat{\sigma}_{T_{i+1}}^2(t) + \Delta \sum_{j=i+1}^{N-1} \xi_t^j$$

$\mathcal{S}_T$  is given by:

$$\begin{aligned} \mathcal{S}_{N\Delta} &= \frac{1}{2\hat{\sigma}^3 T^2} \left( \Delta \sum_{i=0, j>i}^{N-1} \int_{T_i}^{T_{i+1}} \langle d \ln S_t d\xi_t^j \rangle_0 \right. \\ &\quad \left. + \sum_{i=0}^{N-1} \int_{T_i}^{T_{i+1}} (T_{i+1} - t) \langle d \ln S_t d\hat{\sigma}_{T_{i+1}}^2(t) \rangle_0 \right) \end{aligned} \quad (7.113)$$

The derivation of (7.112) utilizes the assumption that  $E [\langle d \ln S_t d\hat{\sigma}_T^2(t) \rangle | \ln S]$  does not depend on  $S$ . We now verify that this holds at order one in  $\nu$  and  $(\beta - 1)$ .

- Consider the first line of (7.113). At order one in  $\nu$ :

$$d\xi_t^j = (2\nu_\Delta) \xi_0^j \alpha_{\theta_\Delta} \left( (1 - \theta_\Delta) e^{-k_1(T_j - t)} dW_t^1 + \theta_\Delta e^{-k_2(T_j - t)} dW_t^2 \right)$$

For the sake of obtaining the covariation at order one in  $\nu$  and  $(\beta - 1)$ , we take  $dS_t = (r - q)S_t dt + \hat{\sigma} S_t dW_t^S$  and get:

$$\begin{aligned} & \langle d\ln S_t \, d\xi_t^j \rangle_0 = \\ & (2\nu_\Delta) \hat{\sigma}^3 \alpha_{\theta_\Delta} \left( (1 - \theta_\Delta) e^{-k_1(T_j - t)} \rho_{SX^1} + \theta_\Delta e^{-k_2(T_j - t)} \rho_{SX^2} \right) dt \end{aligned} \quad (7.114)$$

where we have used that  $\xi_0^j = \hat{\sigma}^2$ .  $\langle d\ln S_t \, d\xi_t^j \rangle_0$  does not depend on  $\ln S_t$ .

- Now turn to the second line of (7.113) and consider the contribution of interval  $[T_i, T_{i+1}]$ . Let us condition the expectation of the covariation with respect to  $\xi_{T_i}^i$ . At order one in  $(\beta - 1)$ :

$$\sigma \left( \frac{S_t}{S_{T_i}} \right) = \sigma_0 (\xi_{T_i}^i) \left( 1 + \frac{n-1}{n} (\beta (\xi_{T_i}^i) - 1) \ln \frac{S_t}{S_{T_i}} \right)$$

We can now employ results derived in the perturbative analysis of the local volatility model in Sections 2.4.5 and 2.5.7 of Chapter 2. The local volatility function is of the form (2.44):

$$\sigma(t, S) = \bar{\sigma}(t) + \alpha(t) \ln \frac{S_t}{F_t} \quad (7.115)$$

with:

$$\bar{\sigma}(t) = \sigma_0 (\xi_{T_i}^i) \left( 1 + \frac{n-1}{n} (\beta (\xi_{T_i}^i) - 1) \ln \frac{F_t}{S_{T_i}} \right) \quad (7.116a)$$

$$\alpha(t) = \sigma_0 (\xi_{T_i}^i) \frac{n-1}{n} (\beta (\xi_{T_i}^i) - 1) \quad (7.116b)$$

We know from Section 2.5.7 that, given a local volatility function of the form in (7.115),  $\langle d\ln S_t \, d\hat{\sigma}_{T_{j+1}}^2(t) \rangle_0$  does not depend on  $S_t$ , at order one in  $\alpha(t)$ .

This covariation is already of order one in  $(\beta (\xi_{T_i}^i) - 1)$  thus need to be calculated at order zero in  $\nu$ : taking the expectation of  $\langle d\ln S_t \, d\hat{\sigma}_{T_{j+1}}^2(t) \rangle_0$  with respect to  $\xi_{T_i}^i$  simply amounts to setting  $\xi_{T_i}^i = \hat{\sigma}^2$ .

Thus the second contribution in (7.113) does not depend on  $S$  either, at order one in  $\nu$  and  $(\beta - 1)$ .

Since for the sake of calculating the covariation in the second piece of (7.113)  $\xi_{T_i}^i$  is frozen, equal to  $\hat{\sigma}$ , we have, using identity (7.112) for the case of a local volatility function of type (7.115):

$$\int_{T_i}^{T_{i+1}} (T_{i+1} - t) \langle d\ln S_t \, d\hat{\sigma}_{T_{i+1}}^2(t) \rangle_0 = 2\hat{\sigma}^3 \Delta^2 \mathcal{S}_\Delta$$

which allows us to rewrite (7.113) as:

$$\mathcal{S}_{N\Delta} = \frac{1}{N}\mathcal{S}_\Delta + \frac{1}{2\hat{\sigma}^3 T^2} \Delta \sum_{i=0, j>i}^{N-1} \int_{T_i}^{T_{i+1}} \langle d\ln S_t d\xi_t^j \rangle_0 \quad (7.117)$$

Using (7.114) we have:

$$\begin{aligned} & \frac{1}{2\hat{\sigma}^3 T^2} \Delta \sum_{i=0, j>i}^{N-1} \int_{T_i}^{T_{i+1}} \langle d\ln S_t d\xi_t^j \rangle_0 \\ &= \frac{1}{2\hat{\sigma}^3 T^2} \Delta 2\nu_\Delta \hat{\sigma}^3 \alpha_{\theta_\Delta} \sum_{i=0, j>i}^{N-1} \int_{T_i}^{T_{i+1}} \\ & \quad \times \left( (1 - \theta_\Delta) e^{-k_1(T_j - t)} \rho_{SX^1} + \theta_\Delta e^{-k_2(T_j - t)} \rho_{SX^2} \right) dt \\ &= \nu_\Delta \alpha_{\theta_\Delta} \frac{1}{N^2} \sum_{i=0, j>i}^{N-1} \frac{1}{\Delta} \int_{T_i}^{T_{i+1}} \left( (1 - \theta_\Delta) e^{-k_1(T_j - t)} \rho_{SX^1} + \theta_\Delta e^{-k_2(T_j - t)} \rho_{SX^2} \right) dt \\ &= \nu_\Delta \alpha_{\theta_\Delta} \left( (1 - \theta_\Delta) \rho_{SX^1} \zeta(k_1 \Delta, N) + \theta_\Delta \rho_{SX^2} \zeta(k_2 \Delta, N) \right) \end{aligned}$$

where we have introduced function  $\zeta(x, N)$  defined by:

$$\zeta(x, N) = \frac{1}{N^2} \sum_{i=0, j>i}^{N-1} \int_i^{i+1} (e^{-x})^{j-u} du = \frac{e^x - 1}{x} \sum_{n=1}^{N-1} \frac{N-n}{N^2} (e^{-x})^n \quad (7.118)$$

The final expression of the ATMF skew for maturity  $T = N\Delta$ , in the discrete two-factor model, at order one in  $\nu$  and  $(\beta - 1)$  is thus:

$$\mathcal{S}_{N\Delta} = \frac{1}{N}\mathcal{S}_\Delta + \frac{1}{2\hat{\sigma}^3 T^2} \Delta \sum_{i=0, j>i}^{N-1} \int_{T_i}^{T_{i+1}} \langle d\ln S_t d\xi_t^j \rangle_0 \quad (7.119a)$$

$$= \frac{1}{N}\mathcal{S}_\Delta + \nu_\Delta \alpha_{\theta_\Delta} \left( (1 - \theta_\Delta) \rho_{SX^1} \zeta(k_1 \Delta, N) + \theta_\Delta \rho_{SX^2} \zeta(k_2 \Delta, N) \right) \quad (7.119b)$$

- Expression (7.119) is an expansion of the ATMF skew at order one in  $(\beta - 1)$  (first piece in (7.119b)) and  $\nu$  (second piece). Note that  $(\beta - 1)$  does not appear explicitly – only the ATMF future skew  $\mathcal{S}_\Delta$  for tenor  $\Delta$  appears in (7.119b). Indeed, from (7.116), the expansion at order one in  $(\beta - 1)$  is really an expansion at order one in  $\alpha(t)$ . Owing to the skew-averaging expression (2.48), page 46, relating  $\mathcal{S}_\Delta$  to  $\alpha(t)$ , ours is equivalently an expansion at order one in  $\mathcal{S}_\Delta$ .

- The ATMF skew is the sum of two components: the forward-smile contribution and the volatility-of-volatility contribution, which can be separately switched on and off by setting  $\mathcal{S}_\Delta$  or  $\nu_\Delta$  equal to 0.

When pricing a cliquet of period  $\Delta$ ,  $\mathcal{S}_\Delta$  controls the forward-smile adjustment  $\delta P_2$  while  $\nu$  controls the volatility-of-volatility adjustment  $\delta P_1$  – see the discussion in Section 3.1.6 of Chapter 3.

- (7.119) makes it plain that the ATMF skew of discrete forward variance models is the sum of two contributions: the first piece in (7.119b) is contributed by the forward skew for tenor  $\Delta$ ,  $\mathcal{S}_\Delta$ , while the covariance of  $S_t$  with forward variances  $\xi_t^i$  is the second source of skew. Imagine switching off volatility of volatility. We then have:

$$\mathcal{S}_{N\Delta} = \frac{1}{N} \mathcal{S}_\Delta$$

Thus the ATMF skew for maturity  $T$  decays like  $\frac{1}{T}$ .

This is understood by noting that with  $\nu = 0$ , log-returns  $\ln \frac{S_{i\Delta}}{S_{(i-1)\Delta}}$  are independent, thus  $\ln \frac{S_{N\Delta}}{S_0}$  is the sum of  $N$  independent, identically distributed random variables. The skewness of  $\ln \frac{S_{N\Delta}}{S_0}$  then scales like  $\frac{1}{\sqrt{N}}$ . In Appendix B of Chapter 5, it is shown that, at order one in the skewness  $s$  of  $\ln S_T$ , the ATMF skew  $\mathcal{S}_T$  for maturity  $T$  is given by expression (5.93), page 194:  $\mathcal{S}_T = \frac{s}{6\sqrt{T}}$ . We thus get:  $\mathcal{S}_{N\Delta} \propto \frac{1}{N}$ . Because these results are derived at order one in  $\mathcal{S}_\Delta$ , they only hold for small values of  $\mathcal{S}_\Delta$ .

- Imagine taking the limit  $\Delta \rightarrow 0$ . Our model becomes a plain continuous forward variance model. There is only one source of skew:  $\mathcal{S}_{N\Delta}$  is generated by the second piece in (7.119b). In expression (7.118) of  $\zeta(x, N)$ , take the limit  $x \rightarrow 0$ ,  $N \rightarrow \infty$  with  $Nx$  fixed, equal to, respectively,  $k_1 T$  or  $k_2 T$ . Converting the sum in (7.118) in an integral, we get:

$$\begin{aligned} \lim_{\substack{x \rightarrow 0, N \rightarrow \infty \\ Nx = kT}} \frac{e^x - 1}{x} \sum_{n=1}^{N-1} \frac{N-n}{N^2} (e^{-x})^n &= \int_0^T \frac{T-t}{T^2} e^{-kt} dt \\ &= \frac{kT - (1 - e^{-kT})}{(kT)^2} \end{aligned}$$

$S_T$  is then given at order one in  $\nu$  by:

$$\mathcal{S}_T = \nu \alpha_\theta \left( (1-\theta) \rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta \rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2} \right) \quad (7.120)$$

where we have simply replaced  $\nu_\theta, \theta_\Delta, \alpha_{\theta_\Delta}$  with  $\nu, \theta, \alpha_\theta$ .

We derive below, in Chapter 8, the expansion of implied volatilities in continuous forward variance models at order two in  $\nu$ . At order one in  $\nu$ , we unsurprisingly recover (7.120) – see formula (8.55), page 330.

## Numerical examples

The numerical results presented below are obtained with  $\Delta = 1$  month, and using:

- parameters in Table 7.8 for the (continuous) two-factor model,<sup>56</sup> which are mapped according to (7.102)
- $\sigma_0()$  and  $\beta()$  chosen so that future smiles of maturity  $\Delta$  exhibit a fixed ATMF skew, with  $S_\Delta = -0.5$  – see Figure 7.18. The latter value of  $S_\Delta$  corresponds to a difference of 5 points of implied volatility for 95% and 105% strikes for maturity  $\Delta$ . We take  $n_i = 3, \forall i$ .
- a flat VS volatility  $\hat{\sigma}$  equal to 20%.

$\nu$	$\theta$	$k_1$	$k_2$	$\rho_{12}$	$\rho_{SX^1}$	$\rho_{SX^2}$
174%	0.245	5.35	0.28	0%	-75.9%	-48.7%

Table 7.8: Parameters of the (continuous) two-factor model.

These parameter values are realistic – they are such that, in the continuous version of the two-factor model:

- volatilities of VS volatilities approximately decay like  $\frac{1}{\sqrt{T}}$  with the volatility of a 3-month VS volatility equal to 100% – see Figure 7.1, page 228.
- the ATMF skew approximately decays like  $\frac{1}{\sqrt{T}}$  as well, with the 95%/105% skew equal to 3 points of volatility for maturity 1 year – see Figures 8.3 and 8.5, pages 331 and 332.

In order to gauge the magnitudes of both contributions to the ATMF skew in (7.119b), let us turn off either the forward skew ( $S_\Delta = 0$ ) or volatility of volatility ( $\nu = 0$ ).

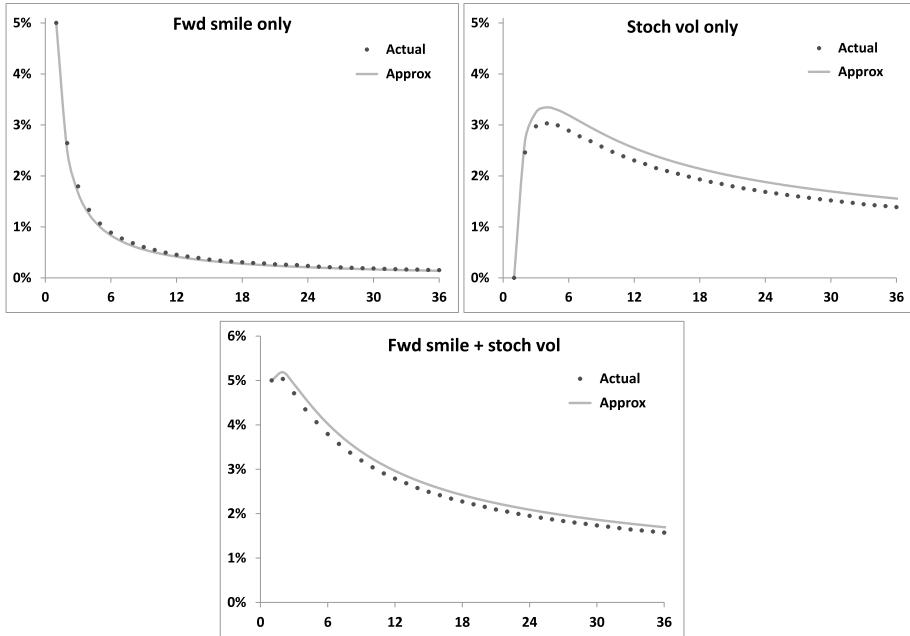
The resulting ATMF skew, expressed as the difference of the implied volatilities for the 95% and 105% strikes, is shown in Figure 7.20, together with the approximate value in (7.119b) and the ATMF skew, when both forward smile and volatility of volatility are turned on.

It is apparent that the decay of the “forward smile” contribution agrees well with the  $1/T$  form in (7.119b). The agreement of approximate and actual values of the stochastic volatility components is somewhat less satisfactory, still the increase and subsequent decrease with maturity is well captured by function  $\zeta(x, N)$ .

While the “approx” curve in the bottom graph of Figure 7.20 is the sum of the components in the top graphs, this is not the case for the “actual” curve; the latter is obtained in a Monte Carlo simulation with both effects turned on. The good agreement of approximate and actual ATMF skews is testament to the fact that order-2 cross terms of the type  $(\beta - 1)\nu$  contribute negligibly.  $S_T$  in Figure 7.20 is

---

<sup>56</sup>Parameters in Table 7.8 are those of Set II in Table 7.1, page 229, which we have used in Sections 7.4 and 7.6. We use these parameters again in the following chapter – see Sections 8.7 and 8.8 for smiles generated by the continuous model thus parametrized.



**Figure 7.20:** Top:  $S_T$ , as the 95%/105% skew, obtained either with  $\nu = 0$  (left) or with  $S_\Delta = 0$  (right) as a function of  $T$  (months), evaluated in a Monte-Carlo simulation of the discrete two-factor model with parameters in Table 7.8 (Actual) and as given by order-one formula (7.119b) (Approx).  $\Delta = 1$  month.

Bottom:  $S_T$  when both forward smile ( $S_\Delta = -0.5$ ) and volatility of volatility ( $\nu = 174\%$ ) are switched on.

non-monotonic as a function of  $T$ , however this depends on the relative magnitude of  $S_\Delta$  and  $\nu_\Delta$  – see Figure 7.21 where  $S_T$  is graphed for 3 different values of  $S_\Delta$ .

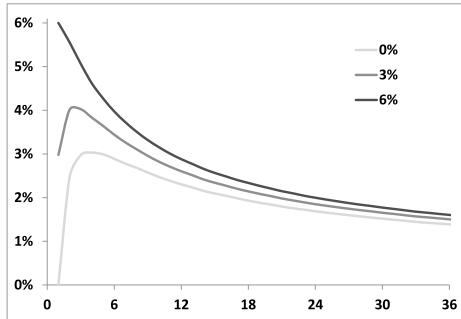
This is illustrated also in Figure 7.22, which shows  $S_T$  for the same parameters as in Figure 7.8, except  $\nu$  has been halved. For this smaller level of volatility of volatility, agreement with formula (7.119b) is excellent.

### 7.8.5 Conclusion

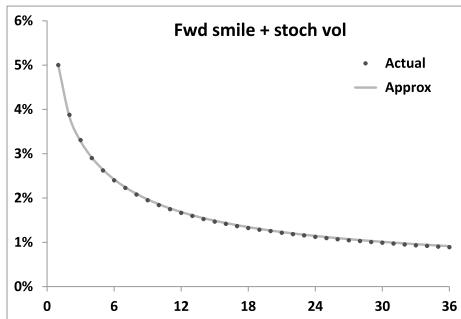
- With respect to their continuous counterparts, discrete forward variance models allow separation of the effects of (a) spot/volatility covariance, (b) future smile for a given time scale  $\Delta$ . In models thus specified, prices of cliques of the form:

$$\sum_i \omega_i f\left(\frac{S_{T_{i+1}}}{S_{T_i}}\right)$$

where  $T_i = i\Delta$ , do not depend anymore on spot/volatility correlations.



**Figure 7.21:**  $S_T$ , as the 95%/105% skew, as a function of  $T$  (months), evaluated in a Monte-Carlo simulation of the discrete two-factor model with parameters in Table 7.8 for 3 different values of  $S_\Delta$ , such that the 1-month 95%/105% ATMF skew is equal to 0, 3% and 6%.



**Figure 7.22:**  $S_T$ , as the 95%/105% skew, as a function of  $T$  (months), evaluated in a Monte-Carlo simulation of the discrete two-factor model with the same parameters as in Figure 7.20 except  $\nu$  has been halved.

One first sets future smiles for maturity  $\Delta$ , then chooses spot/volatility correlations so as to obtain desired levels of covariances of spot and forward VS/ATMF volatilities, or desired levels for the vanilla ATMF skew. Discrete forward variance models are thus naturally suited to the risk-management of cliques, such as accumulators.

- For any choice of time scale  $\Delta$ , simple parameter mappings exist that ensure that instantaneous volatilities of spot or forward-starting VS volatilities in the discrete two-factor model match those of the continuous version of the model.
- Specification of future smiles is very flexible. We give an example of parametrization that allows the user to specify how the ATMF skew for matu-

rity  $\Delta$  depends on the ATMF volatility for the same maturity. Once functions  $\sigma_0()$  and  $\beta()$  are tabulated, simulation of the discrete model is as uncomplicated as in the continuous model.

- Discrete forward variance models are also ideally suited to the risk management of payoffs involving both VIX futures and the S&P 500 index. They can be calibrated exactly to VIX smiles, if one so desires, while preserving full flexibility as to forward smile scenarios for the S&P 500 index.
- The vanilla smile of discrete forward variance models is produced by both forward-smile and volatility-of-volatility components. Order-one formula (7.119b) allows for an assessment of the contribution of each effect to the ATMF skew of vanilla options.

## Chapter's digest

### 7.1 Pricing equation

- The pricing equation of forward variance models is obtained through a replication argument. It reads:

$$\frac{dP}{dt} + (r - q) S \frac{dP}{dS} + \frac{\xi^t}{2} S^2 \frac{d^2 P}{dS^2} + \frac{1}{2} \int_t^T du \int_t^T du' \nu(t, u, u', \xi) \frac{d^2 P}{\delta \xi^u \delta \xi^{u'}} + \int_t^T du \mu(t, u, \xi) S \frac{d^2 P}{dS \delta \xi^u} = rP$$

This SDE admits a probabilistic interpretation. Its solution is given by:

$$P = E[g(S_T) | S_t = S, \xi_t^u = \xi^u]$$

under a dynamics for  $S_t$ ,  $\xi_t^u$  given by:

$$\begin{cases} dS_t &= (r - q)S_t dt + \sqrt{\xi_t^u} S_t dW_t^S \\ d\xi_t^u &= \lambda_t^u dW_t^u \end{cases}$$

with  $\lambda_t^u$  and correlations between  $W_t^S$  and  $W_t^u$  such that:

$$\begin{aligned} \lim_{dt \rightarrow 0} \frac{1}{dt} E_t[d \ln S_t d \xi_t^u] &= \sqrt{\xi_t^u} \lambda_t^u \frac{1}{dt} E_t[dW_t^S dW_t^u] = \mu(t, u, \xi) \\ \lim_{dt \rightarrow 0} \frac{1}{dt} E_t[d \xi_t^u d \xi_t^{u'}] &= \lambda_t^u \lambda_t^{u'} \frac{1}{dt} E_t[dW_t^u dW_t^{u'}] = \nu(t, u, u', \xi) \end{aligned}$$



### 7.3 N-factor models

- Markovian representations of forward variance models are economically obtained by choosing exponential weightings for the driving Brownian motions:

$$d\xi_t^T = \omega \alpha_w \xi_t^T \sum_i w_i e^{-k_i(T-t)} dW_t^i$$

- $N$ -factor models are simulated by evolving, together with the spot process,  $N$  Ornstein-Ühlenbeck processes, which are easily simulated exactly.

- The number of driving factors in a model bears no relationship to the number of hedging instruments required; it simply sets the structure and rank of the break-even covariance matrix of the gamma/theta P&L of a hedged position.



### 7.4 A two-factor model

► Two factors afford sufficient flexibility as to volatilities and correlations of volatilities. The SDE of  $\xi_t^T$  reads:

$$\begin{aligned} d\xi_t^T &= (2\nu)\xi_t^T \alpha_\theta \left( (1-\theta) e^{-k_1(T-t)} dW_t^1 + \theta e^{-k_2(T-t)} dW_t^2 \right) \\ \alpha_\theta &= 1/\sqrt{(1-\theta)^2 + \theta^2 + 2\rho_{12}\theta(1-\theta)} \end{aligned}$$

where  $\nu$  is the volatility of a very short volatility. We introduce driftless processes  $x_t^T$ :

$$dx_t^T = \alpha_\theta \left[ (1-\theta) e^{-k_1(T-t)} dW_t^1 + \theta e^{-k_2(T-t)} dW_t^2 \right]$$

$\xi_t^T$  is given by:

$$\begin{aligned} \xi_t^T &= \xi_0^T f^T(t, x_t^T) \\ f^T(t, x) &= e^{\omega x - \frac{\omega^2}{2} \chi(t, T)} \end{aligned}$$

with  $\omega = 2\nu$  and  $\chi(t, T)$  given in (7.35).

► The two-factor model can be parametrized so that volatilities of spot-starting volatilities approximately match a power-law decay with maturity. We use the following benchmark:

$$\nu_T^B(t) = \sigma_0 \left( \frac{\tau_0}{T-t} \right)^\alpha$$

This benchmark form, for given values of  $\sigma_0, \tau_0, \alpha$ , can be approximately captured in the two-factor model with different sets of parameters.

This additional flexibility is utilized to generate different volatilities of forward-starting volatilities.

► While instantaneous forward variances are lognormal, discrete forward variances are not, thus variance swaptions exhibit a slight positive skew.

► For set parameters, volatilities of volatilities depend on the shape of the variance curve: they are larger for decreasing term structures of VS volatilities.

► The correlation structure of forward volatilities in the two-factor model is poor, as it involves one single time scale:  $\frac{1}{k_1 - k_2}$ . This could motivate the inclusion of additional factors.



### 7.6 Options on realized variance

► An option on realized variance of maturity  $T$  pays a call or put on the realized variance over  $[0, T]$ , measured with daily log-returns.

- The natural hedge instrument is the VS for the residual maturity. A simple model (SM) is built by specifying the dynamics of  $U_t$ , defined by:

$$U_t = \frac{Q_t + (T - t)\hat{\sigma}_T^2(t)}{T}$$

$Q_t$  is the quadratic variation at  $t$ , and  $\hat{\sigma}_T(t)$  is the VS volatility at  $t$  for maturity  $T$ .  $U_t$  has no drift.

- Assuming a lognormal dynamics for  $U_t$  given by:

$$\frac{dU_t}{U_t} = 2R_t \frac{T-t}{T} \nu_T(t) dW_t$$

yields the price of an option on realized variance, in the form of a simple Black-Scholes formula, where  $R_t = \frac{\hat{\sigma}_T^2(t)}{U_t}$  has been take equal to 1.

$$\begin{aligned} P(t, U) &= P_{\text{BS}}(t, U, \sigma_{\text{eff}}, T) \\ \sigma_{\text{eff}}^2 &= \frac{1}{T-t} \int_t^T 4 \left( \frac{T-\tau}{T} \right)^2 \nu_T^2(\tau) d\tau \end{aligned}$$

where  $\nu_T^2(\tau)$  is the volatility at  $\tau$  of a VS volatility of maturity  $T$ .

- Numerical tests whereby the option price computed in the SM is compared to that produced by the two-factor model, parametrized so that the volatilities of volatilities it generates match  $\nu_T(\tau)$ , show that, for the case of a flat term-structure of VS volatilities, the approximation in the SM is adequate.

- When the term structure of VS volatilities is not flat, the approximation  $R_t = 1$  is replaced with  $R_t = \frac{\hat{\sigma}_{\tau T}^2(0)}{\hat{\sigma}_T^2(0)}$ , which produces the following amended expression for  $\sigma_{\text{eff}}$ , at inception:

$$\sigma_{\text{eff}}^2 = \frac{4}{T} \int_0^T \left( \frac{T-\tau}{T} \right)^2 \left( \frac{\hat{\sigma}_{\tau T}^2(0)}{\hat{\sigma}_T^2(0)} \right)^2 \nu_T^2(\tau) d\tau$$

- The vega-hedge portfolio of an option on realized variance comprises, in addition to a VS of maturity  $T$ , a continuum of VSs of intermediate maturities. This vega hedge also functions as a gamma hedge.

- This hedge portfolio can be benchmarked against that produced by the two-factor model. The two hedges are slightly different, because of the sensitivity of volatilities of volatilities to the term structure of VS volatilities, in the two-factor model.

- The fact that variance does not accrue continuously, but is measured using discrete returns, impacts the value of options on realized variance, especially for short maturities. In case volatilities of volatilities vanish, one is still exposed to the intrinsic variance of the variance estimator itself. This effect is taken care of by

using the following expression for  $\sigma_{\text{eff}}^2$  where  $\kappa$  is the (conditional) kurtosis of daily returns.

$$\sigma_{\text{eff}}^2 = \frac{1}{T} \int_0^T 4 \left( \frac{T-\tau}{T} \right)^2 \left( \frac{\widehat{\sigma}_{\tau T}^2(0)}{\widehat{\sigma}_T^2(0)} \right)^2 \nu_T^2(\tau) d\tau + \frac{2+\kappa}{NT}$$

► Upper and lower bounds for prices of vanilla options on realized variance can be derived from the vanilla smile. If breached, a trading strategy consisting of a static position in a realized variance option and a portfolio of vanilla options, together with a dynamic delta position, nets a positive P&L.

► Options on forward realized variance cannot be priced in the SM as the latter only takes as ingredient the term structure of volatilities of spot-starting VS volatilities. What is needed, in addition, is the volatility of forward-starting volatilities, which cannot be backed-out of volatilities of spot-starting volatilities, in model-independent fashion. This is illustrated by pricing these options in the two-factor model with parameter sets that generate almost identical prices for spot-starting options. Prices of forward-starting options are different. They are also higher than prices of variance swaptions.



## 7.7 VIX futures and options

► The smile of VIX futures can be simply modeled by changing the function that maps processes  $x_t^T$  into forward variances  $\xi_t^T$ . We introduce the following simple parametrization:

$$f^T(t, x) = (1 - \gamma_T) e^{\omega_T x - \frac{\omega_T^2}{2} \chi(t, T)} + \gamma_T e^{\beta_T \omega_T x - \frac{(\beta_T \omega_T)^2}{2} \chi(t, T)}$$

Volatility-of-volatility smile parameters  $\gamma^T, \beta^T, \zeta^T$ , as well as forward variances  $\xi_0^T$ , are taken to be piecewise constant, with constant values for all forward variances that underlie a given VIX future. The model is calibrated by choosing  $\gamma^T, \beta^T, \zeta^T$  and  $\xi_0^T$  so that market values of (a) VIX futures, (b) VIX implied volatilities are matched. This is almost a local volatility model for forward variances.

► Different parameter sets of the two-factor model can be employed, resulting in very similar calibration accuracies. What distinguishes these different sets is the different distributions of volatility they generate for VIX futures.

► In case VIX futures are needed in the simulation of the two-factor model, they can be efficiently computed through a two-dimensional Gaussian quadrature. In case of very frequent observations, it is preferable to turn to the discrete forward variance models of Section 7.8.2.

► Options exist on ETNs whose investment strategies consists in maintaining a long position in the first two nearby VIX futures. These options cannot be priced

off VIX smiles in model-independent fashion, as they are very sensitive to the distribution of the volatility of VIX futures.

- Forward S&P 500 VSs can be synthesized using VIX instruments: futures and options. It is possible to set up a trading strategy that arbitrages the difference of these VIX-synthesized forward VSs with respect to forward VSs derived from the S&P 500 VS market.



## 7.8 Discrete forward variance models

► Discrete variance models arise out of the need to control future smiles independently from the correlation of spot and volatilities, and also to model VIX futures directly.

The specification of these models starts with a schedule of discrete dates  $T_i$ , then the model is built in two stages. First, we define a dynamics for discrete forward variances  $\xi_t^i = \hat{\sigma}_{T_i, T_{i+1}}^2(t)$ , then construct a dynamics for  $S_t$  that complies with that of the  $\xi_t^i$ .

► Since the  $\xi^i$  are driftless, as their continuous counterparts  $\xi^T$ , the two-factor model can be employed for the  $\xi^i$ .

When switching from the continuous to the discrete version of the model, we require that some features remain unchanged, so that volatility-of-volatility risks of different payoffs, calling for different schedules, are still priced at the same levels.

Starting from a parameter set for the continuous model, we generate parameters for the discrete model that ensure that instantaneous volatilities and correlations of spot-starting and forward VS volatilities for maturities  $T_i$  match in both models, for forward variances equal to their initial values.

► A benefit of discrete forward variance models is that VIX futures can be directly modeled, thus are readily accessible in a simulation of the model. VIX smiles can be calibrated exactly – the resulting model is in fact a local volatility model.

► Next we specify a dynamics for  $S_t$  that meets the following requirements: (a) the SDE for  $S_t$  complies with the dynamics of the  $\xi^i$ , (b) the density of  $\frac{S_{T_{i+1}}}{S_{T_i}}$  is independent on  $S_{T_i}$  so that payoffs  $\Sigma_i \omega_i f(\frac{S_{T_{i+1}}}{S_{T_i}})$  have zero sensitivity to spot/volatility correlations, (c) the dependence of future skews over intervals  $[T_i, T_{i+1}]$  on the corresponding VS volatilities can be set at will.

► This is achieved by using a path-dependent local volatility for  $S_t$ . We provide a simple parametrization of this local volatility function and give examples of two specifications corresponding to two typical future skew scenarios: future ATMF skews either independent on, or proportional to, ATMF volatilities.

► The ATMF skew of discrete forward variance models is generated by two mechanisms: (a) the local volatility functions  $\sigma^i$ , (b) the correlation of  $S_t$  and forward variances  $\xi^i$ . Working at order one in both the slope of local volatility functions  $\sigma^i$

and volatility of volatility, we express the ATMF skew for maturities  $T_i$  as the sum of two contributions, generated by both effects.

The portion of the ATMF skew generated by local volatility functions  $\sigma^i$  decays as  $\frac{1}{T}$ . Numerical tests confirm the accuracy of the order-one expansion.

**This page intentionally left blank**

# **Chapter 8**

---

## ***The smile of stochastic volatility models***

What is it that determines the shape of the smile generated by stochastic volatility models? We derive an expansion at order two in volatilities of volatilities, which is easily carried out in the forward variance framework introduced in Chapter 7, and characterize the smile near the money.<sup>1</sup> We derive in particular a general approximate expression for the ATMF skew that is accurate and can be used in practice.

We also present an alternative derivation of the order-one correction based on a representation of European option prices in terms of spot/volatility and volatility/volatility gamma P&Ls.

The characterization of the near-the-money smile in terms of the spot/variance and variance/variance covariance functions is used in Chapter 9 to establish a link between static and dynamic properties of stochastic volatility models.

Finally, efficient techniques for generating vanilla smiles in stochastic volatility models are explored in the appendix.

---

### **8.1 Introduction**

Any stochastic volatility model – including models based on the dynamics of the instantaneous variance  $V_t$  – can be written as a forward variance model. The corresponding pricing equation for a European payoff is given in (7.4), page 219:

$$\begin{aligned} \frac{dP}{dt} + (r - q) S \frac{dP}{dS} + \frac{\xi^t}{2} S^2 \frac{d^2 P}{dS^2} \\ + \frac{1}{2} \int_t^T du \int_t^T du' \nu(t, u, u', \xi) \frac{d^2 P}{\delta \xi^u \delta \xi^{u'}} + \int_t^T du \mu(t, u, \xi) S \frac{d^2 P}{dS \delta \xi^u} = rP \end{aligned} \quad (8.1)$$

---

<sup>1</sup>This is based on joint work with Julien Guyon, published in [13].

where the spot/variance and variance/variance covariance functions  $\mu, \nu$  are defined as:

$$\begin{aligned}\mu(t, u, \xi) &= \lim_{dt \rightarrow 0} \frac{1}{dt} E_t \left[ d\ln S_t d\xi_t^u \right] \\ \nu(t, u, u', \xi) &= \lim_{dt \rightarrow 0} \frac{1}{dt} E_t \left[ d\xi_t^u d\xi_t^{u'} \right]\end{aligned}$$

Remember that  $\mu$  and  $\nu$  can depend on the variance curve  $\xi$ , but are not allowed to explicitly depend on  $S$ : this excludes from our scope mixed local-stochastic volatility models.

Obviously the smile generated by a stochastic volatility model is a product of the joint dynamics of  $S_t$  and forward variances  $\xi_t^T$ ; is it possible to pinpoint precisely which functionals of the covariance functions  $\mu$  and  $\nu$  determine the shape of the vanilla smile?

We answer this question by deriving an expansion of vanilla option prices in powers of volatility of volatility. To this end we introduce a dimensionless parameter  $\varepsilon$  which we use to scale  $\mu$  and  $\nu$  according to:

$$\begin{aligned}\mu &\rightarrow \varepsilon \mu \\ \nu &\rightarrow \varepsilon^2 \nu\end{aligned}$$

Once the expansion at the desired order in  $\varepsilon$  is obtained we set  $\varepsilon = 1$ .

---

## 8.2 Expansion of the price in volatility of volatility

Consider a European option of maturity  $T$  whose payoff is  $g(S_T)$ . With no loss of generality we take vanishing interest rate and repo. The option's price  $P(t, S, \xi)$  solves the following backward equation:

$$\frac{dP}{dt} + H_t P = 0 \quad (8.2)$$

$H_t$  is given by:

$$H_t = H_t^0 + \varepsilon \mathcal{W}_t^1 + \varepsilon^2 \mathcal{W}_t^2$$

where operators  $H_t^0$ ,  $\mathcal{W}_t^1$ ,  $\mathcal{W}_t^2$  read:

$$H_t^0 = \frac{\xi_t^t}{2} (\partial_x^2 - \partial_x) \quad (8.3a)$$

$$\mathcal{W}_t^1 = \int_t^T du \mu(t, u, \xi) \partial_{x\xi^u}^2 \quad (8.3b)$$

$$\mathcal{W}_t^2 = \frac{1}{2} \int_t^T du \int_t^T du' \nu(t, u, u', \xi) \partial_{\xi^u \xi^{u'}}^2 \quad (8.3c)$$

where  $x = \ln S$  and  $\partial_x$ ,  $\partial_{x\xi^u}^2$ ,  $\partial_{\xi^u\xi^{u'}}^2$  stand respectively for  $\frac{d}{dx}$ ,  $\frac{d^2}{dx\delta\xi^u}$ ,  $\frac{d^2}{\delta\xi^u\delta\xi^{u'}}$  – note that derivatives with respect to  $\xi^u$  are functional derivatives. The terminal condition for  $P$  is:  $P(T, S, \xi) = g(S)$ . Let us write the expansion of  $P$  in powers of  $\varepsilon$  as:

$$P = P_0 + \varepsilon P_1 + \varepsilon^2 P_2 + \dots \quad (8.4)$$

Rather than inserting expression (8.4) in (8.2), deriving sequential PDEs for the  $P_i$  by equating to zero the contribution of each order in  $\varepsilon$ , and computing the  $P_i$  using the Feynman-Kac representation, we use the time-dependent perturbation technique. Integrating PDE (8.2),  $P$  is given by:

$$P(t, x, \xi) = U_{tT}g$$

where operator  $U_{st}$  with  $s \leq t$  is defined by:

$$U_{st} = \lim_{n \rightarrow \infty} (1 + \delta t H_{t_0})(1 + \delta t H_{t_1}) \cdots (1 + \delta t H_{t_{n-1}}) \quad (8.5)$$

with  $\delta t = \frac{t-s}{n}$  and  $t_i = s + i\delta t$ .  $U_{st}$  can be written as:

$$U_{st} = : \exp \left( \int_s^t H_\tau d\tau \right) :$$

where  $:$  indicates that the operators inside the colons are time-ordered, as in (8.5).  $U_{st}$  satisfies the semi-group property – for  $s \leq r \leq t$ :

$$U_{st} = U_{sr}U_{rt} \quad (8.6)$$

Let us write  $H_t = H_t^0 + \delta H_t$  where  $H_t^0$  is the unperturbed operator corresponding to the Black-Scholes model and  $\delta H_t$  is a perturbation. From (8.5), the expansion of  $U_{st}$  in powers of  $\delta H$  reads:

$$U_{st} = U_{st}^0 + \int_s^t d\tau U_{s\tau}^0 \delta H_\tau U_{\tau t}^0 + \int_s^t d\tau_1 \int_{\tau_1}^t d\tau_2 U_{s\tau_1}^0 \delta H_{\tau_1} U_{\tau_1\tau_2}^0 \delta H_{\tau_2} U_{\tau_2 t}^0 + \dots \quad (8.7)$$

$\delta H_t$  reads:  $\delta H_t = \varepsilon \mathcal{W}_t^1 + \varepsilon^2 \mathcal{W}_t^2$ . Inserting this expression in (8.7) and keeping terms up to order two in  $\varepsilon$  yields:

$$P_0 = U_{tT}^0 g \quad (8.8)$$

$$P_1 = \int_t^T d\tau U_{t\tau}^0 \mathcal{W}_\tau^1 U_{\tau T}^0 g \quad (8.9)$$

$$P_2 = \left( \int_t^T d\tau U_{t\tau}^0 \mathcal{W}_\tau^2 U_{\tau T}^0 + \int_t^T d\tau_1 \int_{\tau_1}^T d\tau_2 U_{t\tau_1}^0 \mathcal{W}_{\tau_1}^1 U_{\tau_1\tau_2}^0 \mathcal{W}_{\tau_2}^1 U_{\tau_2 T}^0 \right) g \quad (8.10)$$

The expression for  $U_{st}^0$  – the so-called free propagator – is:

$$U_{st}^0 = : \exp \left( \int_s^t H_\tau^0 d\tau \right) : = e^{\frac{1}{2} \left( \int_s^t \xi^\tau d\tau \right) (\partial_x^2 - \partial_x)} \quad (8.11)$$

where we have removed the time-ordering symbol in the right-hand side as operators  $H_t^0$  for different values of  $t$  commute.  $P_0$  in (8.8) is the standard Black-Scholes price:  $P(t, x, \xi) = P_{BS}(t, S, \hat{\sigma})$  with  $S = e^x$  and  $\hat{\sigma}^2 = \frac{1}{T-t} \int_t^T \xi^\tau d\tau$ . Before computing  $P_1$  and  $P_2$  observe that:

- as is clear from expression (8.11) for  $U_{st}^0$ ,  $\partial_x$  and  $U_{st}^0$  commute. This is equivalent to saying that  $\partial_x^n P_0$  is a martingale for all  $n$ . Specializing to the cases  $n = 1, 2$  this recovers the well-known property that the delta and dollar gammas  $S \frac{dP}{dS}$  and  $S^2 \frac{d^2P}{dS^2}$  of a European option in the Black-Scholes model are martingales.
- likewise  $\partial_{\xi^u}$  and  $U_{st}^0$  commute unless  $u \in [s, t]$ :

$$\partial_{\xi^u} U_{st}^0 = U_{st}^0 \partial_{\xi^u} + \mathbf{1}_{u \in [s, t]} \frac{1}{2} (\partial_x^2 - \partial_x) U_{st}^0 \quad (8.12)$$

Moreover,  $\partial_{\xi^u}^n g = 0$  as  $g$  is a function of  $x$  only. Applying relationship (8.12) on  $g$  with  $t = T$  and  $s = t$  yields:  $\partial_{\xi^u} U_{tT}^0 g = \frac{1}{2} (\partial_x^2 - \partial_x) U_{tT}^0 g$ . This expresses the already-mentioned property that, for a European option in the Black-Scholes model, vega (left-hand side) and gamma (right-hand side) are related:  $\frac{dP}{d(\sigma^2(T-t))} = \frac{1}{2} S^2 \frac{d^2P}{dS^2}$ .

Using these two rules, the semi-group property (8.6), and the fact that  $\mu$  does not depend on  $x$ , we get:

$$\begin{aligned} P_1 &= \int_t^T d\tau U_{t\tau}^0 \mathcal{W}_\tau^1 U_{\tau T}^0 g = \int_t^T d\tau U_{t\tau}^0 \int_\tau^T du \mu_{\tau,u} \partial_{x\xi^u}^2 U_{\tau T}^0 g \\ &= \int_t^T d\tau U_{t\tau}^0 \int_\tau^T du \mu_{\tau,u} \partial_x \frac{1}{2} (\partial_x^2 - \partial_x) U_{\tau T}^0 g \\ &= \int_t^T d\tau \int_\tau^T du \mu_{\tau,u} \partial_x \frac{1}{2} (\partial_x^2 - \partial_x) U_{t\tau}^0 U_{\tau T}^0 g \\ &= \frac{C_t^{x\xi}(\xi)}{2} \partial_x (\partial_x^2 - \partial_x) P_0 \end{aligned}$$

where we have used the more compact notation  $\mu_{\tau,u} \equiv \mu(\tau, u, \xi)$ . The dimensionless quantity  $C_t^{x\xi}(\xi)$  is given by:

$$C_t^{x\xi}(\xi) = \int_t^T d\tau \int_\tau^T du \mu(\tau, u, \xi) \quad (8.13a)$$

$$= \int_t^T (T - \tau) \langle d\ln S_\tau \, d\hat{\sigma}_T^2(\tau) \rangle \quad (8.13b)$$

Likewise, the first contribution to  $P_2$  in (8.10) is given by:

$$\begin{aligned} \int_t^T d\tau U_{t\tau}^0 \mathcal{W}_\tau^2 U_{\tau T}^0 g &= \frac{1}{2} \int_t^T d\tau U_{t\tau}^0 \int_\tau^T du \int_\tau^T du' \nu_{\tau,u,u'} \partial_{\xi^u \xi^{u'}}^2 U_{\tau T}^0 g \\ &= \frac{1}{8} \int_t^T d\tau U_{t\tau}^0 \int_\tau^T du \int_\tau^T du' \nu_{\tau,u,u'} (\partial_x^2 - \partial_x)^2 U_{\tau T}^0 g \\ &= \frac{C_t^{\xi\xi}(\xi)}{8} (\partial_x^2 - \partial_x)^2 P_0 \end{aligned}$$

where the dimensionless quantity  $C_t^{\xi\xi}(\xi)$  reads:

$$C_t^{\xi\xi}(\xi) = \int_t^T d\tau \int_\tau^T du \int_\tau^T du' \nu(\tau, u, u', \xi) \quad (8.14a)$$

$$= \int_t^T (T - \tau)^2 \langle d\hat{\sigma}_T^2(\tau) d\hat{\sigma}_T^2(\tau) \rangle \quad (8.14b)$$

The second contribution to  $P_2$  involves the spot/variance operator  $\mathcal{W}^1$  twice:

$$\begin{aligned} &\int_t^T d\tau_1 \int_{\tau_1}^T d\tau_2 U_{t\tau_1}^0 \mathcal{W}_{\tau_1}^1 U_{\tau_1\tau_2}^0 \mathcal{W}_{\tau_2}^1 U_{\tau_2 T}^0 g \\ &= \int_t^T d\tau_1 U_{t\tau_1}^0 \int_{\tau_1}^T d\tau_2 \int_{\tau_1}^T du \mu_{\tau_1,u} \partial_{x\xi^u}^2 U_{\tau_1\tau_2}^0 \int_{\tau_2}^T du' \mu_{\tau_2,u'} \partial_{x\xi^{u'}}^2 U_{\tau_2 T}^0 g \\ &= \frac{1}{2} \int_t^T d\tau_1 U_{t\tau_1}^0 \int_{\tau_1}^T du \mu_{\tau_1,u} \int_{\tau_1}^T d\tau_2 \partial_{x\xi^u}^2 U_{\tau_1\tau_2}^0 \int_{\tau_2}^T du' \mu_{\tau_2,u'} \partial_x (\partial_x^2 - \partial_x) U_{\tau_2 T}^0 g \\ &= \frac{1}{2} \int_t^T d\tau_1 U_{t\tau_1}^0 \int_{\tau_1}^T du \mu_{\tau_1,u} \partial_{x\xi^u}^2 \int_{\tau_1}^T d\tau_2 \int_{\tau_2}^T du' \mu_{\tau_2,u'} \partial_x (\partial_x^2 - \partial_x) U_{\tau_1 T}^0 g \\ &= \frac{1}{2} \partial_x^2 (\partial_x^2 - \partial_x) \int_t^T d\tau_1 U_{t\tau_1}^0 \int_{\tau_1}^T du \mu_{\tau_1,u} \partial_{\xi^u} C_{\tau_1}^{x\xi}(\xi) U_{\tau_1 T}^0 g \\ &= \frac{1}{4} \partial_x^2 (\partial_x^2 - \partial_x)^2 \int_t^T d\tau_1 \int_{\tau_1}^T du \mu_{\tau_1,u} C_{\tau_1}^{x\xi}(\xi) U_{t T}^0 g \quad (8.15) \end{aligned}$$

$$\begin{aligned} &+ \frac{1}{2} \partial_x^2 (\partial_x^2 - \partial_x) \left( \int_t^T d\tau_1 \int_{\tau_1}^T du \mu_{\tau_1,u} \frac{\delta C_{\tau_1}^{x\xi}(\xi)}{\delta \xi^u} \right) U_{t T}^0 g \\ &= \frac{C_t^{x\xi}(\xi)^2}{8} \partial_x^2 (\partial_x^2 - \partial_x)^2 P_0 + \frac{D_t(\xi)}{2} \partial_x^2 (\partial_x^2 - \partial_x) P_0 \quad (8.16) \end{aligned}$$

The dimensionless quantity  $D_t(\xi)$  reads:

$$D_t(\xi) = \int_t^T d\tau \int_\tau^T du \mu_{\tau u} \frac{\delta C_\tau^{x\xi}(\xi)}{\delta \xi^u} \quad (8.17a)$$

$$= \int_t^T d\tau \lim_{d\tau \rightarrow 0} \frac{1}{d\tau} E_\tau [d \ln S_\tau dC_\tau^{x\xi}] \quad (8.17b)$$

$$= \int_t^T d\tau \int_\tau^T du (T-u) \frac{1}{d\tau} \langle d \ln S_\tau d \left[ \frac{\langle d \ln S_u d\hat{\sigma}_T^2(u) \rangle}{du} \right] \rangle \quad (8.17c)$$

The alternative expressions in (8.17) for  $D_t(\xi)$  follow from the fact that  $C_\tau^{x\xi}(\xi)$  is a functional of the variance curve. Indeed, using the definition of  $\mu_{\tau u}$ :

$$\begin{aligned} \int_\tau^T du \mu_{\tau u} \frac{\delta C_\tau^{x\xi}(\xi)}{\delta \xi^u} &= \int_\tau^T du \lim_{d\tau \rightarrow 0} \frac{1}{d\tau} E_\tau [d \ln S_\tau d\xi^u] \frac{\delta C_\tau^{x\xi}(\xi)}{\delta \xi^u} \\ &= \lim_{d\tau \rightarrow 0} \frac{1}{d\tau} E_\tau [d \ln S_\tau dC_\tau^{x\xi}] \\ &= \int_\tau^T du \int_u^T du' \lim_{d\tau \rightarrow 0} \frac{1}{d\tau} E_\tau [d \ln S_\tau d\mu_{uu'}] \end{aligned}$$

Note that the definition of  $C_{\tau_1}^{x\xi}(\xi)$  in (8.13) has allowed for the following simplification in (8.15):

$$\int_t^T d\tau_1 \int_{\tau_1}^T du \mu_{\tau_1, u} C_{\tau_1}^{x\xi}(\xi) = - \int_t^T d\tau_1 \frac{dC_{\tau_1}^{x\xi}(\xi)}{d\tau_1} C_{\tau_1}^{x\xi}(\xi) = \frac{1}{2} C_t^{x\xi}(\xi)^2$$

The final expression of  $P$  at order two in  $\varepsilon$  at  $t = 0$  is thus:

$$\begin{aligned} P &= \left[ 1 + \varepsilon \frac{C_0^{x\xi}(\xi_0)}{2} \partial_x (\partial_x^2 - \partial_x) \right. \\ &\quad \left. + \varepsilon^2 \left( \frac{C_0^{\xi\xi}(\xi_0)}{8} (\partial_x^2 - \partial_x)^2 + \frac{C_0^{x\xi}(\xi_0)^2}{8} \partial_x^2 (\partial_x^2 - \partial_x)^2 + \frac{D_0(\xi_0)}{2} \partial_x^2 (\partial_x^2 - \partial_x) \right) \right] P_0 \end{aligned} \quad (8.18)$$

The subscript 0 in  $\xi_0$  indicates that  $C_0^{x\xi}, C_0^{\xi\xi}, D_0$  are evaluated in the unperturbed state, that is using the variance curve observed at time  $t = 0$ .

## Discussion

- The corrections to the Black-Scholes price  $P_0$  are obtained in terms of derivatives of  $P_0$  with respect to  $x = \ln S$ . One can check on (8.18) that the contributions at order 1 and 2 in  $\varepsilon$  are of the form:

$$\sum_{n=3}^{\infty} \alpha_n (\partial_x^n - \partial_x^2) P_0 \quad (8.19)$$

which is in agreement with expression (5.88), page 193, for general perturbations of the lognormal density of  $S_T$  that leave the implied volatility of the log contract unchanged.

In our expansion in powers of  $\varepsilon$ , forward variances  $\xi^T$  are driftless:  $E[\xi_T^T] = \xi_0^T$  at each order. Thus VS implied volatilities – equal to log contract implied volatilities since we are in a diffusive setting – stay fixed as  $\varepsilon$  is varied. This stands in contrast with other types of expansions whose accuracy is marred by the fact that the overall level of implied volatilities in the model shifts as the volatility-of-volatility parameter is increased.

- We have already derived the order-one expansion in the special case of the Heston model – see equation (6.15), page 209 – it is exactly as in (8.18).
- At second order in volatility of volatility,  $P$  only depends on 3 dimensionless model-dependent numbers:  $C_0^{x\xi}(\xi_0)$ ,  $C_0^{\xi\xi}(\xi_0)$  and  $D_0^\mu(\xi_0)$  summarize the joint spot/variance dynamics of the model at hand. While  $C_0^{x\xi}(\xi_0)$  and  $C_0^{\xi\xi}(\xi_0)$  are integrals of the spot/variance and variance/variance covariance functions evaluated on the initial variance curve,  $D_0^\mu(\xi_0)$  involves an extra degree of model-dependence as it depends on the derivative of  $C_t^{x\xi}(\xi)$  with respect to  $\xi^u$ : it incorporates the additional information of how  $\mu(t, u, \xi)$  changes as the variance curve changes.
- Comparison of expression (8.18) with expression (5.88), page 193, shows that the correction to  $P_0$  at order 1 in  $\varepsilon$  is of the same form as that generated at order one by the third order cumulant of the distribution of  $\ln S_T$ . Indeed, the expansion at order one in volatility of volatility can be derived by direct calculation of the third-order cumulant  $\kappa_3$  of  $\ln S_T$ , at this order. The interested reader will find the derivation in [11].
- The observant reader will have spotted in (8.18) what looks like the beginning of the expansion of  $\exp\left(\varepsilon \frac{C_0^{x\xi}(\xi_0)}{2} \partial_x (\partial_x^2 - \partial_x)\right)$ . This is not fortuitous – more on this in Appendix C, page 347.

### 8.3 Expansion of implied volatilities

We now convert the expansion for the price (8.18) into an expansion for implied volatilities – the (tedious but straightforward) derivation can be found in [13].

Remarkably, at order  $\varepsilon^2$ , implied volatilities are exactly quadratic in log-moneyness:<sup>2</sup>

$$\hat{\sigma}(K, T) = \hat{\sigma}_{FTT} + \mathcal{S}_T \ln \left( \frac{K}{F_T} \right) + \frac{\mathcal{C}_T}{2} \ln^2 \left( \frac{K}{F_T} \right) + O(\varepsilon^3) \quad (8.20)$$

The ATMF volatility  $\hat{\sigma}_{FTT}$ , the ATMF skew  $\mathcal{S}_T$  and curvature  $\mathcal{C}_T$  are given by:

$$\begin{aligned} \hat{\sigma}_{FTT} &= \hat{\sigma}_T \left[ 1 + \frac{\varepsilon}{4Q} C^{x\xi} \right. \\ &\quad \left. + \frac{\varepsilon^2}{32Q^3} \left( 12(C^{x\xi})^2 - Q(Q+4)C^{\xi\xi} + 4Q(Q-4)D \right) \right] \end{aligned} \quad (8.21a)$$

$$\mathcal{S}_T = \hat{\sigma}_T \left[ \frac{\varepsilon}{2Q^2} C^{x\xi} + \frac{\varepsilon^2}{8Q^3} \left( 4QD - 3(C^{x\xi})^2 \right) \right] \quad (8.21b)$$

$$\mathcal{C}_T = \hat{\sigma}_T \frac{\varepsilon^2}{4Q^4} \left( 4QD + QC^{\xi\xi} - 6(C^{x\xi})^2 \right) \quad (8.21c)$$

where  $\hat{\sigma}_T$  is the VS volatility for maturity  $T$ ,  $Q = \hat{\sigma}_T^2 T$  and  $C^{x\xi}, C^{\xi\xi}, D$  are compact notations for  $C_0^{x\xi}(\xi_0), C_0^{\xi\xi}(\xi_0)$  and  $D_0(\xi_0)$  defined in (8.13), (8.14), (8.17). As we use these formulas further below, we will set  $\varepsilon = 1$ .

- At order one in  $\varepsilon$ , from (8.21b), the ATMF skew is given by

$$\mathcal{S}_T = \hat{\sigma}_T \frac{C^{x\xi}}{2(\hat{\sigma}_T^2 T)^2} \quad (8.22)$$

where we have set  $\varepsilon = 1$ . Whenever spot and variances are uncorrelated  $\mathcal{S}_T$  vanishes both at order  $\varepsilon$  and  $\varepsilon^2$ , and at all orders – as it should, since it is a well-known result that the smile is symmetric in log-moneyness for uncorrelated spot and variances.<sup>3</sup>

- At order one in  $\varepsilon$ ,  $\mathcal{S}_T$  is simply proportional to the doubly-integrated spot/variance covariance function. At this order, one recovers the correction at order one in the cumulant expansion in Appendix B of Chapter 5, contributed by the third-order cumulant, or equivalently the skewness  $s_T$  of  $\ln S_T$ . The interested reader will find in [11] the derivation of  $s_T$  at order one in  $\varepsilon$ :

$$s_T = \frac{3C^{x\xi}}{(\hat{\sigma}_T^2 T)^{\frac{3}{2}}}$$

Using then formula (5.93) relating the skewness of  $\ln S_T$  to the ATMF skew:

$$\mathcal{S}_T = \frac{s_T}{6\sqrt{T}} \quad (8.23)$$

yields (8.22).

---

<sup>2</sup>The cancellation at order  $\varepsilon^2$  of higher-order terms in log-moneyness was already noted in [69], for the particular case of a one-factor model. Note that a different convention for the normalization of  $\mathcal{C}_T$  is used in [13].

<sup>3</sup>See footnote 8, page 328.

- Using expression (8.13) for  $C^{x\xi}$ , (8.22) can be rewritten, more meaningfully, as:

$$\mathcal{S}_T = \frac{1}{2\hat{\sigma}_T^3 T} \int_0^T \frac{T-\tau}{T} \frac{\langle d\ln S_\tau d\hat{\sigma}_T^2(\tau) \rangle_0}{d\tau} d\tau \quad (8.24)$$

where the 0 subscript indicates that the instantaneous covariation is evaluated with the initial VS term structure. The ATMF skew is the weighted average of the instantaneous covariance of the spot and the VS volatility for the residual maturity at future dates.

The sagacious reader will remember that this exact same relationship was derived in the context of the local volatility model – see formula (2.89), page 62. We show in the following section that (8.24) follows naturally from a representation of option prices in terms of expectations of spot/volatility and volatility/volatility gamma P&Ls.

Specializing to the case of a flat term structure of VS volatilities equal to  $\hat{\sigma}_T$ :

$$\mathcal{S}_T = \frac{1}{\hat{\sigma}_T^2 T} \int_0^T \frac{T-\tau}{T} \frac{\langle d\ln S_\tau d\hat{\sigma}_T(\tau) \rangle_0}{d\tau} d\tau \quad (8.25)$$

- At order one the ATMF volatility is given by:

$$\hat{\sigma}_{FTT} = \hat{\sigma}_T + \frac{Q}{2} \mathcal{S}_T \quad (8.26)$$

In (8.26) we recover, at order one in  $\varepsilon$ , the relationship between ATMF and VS volatilities that we had already derived in (3.30) at order one in  $\mathcal{S}_T$ .

- While the exact smile is arbitrage-free by construction, approximate prices (8.18) or implied volatilities (8.20) are not. Expression (8.20), which is quadratic in log-moneyness, is bound to generate arbitrage for very large or small strikes.

Indeed we know from [67] that asymptotically  $\hat{\sigma}^2(K, T)$  is at most an affine function of log-moneyness – see Section 4.3.1. The presence of arbitrage for far-away strikes can also be assessed on the density of  $\ln S_T$  directly.

Consider the expansion of  $P$  in (8.18). As mentioned in the discussion above, the correction to  $P_0$  is exactly of the form (5.88), obtained in a cumulant expansion of the Gaussian density of  $x = \ln \frac{S_T}{F_T}$ . This corresponds to a perturbation of the density of  $x$ ,  $\delta\rho(x)$ , given by (5.85), which, using (5.83), can also be written as:

$$\frac{\delta\rho(x)}{\rho_0(x)} = \sum_{n=3}^6 \frac{\delta\kappa_n}{(\hat{\sigma}_T\sqrt{T})^n\sqrt{n!}} H_n\left(\frac{x + \frac{\hat{\sigma}_T^2 T}{2}}{\hat{\sigma}_T\sqrt{T}}\right)$$

where  $\rho_0$  is the unperturbed Gaussian density with volatility  $\hat{\sigma}_T$  and  $H_n$  is the Hermite polynomial of degree  $n$ , defined in (5.84).

Expression (8.18) for  $P$  thus corresponds to a density for  $\ln \frac{S_T}{F_T}$  which is the unperturbed Gaussian density multiplied by 1 plus a linear combination of Hermite polynomials in  $\ln \frac{S_T}{F_T}$  of order up to 6.

For sufficiently large values of  $S_T$  this density may – and will – be negative. As we will see shortly, approximation (8.20) is most accurate for strikes near the money. The good agreement of (8.20) with the exact result implies that, in practice, arbitrage in this range of strikes is unlikely.

---

## 8.4 A representation of European option prices in diffusive models

We now give an alternative derivation of the order-one expansion that uses a representation of European option prices in general diffusive models in terms of spot/volatility and volatility/volatility gamma P&Ls.

In Section 2.4.1, page 39, we expressed the price of an option in a model with instantaneous volatility  $\sigma_2$  as the sum of (a) the price in a base model with instantaneous volatility  $\sigma_1$ , (b) the expectation of the integral over the option’s lifetime of the option’s dollar gamma multiplied by the difference ( $\sigma_2^2 - \sigma_1^2$ ).

Imagine however that, in addition to delta-hedging, gamma risk is hedged away by dynamically trading vanilla options – or VSs. Our P&L now arises from the spot/volatility cross-gamma and the volatility gamma, and it should be possible to express the price of a European option as the cost of these two gammas. We now make this notion explicit, for general diffusive models, for European options.

Consider the process  $Q_t$  defined by:

$$\begin{aligned} Q_t &= Q\left(t, S_t, \omega_t = \frac{1}{T-t} \int_t^T \xi^\tau d\tau\right) \\ Q(t, S, \omega) &= e^{-rt} P_{\text{BS}}(t, S, \omega) \end{aligned}$$

where  $P_{\text{BS}}(t, S, \omega)$  is the option’s price in a Black-Scholes model where we use variance  $\omega$  rather than volatility, and  $\xi^\tau$  are VS forward variances. With respect to the reasoning in Section 2.4.1, this amounts to using the VS of maturity  $T$  as a hedge instrument, in addition to  $S$ .

$Q(t, S, \omega)$  is the undiscounted price in a Black-Scholes model with constant variance  $\omega$ , thus obeys the following PDE:

$$\frac{dQ}{dt} + (r - q) S \frac{dQ}{dS} + \frac{\omega}{2} S^2 \frac{d^2Q}{dS^2} = 0 \quad (8.27)$$

Moreover, the relationship between vega and gamma in the Black-Scholes model – see Appendix A of Chapter 5 – reads, in our context:

$$\frac{dQ}{d\omega} = \frac{T-t}{2} S^2 \frac{d^2 Q}{dS^2} \quad (8.28)$$

During  $dt$  the variation of  $Q_t$  is given by:

$$\begin{aligned} dQ_t &= \frac{dQ}{dt} dt + \frac{dQ}{dS} dS_t + \frac{dQ}{d\omega} \left( \int_t^T \frac{\delta\omega}{\delta\xi^\tau} d\xi_t^\tau d\tau + \frac{d\omega}{dt} dt \right) + \frac{1}{2} \frac{d^2 Q}{dS^2} \langle dS_t^2 \rangle \\ &\quad + \frac{d^2 Q}{dS d\omega} \langle dS_t d\omega_t \rangle + \frac{1}{2} \frac{d^2 Q}{d\omega^2} \langle d\omega_t^2 \rangle \end{aligned}$$

where all derivatives are evaluated at  $t, S_t, \xi_t$ .

From the definition of  $\omega$ ,  $\frac{d\omega}{dt} = \frac{\omega - \xi_t^t}{T-t} dt$ . We then have, using (8.28):

$$\frac{dQ}{d\omega} \frac{d\omega}{dt} = \frac{dQ}{d\omega} \frac{\omega - \xi_t^t}{T-t} = \frac{\omega - \xi_t^t}{2} S^2 \frac{d^2 Q}{dS^2}$$

Using now (8.27), we have:

$$\begin{aligned} &\frac{dQ}{dt} dt + \frac{dQ}{dS} dS_t + \frac{dQ}{d\omega} \left( \int_t^T \frac{\delta\omega}{\delta\xi^\tau} d\xi_t^\tau d\tau + \frac{d\omega}{dt} \right) + \frac{1}{2} \frac{d^2 Q}{dS^2} \langle dS_t^2 \rangle \\ &= -(r-q)S_t \frac{dQ}{dS} dt + \frac{dQ}{dS} dS_t + \frac{dQ}{d\omega} \int_t^T \frac{\delta\omega}{\delta\xi^\tau} d\xi_t^\tau d\tau \\ &\quad + \frac{1}{2} S^2 \frac{d^2 Q}{dS^2} \left( (\omega - \xi_t^t) dt - \omega dt + \frac{\langle dS_t^2 \rangle}{S_t^2} \right) \\ &= \frac{dQ}{dS} (dS_t - (r-q)S_t dt) + \frac{dQ}{d\omega} \int_t^T \frac{\delta\omega}{\delta\xi^\tau} d\xi_t^\tau d\tau \end{aligned}$$

where we have used that  $\frac{\langle dS_t^2 \rangle}{S_t^2} = \xi_t^t$ , a property of diffusive models.

The spot/gamma P&L has cancelled out. This is normal; because of the vega/gamma relationship for European options in the Black-Scholes model, the VS used as vega hedge also functions as a gamma hedge. We then have:

$$\begin{aligned} dQ_t &= \frac{dQ}{dS} (dS_t - (r-q)S_t dt) + \frac{dQ}{d\omega} \int_t^T \frac{\delta\omega}{\delta\xi^\tau} d\xi_t^\tau d\tau \\ &\quad + \frac{d^2 Q}{dS d\omega} \langle dS_t d\omega_t \rangle + \frac{1}{2} \frac{d^2 Q}{d\omega^2} \langle d\omega_t^2 \rangle \end{aligned}$$

Let us now take the expectation of  $dQ_t$ . The first two contributions vanish as  $E[dS_t - (r-q)S_t dt] = 0$  and the  $\xi_t^\tau$  are martingales:

$$E[dQ_t | S_t, \xi_t] = \frac{d^2 Q}{dS d\omega} \langle dS_t d\omega_t \rangle + \frac{1}{2} \frac{d^2 Q}{d\omega^2} \langle d\omega_t^2 \rangle$$

Integrating now this expression on  $[0, T]$ :

$$E[Q_T] = Q_0 + E \left[ \int_0^T e^{-rt} \left( \frac{d^2 P_{\text{BS}}}{dS d\omega} \langle dS_t d\omega_t \rangle + \frac{1}{2} \frac{d^2 P_{\text{BS}}}{d\omega^2} \langle d\omega_t^2 \rangle \right) \right]$$

At  $t = T$ ,  $Q_T = e^{-rT} P(T, S_T, \omega_T) = e^{-rT} f(S_T)$  where  $f$  is the option's payoff.  $E[Q_T]$  is thus simply the option's price.

We then have our final representation of the price of a European option in a diffusive model as the sum of the Black-Scholes price with the initial VS volatility, augmented by the expectation of the sum of spot/volatility and volatility/volatility gamma P&Ls, the volatility being the VS volatility for the residual maturity:

$$\begin{aligned} P &= P_{\text{BS}}(0, S_0, \hat{\sigma}_T^2(0)) \\ &+ E \left[ \int_0^T e^{-rt} \left( \frac{d^2 P_{\text{BS}}}{dS d(\hat{\sigma}_T^2)} \langle dS_t d\hat{\sigma}_T^2(t) \rangle + \frac{1}{2} \frac{d^2 P_{\text{BS}}}{(d(\hat{\sigma}_T^2))^2} \langle d\hat{\sigma}_T^2(t) d\hat{\sigma}_T^2(t) \rangle \right) \right] \end{aligned} \quad (8.29)$$

where  $\hat{\sigma}_T^2(t) = \omega_t$  is the square of the VS volatility of maturity  $T$  at time  $t$ .

Unlike representation (2.30), page 40, based on spot/spot gamma P&L, (8.29) involves spot/volatility and volatility/volatility gamma P&Ls.

The above derivation is equivalent to considering the P&L of a delta-hedged, vega-hedged (and also gamma-hedged) position, the vega-hedge instruments being VSSs.

Can we obtain a similar formula using implied volatilities of other payoffs, for example ATMF options rather than VSSs? The answer is no, the reason being that the square of the VS implied volatility is what comes closest to a price – it has zero drift.

#### 8.4.1 Expansion at order one in volatility of volatility

With volatility of volatility switched off, the  $\xi^\tau$  are frozen, and  $\hat{\sigma}_T^2$  has only a drift:  $\langle dS_t d\hat{\sigma}_T^2 \rangle = \langle d\hat{\sigma}_T^2 d\hat{\sigma}_T^2 \rangle = 0$ .

Imagine scaling the volatilities of the  $\xi^\tau$  by a factor  $\varepsilon$ . Then  $\langle dS_t d\hat{\sigma}_T^2 \rangle$  is of order one in  $\varepsilon$ , and  $\langle d\hat{\sigma}_T^2 d\hat{\sigma}_T^2 \rangle$  is of order two. At order one in  $\varepsilon$ ,  $P$  is thus given by:

$$P = P_0 + E \left[ \int_0^T e^{-rt} \frac{d^2 P_{\text{BS}}}{dS d(\hat{\sigma}_T^2)} \langle dS_t d\hat{\sigma}_T^2(t) \rangle \right] \quad (8.30)$$

where the expectation is taken with respect to the density at order zero – that is the lognormal density of a Black-Scholes model with instantaneous volatility  $\sqrt{\xi_0^\varepsilon}$  and  $P_0 = P_{\text{BS}}(0, S_0, \omega_0 = \hat{\sigma}_0^2)$ .

Using  $x = \ln S$  the vega/gamma relationship (8.28) reads:

$$\frac{dP_{\text{BS}}}{d(\hat{\sigma}_T^2)} = \frac{T-t}{2} \left( \frac{d^2 P_{\text{BS}}}{dx^2} - \frac{dP_{\text{BS}}}{dx} \right)$$

Switching from  $S_t$  to  $\ln S_t$ :

$$\begin{aligned} P &= P_0 + E \left[ \int_0^T e^{-rt} \frac{d^2 P_{\text{BS}}}{d \ln S d(\hat{\sigma}^2)} \langle d \ln S_t d \hat{\sigma}_T^2 \rangle \right] \\ &= P_0 + E \left[ \int_0^T e^{-rt} \frac{T-t}{2} \left( \frac{d^3 P_{\text{BS}}}{dx^3} - \frac{d^2 P_{\text{BS}}}{dx^2} \right) \langle d \ln S_t d \hat{\sigma}_T^2 \rangle \right] \end{aligned}$$

Let us assume that  $E [\langle d \ln S_t d \hat{\sigma}_T^2 \rangle | x]$  does not depend on  $x$ .

We can then use the property – see Appendix A of Chapter 5 – that, in the Black-Scholes model  $E [e^{-rt} \frac{d^n P}{dx^n}] = \frac{d^n P_0}{dx^n}$  and we get:

$$P = P_0 + \frac{1}{2} E \left[ \int_0^T (T-t) \langle d \ln S_t d \hat{\sigma}_T^2 \rangle \right] \left( \frac{d^3 P_{\text{BS}}}{dx^3} - \frac{d^2 P_{\text{BS}}}{dx^2} \right) \quad (8.31)$$

This recovers the expansion at order one in (8.18) – we refer the reader to the expression of  $C^{x\xi}$  in (8.13). We can then immediately write down the formula for the ATMF skew:

$$\mathcal{S}_T = \frac{1}{2\hat{\sigma}_T^3(0)T} \int_0^T \frac{T-t}{T} \frac{\langle d \ln S_t d \hat{\sigma}_T^2(t) \rangle_0}{dt} dt \quad (8.32)$$

where the subscript 0 signals that the instantaneous covariation, as a function of forward variances, is evaluated using initial values for the latter. This is exactly formula (8.24). Note that we can substitute in (8.32) the VS volatility with the ATMF volatility as they are identical at order one in volatility of volatility.

It is now clear why, in the derivation of the expansion at order two in Section 8.2, we needed the assumption that  $\mu$  does not depend on  $S$ .

### Local volatility

We have already encountered formula (8.32) in the context of local volatility in Chapter 2 – see Section 2.5.7, page 61. In that context,  $\hat{\sigma}_T(t)$  is the ATMF volatility, rather than the VS volatility, but that is fine, as at order zero in volatility of volatility they are identical.

The reason (8.32) holds in the local volatility model is we have used a local volatility function of type:

$$\sigma(t, S) = \sigma_0 + \alpha(t) \ln \frac{S}{F_t}$$

From equation (2.88), page 61, at order one in  $\alpha(t)$ , the covariance of  $\widehat{\sigma}_{F_T-T}^2$  with  $\ln S$  is:

$$\langle d \ln S_t d \widehat{\sigma}_T^2 \rangle = 2\sigma_0^2 \left( \frac{1}{T-t} \int_t^T \alpha(\tau) d\tau \right) dt$$

It is independent of  $S$ , hence (8.32) applies.

### 8.4.2 Materializing the spot/volatility cross-gamma P&L

At order one in volatility of volatility, and assuming that the covariance of  $\ln S$  and  $\widehat{\sigma}_T^2$  does not depend on  $S$ , the price of a European payoff is given by (8.31). While  $\frac{d^3 P_{BS}}{dx^3} - \frac{d^2 P_{BS}}{dx^2}$  is payoff-dependent, the prefactor involving the weighted average of the spot/volatility covariance is not. Thus, at order one in volatility of volatility, this quantity could be read off any European payoff's market price.

Practically however, backing out of option prices an implied value for the integrated spot/volatility covariance is useful only if the latter can be materialized as a cross-gamma P&L. Does there exist a payoff such that by delta-hedging it and vega-hedging it with VSs we generate as a cross-gamma P&L the integrated spot/volatility covariance in (8.31)? In other words, which is the payoff out of which a measure of the *implied* integrated spot/volatility covariance can be extracted?

This payoff is  $\ln^2(S/S_0)$ . In the Black-Scholes model the price of payoff  $\ln^2(S_T/S_0)$  is given by:

$$P_{BS} = e^{-r\tau} \left[ \ln^2 \frac{S}{S_0} + \left( r - q - \frac{\sigma^2}{2} \right)^2 \tau^2 + 2 \left( r - q - \frac{\sigma^2}{2} \right) \tau \ln \frac{S}{S_0} + \sigma^2 \tau \right]$$

where  $\tau = T - t$ . The derivative with respect to  $\sigma^2$  is given by:

$$\frac{dP_{BS}}{d(\sigma^2)} = -e^{-r(T-t)} (T-t) \left[ \ln \frac{S}{S_0} + \left( r - q - \frac{\sigma^2}{2} \right) (T-t) - 1 \right]$$

By choosing  $S_0 = S_{t=0} e^{(r-q-\frac{\widehat{\sigma}_T^2(0)}{2})T}$ , the VS hedge ratio at  $t = 0$  vanishes. The cross-derivative with respect to  $\ln S$  and  $\sigma^2$  is given by:

$$\frac{d^2 P_{BS}}{d(\sigma^2) d \ln S} = -e^{-r(T-t)} (T-t)$$

Consider now delta-hedging and vega-hedging with VSs a short position in this payoff. Summing spot/volatility cross-gamma P&Ls and discounting them to  $t = 0$  yields the following expression for our *P&L* over  $[0, T]$ :

$$P&L = e^{-rT} \sum_i (T - t_i) \delta \ln S_i \delta(\widehat{\sigma}_T^2(t_i)) \quad (8.33)$$

Up to factor  $e^{-rT}$ , this is exactly the prefactor in (8.31).

At order one in volatility of volatility the market price of payoff  $\ln^2(S_T/S_0)$  minus its Black-Scholes price calculated with the VS volatility at  $t = 0$  thus provides a measure of the implied value of the integrated spot/volatility covariance.<sup>4</sup>

As a European payoff,  $\ln^2(S/S_0)$  can be replicated with vanilla options. The density  $\rho(K)$  of vanilla options of strike  $K$  is given by the second derivative of the payoff with respect to  $S$ :

$$\rho(K) = \frac{2}{K^2} \left( 1 - \ln \frac{K}{S_0} \right)$$

$\rho(K)$  is positive for  $K \ll S_0$  and negative for  $K \gg S_0$ ; the price of  $\ln^2(S_T/S_0)$  is a global measure of the slope of the smile.

---

## 8.5 Short maturities

We resume now our discussion of expansion (8.21) and first consider the limit of vanishing maturities. This special case is worth investigating as the variance curve then collapses to a single object – the instantaneous variance – whose dynamics determines the smile. This allows for a particularly simple characterization of the smile at order two in volatility of volatility.

In addition, while  $\widehat{\sigma}_{F_T T}, \mathcal{S}_T, \mathcal{C}_T$  in (8.21) have been derived in an expansion at order two in volatility of volatility, it turns out that the expressions they provide for the ATM volatility, skew and curvature are exact in the limit  $T \rightarrow 0$ .

Let us take the limit  $T \rightarrow 0$  in equations (8.21). We assume that the covariance functions  $\mu$  and  $\nu$  are smooth in  $t = 0$ . From their expressions in (8.13), (8.14), (8.17),  $C^{x\xi}$  is of order  $T^2$ ,  $C^{\xi\xi}$  of order  $T^3$  and  $D$  of order  $T^3$ .<sup>5</sup> At leading order in  $T$ :

$$\begin{aligned} C^{x\xi} &= \frac{T^2}{2} \mu_0 \\ C^{\xi\xi} &= \frac{T^3}{3} \nu_0 \\ D &= \frac{T^3}{6} \mu_0 \frac{d\mu_0}{d\xi_0^0} \end{aligned}$$

where we have used the compact notation:  $\mu_0 = \mu(0, 0, \xi_0^0)$ ,  $\nu_0 = \nu(0, 0, 0, \xi_0^0)$ .

<sup>4</sup>The total P&L at order two in  $\delta \ln S$  and  $\delta \widehat{\sigma}$  also comprises a contribution from the volatility/volatility gamma. We have:  $\frac{d^2 P_{BS}}{(d\sigma^2)^2} = e^{-rT} \frac{\tau^2}{2}$ . A short position in payoff  $\ln^2(S_T/S_0)$ , delta-hedged and vega-hedged with VSs thus generates, in addition to (8.33), the volatility/volatility gamma P&L  $-e^{-rT} \sum_i \frac{(T-t_i)^2}{2} \delta(\widehat{\sigma}_T^2(t_i))^2$ . This contribution is of second order in volatility of volatility.

<sup>5</sup>Despite what expression (8.17) suggests  $D$  is not of order 4. Unlike a standard derivative, a functional derivative has an additional dimension  $\frac{1}{T}$ :  $\frac{\delta C}{\delta \xi^u}$  is such that  $dC = \int_0^T \frac{\delta C}{\delta \xi^u} d\xi^u du$ .

- In stochastic volatility models – in contrast with jump/Lévy models – the drivers in the dynamics of  $S_t$  are Brownian motions: the non-Gaussian character of  $\ln S_T$  is only generated by the volatility of volatility. As  $T \rightarrow 0$  the distribution of  $\ln S_T$  does become Gaussian, but exactly how fast it becomes Gaussian determines the behavior of the smile for short maturities. Consider the skewness  $s_T$  of  $\ln S_T$  and the ATM skew  $\mathcal{S}_T$ , which are related at order one in volatility of volatility. Using (8.23), (8.21b) and the above expression for  $C^{x\xi}$  we get the following expression for  $s_T$ , at order one in volatility of volatility:

$$s_T = \frac{3}{2} \frac{\mu_0}{\hat{\sigma}_T^3} \sqrt{T} \quad (8.34)$$

This expression is instructive: as  $T \rightarrow 0$ ,  $s_T$  vanishes like  $\sqrt{T}$ . (8.23) then implies that  $\mathcal{S}_T$  tends to a constant. Only if the skewness vanishes faster than  $\sqrt{T}$  does the short-maturity ATM skew tend to zero.

Keeping only terms at leading order in  $T$  and setting  $\varepsilon = 1$  we get:

$$\hat{\sigma}_{S,T=0} = \hat{\sigma}_0 \quad (8.35a)$$

$$\mathcal{S}_0 = \hat{\sigma}_0 \frac{1}{4(\xi_0^0)^2} \mu_0 \quad (8.35b)$$

$$\mathcal{C}_0 = \hat{\sigma}_0 \frac{1}{4(\xi_0^0)^4} \left( \frac{2}{3} \xi_0^0 \mu_0 \frac{d\mu_0}{d\xi_0^0} + \frac{1}{3} \xi_0^0 \nu_0 - \frac{3}{2} \mu_0^2 \right) \quad (8.35c)$$

where  $\hat{\sigma}_0 = \sqrt{\xi_0^0}$ .

- In the limit  $T \rightarrow 0$  the ATM volatility is equal to the VS volatility.

- Remembering that  $\mu_0 = \frac{\langle d\ln S_0 d\xi_0^0 \rangle}{dt}$  we see from (8.35b) that, as  $T \rightarrow 0$ , the short ATM skew tends to a finite value, which is a direct measure of the instantaneous covariance at  $t = 0$  of  $\ln S_t$  and the instantaneous variance. Thus, the short spot/variance covariance can be read off the smile in model-independent fashion. Using volatilities rather than variances yields the following expression:

$$\mathcal{S}_0 = \frac{1}{2\hat{\sigma}_0^2} \frac{\langle d\ln S d\hat{\sigma}_0 \rangle}{dt} \quad (8.36)$$

where, from (8.35a),  $\hat{\sigma}_0$  is the short ATM volatility. Anticipating on the following chapter and using the general definition of the SSR:  $\mathcal{R}_T = \frac{1}{\mathcal{S}_T} \frac{E[d\ln S d\hat{\sigma}_{F_T(S)T}]}{E[(d\ln S)^2]}$  – see page 358 – (8.36) is equivalent to the property:

$$\mathcal{R}_0 = 2$$

- The short curvature  $\mathcal{C}_0$ , on the other hand, depends not only on  $\mu_0$  and  $\nu_0$ , but also involves the quantity  $\frac{d\mu_0}{d\xi_0^0}$ , which quantifies how the short skew varies as the instantaneous variance changes.  $\mu_0 \frac{d\mu_0}{d\xi_0^0}$  can be written in terms of the covariance of  $\ln S$  and the short ATM skew. Using the definition of  $\nu_0$  and expressing everything in terms of volatilities, we get:

$$\mathcal{C}_0 = \frac{1}{4\hat{\sigma}_0} \left( \frac{8}{3} \frac{\langle d\ln S d\mathcal{S}_0 \rangle}{\hat{\sigma}_0 dt} + \frac{4}{3} \frac{\langle d\hat{\sigma}_0 d\hat{\sigma}_0 \rangle}{\hat{\sigma}_0^2 dt} - 8\mathcal{S}_0^2 \right) \quad (8.37)$$

While the short spot/variance covariance can be read off the market smile directly, inverting equation (8.37) does not provide a model-independent value for the short variance of volatility  $\langle d\hat{\sigma}_0 d\hat{\sigma}_0 \rangle$ . In other words it is not possible to read the level of volatility of volatility off the market smile in model-independent fashion. One needs to make an assumption on how the short skew changes as the spot moves; this is quantified in  $\langle d\ln S d\mathcal{S}_0 \rangle / dt$ .

Results (8.36) and (8.37) are practically useful as they characterize near-the-money implied volatilities of general stochastic volatility models in terms of the joint dynamics of financial observables such as  $\ln S$ , the ATM – or VS – volatility  $\hat{\sigma}_0$ , and the short ATM skew  $\mathcal{S}_0$ .

These relationships are specific to the case of vanishing maturities. They are very general as they also hold for local volatility or mixed local-stochastic volatility models. They can be derived many other ways – see for example [43], [44], [76]. One can in fact get exact expressions for short-dated smiles in general stochastic volatility models at order zero in  $T$  – see [7] – and at order one in  $T$  – see [56].

Equation (8.36) which relates the spot/short ATM volatility covariance to the ATM skew in model-independent fashion is equivalent to the statement that

$$\mathcal{R}_{T=0} = 2$$

in diffusive models. See Section 9.11.1 for a comparison of the SSR in local volatility and stochastic volatility models.

We now consider two examples of volatility dynamics.

### 8.5.1 Lognormal ATM volatility – SABR model

Assume that the short ATM volatility  $\hat{\sigma}_0$  is lognormal; denote by  $\nu$  its volatility and  $\rho$  its correlation with  $S$ :

$$dS = \hat{\sigma}_0 S dW^S \quad (8.38a)$$

$$d\hat{\sigma}_0 = \bullet dt + \nu \hat{\sigma}_0 dW^{\hat{\sigma}_0} \quad (8.38b)$$

where we leave the drift unspecified as they are immaterial for  $T \rightarrow 0$ . Using expressions (8.36) and (8.37) we get for  $T \rightarrow 0$ :

$$\mathcal{S}_0 = \frac{\rho}{2} \nu \quad (8.39a)$$

$$\mathcal{C}_0 = \frac{1}{6\hat{\sigma}_0} (2 - 3\rho^2) \nu^2 \quad (8.39b)$$

- In a model where the short ATM volatility  $\hat{\sigma}_0$  is lognormal – such as the lognormal model for forward variances of Section 7.4 – the short skew is constant, independent on  $\hat{\sigma}_0$ , and the curvature is inversely proportional to  $\hat{\sigma}_0$ .
- This has important implications for the pricing of cliques. In a model with lognormal instantaneous variances/volatilities, while the short forward ATM volatility at time  $t$ ,  $\hat{\sigma}_0(t)$ , is random, the level of short forward skew  $\mathcal{S}_0(t)$  is fixed: prices of narrow forward ATM call spreads with short maturities are approximately independent on the level of volatility of volatility.
- The relationship between skew, curvature and volatility of volatility reads:

$$\nu^2 = 3\hat{\sigma}_0 \mathcal{C}_0 + 6\mathcal{S}_0^2 \quad (8.40)$$

The dynamics in (8.38) is the short-maturity limit of the SABR model with  $\beta = 1$ . We leave it to the reader to check that the ATM skew and curvature obtained from the well-known SABR formula for  $T \rightarrow 0$  indeed yield (8.39a) and (8.39b) – see [55].

### 8.5.2 Normal ATM volatility – Heston model

Consider now the case of a normal ATM volatility – to avoid any confusion with the lognormal case denote by  $\sigma$  the normal volatility of volatility:

$$dS = \hat{\sigma}_0 S dW^S \quad (8.41a)$$

$$d\hat{\sigma}_0 = \bullet dt + \sigma dW^{\hat{\sigma}_0} \quad (8.41b)$$

Using (8.36), (8.37) we now get for  $T \rightarrow 0$ :

$$\mathcal{S}_0 = \frac{\rho}{2} \frac{\sigma}{\hat{\sigma}_0} \quad (8.42a)$$

$$\mathcal{C}_0 = \frac{1}{6\hat{\sigma}_0} (2 - 5\rho^2) \left( \frac{\sigma}{\hat{\sigma}_0} \right)^2 \quad (8.42b)$$

- If the short ATM volatility  $\hat{\sigma}_0$  is normal then the short skew is inversely proportional to  $\hat{\sigma}_0$  and the curvature is inversely proportional to  $\hat{\sigma}_0^3$ .
- $\mathcal{S}_0, \mathcal{C}_0, \sigma$  are related through:

$$\left( \frac{\sigma}{\hat{\sigma}_0} \right)^2 = 3\hat{\sigma}_0 \mathcal{C}_0 + 10\mathcal{S}_0^2 \quad (8.43)$$

Compare results in the normal model with those in the lognormal model; imagine that  $\nu$  and  $\sigma$  are such that the instantaneous volatility of  $\hat{\sigma}_0$  at  $t = 0$  is identical in both models:  $\nu = \frac{\sigma}{\hat{\sigma}_0}$ . Comparison of (8.39a, 8.39b) and (8.42a, 8.42b), shows that, while the values of  $S_0$  in both models are identical, values of  $C_0$  are not: the smile curvature depends not only on the spot/variance covariances at  $t = 0$  but also on their dependence on the level of volatility. As mentioned in the comment that follows (8.37), extracting the level of volatility of volatility out of market smiles requires a modeling assumption. This is clearly demonstrated in the comparison of (8.40) and (8.43).

- The dynamics in (8.41) is in fact the short-maturity limit of the Heston model – the reader can check that expression (8.42a) for  $S_0$  agrees with (6.18b). Note that  $\sigma$  in Section 6.1 denotes the normal volatility of the instantaneous *variance* – twice the volatility of the instantaneous *volatility*. (8.42a) shows that the Heston model embeds a structural connection between the level of short forward ATM volatility and skew: cliques of narrow forward ATM call spreads will exhibit a non-vanishing vega. This only occurs because of the hard-wired relationship  $S_0(t) \propto \frac{1}{\hat{\sigma}_0(t)}$ .

### 8.5.3 Vanishing correlation – a measure of volatility of volatility

Imagine there is no correlation between  $S$  and the short ATM volatility. Then the ATM skew vanishes –  $S_0 = 0$  – and from (8.37) the smile curvature is given by:

$$C_0 = \frac{1}{3\hat{\sigma}_0} \frac{\langle d\hat{\sigma}_0 d\hat{\sigma}_0 \rangle}{\hat{\sigma}_0^2 dt} \quad (8.44)$$

This is correct at second order in volatility of volatility – the first non-trivial contribution when the spot/volatility correlation vanishes. The ATM smile curvature is then a direct measure of the volatility of the short-maturity ATM volatility. Since the spot/volatility correlation vanishes,  $\mu(t, u, \xi) = 0$  and the condition that  $\mu$  not depend on  $S$  is not needed anymore.

(8.44) is thus general – it holds in all diffusive models – no assumption has been made about  $\nu(t, u, u', \xi)$ .

## 8.6 A family of one-factor models – application to the Heston model

Here, we illustrate how our framework equally applies to first-generation models built on the dynamics of the instantaneous variance  $V_t$  – such as those examined in

[69]. Consider a one-factor mean-reverting model of the following type:

$$\begin{cases} dS_t = \sqrt{V_t} S_t dW_t^S \\ dV_t = -k(V_t - V^0) dt + \sigma V_t^\varphi dW_t^V \end{cases} \quad (8.45)$$

where the correlation between the Brownian motions  $W^S$  and  $W^V$  is  $\rho$ .

The Heston model, covered in Chapter 6, corresponds to  $\varphi = \frac{1}{2}$ .

Forward variance  $\xi_t^T$  is given by:  $\xi_t^T = E_t[V_T]$ . Taking the conditional expectation of both sides of (8.45) and integrating with respect to  $t$  yields:

$$\xi_t^T = E_t[V_T] = V^0 + (V_t - V^0) e^{-k(T-t)} \quad (8.46)$$

The (driftless) dynamics of  $\xi_t^T$  is:

$$d\xi_t^T = e^{-k(T-t)} \sigma \xi_t^{t \varphi} dW_t^V \quad (8.47)$$

As is typical of first-generation stochastic volatility models,  $k$  determines both the term structure of forward variances (8.46) and their volatilities (8.47).

The spot/variance and variance/variance covariance functions are given by:

$$\mu(t, u, \xi) = \rho \sigma e^{-k(u-t)} \xi_t^{t \varphi + \frac{1}{2}} \quad (8.48a)$$

$$\nu(t, u, u', \xi) = \sigma^2 e^{-k(u-t)} e^{-k(u'-t)} \xi_t^{t 2\varphi} \quad (8.48b)$$

From these expressions  $C^{x\xi}, C^{\xi\xi}, D$  are easily computed and can be inserted in (8.21) to obtain the smile at order two in  $\sigma$ : the reader can check that one recovers the expressions derived in [69]. In particular, using (8.35b), the short ATMF skew is given by:

$$\mathcal{S}_0 = \frac{\rho \sigma}{4} V_{t=0}^{(\varphi-1)} \quad (8.49)$$

Take the particular case of the Heston model –  $\varphi = \frac{1}{2}$  – and consider the expansion of the option price at order one. From (8.18), it is given by:

$$P = P_0 + \varepsilon \frac{C_0^{x\xi}(\xi_0)}{2} (\partial_x^3 - \partial_x^2) P_0$$

Going back to expression (6.15) in Chapter 6, the interpretation of the prefactor is now clear: it is the doubly-integrated spot/variance covariance function  $C_0^{x\xi}$ .

## 8.7 The two-factor model

We now consider the model introduced in Section 7.4, defined by the following SDEs:

$$\begin{cases} dS_t = \sqrt{\xi_t^S} S_t dW_t^S \\ d\xi_t^S = 2\nu \xi_t^S \alpha_\theta \left( (1-\theta)e^{-k_1(T-t)} dW_t^1 + \theta e^{-k_2(T-t)} dW_t^2 \right) \end{cases}$$

where  $\nu$  is the volatility of a VS volatility of vanishing maturity,  $\alpha_\theta = 1/\sqrt{(1-\theta)^2 + \theta^2 + 2\rho_{12}\theta(1-\theta)}$  and  $\rho_{12}$  is the correlation between  $W^1$  and  $W^2$ . We denote by  $\rho_{SX^1}$  ( $\rho_{SX^2}$ ) the correlation between  $W^S$  and  $W^1$  ( $W^2$ ). We would like to answer the following questions:

- Is the accuracy of the order-two approximation (8.21) sufficient for practical purposes?
- Does the two-factor model afford sufficient flexibility as to the type of smiles it is able to generate? Can we obtain a term-structure of ATMF skews that is consistent with typical index smiles?<sup>6</sup>

Among the parameters of the two-factor model, the subset  $\nu, \theta, k_1, k_2, \rho_{12}$  determines the dynamics of the VS volatilities in the model. Once these parameters are set, the dynamics of VS volatilities is set; we can then select the additional parameters  $\rho_{SX^1}$  and  $\rho_{SX^2}$  to generate the desired spot/volatility dynamics and the ensuing vanilla smile. We use in our tests below Set II parameters – see Table 8.1.

$\nu$	$\theta$	$k_1$	$k_2$	$\rho_{12}$
174%	0.245	5.35	0.28	0%

Table 8.1: Numerical values of parameters used in tests of the two-factor model.

Set II generates a term structure of instantaneous volatilities of VS volatilities displayed in Figure 7.1, page 228, which reproduces with good accuracy the power-law dependence (7.40) with  $\alpha = 0.4$  and a volatility of a 3-month VS volatility equal to 100%. Volatilities of VS volatilities in the two-factor model are given by (7.39), page 227.

$\mu(t, u, \xi)$  and  $\nu(t, u, u', \xi)$  are given by:

$$\mu(t, u, \xi) = 2\nu\xi^u \sqrt{\xi^t} \alpha_\theta \left[ \rho_{SX^1}(1-\theta)e^{-k_1(u-t)} + \rho_{SX^2}\theta e^{-k_2(u-t)} \right] \quad (8.50)$$

$$\begin{aligned} \nu(t, u, u', \xi) = & 4\nu^2 \xi^u \xi^{u'} \alpha_\theta^2 \left[ (1-\theta)^2 e^{-k_1(u+u'-2t)} + \theta^2 e^{-k_2(u+u'-2t)} \right. \\ & \left. + \rho_{12}\theta(1-\theta) \left( e^{-k_1(u-t)} e^{-k_2(u'-t)} + e^{-k_2(u-t)} e^{-k_1(u'-t)} \right) \right] \end{aligned} \quad (8.51)$$

$C^{x\xi}, C^{\xi\xi}, D$  which are multiple integrals of  $\mu$  and  $\nu$  can be efficiently calculated numerically by Gaussian quadrature.<sup>7</sup> Because  $\mu$  and  $\nu$  are smooth functions of  $t, u, u'$ , very few points are needed. In the special case of a flat term structure of VS

<sup>6</sup>This is an important question for pricing the forward-smile risk of cliquets.

<sup>7</sup>Whenever, as is natural, the variance curve is generated by interpolating  $T\hat{\sigma}_T^2$  as a piecewise affine function of  $T$ , forward variances  $\xi^T$  are piecewise constant.  $C^{x\xi}, C^{\xi\xi}, D$  can be computed analytically.

volatilities,  $\xi_t^u$  does not depend on  $u$  and the integrations can be done analytically. The analytical expressions of  $C^{x\xi}$ ,  $C^{\xi\xi}$ ,  $D$  – which we use in our tests below – can be found in [13].

### 8.7.1 Uncorrelated case

Consider first the case when the correlation of forward variances  $\xi_t^u$  with  $S_t$  vanishes:  $\rho_{SX^1} = \rho_{SX^2} = 0$ , thus  $\mu(\tau, u, \xi) = 0$  and  $C^{x\xi} = D = 0$ . Expressions (8.21) become:

$$\begin{aligned}\hat{\sigma}_{F_T T} &= \hat{\sigma}_T \left( 1 - \frac{\varepsilon^2}{32Q^2} (Q + 4) C^{\xi\xi} \right) \\ S_T &= 0 \\ C_T &= \hat{\sigma}_T \frac{\varepsilon^2}{4Q^3} C^{\xi\xi}\end{aligned}$$

As mentioned before,  $S_T$  vanishes both at order one and two in  $\varepsilon$  – as it should: it is well known that for uncorrelated spot and variances, the smile is symmetric in log-moneyness.<sup>8</sup> Besides, the order-one contributions to  $\hat{\sigma}_{F_T T}$  and  $C_T$  vanish altogether.

Figure 8.1 shows a comparison of exact and approximate smiles for four different maturities, for a flat term structure of VS volatilities at 20%. “Exact” smiles are obtained by Monte Carlo simulation using the gamma/theta technique – see Section 8.10, page 336. As already mentioned, the order-two approximation is bound to differ markedly from the exact result for far-away strikes. Note however how accurate it is for near-the-money strikes.

### 8.7.2 Correlated case – the ATMF skew and its term structure

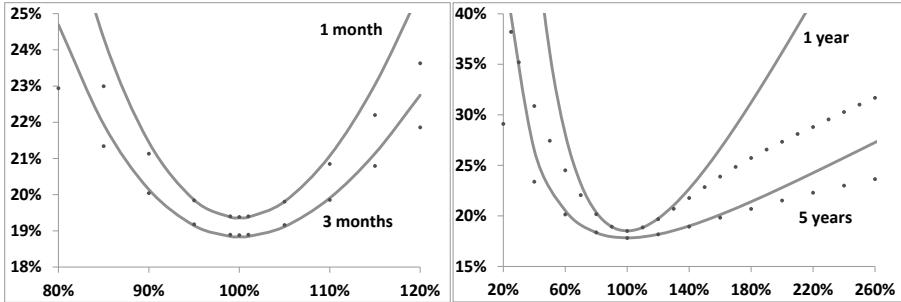
We now test the order-two approximation in the correlated case. We still use Set II parameters and take  $\rho_{SX^1} = -75.9\%$ ,  $\rho_{SX^2} = -48.7\%$  – more on the choice of these correlations below. The full list of parameters appears in Table 8.2 below.

This exact same set has been used in the tests of Section 7.8.4 of Chapter 7, devoted to the vanilla smile of discrete forward variance models.

Exact and approximate smiles at order one and two in volatility of volatility are shown in Figure 8.2. As is apparent, while smiles at order one and two in  $\varepsilon$  have

---

<sup>8</sup>Given a strike  $K$ , denote by  $K^*$  the strike with opposite log-moneyness:  $K^* = F_T^2/K$ . If forward variances are uncorrelated with  $S_t$ , conditional on the path of  $\xi_t^u$ ,  $t \in [0, T]$ ,  $S_T$  is lognormally distributed with a volatility  $\sigma$  given by:  $\sigma^2 T = \int_0^T \xi_t^u dt$ . The price of a call option of strike  $K$  is then equal to  $\int C_K^{BS}(\sigma) \rho(\sigma) d\sigma$  where  $C_K^{BS}(\sigma)$  is the Black-Scholes formula with volatility  $\sigma$  and  $\rho(\sigma)$  is the density of  $\sigma$ . Denote by  $\hat{\sigma}_K$  the implied volatility for strike  $K$ :  $C_K^{BS}(\hat{\sigma}_K) = \int C_K^{BS}(\sigma) \rho(\sigma) d\sigma$ . Consider now a put option of strike  $K^*$  and use the following relationship between prices of call and put options with opposite log-moneyness in the Black-Scholes model:  $P_{K^*}^{BS} = \frac{F_T}{K} C_K^{BS}$ . By definition of  $\hat{\sigma}_{K^*}$ ,  $P_{K^*}^{BS}(\hat{\sigma}_{K^*}) = \int P_{K^*}^{BS}(\sigma) \rho(\sigma) d\sigma = \int \frac{F_T}{K} C_K^{BS}(\sigma) \rho(\sigma) d\sigma = \frac{F_T}{K} C_K^{BS}(\hat{\sigma}_K) = P_K^{BS}(\hat{\sigma}_K)$ . Thus  $\hat{\sigma}_{K^*} = \hat{\sigma}_K$ .



**Figure 8.1:** Exact (dots) and approximate (continuous line) smiles of the two-factor model with Set II parameters and uncorrelated spot and variances, for four different maturities: 1 month, 3 months, 1 year, 5 years. The term structure of VS volatilities is flat at 20%. The algorithm used is that in Section A.2 of Appendix A.

$\nu$	$\theta$	$k_1$	$k_2$	$\rho_{12}$	$\rho_{SX^1}$	$\rho_{SX^2}$
174%	0.245	5.35	0.28	0%	-75.9%	-48.7%

Table 8.2: Numerical values of parameters of the two-factor model.

similar shapes, the overall volatility level – and especially the ATMF volatility – is much better captured at order two in  $\varepsilon$ .

The accuracy of expression (8.21b) for the ATMF skew  $S_T$ , however, is excellent already at order one in  $\varepsilon$  – as highlighted in [9]. At this order (8.21b) simplifies to:

$$S_T^{\text{order } 1} = \hat{\sigma}_T \frac{\varepsilon}{2Q^2} C^{x\xi} \quad (8.52)$$

In the two-factor model, for a flat term structure of forward variances equal to  $\xi_0$ ,  $\mu(t, u, \xi)$  is given by:

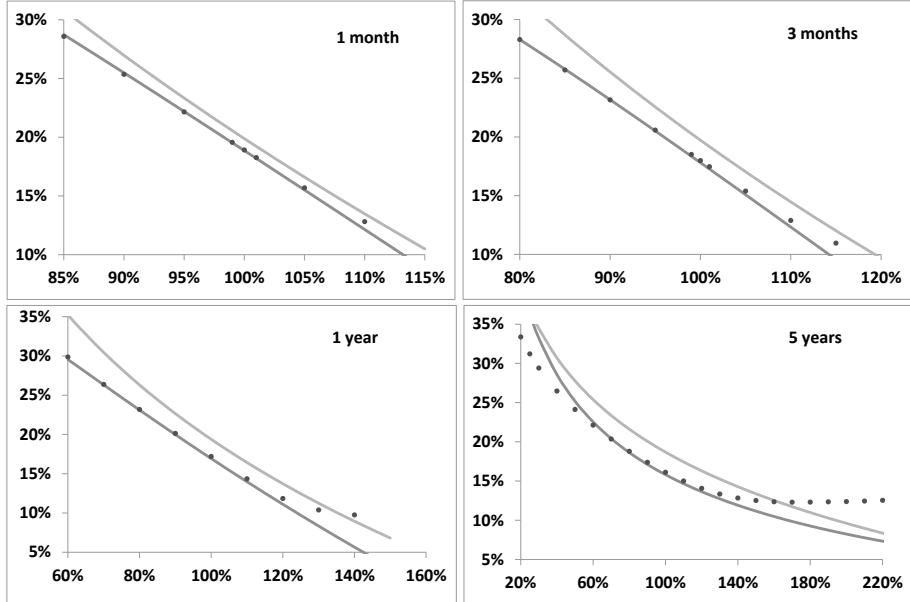
$$\mu(t, u, \xi) = 2\nu\xi_0^{\frac{3}{2}}\alpha_\theta \left[ (1-\theta)\rho_{SX^1}e^{-k_1(u-t)} + \theta\rho_{SX^2}e^{-k_2(u-t)} \right] \quad (8.53)$$

where  $\alpha_\theta = ((1-\theta)^2 + \theta^2 + 2\rho_{12}\theta(1-\theta))^{-\frac{1}{2}}$ . Plugging this in expression (8.13) for  $C^{x\xi}$  yields, after setting  $\varepsilon = 1$ :

$$S_T^{\text{order } 1} = \frac{\nu\alpha_\theta}{\hat{\sigma}_T^3 T^2} \int_0^T dt \sqrt{\xi_0^3} \int_t^T du \xi_0^u \left[ (1-\theta)\rho_{SX^1}e^{-k_1(u-t)} + \theta\rho_{SX^2}e^{-k_2(u-t)} \right] \quad (8.54)$$

where  $\hat{\sigma}_T = \sqrt{\frac{1}{T} \int_0^T \xi_0^t dt}$ .

In the case a flat term structure of forward variances/VS volatilities, the double integrals in (8.54) can be done analytically and we get:



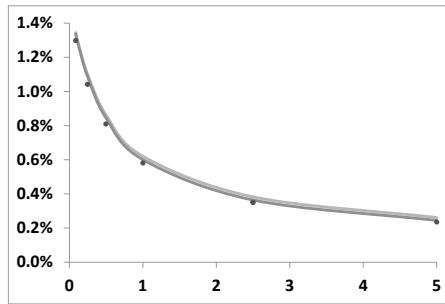
**Figure 8.2:** Exact (dots) as well as approximate smiles at order one (light line) and order two (dark line) in  $\varepsilon$ , for parameters in Table 8.2, and a flat term structure of VS volatilities at 20%. The algorithm used is that in Section A.2 of Appendix A.

$$\mathcal{S}_T^{\text{order } 1} = \nu \alpha_\theta \left[ (1 - \theta) \rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta \rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2} \right] \quad (8.55)$$

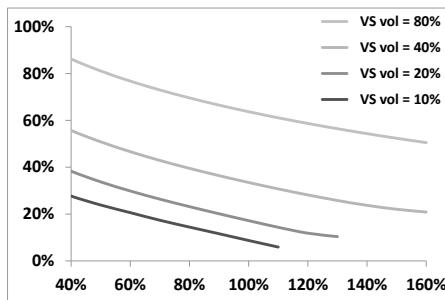
The exact ATMF skew, together with the order-two (8.21b) and order-one expression (8.55) is displayed in Figure 8.3. Figure 8.3 actually shows the difference of the implied volatilities for strikes  $0.99F_T$  and  $1.01F_T$ , approximately equal to  $-0.02\mathcal{S}_T$ . The order-two contribution only marginally improves on the order-one result, which is remarkably accurate.

Observe that in the two-factor model, expression (8.55) for  $\mathcal{S}_T^{\text{order } 1}$  does not involve the level of VS volatility: at order one in  $\varepsilon$ , the ATMF skew is unchanged if VS volatilities are rescaled by a common factor.<sup>9</sup> This can be traced to the fact that forward variances in the two-factor model are lognormal. We have already observed this property in the short-maturity limit – see expression (8.39a) for  $\mathcal{S}_0$ . Because of

<sup>9</sup>Expression (8.55) is obtained for a flat term structure of VS volatilities. In the case of a sloping term structure the double integrals in  $C^{x\xi}$  cannot be done analytically – the expression for  $\mathcal{S}_T^{\text{order } 1}$  is more complicated than (8.55) and does depend on the term structure of VS volatilities. However,  $\mathcal{S}_T^{\text{order } 1}$  is unchanged in a global rescaling of VS volatilities.



**Figure 8.3:** Exact (dots) and approximate values of  $\widehat{\sigma}_{0.99F_T} - \widehat{\sigma}_{1.01F_T}$  at order one (light line) and two (dark line) in  $\varepsilon$ , for maturities from 1 month to 5 years. Parameters in Table 8.2 have been used, with a flat term structure of VS volatilities at 20%.



**Figure 8.4:** Smiles for various levels of VS volatilities, for a one-year maturity, computed in a Monte Carlo simulation of the two-factor model. The term structure of VS volatilities is flat and parameters are those of Table 8.2. As is manifest, ATMF skew levels are practically independent on the level of VS volatility.

the accuracy of  $\mathcal{S}_T^{\text{order } 1}$  we expect this behavior to persist in the exact smile: this is illustrated in Figure 8.4, which shows implied volatilities for different levels of VS volatilities, for a one-year maturity.

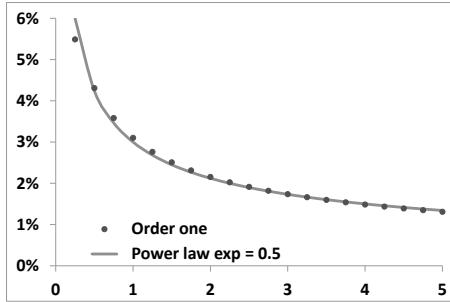
Figure 8.4 should be contrasted with Figure 6.4, page 211, for the case of the Heston model. As already mentioned, the fact that VS volatilities are approximately normal – rather than lognormal – in the Heston model implies that, at order one in  $\varepsilon$ , the short ATMF skew is inversely proportional to the short VS volatility.

### Term structure of the ATMF skew

How should we choose  $\rho_{SX^1}$ ,  $\rho_{SX^2}$ ? Once other parameters are set,  $\rho_{SX^1}$ ,  $\rho_{SX^2}$  will determine both the vanilla smile *and* future smiles. It is necessary that the two-factor model be at least able to generate smiles that are comparable to historically

observed smiles, in particular with respect to the term structure of the ATMF skew. This is especially important when pricing cliques: see the discussion in Section 3.1. Typically equity index smiles display a term structure of the ATMF skew that is well approximated by a power law with an exponent usually around  $\frac{1}{2}$  – see examples in Figure 6.5.

The values for  $\rho_{SX^1}, \rho_{SX^2}$  in Table 8.2 are such that they generate a term structure for the ATMF skew that is approximately a power law with exponent  $\frac{1}{2}$ , with  $\widehat{\sigma}_{0.95F_T} - \widehat{\sigma}_{1.05F_T} = 3\%$  for  $T = 1$  year. This is illustrated in Figure 8.5.



**Figure 8.5:** The ATMF skew measured as the difference of implied volatilities for strikes  $0.95F_T$  and  $1.05F_T$  given by expression (8.55) for  $S_T^{\text{order } 1}$  (continuous line) and by a power law benchmark with exponent  $\frac{1}{2}$  and  $\widehat{\sigma}_{0.95F_T} - \widehat{\sigma}_{1.05F_T} = 3\%$  for  $T = 1$  year (dots), as a function of maturity (years). Maturities run from 3 months to 5 years. Parameters are those of Table 8.2.

We could as well have chosen other values for  $\rho_{SX^1}, \rho_{SX^2}$  such that the ATMF skew in the two-factor model generates a power-law-like dependence with a different exponent.

Our freedom is however limited. Indeed, the other parameters in the model – notably  $k_1, k_2, \rho_{12}$  – are already set and the triplet  $\rho_{12}, \rho_{SX^1}, \rho_{SX^2}$  must make up a valid correlation matrix. This is the case if  $\rho_{SX^2}$  is defined as:

$$\rho_{SX^2} = \rho_{12}\rho_{SX^1} + \chi\sqrt{1 - \rho_{12}^2}\sqrt{1 - \rho_{SX^1}^2} \quad (8.56)$$

where  $\chi \in [-1, 1]$ .

Still, it is usually possible to cover the range of skew decays that are observed practically. For example, taking  $\rho_{SX^1} = -56\%$ ,  $\rho_{SX^2} = -68\%$  (resp.  $\rho_{SX^1} = -95\%$ ,  $\rho_{SX^2} = -31\%$ ) approximately generates a power law decay for  $S_T$  with an exponent 0.4 (resp. 0.6), with  $\widehat{\sigma}_{0.95F_T} - \widehat{\sigma}_{1.05F_T} = 3\%$  for  $T = 1$  year.

## 8.8 Conclusion

The expansion of implied volatilities at order two is accurate for near-the-money strikes. It can be used for calibrating near-the-money implied volatilities or whenever vanilla implied volatilities are needed as observables, for example in Longstaff-Schwartz algorithms.

Its accuracy deteriorates for longer maturities and larger volatilities of volatilities – note in this respect that we have used realistic levels of volatility of volatility in our tests. The term structure of volatilities of VS volatilities generated by Set II appears in Figure 7.1: a 3-month volatility has a (lognormal) volatility of about 100% while volatilities of 1-year and 5-year VS volatilities are about 50% and 30%, respectively.

With parameters in Table 8.2, 5-year ATMF implied volatilities for a flat VS term structure at 20%, which appear in Figure 8.2, are 16.0% (Monte Carlo simulation) and 15.6% (order two expansion). If we now double  $\nu$ , the 5-year ATMF volatilities are now 9.8% (Monte Carlo simulation) and 5.1% (order two expansion) – these are however unreasonably large levels of volatility.<sup>10</sup>

Finally, expression (8.55) for the ATMF skew at order one in  $\varepsilon$  is remarkably accurate. Values of  $\rho_{SX^1}, \rho_{SX^2}$  can be chosen so that the term structure of the ATMF skew is consistent with actual term structures of ATMF skews of equity indexes.

In (8.55) the ATMF skew is given by the product of volatility of volatility and spot/volatility correlations. Rescaling  $\rho_{SX^1}, \rho_{SX^2}$  by the same constant and adjusting  $\nu$  so that products  $\rho_{SX^1}\nu, \rho_{SX^2}\nu$  are unchanged leaves the ATMF skew unchanged.

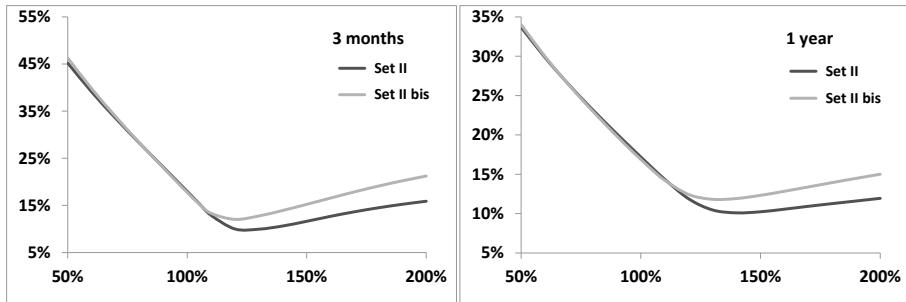
One expects the two smiles to differ only by their ATMF volatility and curvature. This is true locally, but not for the global smile.

This is illustrated in Figure 8.6; we have generated smiles for the 3-month and 1-year maturities in the two-factor model with Set II parameters, and with Set II parameters, but with  $\rho_{SX^1}, \rho_{SX^2}$  multiplied by 90% and  $\nu$  divided by 90% (Set II bis). This rescaling of  $\rho_{SX^1}, \rho_{SX^2}, \nu$  indeed leaves the ATMF skew unchanged, but affects out-of-the-money volatilities asymmetrically .

## 8.9 Forward-start options – future smiles

So far we have considered spot-starting smiles. Consider the simple case of an option paying at time  $T_1$  a payout  $f\left(\frac{S_{T_2}}{S_{T_1}}\right)$ . As explained in Sections 3.1.5 and 3.1.6

<sup>10</sup>Unrealistic levels of volatility of volatility may be needed to generate the inordinately large values of forward skew that one implies at times from market prices of cliques. This drawback is typical of continuous forward variance models – discrete forward variance models, covered in Chapter 7.8, are immune to it.



**Figure 8.6:** 3-month and 1-year smiles in the two-factor model with Set II parameters – see Table 8.2, page 329 – and Set II parameters with  $\rho_{SX^1}, \rho_{SX^2}$  multiplied by 0.9 and  $\nu$  divided by 0.9 (Set II bis). The algorithm used is that in Section A.2 of Appendix A.

the price of such an option incorporates a volatility-of-volatility contribution,  $\delta P_1$  and forward smile contribution  $\delta P_2$ .

$\delta P_2$  quantifies the forward-smile risk. It prices the difference of the market price at  $t = T_1$  of the then-vanilla payoff  $f$  of maturity  $T_2$  and its price as given in the Black-Scholes model with the log-contract implied volatility  $\widehat{\sigma}_{T_1 T_2}(T_1)$ .

In Section 3.2 we have analyzed how the local volatility models handles forward-smile risk. Calibration of the local volatility model on the market smile at  $t = 0$  likely results in a mispricing of  $\delta P_2$  – and of  $\delta P_1$  as well. What about stochastic volatility models?

Typically – that is unless we make model parameters explicitly time-dependent – stochastic volatility models are time-homogeneous, thus the instantaneous volatility of  $\widehat{\sigma}_{T_1 T_2}(t)$  is a function of  $T_1 - t$ : if our model is properly parametrized,  $\delta P_1$  is priced correctly. Note in that respect that, already with the simple two-factor model, because a given term-structure of instantaneous volatilities of spot-starting volatilities can be matched with different sets of parameters – see Figure 7.1, page 228 – we can use this flexibility to adjust the level of volatilities of forward volatilities somewhat.

Future smiles in stochastic volatility models are similar to spot-starting smiles, but may depend on volatility levels prevailing at future dates. For example, in the Heston model, the short ATMF skew is inversely proportional to the short ATMF volatility. Generally, for the sake of pricing  $\delta P_2$  – and unless we have strong reasons not to do so – it is preferable not to hard-wire in the model dependencies between risks of a different nature, for example forward volatility and forward-smile risks.

In this respect, the two-factor model is attractive: future ATMF skews do not depend on the level of VS volatilities.<sup>11</sup> For flat future term-structures of VS volatili-

<sup>11</sup>They do depend somewhat on the term structure of VS volatilities, but are invariant in a rescaling of VS volatilities.

ties, at order one in volatility of volatility, they are given by (8.55) where  $T$  is the residual maturity.

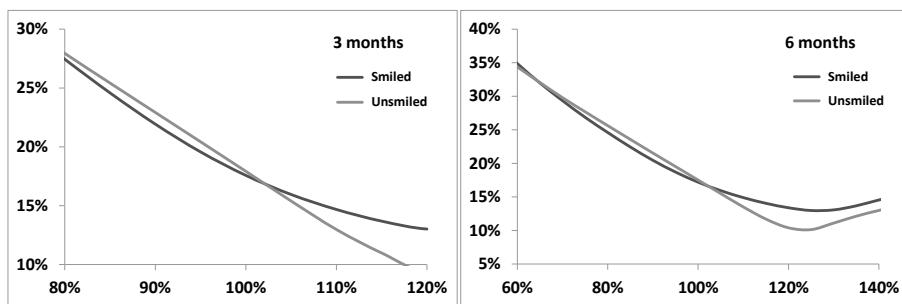
Finally, we refer the reader to the discussion on page 104 for why we do not use the notion of “forward smile”.

## 8.10 Impact of the smile of volatility of volatility on the vanilla smile

Imagine we use the two-factor model in the version of Section 7.7.1 with piecewise-constant volatility-of-volatility parameters  $\gamma_T, \beta_T, \zeta_T$  calibrated so that market VIX smiles are matched. How does the smile of volatility of volatility impact the vanilla smile?

The expressions of implied volatilities in the order-two expansion (8.21) involve  $C_0^{\alpha\xi}(\xi_0), C_0^{\xi\xi}(\xi_0), D_0(\xi_0)$ .  $C_0^{\alpha\xi}(\xi_0)$  and  $C_0^{\xi\xi}(\xi_0)$ , defined in (8.13) and (8.14), only depend on the covariance functions  $\mu$  and  $\nu$  evaluated on the initial variance curve  $\xi_0$ . In contrast,  $D_0(\xi_0)$  – defined in (8.17) – depends on the derivative of  $\mu$  with respect to  $\xi$ , hence is sensitive to the smile of volatility of volatility.  $D_0(\xi_0)$  contributes at order  $\varepsilon^2$ : at order one in volatility of volatility the vanilla smile is unaffected by the smile of volatility of volatility.

VIX smiles are positively sloping: a model calibrated to VIX smiles will generate lower (more negative) values for  $\frac{\delta\mu}{\delta\xi^T}$ , hence a larger (more positive) value for  $D_0(\xi_0)$  – remember that  $\mu$  is negative. From (8.21) we then expect that the ATMF skew will be weaker (less negative) while the ATMF curvature will be larger. This is demonstrated in Figure 8.7.



**Figure 8.7:** Vanilla smiles produced by the two-factor model for different values of the volatility-of-volatility parameters.

We have used Set II parameters along with the following constant values for  $\gamma_T, \beta_T, \zeta_T$ :

$$\gamma_T = 50\%, \beta_T = 15\%, \zeta_T = 100\% \quad (8.57)$$

These values are typical of VIX smiles – see the values of  $\gamma_T, \beta_T, \zeta_T$  calibrated on the VIX smile of June 14, 2011 in Figure 7.9. We have chosen  $\zeta_T = 100\%$  so that instantaneous volatilities of forward variances at  $t = 0$  match those of the lognormal version of the model ( $\gamma^T = \beta^T = 0, \zeta_T = 100\%$ ): the difference in the vanilla smile is then only generated by the smile of volatility of volatility.

Figure 8.7 shows the vanilla smile for 3-month and 6-month maturities both in the smiled (parameters in (8.57)) and unsmiled ( $\gamma^T = \beta^T = 0, \zeta_T = 100\%$ ) version of the two-factor model. The predictions from the order-two expansion regarding ATMF skew and curvature are indeed verified.

---

## Appendix A – Monte Carlo algorithms for vanilla smiles

The “exact” results in Figures 8.1 and 8.2 are computed with a Monte Carlo simulation: time is discretized,  $S_t$  is simulated with a simple Euler scheme and  $X_t^1, X_t^2$  are simulated exactly – see Section 7.3.1.

The standard technique of averaging vanilla option’s payoffs over all Monte Carlo paths produces price estimates that are in practice too noisy. We now present more efficient techniques.

### A.1 The mixing solution

Consider an  $N$ -factor stochastic volatility model of type (7.11) – the spot/volatility joint dynamics reads:

$$\begin{cases} dS_t = (r - q) S_t dt + \sqrt{\xi_t^t} S_t dW_t^S \\ d\xi_t^T = \omega \xi_t^T \sum_i w_i e^{-k_i(T-t)} dW_t^i \end{cases}$$

where we have assumed zero interest rate and repo without loss of generality – otherwise simply replace  $S_0$  with the forward for maturity  $T$  in what follows.

Assume first that  $W^S$  is uncorrelated with the  $W^i$ . We can simulate the  $W^i$  first, thus generating  $\xi_t^t$  for  $t \in [0, T]$ , then simulate  $S_t$  independently. Conditional on the path of  $\xi_t^t$ ,  $S_t$  is lognormal with deterministic instantaneous volatility  $\sqrt{\xi_t^t}$ .

Rather than simulating  $S_t$  we can thus compute analytically the expectation on  $W^S$ :

$$E_{W^S} [f(S_T) | W^i] = P_{BS}(0, S_0, \hat{\sigma}^*)$$

where  $f(S)$  is the payoff of our vanilla option and  $P_{BS}(t, S, \hat{\sigma})$  is the corresponding Black-Scholes formula for maturity  $T$ . The effective volatility  $\hat{\sigma}^*$  is a function of the

path of  $\xi_t^i$  – that is of the  $W^i$  – and is defined by:

$$\hat{\sigma}^* = \sqrt{\frac{1}{T} \int_0^T \xi_t^i dt}$$

The price  $P$  of the vanilla option in the stochastic volatility model is thus given by:

$$P = E_{W^i, W^S} [f(S_T)] = E_{W^i} [P_{BS}(0, S_0, \hat{\sigma}^*)] \quad (8.58)$$

Consider now the correlated case. We split  $W_t^S$  into two pieces: a portion  $\lambda W_t^{\parallel}$  that is correlated with the  $W_t^i$  and an uncorrelated portion  $\sqrt{1 - \lambda^2} W_t^{\perp}$ :

$$W_t^S = \lambda W_t^{\parallel} + \sqrt{1 - \lambda^2} W_t^{\perp} \quad (8.59)$$

We have:

$$\begin{aligned} d \ln S_t &= -\frac{\xi_t^i}{2} dt + \sqrt{\xi_t^i} dW_t^S \\ &= \left[ -\frac{\lambda^2}{2} \xi_t^i dt + \lambda \sqrt{\xi_t^i} dW_t^{\parallel} \right] + \left[ -\frac{1 - \lambda^2}{2} \xi_t^i dt + \sqrt{1 - \lambda^2} \sqrt{\xi_t^i} dW_t^{\perp} \right] \end{aligned}$$

Conditional on the paths of Brownian motions  $W_t^i$ ,  $S_T$  can be rewritten as:

$$S_T = S_0^* e^{-\frac{(\sigma^*)^2 T}{2} + \hat{\sigma}^* \sqrt{T} Z} \quad (8.60)$$

where  $Z$  is a standard normal variable and the effective initial spot  $S_0^*$  and effective volatility  $\sigma^*$  are a function of the paths of the  $W^i$ :

$$\ln S_0^* = \ln S_0 - \frac{\lambda^2}{2} \int_0^T \xi_t^i dt + \lambda \int_0^T \sqrt{\xi_t^i} dW_t^{\parallel} \quad (8.61a)$$

$$\hat{\sigma}^{*2} = (1 - \lambda^2) \frac{1}{T} \int_0^T \xi_t^i dt \quad (8.61b)$$

$P$  is computed as:

$$P = E_{W^i} [P_{BS}(0, S_0^*, \hat{\sigma}^*)]$$

Thus, in the mixing solution we only need to:

- simulate the variance degrees of freedom
- accrue the two integrals  $\int_0^T \xi_t^i dt$  and  $\lambda \int_0^T \sqrt{\xi_t^i} dW_t^{\parallel}$

The mixing technique was originally published by G.A. Willard – see [85]. See also [79] for the uncorrelated case. We borrow the “mixing solution” denomination from Alan Lewis – see [69].

### The two-factor model

Specializing to our two-factor model, we have:

$$\lambda W_t^{\parallel} = \frac{\rho_{SX^1} - \rho_{12}\rho_{SX^2}}{1 - \rho_{12}^2} W_t^1 + \frac{\rho_{SX^2} - \rho_{12}\rho_{SX^1}}{1 - \rho_{12}^2} W_t^2 \quad (8.62a)$$

$$\lambda = \sqrt{\frac{\rho_{SX^1}^2 + \rho_{SX^2}^2 - 2\rho_{12}\rho_{SX^1}\rho_{SX^2}}{1 - \rho_{12}^2}} \quad (8.62b)$$

The mixing solution consists in analytically integrating on  $W_t^{\perp}$ . We thus expect maximum efficiency when spot and forward variances have low correlation – see Section A.4 below for practical tests.

## A.2 Gamma/theta P&L accrual

We now examine techniques that involve the simulation of  $S_t$ . As mentioned above, simply evaluating the vanilla option's payoff produces a noisy estimate of the price.

One can use the final spot price  $S_T$  as a control variate. Still a better idea would be to use as control variate the sum of delta P&Ls

$$\sum_i \frac{dP_{BS}}{dS}(t_i, S_i, \hat{\sigma}) \left( S_{i+1} - e^{(r-q)(t_{i+1}-t_i)} S_i \right) \quad (8.63)$$

where delta is computed in the Black-Scholes model, with an arbitrary implied volatility  $\hat{\sigma}$  – after all, delta-hedging aims at reducing the variance of the final P&L by neutralizing the order-one contribution of the spot variation.<sup>12</sup> While efficient, this solution is costly as computing  $\frac{dP_{BS}}{dS}$  entails evaluating the cumulative distribution function of the standard normal distribution at each step of the simulation.

The technique we use is based on representation (2.30), page 40. There, local volatility is used as a base model for computing the price  $P_2$  of a European option in a model whose instantaneous volatility is  $\sigma_2$ . Here we choose as base model the Black-Scholes model with volatility  $\hat{\sigma}$ ; (2.30) becomes:

$$P_2(0, S_0, \bullet) = P_{BS}(0, S_0, \hat{\sigma}) + E_2 \left[ \int_0^T e^{-rt} \frac{S_t^2}{2} \frac{d^2 P_{BS}}{dS^2} (\sigma_{2t}^2 - \hat{\sigma}^2) dt \right] \quad (8.64)$$

where  $E_2$  denotes that the expectation is taken with respect to the dynamics generated by the stochastic volatility model at hand, whose instantaneous volatility process is  $\sigma_2$ .

Equation (8.64) expresses the price of a European option in an arbitrary stochastic volatility model as its price in the Black-Scholes model with implied volatility  $\hat{\sigma}$  augmented by the (discounted) expectation of the integrated gamma/theta P&L evaluated with the Black-Scholes gamma – a natural representation from a trading

---

<sup>12</sup>(8.63) has vanishing expectation.

point of view. We call  $\widehat{\sigma}$  the risk-management volatility. We still need to simulate the spot process, but  $S_t$  is only used to compute the Black-Scholes gamma.

Expression (8.64) is practically useful as computing gamma in the Black-Scholes model amounts to evaluating one exponential:

$$\begin{aligned} S^2 \frac{d^2 P_{BS}}{dS^2} &= S e^{-q(T-t)} \frac{1}{\sqrt{2\pi\widehat{\sigma}^2(T-t)}} e^{-\frac{d_1^2}{2}} \\ d_1 &= \frac{1}{\widehat{\sigma}\sqrt{T-t}} \ln \frac{Se^{(r-q)(T-t)}}{K} + \frac{\widehat{\sigma}\sqrt{T-t}}{2} \end{aligned}$$

This cost is more than offset by the increased accuracy brought about by what is, in effect, a perfect delta hedge.

In our simulations we use for  $\widehat{\sigma}$  the VS volatility for maturity  $T$ .  $X_t^1, X_t^2, S_t$  are simulated at discrete times  $t_i$ : over each path, the second piece in the right-hand side of (8.64) is evaluated as:

$$\sum_i e^{-rt_i} \frac{S_i^2}{2} \frac{d^2 P_{BS}}{dS^2}(t_i, S_i, \widehat{\sigma}) (\xi_{t_i}^{t_i} - \widehat{\sigma}^2) \Delta \quad (8.65)$$

where  $\Delta$  is the time step. Observe that (8.65) involves the instantaneous *implied* quadratic variation  $\xi_{t_i}^{t_i} \Delta$  rather than its *realized* value  $(S_{i+1}/S_i - 1)^2$  – this also contributes to the accuracy of this technique.

### Dynamic adjustment of the implied volatility

Choosing for  $\widehat{\sigma}$  the VS volatility of maturity  $T$  is somewhat arbitrary. We could choose a value for  $\widehat{\sigma}$  so that, on average, the difference  $\xi_{t_i}^{t_i} - \widehat{\sigma}^2$  is as small as possible. Still, on each path  $\xi_t^t$  may be very different from  $\widehat{\sigma}^2$ , resulting in large gamma/theta P&Ls and consequently a large variance for (8.65), if  $S_t$  happens to be in the vicinity of the option's strike.

Trading intuition suggests that we should dynamically readjust our implied volatility  $\widehat{\sigma}$  so that  $\widehat{\sigma}^2$  remains close enough to  $\xi_t^t$ . Imagine switching at time  $t$  from  $\widehat{\sigma}_1$  to  $\widehat{\sigma}_2$ . Write equation (8.64) at time  $t$  for volatilities  $\widehat{\sigma}_1$  and  $\widehat{\sigma}_2$  and subtract one from the other.  $P_{\widehat{\sigma}}(t)$  cancels out and we get:

$$\begin{aligned} E_{\widehat{\sigma}_t} \left[ \int_t^T e^{-r(u-t)} \frac{S_u^2}{2} \frac{d^2 P_{BS}}{dS^2}(u, S_u, \widehat{\sigma}_1) (\bar{\sigma}_u^2 - \widehat{\sigma}_1^2) du \right] \\ = E_{\widehat{\sigma}_t} \left[ \int_t^T e^{-r(u-t)} \frac{S_u^2}{2} \frac{d^2 P_{BS}}{dS^2}(u, S_u, \widehat{\sigma}_2) (\bar{\sigma}_u^2 - \widehat{\sigma}_2^2) du \right] \\ + \left[ P_{BS}(t, S_t, \widehat{\sigma}_2) - P_{BS}(t, S_t, \widehat{\sigma}_1) \right] \end{aligned}$$

This equation expresses that we are allowed to switch at time  $t$  from  $\widehat{\sigma}_1$  to  $\widehat{\sigma}_2$ , provided we supplement the gamma/theta P&L with the difference  $P_{BS}(t, S_t, \widehat{\sigma}_2) - P_{BS}(t, S_t, \widehat{\sigma}_1)$ , which, from a trading point of view, is the P&L generated by remarking our vanilla option to volatility  $\widehat{\sigma}_2$ .

Denote by  $\tau_k$ ,  $k = 1 \dots n$ , the dates at which we switch from volatility  $\widehat{\sigma}_{k-1}$  to volatility  $\widehat{\sigma}_k$  – these dates can be set path by path dynamically, in the course of the simulation. Set  $\tau_{n+1} = T$ . The final expression for the option's price in our stochastic volatility model then reads:

$$\begin{aligned} P_{\bar{\sigma}}(t=0) &= P_{BS}(0, S_0, \widehat{\sigma}_0) \\ &+ \sum_{k=1}^n E_{\bar{\sigma}_t} \left[ \int_{\tau_k}^{\tau_{k+1}} e^{-ru} \frac{S_u^2}{2} \frac{d^2 P_{BS}}{dS^2}(u, S_u, \widehat{\sigma}_k) (\bar{\sigma}_u^2 - \widehat{\sigma}_k^2) du \right] \\ &+ \sum_{k=1}^n e^{-r\tau_k} \left[ P_{BS}(\tau_k, S_{\tau_k}, \widehat{\sigma}_k) - P_{BS}(\tau_k, S_{\tau_k}, \widehat{\sigma}_{k-1}) \right] \end{aligned}$$

What is the optimal strategy for choosing times  $\tau_k$ ? As the computational cost of evaluating  $P_{BS}$  is appreciable, we should readjust our risk-management volatility only when (a) the instantaneous volatility  $\bar{\sigma}_t$  is significantly different from  $\widehat{\sigma}_k$ , (b) the dollar gamma is large.

For vanilla options, the dollar gamma becomes largest near the option's maturity, for spot values in the neighborhood of the option's strike. As we near the option's maturity it is probably preferable to risk-manage the option at zero implied volatility, so that contributions are only generated by returns that cross the option's strike.

### A.3 Timer option-like algorithm

With respect to straight evaluation of the final payoff, accruing the gamma/theta P&L produces a less noisy estimate as it corresponds to delta-hedging our option. Can we take this idea one step further and get rid of the gamma/theta P&L as well? This is achieved by risk-managing our option in timer-wise fashion – we kindly ask the reader to read Appendix A of Chapter 5, page 180, before proceeding further.

Start at time  $t_0 = 0$  with a quadratic variation budget  $\mathcal{Q}_0$  and zero initial quadratic variation of  $\ln S_t$ :  $Q_{t=0} = 0$ . As time advances  $\mathcal{Q}_0 - Q_t$  decreases. When, at time  $t = \tau_1$   $\mathcal{Q}_0 - Q_t$  falls below a set threshold we add to the initial budget  $\mathcal{Q}_0$  so that at  $t = \tau_1^+$  it is equal to  $\mathcal{Q}_1$ . Remarking our budget from  $\mathcal{Q}_0$  to  $\mathcal{Q}_1$  generates negative mark-to-market P&L for us.

We proceed likewise, incrementing the quadratic variation budget at (random) times  $\tau_i$ ,  $i = 1 \dots n$ , whenever  $\mathcal{Q}_{i-1} - Q_{\tau_i}$  falls below the threshold, and recording the corresponding mark-to-market P&Ls, until maturity. At  $T$ , we pay to the client the option's intrinsic value; any remaining quadratic variation budget generates positive P&L for us.

$P_{\bar{\sigma}}(t = 0)$  is thus given by:

$$\begin{aligned} P_{\bar{\sigma}}(t = 0) &= e^{-rT} \mathcal{P}_{BS}(S_0 e^{(r-q)(T-t)}, 0; \mathcal{Q}_0) \\ &+ e^{-rT} \sum_{i=1}^n E_{\bar{\sigma}_t} \left[ \mathcal{P}_{BS}(F_{\tau_i}^T(S_{\tau_i}), Q_{\tau_i}; \mathcal{Q}_i) - \mathcal{P}_{BS}(F_{\tau_i}^T(S_{\tau_i}), Q_{\tau_i}; \mathcal{Q}_{i-1}) \right] \\ &- e^{-rT} E_{\bar{\sigma}_t} \left[ \mathcal{P}_{BS}(S_T, Q_T; \mathcal{Q}_n) - f(S_T) \right] \end{aligned} \quad (8.66)$$

where  $F_{\tau_i}^T(S_{\tau_i}) = S_{\tau_i} e^{r(T-\tau_i)}$  is the forward for maturity  $T$  at time  $\tau_i$  for spot  $S_{\tau_i}$ ,  $\mathcal{Q}_i$  is the quadratic variation budget at  $t = \tau_i$  and  $f(S)$  is the option's payoff.

$\mathcal{P}_{BS}(S, Q; \mathcal{Q})$ , defined in (5.68), page 183, is the Black-Scholes price of our European option, with vanishing interest rate and repo, as a function of the quadratic variation budget  $\mathcal{Q}$  and quadratic variation  $Q$ :

$$\mathcal{P}_{BS}(S, Q; \mathcal{Q}) = E \left[ f \left( S e^{-\frac{\mathcal{Q}-Q}{2} + \sqrt{\mathcal{Q}-Q} Z} \right) \right]$$

where  $Z$  is a standard normal variable.

Let us prove that (8.66) is indeed correct. We show on page 186 that  $P(t, S_t, Q_t)$ , given by:

$$P(t, S_t, Q_t) = e^{-r(T-t)} \mathcal{P}_{BS}(S_t e^{(r-q)(T-t)}, Q_t; \mathcal{Q})$$

is, by construction, a discounted martingale. Indeed:

$$E[dP] = \left( \frac{dP}{dt} + (r - q)S \frac{dP}{dS} \right) dt + \left( \frac{S^2}{2} \frac{d^2P}{dS^2} + \frac{dP}{dQ} \right) \bar{\sigma}_t^2 dt \quad (8.67)$$

Condition (5.72a) ensures that the second piece in the right-hand side of (8.67) vanishes, thus  $P$  is not sensitive to realized volatility. Condition (5.72b) then implies that  $E[dP] = rPdt$ :  $P(t, S_t, Q_t)$  is a discounted martingale. Thus

$$\mathcal{P}_{BS}(F^T(S_t), Q_t; \mathcal{Q}) = e^{r(T-t)} P(t, S_t, Q_t)$$

is a martingale.

In our algorithm  $\mathcal{Q}$  is a process that starts from  $\mathcal{Q}_0$  at  $t = 0$  and is piecewise constant, jumping from  $\mathcal{Q}_{i-1}$  to  $\mathcal{Q}_i$  at times  $\tau_i$ . Define  $\mathcal{P}_t$  as:  $\mathcal{P}_t = \mathcal{P}_{BS}(F^T(S_t), Q_t; \mathcal{Q}_t)$ :  $\mathcal{P}_t$  is a martingale on each interval  $[\tau_{i-1}, \tau_i]$  and its discontinuity at times  $\tau_i$  is :

$$\mathcal{P}_{\tau_i^+} - \mathcal{P}_{\tau_i^-} = \mathcal{P}_{BS}(F^T(S_{\tau_i}), Q_{\tau_i}; \mathcal{Q}_i) - \mathcal{P}_{BS}(F^T(S_{\tau_i}), Q_{\tau_i}; \mathcal{Q}_{i-1})$$

Taking expectations, we get the identity:

$$\mathcal{P}_0 = E_{\bar{\sigma}_t}[\mathcal{P}_T] - \sum_{i=1}^n E_{\bar{\sigma}_t} \left[ \mathcal{P}_{BS}(F^T(S_{\tau_i}), Q_{\tau_i}; \mathcal{Q}_i) - \mathcal{P}_{BS}(F^T(S_{\tau_i}), Q_{\tau_i}; \mathcal{Q}_{i-1}) \right]$$

This, together with the identity  $P_{\bar{\sigma}}(t = 0) = e^{-rT} E_{\bar{\sigma}_t}[f(S_T)]$  yields (8.66).

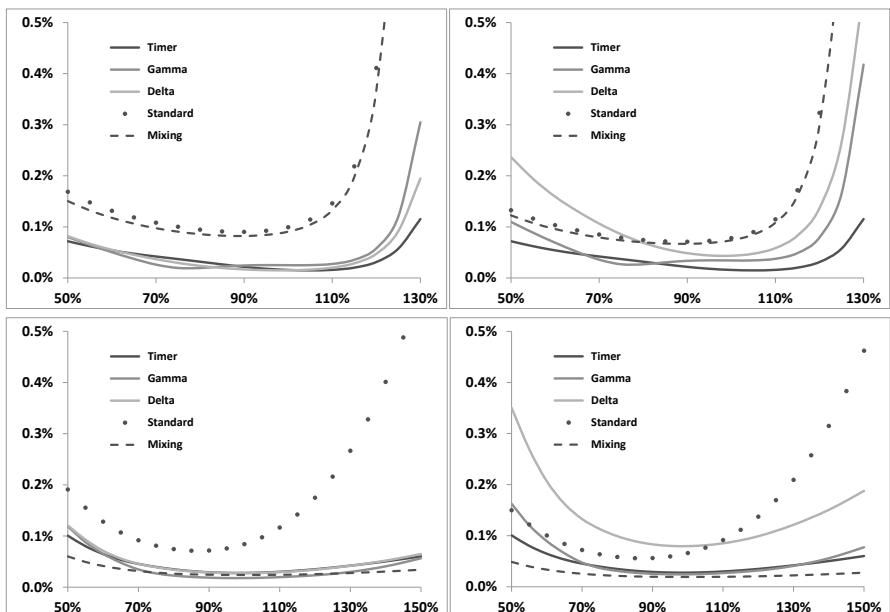
The algorithm that (8.66) expresses is very simple: simulate paths of  $S$ , keeping track of the realized quadratic variation, accumulating the Black-Scholes mark-to-market P&Ls in the second line of (8.66) as they occur, as well as the final P&L at maturity.

The strategy for setting the threshold level and the budget increment when it is hit can be optimized. On one hand, incrementing the budget by small quantities will generate small P&Ls at times  $\tau_k$  and also a small P&L at maturity. On the other hand too many evaluations of these P&Ls slow down the algorithm.

Even without any optimization, this is a very efficient algorithm for computing smiles in stochastic volatility models.

#### A.4 A comparison

The relative accuracies of the techniques discussed above are illustrated in Figure 8.8, where we have used Set II parameters – see Table 8.2, page 329 – which were used to generate smiles in Figure 8.2. The bottom graphs use Set II parameters, but with  $\rho_{SX^1} = \rho_{SX^2} = 0$ . The VS volatilities are flat at 20%.



**Figure 8.8:** Errors in implied volatilities of 1-year vanilla options, in the two-factor model with Set II parameters (top) and Set II parameters with  $\rho_{SX^1} = \rho_{SX^2} = 0$  (bottom), for different Monte Carlo algorithms. A VS volatility at 20% has been used.

The left-hand graphs show the errors of one-year implied volatilities, computed as one standard deviation of Monte Carlo prices of vanilla options, divided by the options' vegas to convert the errors in volatility units, with 100000 paths.

“Mixing” denotes the mixing solution, “Standard” denotes the standard technique of evaluating the final payoff, “Delta” refers to the same technique, but where the option is delta-hedged with the Black-Scholes delta computed with  $\hat{\sigma} = 20\%$ , “Gamma” denotes the estimator in (8.64) where we use  $\hat{\sigma} = 20\%$ . “Timer” denotes the estimator (8.66); we use as initial budget  $\mathcal{Q}_0 = \frac{1}{2}\hat{\sigma}^2 T$ ; and budgets  $\mathcal{Q}_k$  are a function of the VS volatility for the residual maturity:  $\mathcal{Q}_k = \frac{1}{2}\hat{\sigma}_T^2(\tau_k)(T - \tau_k)$ .  $\mathcal{Q}$  is readjusted at time  $\tau_k$  when the remaining budget falls below  $0.03 \bullet \xi_{\tau_k}^{\tau_k}$ .

Since the computational costs of these algorithms are different, the same errors appear in the right-hand graphs, rescaled according to  $\varepsilon_{\text{rescaled}} = \varepsilon \sqrt{\frac{T}{T_{\text{timer}}}}$ , where  $\varepsilon$  is error of the algorithm considered and  $T$  is its computational cost. The rescaled errors then correspond to a fixed computational cost – equal to that of the “Timer” algorithm – rather than a fixed number of Monte Carlo simulations.<sup>13</sup>

As expected, the mixing solution outperforms other techniques in the case of vanishing spot/volatility correlation (bottom graphs). Its effectiveness is greatly reduced in the correlated case (top graphs) to the point where it is barely more accurate than the standard technique, even though spot/volatility correlation levels can hardly be considered extreme. The value of  $\lambda$  in (8.62b) is 90.2%.

As is clear from the left-hand graphs the “Gamma” and “Timer” algorithms outperform the standard technique and have comparable accuracies for near-the-money strikes. The “Delta” technique, though roughly as accurate as the “Gamma” algorithm, is hampered by its computational cost.

It should be mentioned that this discussion is relevant to situations when standard random numbers are used in our Monte Carlo simulation. In this case, the standard deviation of our Monte Carlo estimator is indeed related to the second moment of the random variable whose expectation we are evaluating.

This relationship no longer holds when one uses quasi-random sequences – which is usually the case in practice. It becomes then difficult to assess the accuracy of our Monte Carlo estimate; one typically compares the estimate at hand with a benchmark obtained with a very large number of paths.

The mixing solution produces, by construction, a price estimate for each path that is strictly positive. It is thus possible to imply for all strikes strictly positive and arbitrage-free volatilities – a valuable benefit, even though these volatilities may be inaccurate. This is not the case for other techniques. For example, for far-away strikes, in the “Standard” technique, path contributions vanish, while with the “Timer” and “Gamma” techniques, they may even be negative. Despite this, the “Gamma” and “Timer” algorithms as well as the mixing solution are all good default algorithms for generating vanilla smiles.

---

<sup>13</sup>The Monte Carlo estimate of the option’s price is  $F = \frac{1}{n} \sum_i f_i$  where  $n$  is the number of paths used, and  $f_i$  is the contribution of path  $i$ . When using standard random numbers, the  $f_i$  are independent and we have  $\text{Stdev}(F) = \frac{1}{\sqrt{n}} \text{Stdev}(f)$ , hence the formula for the rescaling.

## A.5 Dividends

In the presence of cash-amount dividends, both the “Timer”, “Gamma” and mixing solution techniques cannot be used as is, as they make use of the Black-Scholes formula for the vanilla option’s price and gamma.

The “Gamma” and “Timer” algorithms can be amended so that they still work with cash dividends. Take as base model the Black-Scholes model with no dividends – better, take as base model a Black-Scholes model with effective proportional dividends such that forwards for all maturities are matched.

Denote by  $d_k$  the dividend falling at time  $t_k$ , which may generally be a function of  $S_{t_k^-}$ , and  $y_k^*$  the corresponding effective yield such that forwards for all maturities are matched.  $y_k^*$  is defined by:

$$y_k^* F^{\tau_k^-}(S_0) = F^{\tau_k^-}(S_0) - F^{\tau_k^+}(S_0)$$

where  $F^\tau(S)$  is the forward for maturity  $\tau$ , for the initial spot value  $S$ . We denote by  $P_{BS}^*$  prices computed in the Black-Scholes model with proportional dividends  $y_k^*$  – they are still given by the standard Black-Scholes formula.

Going through the derivation on page 40 that led to (2.30) we get an additional contribution to the right-hand side of (8.64), generated by jumps of  $S_t$  at dividend dates:

$$\begin{aligned} E_{\bar{\sigma}_t} \left[ \sum_k e^{-rt_k} \left( P_{BS}^*(t_k^+, S_{t_k^-} - d_k(S_{t_k^-})) - P_{BS}^*(t_k^+, (1 - y_k^*)S_{t_k^-}) \right) \right] \\ = E_{\bar{\sigma}_t} \left[ \sum_k e^{-rt_k} \left( P_{BS}^* \left( t_k^-, \frac{S_{t_k^-} - d_k(S_{t_k^-})}{1 - y_k^*} \right) - P_{BS}^*(t_k^-, S_{t_k^-}) \right) \right] \end{aligned}$$

where the second line follows from the fact that  $P_{BS}^*(t, S)$  is such that  $P_{BS}^*(t_k^+, (1 - y_k^*)S_{t_k^-}) = P_{BS}^*(t_k^-, S_{t_k^-})$  by construction.

Thus, in the “Gamma” algorithm, (8.65) is supplemented with :

$$\sum_k e^{-rt_k} \left[ P_{BS}^* \left( t_k^-, \frac{S_{t_k^-} - d_k(S_{t_k^-})}{1 - y_k^*}, \sigma_0 \right) - P_{BS}^*(t_k^-, S_{t_k^-}, \sigma_0) \right]$$

while in the “Timer” algorithm, (8.66) is supplemented with:

$$\sum_k e^{-rt_k} \left[ \mathcal{P}_{BS}^* \left( F_{t_k}^T \left( \frac{S_{t_k^-} - d_k(S_{t_k^-})}{1 - y_k^*} \right), Q_{t_k}; \mathcal{Q}(t_k) \right) - \mathcal{P}_{BS}^* \left( F_{t_k}^T(S_{t_k^-}), Q_{t_k}; \mathcal{Q}(t_k) \right) \right]$$

where  $\mathcal{Q}(t_k)$  is the quadratic variation budget at time  $t_k$  and  $F_t^T(S)$  is the forward at time  $t$ , spot  $S$ , for maturity  $T$ , in the proportional dividend model.

The computational cost of evaluating a Black-Scholes price is equivalent to computing two deltas. This technique can thus be employed when there are few dividends, that is for stocks. What should we do when there are numerous dividends, as in indexes?

### A.5.1 An efficient approximation

We now make use of an approximation for vanilla option prices in the Black-Scholes model introduced in Section 2.3.1, page 34, published by Michael Bos and Stephen Vandermarck – see [16].

In this approximation dividends are converted into two effective dividends falling at  $t = 0$  and at maturity, thus resulting in a negative adjustment of the initial spot value and a positive adjustment of the strike.

For the sake of pricing a vanilla option of strike  $K$ , maturity  $T$  in the Black-Scholes model, the regular Black-Scholes formula is used, with  $S, K$  replaced with  $\alpha(T)S - \delta S(T)$  and  $K + \delta K(T)$ .

We recall here the expressions of  $\alpha(T)$ ,  $\delta S(T)$ ,  $\delta K(T)$  already given on page 37. Let  $y_i$  and  $c_i$  be the yield and cash-amount of the dividend falling at time  $t_i$ :  $S_{t_i^+} = (1 - y_i)S_{t_i^-} - c_i$ .  $\alpha(T)$ ,  $\delta S(T)$ ,  $\delta K(T)$  read:

$$\left\{ \begin{array}{l} \alpha(T) = \prod_{t_i < T} (1 - y_i) \\ \delta S(T) = \sum_{t_i < T} \frac{T - t_i}{T} c_i^* e^{-(r-q)t_i} \\ \delta K(T) = \sum_{t_i < T} \frac{t_i}{T} c_i^* e^{(r-q)(T-t_i)} \end{array} \right.$$

where the effective cash amounts  $c_i^*$  are given by:  $c_i^* = c_i \prod_{t_i < t_j < T} (1 - y_j)$ .

Our (heuristic) recipe is:

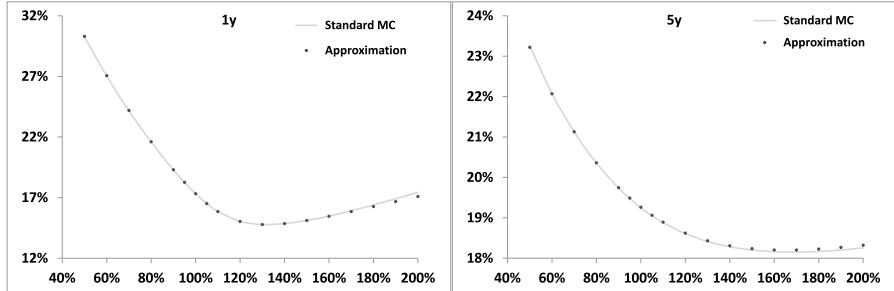
- price vanilla options in the stochastic volatility model using effective spot and strike values  $\alpha(T)S - \delta S(T)$  and  $K + \delta K(T)$  and no dividends,
- imply Black-Scholes volatilities using these effective values as well.

As there are no cash-amount dividends anymore, both the mixing solution, gamma/theta and “Timer” algorithms can be used with no alteration. This approximation for implied volatilities is accurate, both for stock and index smiles – see Figure 8.9 for an example.<sup>14</sup>

## Appendix B – local volatility function of stochastic volatility models

Given the smile of a stochastic volatility model, one may need to determine the corresponding local volatility function, for example for the sake of comparing

<sup>14</sup>Besides, in case the stochastic volatility model degenerates into the Black-Scholes model – for example with vanishing volatility of volatility – the exact implied volatilities are recovered.



**Figure 8.9:** 1-year and 5-year smiles as generated by the two-factor model, either in a standard Monte Carlo simulation or using the approximation in Section A.5.1. Market data for the Euro Stoxx 50 index as of October 1, 2014 have been used, with zero repo and interest rate. The two-factor model parameters are:  $\nu = 263\%$ ,  $\theta = 11.5\%$ ,  $k_1 = 10.28$ ,  $k_2 = 0.42$ ,  $\rho = 40\%$ ,  $\rho_{SX^1} = -71.8\%$ ,  $\rho_{SX^2} = -21.1\%$ .

option prices computed with a stochastic volatility model and a local volatility model calibrated to the same smile. We carry out such a test in Chapter 11.

One may of course generate a vanilla smile using one of the techniques presented above, and then determine the local volatility function with the Dupire formula (2.3) using prices or with (2.19) using implied volatilities.

The local volatility function can however be obtained directly. From (2.6), page 28, the square of the local volatility function is equal to the expectation of the instantaneous variance conditional on the spot value:

$$\sigma(t, S)^2 = E[\xi_t^t | S_t = S] = \frac{E[\xi_t^t \delta(S_t - S)]}{E[\delta(S_t - S)]} \quad (8.68)$$

From Section A.1, conditional on the paths of the Brownian motions driving the instantaneous variance  $\xi_t^t$ ,  $S_t$  is lognormally distributed. From equations (8.60) and (8.61):

$$S_t = S_0^* e^{-\frac{\hat{\sigma}_t^{*2} t}{2} + \hat{\sigma}^* \sqrt{t} Z}$$

where  $Z$  is a standard normal variable and the effective spot and volatility  $S_0^*$ ,  $\sigma^*$  are given by:

$$\begin{aligned} \ln S_{0t}^* &= \ln S_0 - \frac{\lambda^2}{2} \int_0^t \xi_u^u du + \lambda \int_0^t \sqrt{\xi_u^u} dW_u^\parallel \\ \hat{\sigma}_t^{*2} &= (1 - \lambda^2) \frac{1}{t} \int_0^t \xi_u^u du \end{aligned}$$

where  $\lambda$ ,  $W_u^\parallel$  are defined in (8.59). We use  $t$  subscripts for  $S_{0t}^*$  and  $\hat{\sigma}_t^*$  as they are processes. Conditional on the paths of Brownian motions driving the instantaneous

variance, the density  $\rho^*$  of  $S_t$  is thus given by:

$$\rho^*(S_t) = \frac{1}{S_t} \frac{1}{\sqrt{2\pi \hat{\sigma}_t^{*2} t}} e^{-\frac{1}{2\hat{\sigma}_t^{*2} t} \left( \ln \frac{S_t}{S_{0t}^*} + \frac{\hat{\sigma}_t^{*2} t}{2} \right)^2}$$

Calculating the expectations in the right-hand side in (8.68) yields:

$$\sigma(t, S)^2 = \frac{E_\xi \left[ \xi_t^t \frac{1}{\sqrt{2\pi \hat{\sigma}_t^{*2} t}} e^{-\frac{1}{2\hat{\sigma}_t^{*2} t} \left( \ln \frac{S}{S_{0t}^*} + \frac{\hat{\sigma}_t^{*2} t}{2} \right)^2} \right]}{E_\xi \left[ \frac{1}{\sqrt{2\pi \hat{\sigma}_t^{*2} t}} e^{-\frac{1}{2\hat{\sigma}_t^{*2} t} \left( \ln \frac{S}{S_{0t}^*} + \frac{\hat{\sigma}_t^{*2} t}{2} \right)^2} \right]} \quad (8.69)$$

where the  $\xi$  subscripts denote that the expectations are taken with respect to the variance degrees of freedom only.

The recipe for computing  $\sigma(t, S)$  in a Monte Carlo simulation is thus:

- Simulate paths for the instantaneous variance, accruing the integrals  $\int_0^t \xi_u^u du$  and  $\int_0^t \sqrt{\xi_u^u} dW_u^\parallel$ . For each path, store the triplet  $(\xi_t^t, S_{0t}^*, \hat{\sigma}_t^*)$  at times  $t$  of interest.
- Compute the expectations in (8.69) as averages over these paths for values of  $S$  of interest.

Formula (8.69) was published in [66].

## Appendix C – partial resummation of higher orders

Carrying out expansion (8.18) to higher orders in  $\varepsilon$  is tedious but straightforward. Can we identify, at each order in  $\varepsilon$ , a subset of terms that could be analytically calculated and resummed? A hint that this may be possible is provided by the contribution of the spot/volatility covariance function evaluated on the initial variance curve,  $C_0^{x\xi}(\xi_0)$ , to the second-order expansion (8.18):

$$\left( \varepsilon \frac{C_0^{x\xi}(\xi_0)}{2} \partial_x (\partial_x^2 - \partial_x) + \varepsilon^2 \frac{C_0^{x\xi}(\xi_0)^2}{8} \partial_x^2 (\partial_x^2 - \partial_x)^2 \right) P_0 \quad (8.70)$$

This looks like the beginning of the expansion of  $\exp \left( \varepsilon \frac{C_0^{x\xi}(\xi_0)}{2} \partial_x (\partial_x^2 - \partial_x) \right) P_0$ . The term in  $\varepsilon^2$  in (8.70) is generated by the order-two contribution from  $\mathcal{W}^1$ :

$$\int_t^T d\tau_1 \int_{\tau_1}^T d\tau_2 U_{t\tau_1}^0 \mathcal{W}_{\tau_1}^1 U_{\tau_1\tau_2}^0 \mathcal{W}_{\tau_2}^1 U_{\tau_2 T}^0 g$$

with  $\mathcal{W}_t^1$  given by:

$$\mathcal{W}_t^1 = \int_t^T du \mu(t, u, \xi) \partial_{x\xi^u}^2$$

where operator  $\partial_{x\xi^u}^2$  in  $\mathcal{W}_{\tau_1}^1$  is not allowed to act on  $\mu(\tau_2, u, \xi)$ .<sup>15</sup> This amounts to replacing the spot/variance covariance function with its value computed on the initial variance curve:  $\mu(t, u, \xi) \equiv \mu(t, u, \xi_0)$ .

Let us do the same with the variance/variance covariance function:

$\nu(t, u, u', \xi) \equiv \nu(t, u, u', \xi_0)$ . Since covariance functions do not depend on forward variances anymore, operator  $\partial_{\xi^u}$  can be replaced with  $\frac{1}{2}(\partial_x^2 - \partial_x)$  and the pricing equation (8.1) is replaced with:

$$\begin{aligned} & \partial_t P + \frac{\xi^t}{2} (\partial_x^2 - \partial_x) P \\ & + \frac{\varepsilon}{2} \int_t^T du \mu_t^u \partial_x (\partial_x^2 - \partial_x) P + \frac{\varepsilon^2}{8} \int_t^T \int_t^T du du' \nu_t^{uu'} (\partial_x^2 - \partial_x)^2 P = 0 \end{aligned} \quad (8.71)$$

where we assume zero rates and repo, we use the compact notation  $\mu_t^u \equiv \mu(t, u, \xi_0)$ ,  $\nu_t^{uu'} = \nu(t, u, u', \xi_0)$ , and  $\mu_t^u$  and  $\nu_t^{uu'}$  have been rescaled respectively by  $\varepsilon$  and  $\varepsilon^2$ . The solution of (8.71) reads:

$$\begin{aligned} P &= \exp \left( \frac{1}{2} \left[ \int_0^T \xi_0^t dt \right] (\partial_x^2 - \partial_x) + \frac{\varepsilon}{2} \left[ \int_0^T dt \int_t^T du \mu_t^u \right] \partial_x (\partial_x^2 - \partial_x) \right. \\ &\quad \left. + \frac{\varepsilon^2}{8} \left[ \int_0^T dt \int_t^T \int_t^T du du' \nu_t^{uu'} \right] (\partial_x^2 - \partial_x)^2 \right) g \end{aligned} \quad (8.72)$$

$$= e^{\frac{Q}{2}(\partial_x^2 - \partial_x) + \varepsilon \frac{C_0^x \xi(\xi_0)}{2} \partial_x (\partial_x^2 - \partial_x) + \varepsilon^2 \frac{C_0^x \xi(\xi_0)}{8} (\partial_x^2 - \partial_x)^2} g \quad (8.72)$$

$$= e^{\varepsilon \frac{C_0^x \xi(\xi_0)}{2} \partial_x (\partial_x^2 - \partial_x) + \varepsilon^2 \frac{C_0^x \xi(\xi_0)}{8} (\partial_x^2 - \partial_x)^2} P_0 \quad (8.73)$$

where  $Q = \int_0^T \xi^t dt$ . The reader can check by direct substitution that  $P$  in (8.72) indeed solves the PDE:

$$\begin{aligned} & \frac{dP}{dt} + \frac{\xi^t}{2} S^2 \frac{d^2 P}{dS^2} \\ & + \frac{1}{2} \int_t^T du \int_t^T du' \nu(t, u, u', \xi_0) \frac{d^2 P}{\delta \xi^u \delta \xi^{u'}} + \int_t^T du \mu(t, u, \xi_0) S \frac{d^2 P}{dS \delta \xi^u} = 0 \end{aligned} \quad (8.74)$$

Expanding (8.73) at order two in  $\varepsilon$ , we recover (8.18) but for the last term, which involves the derivative of  $\mu$  with respect to forward variances – see the definition of  $D_0(\xi_0)$  in (8.17).

---

<sup>15</sup> Action of  $\partial_{x\xi^u}^2$  on  $\mu(\tau_2, u, \xi)$  generates the term  $\varepsilon^2 \frac{D_0(\xi_0)}{2} \partial_x^2 (\partial_x^2 - \partial_x) P_0$  in (8.18).

$P$  is easily computed through a Laplace transform. Define

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

set  $P = Se^{-q(T-t)} f(t, x)$  and introduce the Laplace transform  $F(t, p)$  of  $f$ :

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

The Laplace transform  $G(p)$  of  $g$  is given, for a call option by:

$$G(p) = \int_{-\infty}^{+\infty} e^{-px} (1 - e^{-x})^+ dx = \int_0^{\infty} e^{-px} (1 - e^{-x}) dx = \frac{1}{p(p+1)}$$

It is defined for  $\text{Re}(p) > 0$ . For a put option  $G(p)$  is identical, except it is defined for  $\text{Re}(p) < -1$ . From expression (8.72)  $F(t=0, p)$  is then given by:

$$F(0, p) = \frac{1}{p(p+1)} e^{\frac{Q}{2}p(1+p) + \varepsilon \frac{C_0^{x\xi}(\xi_0)}{2}p(1+p)^2 + \varepsilon^2 \frac{C_0^{\xi\xi}(\xi_0)}{8}p^2(1+p)^2} \quad (8.75)$$

Inverting  $F(0, p)$  then yields  $P$ .

It turns out that, practically, this approximation does not work well and is worse than the order-two expansion in  $\varepsilon$ . In other words, resumming analytically all terms in the expansion of  $P$  in  $\varepsilon$  that do not involve derivatives of  $\mu$  and  $\nu$  with respect to forward variances is not sufficiently accurate. This is probably due to the fact that expression (8.75) for  $F(0, p)$  does not correspond to a legitimate density.

Consider a level  $S_T$  for the spot at time  $T$  and define the log-return

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

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

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

where  $\rho_z(z)$  is the density of  $z$ . Let us introduce the cumulant-generating function  $L(q)$  of  $\rho_z$ :

$$\begin{aligned} e^{L(q)} &= \int_{-\infty}^{+\infty} e^{-qz} \rho_z(z) dz = \int_{-\infty}^{+\infty} e^{-qz} (e^x (\partial_x + \partial_x^2) f)_{x=-z} dz \\ &= q(1+q) F(0, -(1+q)) \end{aligned}$$

We thus get:

$$L(q) = \frac{Q}{2}q(1+q) - \varepsilon \frac{C_0^{x\xi}(\xi_0)}{2}q^2(1+q) + \varepsilon^2 \frac{C_0^{\xi\xi}(\xi_0)}{8}q^2(1+q)^2$$

The conditions that:

- $\rho_z$  integrate to one ( $L(0) = 0$ ),
- the forward for maturity  $T$  is matched:  $E[S_T] = S_0$  ( $L(-1) = 0$ )
- the VS volatility for maturity  $T$  be matched ( $\left.\frac{dL}{dq}\right|_{q=0} = \frac{Q}{2}$ )

are obviously satisfied.<sup>16</sup>

The fact that  $L$  is a polynomial is a problem, however: a theorem by Marcinkiewicz [72] states that if a cumulant-generating function is a polynomial, its order cannot be greater than 2: the density corresponding to solutions of (8.74) is not positive.<sup>17</sup>

---

<sup>16</sup>See also the discussion in Appendix B of Chapter 5.

<sup>17</sup>This drawback is shared by the density generated by our original expansion at order two in  $\varepsilon$ ; the latter however is a small perturbation of a Gaussian density – see the discussion in Section 8.3.

## Chapter's digest

### 8.2 Expansion of the price in volatility of volatility

► We consider forward variance models and derive the following expansion of vanilla option prices at order two in volatility of volatility:

$$\begin{aligned} P = & \left[ 1 + \varepsilon \frac{C_0^{x\xi}(\xi_0)}{2} \partial_x (\partial_x^2 - \partial_x) \right. \\ & \left. + \varepsilon^2 \left( \frac{C_0^{\xi\xi}(\xi_0)}{8} (\partial_x^2 - \partial_x)^2 + \frac{C_0^{x\xi}(\xi_0)^2}{8} \partial_x^2 (\partial_x^2 - \partial_x)^2 + \frac{D_0(\xi_0)}{2} \partial_x^2 (\partial_x^2 - \partial_x) \right) \right] P_0 \end{aligned}$$

in terms of three dimensionless quantities that are readily calculated in any model of interest:

$$\begin{aligned} C_t^{x\xi}(\xi) &= \int_t^T (T-\tau) \langle d\ln S_\tau d\hat{\sigma}_T^2(\tau) \rangle \\ C_t^{\xi\xi}(\xi) &= \int_t^T (T-\tau)^2 \langle d\hat{\sigma}_T^2(\tau) d\hat{\sigma}_T^2(\tau) \rangle \\ D_t(\xi) &= \int_t^T d\tau \int_\tau^T du (T-u) \frac{1}{d\tau} \langle d\ln S_u \frac{\langle d\ln S_u d\hat{\sigma}_T^2(u) \rangle}{du} \rangle \end{aligned}$$

Our derivation requires that spot/volatility covariances not depend on  $S$ .



### 8.3 Expansion of implied volatilities

► This expansion translates into the following expansion of implied volatilities:

$$\hat{\sigma}(K, T) = \hat{\sigma}_{FTT} + \mathcal{S}_T \ln \left( \frac{K}{F_T} \right) + \frac{\mathcal{C}_T}{2} \ln^2 \left( \frac{K}{F_T} \right)$$

with:

$$\begin{aligned} \hat{\sigma}_{FTT} &= \hat{\sigma}_T \left[ 1 + \frac{\varepsilon}{4Q} C^{x\xi} + \frac{\varepsilon^2}{32Q^3} \left( 12(C^{x\xi})^2 - Q(Q+4)C^{\xi\xi} + 4Q(Q-4)D \right) \right] \\ \mathcal{S}_T &= \hat{\sigma}_T \left[ \frac{\varepsilon}{2Q^2} C^{x\xi} + \frac{\varepsilon^2}{8Q^3} \left( 4QD - 3(C^{x\xi})^2 \right) \right] \\ \mathcal{C}_T &= \hat{\sigma}_T \frac{\varepsilon^2}{4Q^4} \left( 4QD + QC^{\xi\xi} - 6(C^{x\xi})^2 \right) \end{aligned}$$

where  $\hat{\sigma}_T$  is the VS volatility for maturity  $T$ ,  $Q = \hat{\sigma}_T^2 T$ ,  $C^{x\xi} \equiv C_0^{x\xi}(\xi_0)$ ,  $C^{\xi\xi} \equiv C_0^{\xi\xi}(\xi_0)$ ,  $D = D_0(\xi_0)$ .

- At order one in volatility of volatility, the ATMF skew is given by:

$$\begin{aligned} S_T &= \hat{\sigma}_T \frac{C^{x\zeta}}{2(\hat{\sigma}_T^2 T)^2} \\ &= \frac{1}{2\hat{\sigma}_T^3 T} \int_0^T \frac{T-\tau}{T} \frac{\langle d\ln S_\tau \, d\hat{\sigma}_T^2(\tau) \rangle_0}{d\tau} d\tau \end{aligned}$$

•••••

#### 8.4 A representation of European option prices in diffusive models

- The expression of the ATMF skew at order one in volatility of volatility can also be obtained from the following general expression of European option prices in diffusive models:

$$\begin{aligned} P &= P_{BS} (0, S_0, \hat{\sigma}_T^2(0)) \\ &+ E \left[ \int_0^T e^{-rt} \left( \frac{d^2 P_{BS}}{dS d(\hat{\sigma}_T^2)} \langle dS_t \, d\hat{\sigma}_T^2(t) \rangle + \frac{1}{2} \frac{d^2 P_{BS}}{(d(\hat{\sigma}_T^2))^2} \langle d\hat{\sigma}_T^2(t) \, d\hat{\sigma}_T^2(t) \rangle \right) \right] \end{aligned}$$

- The European payoff that materializes the uniformly weighted spot/volatility covariance, at order one in volatility of volatility, is  $\ln^2(S/S_0)$ .

•••••

#### 8.5 Short maturities

- At short maturities, the ATMF skew and curvature are given, at leading order, by the following general expressions:

$$\begin{aligned} S_0 &= \frac{1}{2\hat{\sigma}_0^2} \frac{\langle d\ln S d\hat{\sigma}_0 \rangle}{dt} \\ C_0 &= \frac{1}{4\hat{\sigma}_0} \left( \frac{8}{3} \frac{\langle d\ln S dS_0 \rangle}{\hat{\sigma}_0 dt} + \frac{4}{3} \frac{\langle d\hat{\sigma}_0 d\hat{\sigma}_0 \rangle}{\hat{\sigma}_0^2 dt} - 8S_0^2 \right) \end{aligned}$$

where  $\hat{\sigma}_0$  is the short ATM volatility,  $S_0$  the short ATM skew, and  $C_0$  the short ATM curvature. While the short ATM skew is a model-independent measure of the instantaneous covariance of spot and ATM volatility, the short curvature is not a direct measure of volatility of volatility, as the covariance of spot and ATM skew contributes as well.

- Consider a lognormal model for the short ATM volatility, whose lognormal volatility of volatility is  $\nu$ . We have:

$$\begin{aligned} S_0 &= \frac{\rho}{2} \nu \\ C_0 &= \frac{1}{6\hat{\sigma}_0} (2 - 3\rho^2) \nu^2 \end{aligned}$$

The short ATM skew is independent on the level of ATM volatility, while the curvature is inversely proportional to the ATM volatility.

This behavior is typical of the SABR model.

► Consider a normal model for the short ATM volatility, whose normal volatility of volatility is  $\sigma$ . We have:

$$\begin{aligned} \mathcal{S}_0 &= \frac{\rho}{2} \frac{\sigma}{\hat{\sigma}_0} \\ \mathcal{C}_0 &= \frac{1}{6\hat{\sigma}_0} (2 - 5\rho^2) \left( \frac{\sigma}{\hat{\sigma}_0} \right)^2 \end{aligned}$$

The short ATM skew is inversely proportional to the ATM volatility, while the curvature is inversely proportional to the cube of the ATM volatility.

This behavior is typical of the Heston model.

► In case the short ATM volatility is uncorrelated with  $S$ , the ATM skew vanishes and the ATM curvature is given by:

$$\mathcal{C}_0 = \frac{1}{3\hat{\sigma}_0} \frac{\langle d\hat{\sigma}_0 d\hat{\sigma}_0 \rangle}{\hat{\sigma}_0^2 dt}$$



## 8.7 The two-factor model

► At order one in volatility of volatility, the ATMF skew of the two-factor model is given by:

$$\mathcal{S}_T^{\text{order } 1} = \frac{\nu\alpha\theta}{\hat{\sigma}_T^3 T^2} \int_0^T dt \sqrt{\xi_0^t} \int_t^T du \xi_0^u \left[ (1-\theta)\rho_{SX^1} e^{-k_1(u-t)} + \theta\rho_{SX^2} e^{-k_2(u-t)} \right]$$

which, for a flat term structure of VS volatilities, translates into:

$$\mathcal{S}_T^{\text{order } 1} = \nu\alpha\theta \left[ (1-\theta)\rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta\rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2} \right]$$

$\mathcal{S}_T^{\text{order } 1}$  does not depend on the level of VS volatility. These approximate expressions for the ATMF skew are accurate and can be used for choosing  $\rho_{SX^1}$  and  $\rho_{SX^2}$  so as to generate the desired decay for the ATMF skew. In particular, it is possible to approximately produce the typical power-law decay of equity index skews.



## 8.8 Conclusion

► The expansion at order two in volatility of volatility is adequate for near-the-money strikes. Its accuracy deteriorates as one moves to out-of-the-money strikes and longer maturities.

While the order-one expansion is sufficient, in the two-factor model, for generating an accurate estimation of the ATMF skew, for the sake of approximating the ATMF volatility, the second order is needed.

At order one in volatility of volatility, the ATMF skew is determined by the covariance of spot and VS volatilities; changing the spot/volatility correlation while rescaling accordingly volatilities of volatilities so that spot/volatility covariances are unchanged indeed leaves the ATMF skew unchanged.

The effect on out-of-the-money implied volatilities is asymmetrical. Typically, for equity index smiles, implied volatilities for low strikes are roughly unchanged, while those of high strikes are impacted.



### 8.9 Forward-start options – future smiles

► Unlike the local volatility model, stochastic volatility models afford more control on the pricing of the risks of forward-start options: volatility-of-volatility and forward-smile risks. In time-homogeneous stochastic volatility models, future smiles are predictable and similar to spot-starting smiles. In the two-factor model, the level of ATMF skew, be it spot-starting of future, is approximately fixed, independent on the level of ATMF volatility.



### 8.10 Impact of the smile of volatility of volatility on the vanilla smile

► We use volatility-of-volatility smile parameters derived from calibration on VIX smiles. An upward-sloping volatility-of-volatility smile makes the vanilla smile of the underlying more convex near the money and less steep. This numerical result is supported by the order-two expansion.



## Appendix A – Monte Carlo algorithms for vanilla smiles

► We provide three efficient techniques for generating vanilla smiles in stochastic volatility models.

► The Brownian motion driving  $S_t$  can be written as  $W_t^S = \lambda W_t^{\parallel} + \sqrt{1 - \lambda^2} W_t^{\perp}$ , where  $W_t^{\perp}$  is uncorrelated to the Brownian motions driving forward variances.

The mixing solution consists in conditioning with respect to the Brownian motions driving forward variances: integration on  $W_t^{\perp}$  is analytic.

In the Monte Carlo simulation, forward variances are simulated and two integrals are accrued:  $I_1 = \int_0^T \xi_t^S dt$  and  $I_2 = \int_0^T \sqrt{\xi_t^S} \lambda dW_t^{\parallel}$ . The vanilla option price is obtained as:  $P = E_{W^{\parallel}}[P_{BS}(0, S_0^*, \hat{\sigma}^*)]$  where the expectation is taken on the

paths of the Brownian motions driving forward variances, where the effective spot and volatility  $S_0^*$  and  $\hat{\sigma}^*$  are given by:

$$\begin{aligned}\ln S_0^* &= \ln S_0 - \frac{\lambda^2}{2} I_1 + I_2 \\ \hat{\sigma}^{*2} &= (1 - \lambda^2) I_1\end{aligned}$$

For the two-factor model, we have:

$$\begin{aligned}\lambda W_t^{\parallel} &= \frac{\rho_{SX^1} - \rho_{12}\rho_{SX^2}}{1 - \rho_{12}^2} W_t^1 + \frac{\rho_{SX^2} - \rho_{12}\rho_{SX^1}}{1 - \rho_{12}^2} W_t^2 \\ \lambda &= \sqrt{\frac{\rho_{SX^1}^2 + \rho_{SX^2}^2 - 2\rho_{12}\rho_{SX^1}\rho_{SX^2}}{1 - \rho_{12}^2}}\end{aligned}$$

A set of maturities can be priced at once, by storing the values of  $I_1, I_2$  for maturities of interest.

► In the gamma/theta accrual method, the spot and forward variances are simulated, but rather than evaluating the vanilla payoff, one accrues the gamma/theta P&L, calculated with a chosen risk-management volatility  $\hat{\sigma}$ . This corresponds in essence to using the delta P&L as a control variate – without calculating delta.

The contribution of a given path to the vanilla price estimator is:

$$P_{BS}(0, S_0, \hat{\sigma}) + \sum_i e^{-rt_i} \frac{S_i^2}{2} \frac{dP_{BS}}{dS}(t_i, S_i, \hat{\sigma}) (\xi_{t_i}^{t_i} - \hat{\sigma}^2) \Delta$$

This algorithm can be optimized by dynamically adjusting the risk-management volatility  $\hat{\sigma}$  to match current levels of realized volatility, in the course of the simulation.

► In the timer-like algorithm, we get rid of the gamma/theta P&L altogether. We start at time  $t_0 = 0$  with a quadratic variation budget  $\mathcal{Q}_0$  which gets eroded by the realized quadratic variation along the simulated path.

Whenever the remaining budget falls below a given threshold, we add to it and the path estimator accrues a mark-to-market P&L given by the difference of two Black-Scholes prices. At maturity, the remaining time value is subtracted from the path estimator.

The resulting path estimator only involves mark-to-market P&Ls.



## Appendix B – local volatility function of stochastic volatility models

► The local volatility function corresponding to the vanilla smile of a given stochastic volatility model can be very efficiently obtained in a Monte Carlo simulation, without calculating vanilla option prices.

**This page intentionally left blank**

# Chapter 9

---

## *Linking static and dynamic properties of stochastic volatility models*

A stochastic volatility model can be assessed by studying its dynamic properties: the volatilities and correlations of volatilities, be they spot-starting or forward-starting and their joint dynamics with the spot. One can also choose to focus on the (static) smile it produces and examine the strike and maturity dependence of the implied volatilities it generates.

While static and dynamic features of a model are both determined by the joint dynamics of spot and forward variances in the model, how strong is the relationship between both? Which elements of this connection are specific to stochastic volatility models? Would it be possible to parametrize differently a given model so that the vanilla smile is (approximately) unchanged while the dynamics is different?

We establish a link between the rate at which the ATMF skew decays with maturity and the SSR, already introduced in the context of the local volatility model, whose definition we extend to the case of stochastic volatility models. Our analysis will allow us to split stochastic volatility models into two classes.<sup>1</sup>

---

### 9.1 The ATMF skew

The ATMF skew vanishes for vanishing volatility of volatility. At order one it is given by (8.52). Setting  $\varepsilon = 1$  and using the definition of  $C^{x\xi}$  in (8.13) yields:

$$\mathcal{S}_T = \frac{1}{2\sqrt{T}} \frac{1}{\left(\int_0^T \xi_0^\tau d\tau\right)^{3/2}} \int_0^T d\tau \int_\tau^T \mu(\tau, u, \xi_0) du \quad (9.1)$$

which can be rewritten as in (8.24), page 315:

$$\mathcal{S}_T = \frac{1}{2\hat{\sigma}_T^3 T} \int_0^T \frac{T-\tau}{T} \frac{\langle d\ln S_\tau \, d\hat{\sigma}_T^2(\tau) \rangle_0}{d\tau} d\tau \quad (9.2)$$

---

<sup>1</sup>These results were first published in [11].

## 9.2 The Skew Stickiness Ratio (SSR)

The dynamics of a given model is reflected in the covariance of the spot and forward variances. Conditional on a move of the spot, different models generate different scenarios for implied volatilities. This can be quantified by focusing on the ATMF volatility and computing the regression coefficient of  $\delta\hat{\sigma}_{F_T T}$  with respect to  $\delta \ln S$ .

It seems reasonable and natural to normalize this regression coefficient by the ATMF skew. We thus introduce the Skew Stickiness Ratio (SSR)  $\mathcal{R}_T$ , defined by:

$$\mathcal{R}_T = \frac{1}{\mathcal{S}_T} \frac{E[d \ln S \, d\hat{\sigma}_{F_T(S)T}]}{E[(d \ln S)^2]} \quad (9.3)$$

where the notation  $\hat{\sigma}_{F_T(S)T}$  emphasizes that the strike whose implied volatility we consider is not fixed. Unless necessary we will use the lighter notation  $\hat{\sigma}_{F_T T}$ .

$\mathcal{R}_T$  is dimensionless – its value is known for some classes of models:

- In jump-diffusion or Lévy models with independent stationary increments for  $\ln S$  – such as the model used in Section 5.3.2 – implied volatilities are a function of moneyness  $\frac{K}{S}$  only: as  $S$  moves  $\hat{\sigma}_{F_T T}$  is unchanged:  $\mathcal{R}_T = 0, \forall T$ .
- We have already encountered the SSR in the context of the local volatility model. Because implied volatilities are *functions* of  $(t, S)$ , formula (9.3) simplifies to:

$$\mathcal{R}_T = \frac{1}{\mathcal{S}_T} \frac{d\hat{\sigma}_{F_T(S)T}}{d \ln S} = \frac{1}{\mathcal{S}_T} \left( \frac{d\hat{\sigma}_{KT}}{d \ln K} \Big|_{F_T} + \frac{d\hat{\sigma}_{KT}}{d \ln S} \Big|_{K=F_T} \right)$$

which agrees with the definition used for the SSR in Section 2.5.2. We know from Sections 2.5.3.1 and 2.4.6 that for time-independent local volatility functions for all maturities, or for general local volatility functions in the limit  $T \rightarrow 0$ ,  $\mathcal{R}_T = 2$ . For equity index smiles,  $\mathcal{R}_T$  starts from 2 for short maturities and typically reaches values above 2 for long maturities – see for example Figure 2.4, page 59, and Figure 9.9, page 380.

We now compute  $\mathcal{R}_T$  in a general stochastic volatility model at lowest order in volatility of volatility.  $\mathcal{S}_T$  vanishes for vanishing volatility of volatility and starts with a term of order one. To get  $\mathcal{R}_T$  at lowest order we thus need to compute the covariance  $E[dS \, d\hat{\sigma}_{F_T T}]$  at order one. From (8.21a), the difference between ATMF and VS volatilities is of order one. For the purpose of computing  $E[dS \, d\hat{\sigma}_{F_T T}]$  at order one, we can thus conveniently replace the ATMF volatility with the VS volatility:

$$E[dS \, d\hat{\sigma}_{F_T T}] \simeq E[dS \, d\hat{\sigma}_T]$$

From the definition of the VS volatility  $\widehat{\sigma}_T^2(t) = \frac{1}{T-t} \int_t^T \xi_t^u du$  we have:

$$\begin{aligned} E[d \ln S_t d\widehat{\sigma}_T(t)] &= \frac{1}{2(T-t)\widehat{\sigma}_T(t)} \int_t^T E[d \ln S_t d\xi_t^u] du \\ &= \frac{1}{2(T-t)\widehat{\sigma}_T(t)} \int_t^T \mu(t, u, \xi_0) du dt \end{aligned}$$

Setting  $t = 0$  and dividing by  $E[(d \ln S)^2]$  which is equal to  $\xi_0^0 dt$ , and by  $S_T$ , we get the following expression for  $\mathcal{R}_T$  at lowest non-trivial order in volatility of volatility:

$$\mathcal{R}_T = \frac{\int_0^T \xi_0^\tau d\tau}{T\xi_0^0} \frac{T \int_0^T \mu(0, u, \xi_0) du}{\int_0^T d\tau \int_\tau^T \mu(\tau, u, \xi_0) du} \quad (9.4)$$


---

### 9.3 Short-maturity limit of the ATMF skew and the SSR

Let us take the limit  $T \rightarrow 0$  in expression (9.1). We get:

$$\mathcal{S}_0 = \lim_{T \rightarrow 0} \frac{1}{2\sqrt{T}} \frac{1}{\left(\int_0^T \xi_0^\tau d\tau\right)^{3/2}} \int_0^T d\tau \int_\tau^T \mu(\tau, u, \xi_0) du = \frac{\mu(0, 0, \xi_0)}{4(\xi_0^0)^{3/2}}$$

This recovers (8.36): at order one in volatility of volatility the short ATMF skew is a direct measure of the instantaneous spot/volatility covariance. Turning now to the SSR:

$$\begin{aligned} \mathcal{R}_0 &= \lim_{T \rightarrow 0} \frac{\int_0^T \xi_0^\tau d\tau}{T\xi_0^0} \frac{T \int_0^T \mu(0, u, \xi_0) du}{\int_0^T d\tau \int_\tau^T \mu(\tau, u, \xi_0) du} = \lim_{T \rightarrow 0} \frac{T \int_0^T du}{\int_0^T d\tau \int_\tau^T du} \\ &= 2 \end{aligned}$$

Thus, in stochastic volatility models, the short limit of the SSR is 2, as in the local volatility model. This “2” is the 2 in the denominator of (8.36).

---

### 9.4 Model-independent range of the SSR

Let us assume that the term structure of VS volatilities is flat –  $\xi_0^\tau \equiv \xi_0$  – and that the model at hand is time-homogeneous, that is the spot/variance covariance function  $\mu(\tau, u, \xi)$  only depends on  $u - \tau$ . We now use the economical notation  $\mu(u - \tau)$ :

$$\mu(\tau, u, \xi_0) = \lim_{d\tau \rightarrow 0} \frac{1}{d\tau} E[d \ln S_\tau d\xi_\tau^u] \equiv \mu(u - \tau)$$

One of the integrals in (9.1) can be done analytically and we get the following simple expressions for  $\mathcal{S}_T$  and  $\mathcal{R}_T$ :

$$\mathcal{S}_T = \frac{1}{2\xi_0^{3/2}T^2} \int_0^T (T-t)\mu(t)dt \quad (9.5)$$

$$\mathcal{R}_T = \frac{\int_0^T \mu(t)dt}{\int_0^T (1-\frac{t}{T})\mu(t)dt} \quad (9.6)$$

These expressions for  $\mathcal{S}_T$  and  $\mathcal{R}_T$  only involve the spot/variance covariance function  $\mu$  – can we derive general properties without assuming a specific form for  $\mu(t)$ ?

Let us make the natural assumption that  $\mu(t)$  decays monotonically towards zero as  $t \rightarrow \infty$ . Define  $g(\tau)$  as:

$$g(\tau) = \int_0^\tau \mu(t)dt \quad (9.7)$$

$\mathcal{S}_T$  and  $\mathcal{R}_T$  can be rewritten as:

$$\mathcal{S}_T = \frac{1}{2\xi_0^{3/2}T^2} \int_0^T g(\tau)d\tau \quad \mathcal{R}_T = \frac{g(T)}{\frac{1}{T} \int_0^T g(\tau)d\tau} \quad (9.8)$$

$g(\tau) = 0$  for  $\tau = 0$  and is either increasing concave if  $\mu(t) \geq 0$ , or decreasing convex if  $\mu(t) \leq 0$ .

- $\mathcal{R}_T$  is the ratio of  $g(T)$  to its average value over  $[0, T]$ , thus  $\mathcal{R}_T \geq 1$ .
- $g(\tau)$  is either positive and concave, or negative and convex. Thus,  $\frac{g(\tau)}{g(T)} \geq \frac{\tau}{T}$ . This yields:  $\mathcal{R}_T = \frac{1}{\frac{1}{T} \int_0^T \frac{g(\tau)}{g(T)} d\tau} \leq \frac{1}{\frac{1}{T} \int_0^T \frac{\tau}{T} d\tau} = 2$ .

Thus, for a time-homogeneous model such that its spot/variance covariance function decays monotonically, and for a flat term structure of VS volatilities, we get, at order one in volatility of volatility, the following model-independent range for  $\mathcal{R}_T$ :

$$\mathcal{R}_T \in [1, 2] \quad (9.9)$$

In diffusive stochastic volatility models – with the assumptions we have made – the SSR cannot go below 1. To get lower values for  $\mathcal{R}_T$  one presumably needs to incorporate a jump or Lévy component in the process for  $\ln S_t$ .

(9.9) strictly holds for a flat term structure of VS volatilities. Glancing again at the definition of the SSR in (9.3) we can see that the numerator is proportional to the

instantaneous volatility of  $S_t$  while the denominator is proportional to its square, thus the SSR is inversely proportional to the short VS volatility.

It can be made artificially small or large by shifting the short end of the variance curve – one should bear this in mind when assessing the SSR of a given market smile.

An example of the impact of the term structure of VS volatilities is discussed on page 366.

For  $T \rightarrow 0$ , assuming that  $\mu(t)$  is smooth as  $t \rightarrow 0$ ,  $g(\tau) = \tau\mu(0)$ : expression (9.8) for  $\mathcal{R}_T$  again yields:

$$\mathcal{R}_0 = 2$$


---

## 9.5 Scaling of ATMF skew and SSR – a classification of models

To investigate further the connection between  $\mathcal{S}_T$  and  $\mathcal{R}_T$  we need a characterization of the rate of decay of  $\mu(t)$ . Let us assume that for  $t \rightarrow \infty$   $\mu(t)$  decays with a characteristic exponent  $\gamma$ :

$$\mu(t) \propto \frac{1}{t^\gamma} \quad (9.10)$$

Consider  $g(\tau) = \int_0^\tau \mu(t)dt$ . For large  $\tau$ , it is equal to  $C + \alpha\tau^{1-\gamma}$ , where  $C, \alpha$  are constants. If  $\gamma > 1$  it tends towards  $C$  while for  $\gamma < 1$  it is equivalent to  $\alpha\tau^{1-\gamma}$ .

Now turn to  $\int_0^T g(\tau)d\tau$ : for large  $T$  it is equal to  $B + CT + \frac{\alpha}{2-\gamma}T^{2-\gamma}$ . For  $T \rightarrow \infty$  this quantity is equivalent to  $CT$  if  $\gamma > 1$  while it is equivalent to  $\frac{\alpha}{2-\gamma}T^{2-\gamma}$  if  $\gamma < 1$ . As a result we get, using formulas (9.8), two types of behavior for  $\mathcal{S}_T$  and  $\mathcal{R}_T$ , which leads to a division of stochastic volatility models into two classes.

For long maturities:

- (Type I) If  $\gamma > 1$ :

$$\mathcal{S}_T \propto \frac{1}{T} \text{ and } \lim_{T \rightarrow \infty} \mathcal{R}_T = 1 \quad (9.11)$$

- (Type II) If  $\gamma < 1$ :

$$\mathcal{S}_T \propto \frac{1}{T^\gamma} \text{ and } \lim_{T \rightarrow \infty} \mathcal{R}_T = 2 - \gamma \quad (9.12)$$

We leave it to the reader to check that exponential decay of  $\mu(t)$  produces Type I behavior.

Both Type I and Type II scalings are compactly summarized in the following relationship. For  $T \rightarrow \infty$ :

$$\mathcal{S}_T \propto \frac{1}{T^{2-\mathcal{R}_\infty}} \quad (9.13)$$

## 9.6 Type I models – the Heston model

The fact that for a fast-decaying spot/variance covariance function  $S_T$  decays like  $\frac{1}{T}$  is not unexpected. Indeed in this case, in the limit of long time intervals, increments of  $\ln S_t$  become independent and identically distributed. Cumulants of  $\ln S_t$  then scale linearly with  $T$ , thus the skewness  $s_T$  of  $\ln S_T$  scales like  $\frac{1}{\sqrt{T}}$ . Consequently, at order one in volatility of volatility, or equivalently at order one in  $s_T$ , the skew/skewness relationship (8.23) implies that the ATMF skew scales like  $\frac{1}{T}^2$ .<sup>2</sup>

The Heston model provides an example of Type I behavior.  $\mu(t)$  is exponentially decaying – see expressions (8.48) – hence we expect Type I scaling (9.35). Let us verify this by using expressions (6.19) for  $\widehat{\sigma}_{F_T T}$  and  $S_T$  at order one in volatility of volatility for long maturities.

The ATMF skew is given by (6.19a):

$$S_T = \frac{\rho\sigma}{2kT} \frac{1}{\sqrt{V^0}} \quad (9.14)$$

$S_T$  indeed has Type I scaling. Consider now the SSR.  $\widehat{\sigma}_{F_T T}$  in (6.19a) is a function of  $V$  – its covariance with  $d \ln S$  is thus given by:

$$\begin{aligned} \frac{E[d\widehat{\sigma}_{F_T T} d \ln S_t]}{E[(d \ln S_t)^2]} &= \frac{\partial \widehat{\sigma}_{F_T T}}{\partial V} \frac{E[dV d \ln S]}{V dt} \\ &= \frac{1}{2kT\sqrt{V^0}} \frac{E[dV d \ln S]}{V dt} \\ &= \frac{1}{2kT\sqrt{V^0}} \rho\sigma \end{aligned} \quad (9.15)$$

where in  $\frac{\partial \widehat{\sigma}_{F_T T}}{\partial V}$  we have kept the contribution at zeroth order in volatility of volatility to get the covariance of  $d\widehat{\sigma}_{F_T T}$  and  $d \ln S_t$  at order one. Dividing (9.14) by (9.15) yields  $\mathcal{R}_T = 1$ : we have confirmed by hand that the Heston model is indeed of Type I – this was already observed in [8].

<sup>2</sup>This  $\frac{1}{T}$  scaling of the ATMF skew is also shared by jump-diffusion and Lévy models, at order one in the skewness of  $\ln S_T$ , as in these models, increments of  $\ln S_t$  are indeed independent and identically distributed. The skewness of  $\ln S_T$  exactly scales like  $\frac{1}{\sqrt{T}}$ , and at order one in this skewness, the skew scales like  $\frac{1}{T}$ .

However, while  $\lim_{T \rightarrow \infty} \mathcal{R}_T = 1$  in Type I models, the behavior of  $\mathcal{R}_T$  in jump/Lévy models is different. Because implied volatilities are a function of  $K/S$  only,  $\mathcal{R}_T = 0$ ,  $\forall T$ .

## 9.7 Type II models

In Type I models, the long-maturity scaling of the ATMF skew and the limit of the SSR are fixed and do not depend on  $\gamma$ . Depending on the size of  $\gamma$ , the long-maturity regime sets in for shorter or longer maturities, but the limiting behavior of the ATMF skew and of the SSR is universal and bears no trace of the precise underlying dynamics.

Type II models are richer as both the scaling of  $\mathcal{S}_T$  and the limit of  $\mathcal{R}_T$  are non-trivial and reflect the characteristic exponent of the decay of the spot/variance covariance function  $\mu(t)$ .

Moreover, the ATMF skew of market smiles typically decays algebraically, with an exponent around  $-\frac{1}{2}$ . This calls for a Type II model; hence the following natural questions:

- Is it practically possible to build a Type II model?
- Is the *dynamics* of market smiles consistent with Type II?

### Type II scaling in the two-factor model

Consider an  $N$ -factor model of the type studied in Section 7.3:

$$\begin{cases} dS_t = (r - q)S_t dt + \sqrt{\xi_t^S} S_t dW_t^S \\ d\xi_t^T = \omega \xi_t^T \sum_i w_i e^{-k_i(T-t)} dW_t^i \end{cases}$$

Assuming a flat term structure of VS volatilities the model is time-homogeneous and the spot/variance covariance function reads:

$$\mu(\tau) = \omega \xi_0^{\frac{3}{2}} \sum_i w_i \rho_{iS} e^{-k_i \tau}$$

where  $\rho_{iS}$  is the correlation between  $W^S$  and  $W^i$ . Inserting this expression into formulas (9.5) and (9.6) yields the following expressions for the ATMF skew and the SSR:

$$\mathcal{S}_T = \frac{\omega}{2} \sum_i w_i \rho_{iS} \frac{k_i T - (1 - e^{-k_i T})}{(k_i T)^2} \quad (9.16a)$$

$$\mathcal{R}_T = \frac{\sum_i w_i \rho_{iS} \frac{1 - e^{-k_i T}}{k_i T}}{\sum_i w_i \rho_{iS} \frac{k_i T - (1 - e^{-k_i T})}{(k_i T)^2}} \quad (9.16b)$$

$\mu(\tau)$  is a linear combination of exponentials. As  $\tau \rightarrow \infty$ ,  $\mu(\tau) \propto e^{-\min_i k_i T}$ . Thus for  $T \rightarrow \infty$  this model is of type I:  $\mathcal{S}_T \propto \frac{1}{T}$  and  $\lim_{T \rightarrow \infty} \mathcal{R}_T = 1$ ; this can be

checked explicitly on expressions (9.16). The fact that  $\mu(\tau)$  is a linear combination of exponentials is dictated by the property that this form allows for a Markov representation of the model with a number of state variables – besides  $S_t$  – equal to the number of exponentials – see Section 7.1.

We could try and use a model whose dynamics for forward variances reads:

$$d\xi_t^T = \omega \xi_t^T \frac{1}{(T-t+\theta)^\gamma} dW \quad (9.17)$$

where  $\theta$  is a (small) offset parameter. In this model  $\mu(\tau)$  has the desired power law scaling for large  $\tau$ :  $\mu(\tau) \propto \frac{1}{\tau^\gamma}$ . Unfortunately we lose the Markov representation: it is no longer possible to express the set of forward variances  $\xi^T$  as a function of a finite number of state variables. Such a model is not usable in practice.

Luckily, by suitably choosing model parameters it is possible to get Type II scaling in an  $N$ -factor model – in fact a two-factor model – over a range of maturities that is sufficient for practical purposes, even though Type I scaling eventually kicks in for (very) long maturities.

Let us use the two-factor model of Section 8.7 with the parameters listed in Table 8.2, page 329. Remember that (a)  $\nu, \theta, k_1, k_2, \rho$  have been selected so that the volatilities of VS volatilities decay approximately as a power law with exponent 0.4, (b)  $\rho_{SX^1}$  and  $\rho_{SX^2}$  have been chosen so that the ATMF skew decays approximately as a power law with exponent  $\frac{1}{2}$ . This is illustrated in Figure 8.5.

At order one in volatility of volatility the ATMF skew is given by expression (8.54), page 329. The SSR is calculated using expression (9.4) and we have:

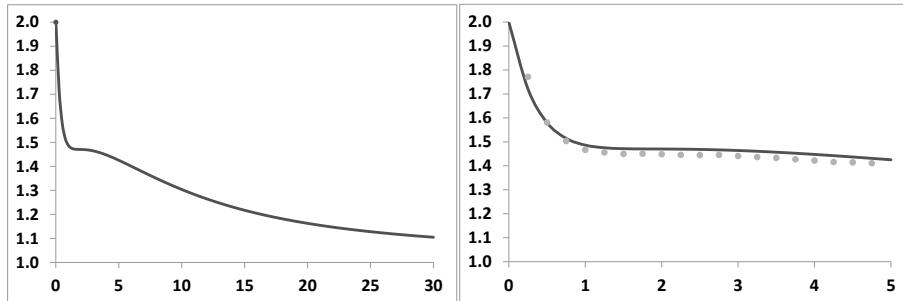
$$\mathcal{S}_T = \frac{\nu \alpha_\theta}{\hat{\sigma}_T^3 T^2} \int_0^T dt \sqrt{\xi_0^t} \int_t^T \xi_0^u \left[ (1-\theta) \rho_{SX^1} e^{-k_1(u-t)} + \theta \rho_{SX^2} e^{-k_2(u-t)} \right] du \quad (9.18)$$

$$\mathcal{R}_T = \frac{1}{\sqrt{\xi_0^0}} \frac{\hat{\sigma}_T^2 T \int_0^T \xi_0^t \left[ (1-\theta) \rho_{SX^1} e^{-k_1 t} + \theta \rho_{SX^2} e^{-k_2 t} \right] dt}{\sqrt{\xi_0^T} \left( \int_0^T dt \sqrt{\xi_0^t} \left[ (1-\theta) \rho_{SX^1} e^{-k_1(u-t)} + \theta \rho_{SX^2} e^{-k_2(u-t)} \right] du \right)} \quad (9.19)$$

For a flat term structure of VS volatilities:

$$\mathcal{S}_T = \nu \alpha_\theta \left[ (1-\theta) \rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta \rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2} \right] \quad (9.20)$$

$$\mathcal{R}_T = \frac{(1-\theta) \rho_{SX^1} \frac{1-e^{-k_1 T}}{k_1 T} + \theta \rho_{SX^2} \frac{1-e^{-k_2 T}}{k_2 T}}{(1-\theta) \rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta \rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2}} \quad (9.21)$$



**Figure 9.1:** The SSR of the two-factor model, computed with formula (9.20) and parameters in Table 8.2, page 329. Left: maturities up to 30 years – right: zoom on maturities less than 5 years. The dots are the result of a Monte Carlo evaluation of the SSR – see Section 9.8 below.

$\mathcal{R}_T$ , computed using (9.21), appears in Figure 9.1, as a function of  $T$ .

Consider first the left-hand graph. The short-maturity limit of the SSR is 2, as shown in Section 9.3. As  $T \rightarrow \infty$ , the SSR tends to 1 – as it should. Notice however the shoulder for a value of the SSR around 1.5, which appears in more detail in the right-hand graph. The SSR for maturities from 1 to 5 years is stable around 1.5. For this range of maturities, the model obeys the Type II scaling rules in (9.36): the value of the SSR is 2 minus the characteristic exponent of the decay of the ATMF skew – in our case  $\frac{1}{2}$ .

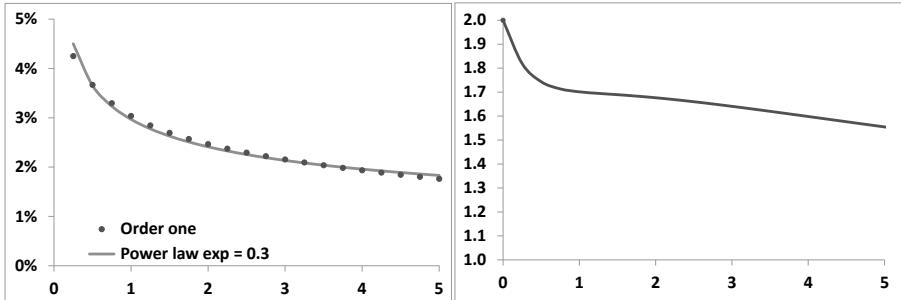
Note in the right-hand graph how the approximate expression (9.16b), obtained at order one in volatility of volatility, agrees well with the actual value of the SSR.

Another example is shown in Figure 9.2: here we have chosen  $\rho_{SX^1}$ ,  $\rho_{SX^2}$  so that the ATMF skew decays approximately with a characteristic exponent equal to 0.3 for maturities up to 5 years (left-hand graph). As the right-hand side graph shows, we get an SSR around 1.7, i.e.  $2 - 0.3$  over this range.<sup>3</sup>

Formula (9.16b) for the SSR has been derived for a flat term structure of VS volatilities. When this is not the case, the prefactor  $\xi^u \sqrt{\xi^t}$  in expression (8.50), page 327, for  $\mu(t, u, \xi)$  in the two-factor model is not constant and, at order one in volatility of volatility, both the numerator and denominator of  $\mathcal{R}_T$  have to be calculated by numerical integration.

The SSR thus depends on the term structure of VS volatilities. For example, for a VS term structure that increases (decreases) from 20% to 25% for  $T = 1$  year, the SSR for this maturity is 1.46 (1.50). Let us take  $T = 5$  years; for a VS term structure that increases (decreases) from 20% to 30%, the SSR is 1.43 (1.49). These values

<sup>3</sup>With parameters  $\nu, \theta, k_1, k_2, \rho$  already set so as to generate the desired term structure of volatilities of VS volatilities,  $\rho_{SX^1}, \rho_{SX^2}$  are the only handles left to control the term structure of (a) the ATMF skew, (b) the SSR. Introducing more factors allows for more flexibility, but in the author's experience two factors are sufficient for capturing typical market ATMF skews and SSR.



**Figure 9.2:** Left: The ATM skew measured as the difference of implied volatilities for strikes  $0.95F_T$  and  $1.05F_T$  at order one in volatility of volatility (dots, formula (9.16a)) and a power law benchmark with exponent 0.3 (line). Right: the SSR at order one for maturities up to 5 years computed with expression (9.16b). The parameters are those of Table 8.1, with  $\rho_{SX^1} = -35\%$ ,  $\rho_{SX^2} = -83\%$ .

are obtained in a Monte Carlo simulation of the two-factor model – see Section 9.8 below – with the same parameters as in Figure 9.1. For all practical purposes, the accuracy of formula (9.16b) for  $\mathcal{R}_T$  is thus adequate as long as VS term structures are not too steep.

The definition of the SSR in (9.3) involves in the denominator the instantaneous variance of  $S$  at  $t = 0$ . In case short VS volatilities are, say, substantially larger than longer-dated ones, we get a smaller value for the SSR than that given by expression (9.16b). For the purpose of evaluating the SSR at  $t = 0$ , approximation (9.16b) is thus incorrect.

However, it is the average level of spot/volatility covariance generated by the model up to the maturity of our exotic option, rather than its particular value at  $t = 0$ , that is practically relevant. It is then advisable to replace the short-dated variance in the denominator of (9.3) with the square of the VS volatility for maturity  $T$ . Approximation (9.16b) is then almost exact.

In conclusion, while the two-factor model is strictly speaking a Type I model, by suitably choosing parameters we obtain Type II scaling for a range of maturities that is practically relevant. For these maturities we are able to generate a decay of the ATM skew with an exponent less than 1 and the relationship between  $\mathcal{S}_T$  and  $\mathcal{R}_T$  in (9.13) is approximately obeyed.

### Type II scaling in reality

We are able to obtain Type II scaling in a two-factor model, but are actual market smiles consistent with Type II behavior? This is characterized in (9.36) by two features: one static, one dynamic:

- The ATM skew decays with a non-trivial exponent:  $\mathcal{S}_T \propto \frac{1}{T^\gamma}$  with  $\gamma < 1$

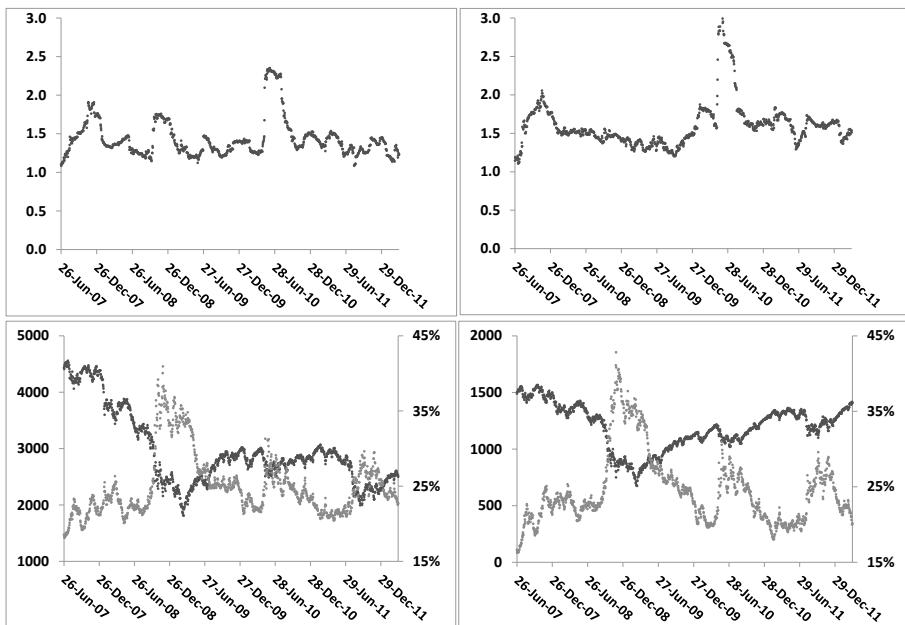
- The SSR is different than 1 and is related to  $\gamma$ : for long maturities  $\mathcal{R}_T \rightarrow 2 - \gamma$

The decay of the ATM skew of market smiles is consistent with Type II, but what about the SSR? From the expression of the SSR in (9.3) we define the realized SSR as:

$$\mathcal{R}_T^r = \frac{\sum_i \ln \frac{S_{i+1}}{S_i} (\hat{\sigma}_{T,i+1} - \hat{\sigma}_{T,i})}{\sum_i S_{T,i} (\ln \frac{S_{i+1}}{S_i})^2} \quad (9.22)$$

where  $\hat{\sigma}_{T,i}$  (resp.  $S_{T,i}$ ) is the ATM implied volatility (resp. ATMF skew) at time  $i$  for residual maturity  $T$ .

$\mathcal{R}_T^r$  for  $T = 2$  years is shown in Figure 9.3 for the Euro Stoxx 50 and S&P 500 indexes. We have used for simplicity ATM rather than ATMF volatilities.



**Figure 9.3:** Top: 3-months realized SSR ( $\mathcal{R}_T^r$ ) for the Euro Stoxx 50 (left) and S&P 500 (right) indexes, for  $T = 2$  years. Bottom: the underlying (dark dots, left axis) and 2-year ATM implied volatility (lighter dots, right axis).

For both indexes, the realized SSR hovers around 1.5, with occasional spikes generated by simultaneous and opposite moves of  $S$  and  $\hat{\sigma}_{F_T(S)T}$ . The magnitude of the spikes and of the fluctuations around 1.5 is not relevant; it would be larger (resp. smaller) had we estimated  $\mathcal{R}_T^r$  on a shorter (resp. longer) window than 3 months – the duration of the spikes is indeed about 3 months. Notice also how similar ATM implied volatilities are for both indexes. Realized values for the SSR for longer-dated maturities – say  $T = 5$  years – are comparable to the 2-year case.

In conclusion, the dynamics of equity market smiles is consistent with Type II behavior:  $\mathcal{S}_T \propto \frac{1}{T^\gamma}$  with  $\gamma < 1$  and for large  $T$ ,  $\gamma + \mathcal{R}_T$  is approximately equal to 2.

---

## 9.8 Numerical evaluation of the SSR

At order one in volatility of volatility, for the sake of computing the SSR the ATMF volatility can be substituted with the VS volatility and this leads to the analytic expression (9.16a) for a flat term structure of VS volatilities.

Generally, however, the SSR can be easily computed numerically. In the two-factor model, the ATMF volatility is a function of  $X^1, X^2$ . At  $t = 0$ :

$$\widehat{\sigma}_{F_T T} \equiv \widehat{\sigma}_{F_T T}(X_0^1, X_0^2)$$

Expanding at first order in  $dX^1, dX^2$ :

$$d\widehat{\sigma}_{F_T T} = \frac{d\widehat{\sigma}_{F_T T}}{dX^1} dX^1 + \frac{d\widehat{\sigma}_{F_T T}}{dX^2} dX^2$$

From the definition of the SSR in (9.3) and using that  $E[(d \ln S_t)^2] = \xi_0^0 dt$ ,  $E[d \ln S_t dX_t^1] = \rho_{SX^1} \sqrt{\xi_0^0} dt$  and likewise for  $X^2$ , we get:

$$\mathcal{R}_T = \frac{1}{\mathcal{S}_T} \frac{E[d \ln S_t d\widehat{\sigma}_{F_T T}]}{E[(d \ln S_t)^2]} = \frac{1}{\mathcal{S}_T} \frac{1}{\sqrt{\xi_0^0}} \left( \frac{d\widehat{\sigma}_{F_T T}}{dX^1} \rho_{SX^1} + \frac{d\widehat{\sigma}_{F_T T}}{dX^2} \rho_{SX^2} \right)$$

Thus  $\mathcal{R}_T$  can be simply evaluated numerically by computing  $\widehat{\sigma}_{F_T T}$  with two different initial values for  $(X, Y)$ :

$$\mathcal{R}_T \simeq \frac{1}{\mathcal{S}_T} \frac{1}{\sqrt{\xi_0^0}} \frac{\widehat{\sigma}_{F_T T}(X_0^1 + \varepsilon \rho_{SX^1}, X_0^2 + \varepsilon \rho_{SX^2}) - \widehat{\sigma}_{F_T T}(X_0^1, X_0^2)}{\varepsilon}$$

with  $\varepsilon$  a small offset. Typically we take  $X_0^2 = X_0^1 = 0$ .

In stochastic volatility models defined by the dynamics of the instantaneous variance  $V_t$ , the SSR is simply computed by shifting  $V$  as volatilities for a fixed moneyness are a function of  $V$ . For example, in the Heston model, using the notations of Chapter 6,  $\mathcal{R}_T$  is simply given by:

$$\mathcal{R}_T \simeq \frac{1}{\mathcal{S}_T} \frac{\widehat{\sigma}_{F_T T}(V + \varepsilon \rho \sigma) - \widehat{\sigma}_{F_T T}(V)}{\varepsilon}$$


---

## 9.9 The SSR for short maturities

In both Type I and Type II models the short-maturity limit of the SSR is 2.

In a stochastic volatility model, in the limit  $T \rightarrow 0$ , the VS and ATM volatilities are identical. We will thus use the notation  $\hat{\sigma}_0$  for both.

From the expression of the SSR in (9.3), in the special case of vanishing maturities, we can derive two definitions of the *realized* SSR according to whether we choose:

- the realized value of the instantaneous variance  $(\ln \frac{S_{i+1}}{S_i})^2$
- or the implied value of the realized variance  $\hat{\sigma}_{T,i}^2 \Delta t$

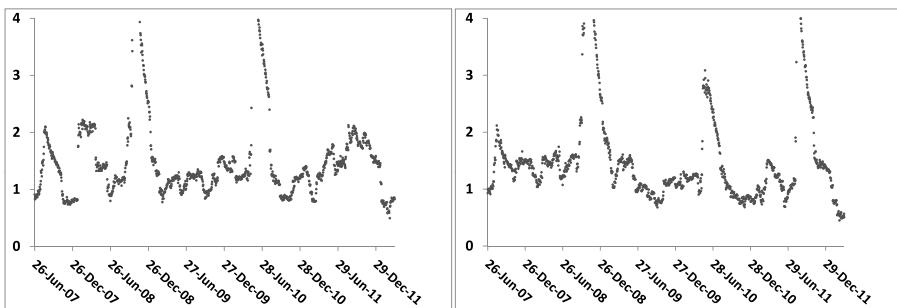
in the denominator of (9.3).

For reasons that will be clearer shortly, we choose the second convention. For small  $T$  we define  $\mathcal{R}_T^{r,\text{short}}$  as:

$$\mathcal{R}_T^{r,\text{short}} = \frac{\sum_i \ln \frac{S_{i+1}}{S_i} (\hat{\sigma}_{T,i+1} - \hat{\sigma}_{T,i})}{\Delta t \sum_i S_{T,i} \hat{\sigma}_{T,i}^2} \quad (9.23)$$

where  $\Delta t$  is the duration of one (trading) day.

$\mathcal{R}_T^r$  for  $T = 1$  month is shown in Figure 9.4 for the Euro Stoxx 50 and S&P 500 indexes.



**Figure 9.4:** 3-months realized SSR ( $\mathcal{R}_T^r$ ) for the Euro Stoxx 50 (left) and S&P 500 (right) indexes, for  $T = 1$  month.

The average value of  $\mathcal{R}_T^r$  is substantially lower than the model-independent “implied” value of 2. Does this point to a discrepancy between the dynamics of stochastic models and the dynamics of market smiles?

Consider the example of realized and implied volatilities – typically, for equity indexes, VS implied volatilities are systematically higher on average than realized volatilities. This in itself does not signal a modeling inconsistency. What is important is that the difference between both volatilities can be materialized as the P&L of an option position – here a VS.

The natural question is thus: is it possible to materialize the difference  $2 - \mathcal{R}_T^r$  as the P&L of an option position?

## 9.10 Arbitraging the realized short SSR

The SSR expresses the implied spot/volatility covariance in units of the ATM skew. Arbitraging  $\mathcal{R}_T^{r,\text{short}}$  entails being able to materialize this covariance as a P&L. In what follows we set  $T = 1$  month and assume zero interest rate and repo for simplicity.

To single out the realized spot/volatility covariance as a carry P&L we risk-manage our option position using the lognormal model of Section 8.5.1. In the limit  $T \rightarrow 0$  the variance curve collapses to a single volatility  $\hat{\sigma}_0$ .

Since any stochastic volatility model calibrated to the market smile yields the same implied value for the instantaneous covariance of  $S$  and  $\hat{\sigma}_0$  we could use any dynamics for  $\hat{\sigma}_0$ . However, in the lognormal model the short ATM skew is independent on  $\hat{\sigma}_0$ , a property that is also approximately shared by market smiles.

Choosing the lognormal model for risk-managing our option position then reduces chances that the mark-to-market P&L generated by remarking the ATM skew to market is large.

### 9.10.1 Risk-managing with the lognormal model

Consider the limit  $T \rightarrow 0$  in the pricing equation (8.1):  $P$  becomes a function of  $t, S, \hat{\sigma}_0$ :  $P(t, S, \xi) \equiv P(t, S, \hat{\sigma}_0)$ . In the lognormal model for the instantaneous volatility the pricing equation is given by:

$$\frac{dP}{dt} + \frac{\hat{\sigma}_0^2}{2} S^2 \frac{d^2 P}{dS^2} + \frac{\nu^2}{2} \hat{\sigma}_0^2 \frac{d^2 P}{d\hat{\sigma}_0^2} + \rho\nu\hat{\sigma}_0^2 S \frac{d^2 P}{dS d\hat{\sigma}_0} = 0 \quad (9.24)$$

where  $\nu$  is the (lognormal) volatility of  $\hat{\sigma}_0$  and  $\rho$  the correlation of  $\hat{\sigma}_0$  with  $S$ .

We have derived in Section 8.5.1 the smile at order two in volatility of volatility, in the limit  $T \rightarrow 0$ . At this order, expressions (8.39a) and (8.39b) for the ATM skew and curvature yield the following expression for the price of a vanilla option of strike  $K$ , maturity  $T$ :

$$P = P_{BS}(t, S, \hat{\sigma}(x), K, T) \quad (9.25a)$$

$$\hat{\sigma}(x) = \hat{\sigma}_0 + \mathcal{S}x + \frac{\mathcal{C}}{2}x^2 \quad (9.25b)$$

where  $P_{BS}$  is the Black-Scholes formula,  $x = \ln(\frac{K}{S})$  and  $\mathcal{S}, \mathcal{C}$  are given by:

$$\mathcal{S} = \frac{\rho\nu}{2} \quad \mathcal{C} = \frac{2 - 3\rho^2}{6\hat{\sigma}_0}\nu^2$$

For a market smile of the form (9.25b)  $\rho, \nu$  are given by:

$$\rho = \frac{2\mathcal{S}}{\sqrt{3\hat{\sigma}_0\mathcal{C} + 6\mathcal{S}^2}} \quad \nu = \sqrt{3\hat{\sigma}_0\mathcal{C} + 6\mathcal{S}^2} \quad (9.26)$$

In the limit  $T \rightarrow 0$ ,  $\mathcal{S}$  and  $\mathcal{C}$ , derived at order two in volatility of volatility, do not depend on  $T$  and the ATM implied volatility is identical to the VS volatility. This identity is exact; indeed consider a diffusive model whose instantaneous volatility at  $t = 0$  is  $\sigma$ . Using expression (8.64) with  $\widehat{\sigma} = 0$ , the price of a European option whose payoff is  $f(S)$  is given, at order one in  $T$  by:  $P = f(S_0) + \frac{\sigma^2 T}{2} S_0^2 \frac{d^2 f}{dS_0^2}$  where  $S_0$  is the spot value at  $t = 0$ . Denote now by  $\widehat{\sigma}$  the implied volatility of this option. At order one in  $T$ ,  $\widehat{\sigma}$  is such that  $P = f(S_0) + \frac{\widehat{\sigma}^2 T}{2} S_0^2 \frac{d^2 f}{dS_0^2}$ . This yields  $\widehat{\sigma} = \sigma$ : the implied volatilities of all European payoffs  $f$  such that  $\frac{d^2 f}{dS_0^2} \neq 0$  are identical, and equal to the instantaneous volatility.<sup>4</sup>

Consider now a long position in a delta-hedged European option, risk-managed with the lognormal model. We will calibrate  $\mathcal{S}, \mathcal{C}$  – or equivalently  $\rho, \nu$  – to the market smile near the money on a daily basis. Denote by  $\Pi(t, S, \widehat{\sigma}_0, \mathcal{S}, \mathcal{C})$  the value of our European option position. We do not consider for now the mark-to-market P&L generated by a change of  $\mathcal{S}, \mathcal{C}$  – or equivalently of  $\rho, \nu$  – and instead focus on the carry P&L. This P&L during  $\delta t$ , at order one in  $\delta t$  and two in  $\delta S$  and  $\widehat{\sigma}_0$ , is given by:

$$P\&L = \frac{d\Pi}{dt} \delta t + \frac{d\Pi}{d\widehat{\sigma}_0} \delta \widehat{\sigma}_0 + \frac{1}{2} \frac{d^2\Pi}{dS^2} \delta S^2 + \frac{1}{2} \frac{d^2\Pi}{d\widehat{\sigma}_0^2} (\delta \widehat{\sigma}_0)^2 + \frac{d^2\Pi}{dS d\widehat{\sigma}_0} \delta S \delta \widehat{\sigma}_0 \quad (9.27)$$

While the delta hedge removes the term in  $\delta S$ , there remains a contribution in  $\delta \widehat{\sigma}_0$  as we are not vega-hedged.

Using (9.24), at order one in  $\delta t$  and two in  $\delta S$  and  $\widehat{\sigma}_0$  our P&L is given by:

$$P\&L = \frac{d\Pi}{d\widehat{\sigma}_0} \delta \widehat{\sigma}_0 \quad (9.28a)$$

$$+ \frac{1}{2} S^2 \frac{d^2\Pi}{dS^2} \left( \left( \frac{\delta S}{S} \right)^2 - \widehat{\sigma}_0^2 \delta t \right) + \frac{1}{2} \widehat{\sigma}_0^2 \frac{d^2\Pi}{d\widehat{\sigma}_0^2} \left( \left( \frac{\delta \widehat{\sigma}_0}{\widehat{\sigma}_0} \right)^2 - \nu^2 \delta t \right) \quad (9.28b)$$

$$+ S \widehat{\sigma}_0 \frac{d^2\Pi}{dS d\widehat{\sigma}_0} \left( \frac{\delta S}{S} \frac{\delta \widehat{\sigma}_0}{\widehat{\sigma}_0} - \rho \nu \widehat{\sigma}_0 \delta t \right) \quad (9.28c)$$

From (9.25b) the ATMF skew  $\mathcal{S}$  is  $\frac{\rho\nu}{2}$  – this is equivalent to the statement  $\mathcal{R}_{T=0} = 2$ . The last piece (9.28c) can thus be rewritten as:

$$S \widehat{\sigma}_0 \frac{d^2\Pi}{dS d\widehat{\sigma}_0} \left( \frac{\delta S}{S} \frac{\delta \widehat{\sigma}_0}{\widehat{\sigma}_0} - \rho \nu \widehat{\sigma}_0 \delta t \right) = S \widehat{\sigma}_0 \frac{d^2\Pi}{dS d\widehat{\sigma}_0} \left( \frac{\delta S}{S} \frac{\delta \widehat{\sigma}_0}{\widehat{\sigma}_0} - 2S \widehat{\sigma}_0 \delta t \right) \\ = S \widehat{\sigma}_0^2 \mathcal{S} \frac{d^2\Pi}{dS d\widehat{\sigma}_0} \left( \frac{\delta S}{S} \frac{\delta \widehat{\sigma}_0}{\widehat{\sigma}_0} - 2 \right) \delta t \quad (9.29a)$$

$$= S \widehat{\sigma}_0^2 \mathcal{S} \frac{d^2\Pi}{dS d\widehat{\sigma}_0} \left( \mathcal{R}_{T=0}^{r,\text{short}} - 2 \right) \delta t \quad (9.29b)$$

---

<sup>4</sup>This is not the case for vanilla options with strikes different than  $S_0$  as for these options  $\frac{d^2 f}{dS_0^2} = 0$  – the contribution of order one in  $T$  vanishes.

From (9.29a) it is apparent that, in defining the realized SSR, the short implied – rather than the realized – variance should be used in the denominator, in the definition of  $\mathcal{R}_T^{r,\text{short}}$ , hence expression (9.23) for  $\mathcal{R}_T^{r,\text{short}}$ .

### 9.10.2 The realized skew

The P&L above can be equivalently expressed as a difference between the implied skew  $\mathcal{S}$  and the instantaneous *realized* skew  $\mathcal{S}^r$  which we define as:

$$\mathcal{S}^r = \frac{1}{2\hat{\sigma}_0\delta t} \frac{\delta S}{S} \frac{\delta\hat{\sigma}_0}{\hat{\sigma}_0} = \left( \frac{\mathcal{R}_{T=0}^{r,\text{short}}}{2} \right) \mathcal{S} \quad (9.30)$$

Expression (9.30) shows that the relative mismatch of realized to implied skew is equal to that of realized to implied SSR. The cross gamma/theta P&L 9.29b reads:

$$2S\hat{\sigma}_0^2 \frac{d^2\Pi}{dSd\hat{\sigma}_0} (\mathcal{S}^r - \mathcal{S}) \delta t \quad (9.31)$$

Note that the P&L in (9.28) is not the total P&L incurred on our option position. Additional P&L is generated by daily recalibration of  $\mathcal{S}$  and  $\mathcal{C}$ . Only if this P&L is small and if the contributions in (9.28a) and (9.28b) are vanishing or negligible are we able to isolate the P&L of interest (9.29b) or equivalently (9.31).

### 9.10.3 Splitting the theta into three pieces

Expression (9.25a) for  $P$  is correct at order two in volatility of volatility:  $P$  does not exactly solve (9.24). Can we still use (9.25a) for P&L accounting? (9.24) expresses that the theta  $\frac{dP}{dt}$  can be broken up in three pieces which match each of the second-order gamma contributions in (9.28).

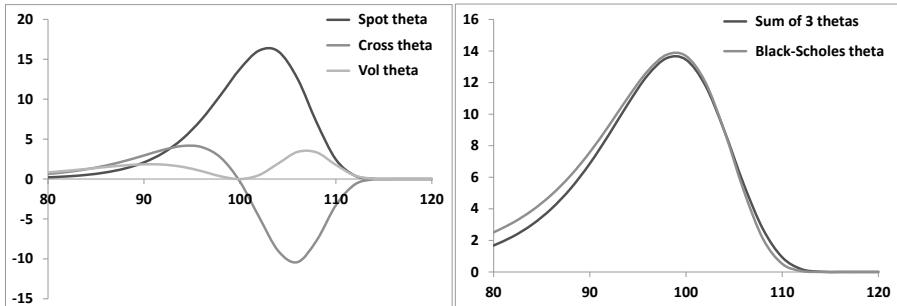
In our model theta is equal to the Black-Scholes theta since  $\hat{\sigma}(x)$  in (9.25b) does not depend on  $T$ . In the Black-Scholes model all of the theta is ascribed to the spot gamma – the second piece in the right-hand side of (9.28) – with a break-even level, the implied volatility, which is strike-dependent. In contrast, in the stochastic volatility model we are using, this theta is distributed across three gammas with break-even levels that are *not* strike-dependent. Checking whether (9.24) holds amounts to checking how well the following equality holds

$$\frac{\hat{\sigma}(x)^2}{2} S^2 \frac{d^2P_{BS}}{dS^2} = \frac{\hat{\sigma}_0^2}{2} S^2 \frac{d^2P}{dS^2} + \frac{\nu^2}{2} \hat{\sigma}_0^2 \frac{d^2P}{d\hat{\sigma}_0^2} + \rho\nu\hat{\sigma}_0^2 S \frac{d^2P}{dSd\hat{\sigma}_0} \quad (9.32)$$

where the three contributions in the right-hand side are called, respectively: spot theta, vol theta, cross theta.

Let us take the typical example of a one-month maturity smile with  $\hat{\sigma}_0 = 20\%$ ,  $\mathcal{S} = -0.7$ ,  $\mathcal{C} = 0.4$ . With these parameters, the implied volatilities of the 90%, 100%, 110% strikes are, respectively, 27.6%, 20%, 13.5%. From (9.26) we have:  $\rho = -78.5\%$ ,  $\nu = 178.3\%$ .

Figure 9.5 shows on the left the three thetas as a function of  $K$ , for  $S = 100$ . The sum of these three thetas, along with the Black-Scholes theta – the left-hand side of (9.32) where a different implied volatility is used for each strike – is shown on the right.



**Figure 9.5:** Left: the three terms in the right-hand side of (9.32) as a function of  $K$ , for  $S = 100$ . Right: the sum of the three thetas compared to the Black-Scholes theta  $\frac{\hat{\sigma}(x)^2}{2} S^2 \frac{d^2 P_{BS}}{dS^2}$ .

As we can see our approximation applied to the lognormal model does a decent job at splitting the Black-Scholes theta in three pieces, whose break-even levels are independent on the option's strike. The three thetas do not exactly add up to the Black-Scholes theta for strikes far out of the money, but the agreement is satisfactory for strikes between 95% and 105%, which is the range we use in our tests.

As is clear from Figure 9.5, the spot theta dominates by far – as our objective is to isolate the cross theta, we need to ensure that the spot theta of our position vanishes. The vol theta, on the other hand, is much smaller and almost cancels for a spread position.

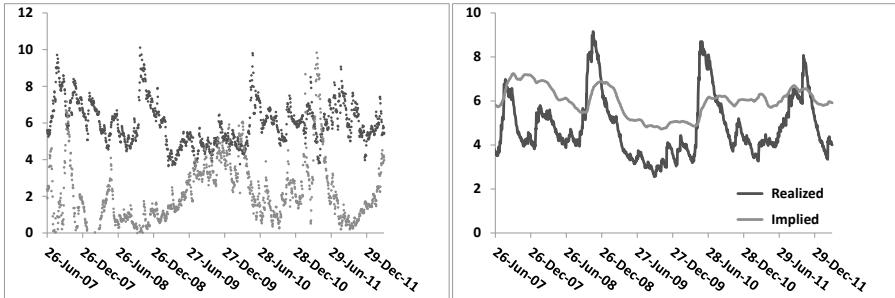
#### 9.10.4 Backtesting on the Euro Stoxx 50 index

We now use historical implied volatilities of the Euro Stoxx 50 index from April 2007 to March 2012 to backtest the following dynamical option trading strategy. We sell one-month options of strike 95% and buy the appropriate number of one-month options of strike 105% so that the spot gamma vanishes. We delta-hedge this position until the next (trading) day, when it is unwound and a new position is started. In order to approximately maintain a constant level of cross-gamma, we trade a constant notional – 100€ – of the 95%-strike option. For typical levels of 1-month market skew, for one 95% option sold, we need to buy about 0.5 options of strike 105%.

We use daily one-month implied volatilities for strikes 95%, 100%, 105% to determine the skew  $\mathcal{S}$  and curvature  $\mathcal{C}$  and back out  $\rho$  and  $\nu$  using (9.26). These, together with the ATM volatility  $\hat{\sigma}_0$ , are fed to the lognormal model which we use

(a) to determine the ratio of the 105% options to 95% options, (b) to compute the delta.

Calibrated values of  $\mathcal{S}$  (multiplied by  $-10$ ) and  $\mathcal{C}$  appear in the left-hand graph of Figure 9.6.  $\mathcal{C}$  is very noisy – indeed over the strike range [95%, 105%], the one-month smile is almost a straight line: in our tests we thus set the product  $\mathcal{C}$  equal to  $\frac{2}{\sigma_0}$  – the precise value of  $\mathcal{C}$  used hardly affects our results since we only use 95% and 105% strikes.



**Figure 9.6:** Left: daily values of skew  $\mathcal{S}$  (darker dots) and curvature  $\mathcal{C}$  (lighter dots) of one-month smiles of the Euro Stoxx 50 index. Right: 3-months exponentially weighted moving averages of the implied ATM skew  $\mathcal{S}$  and its realized counterpart  $\mathcal{S}^{\text{real}}$  for a one-month maturity – see (9.30). Values of  $\mathcal{S}$  in both graphs, and of  $\mathcal{S}^{\text{real}}$  have been multiplied by  $-10$  to correspond approximately to the 95%/105% skew in volatility points.

The right-hand graph shows the implied and realized 1-month skew for the Euro Stoxx 50 index. In our historical sample, the average value of the *implied* 95%/105% skew is 6 points of volatility, while the average value of its *realized* counterpart – defined in (9.30) – is 4.8 points; this is 20% less than its implied counterpart.

Consequently, from (9.30), the average value of the realized SSR is about 20% lower than its implied value of 2; this agrees with the average value of realized SSR in Figure 9.4.

Our daily P&L between day  $i$  and day  $i + 1$  is very plainly given by:

$$\begin{aligned} P\&L_{\text{Total}} &= [\Pi_i(t_{i+1}, S_{i+1}, \hat{\sigma}_{0i+1}, \mathcal{S}_{i+1}, \mathcal{C}_{i+1}) - \Pi_i(t_i, S_i, \hat{\sigma}_{0i}, \mathcal{S}_i, \mathcal{C}_i)] \\ &\quad - \frac{d\Pi_i}{dS_i}(S_{i+1} - S_i) \end{aligned}$$

where  $\Pi_i$  is the market value of the option portfolio purchased at time  $t_i$ . This P&L can be broken down in three contributions:

$$P\&L_{\text{Total}} = P\&L_{\text{Carry}}^{\text{vega-hedged}} + P\&L_{\text{Vega}} + P\&L_{\text{MtM}}^{\mathcal{S}, \mathcal{C}}$$

where:

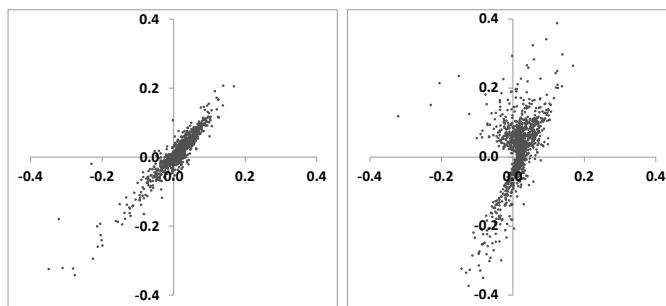
$$\begin{aligned} P\&L_{\text{Carry}}^{\text{vega-hedged}} &= [\Pi_i(t_{i+1}, S_{i+1}, \hat{\sigma}_{0i+1}, \mathcal{S}_i, \mathcal{C}_i) - \Pi_i(t_i, S_i, \hat{\sigma}_{0i}, \mathcal{S}_i, \mathcal{C}_i)] \\ &\quad - \frac{d\Pi_i}{dS_i}(S_{i+1} - S_i) - \frac{d\Pi_i}{d\hat{\sigma}_{0i}}(\hat{\sigma}_{0i+1} - \hat{\sigma}_{0i}) \\ P\&L_{\text{Vega}} &= \frac{d\Pi_i}{d\hat{\sigma}_{0i}}(\hat{\sigma}_{0i+1} - \hat{\sigma}_{0i}) \\ P\&L_{\text{MtM}}^{S,C} &= \Pi_i(t_{i+1}, S_{i+1}, \hat{\sigma}_{0i+1}, \mathcal{S}_{i+1}, \mathcal{C}_{i+1}) - \Pi_i(t_{i+1}, S_{i+1}, \hat{\sigma}_{0i+1}, \mathcal{S}_i, \mathcal{C}_i) \end{aligned}$$

$P\&L_{\text{MtM}}^{S,C}$  is the mark-to-market P&L generated by recalibrating  $\mathcal{S}$  and  $\mathcal{C}$  to the market smile at time  $t_{i+1}$ .

At order one in  $(t_{i+1} - t_i)$  and order two in  $(S_{i+1} - S_i)$  and  $(\hat{\sigma}_{0i+1} - \hat{\sigma}_{0i})$ ,  $P\&L_{\text{Carry}}^{\text{Vega-hedged}}$  is the sum of the three gamma/theta contributions in (9.28b) and (9.28c).

Our option position has vanishing spot gamma by construction. Moreover, inspection of the left-hand graph of Figure 9.5 suggests that the volatility gamma/theta P&L is small: ideally P&L (9.28b) will be negligible so that  $P\&L_{\text{Carry}}^{\text{Vega-hedged}}$  closely tracks the quantity of interest, that is the cross-gamma/theta P&L (9.28c) which we aim to single out. Our P&L is however polluted by the contributions of  $P\&L_{\text{MtM}}^{S,C}$  and  $P\&L_{\text{Vega}}$ .

How well  $(P\&L_{\text{Carry}}^{\text{vega-hedged}} + P\&L_{\text{MtM}}^{S,C})$ , that is  $(P\&L_{\text{Total}} - P\&L_{\text{Vega}})$  correlates to the cross-gamma/theta P&L is assessed in the left-hand graph of Figure 9.7.



**Figure 9.7:** Scatter plots of daily values of  $P\&L_{\text{Total}} - P\&L_{\text{Vega}}$  (left-hand graph) and  $P\&L_{\text{Total}}$  (right-hand graph) as a function of P&L (9.28c).

$P\&L_{\text{MtM}}^{S,C}$  contributes in fact little noise; the reason we do not get a straight line is mostly due to higher-order terms in  $P\&L_{\text{Carry}}^{\text{vega-hedged}}$ . Still, the total daily P&L corrected for the vega contribution captures the daily cross gamma/theta P&L with acceptable accuracy.

Inclusion of the vega contribution – see right-hand graph – reduces the correlation of both P&Ls but, as we will now see, the cumulative vega P&L is small enough that our strategy's P&L is still mostly attributable to the cross gamma/theta P&L.

Figure 9.8 shows the cumulative P&Ls of our arbitrage strategy. That the bulk of the P&L is contributed by the cross gamma/theta P&L is manifest.

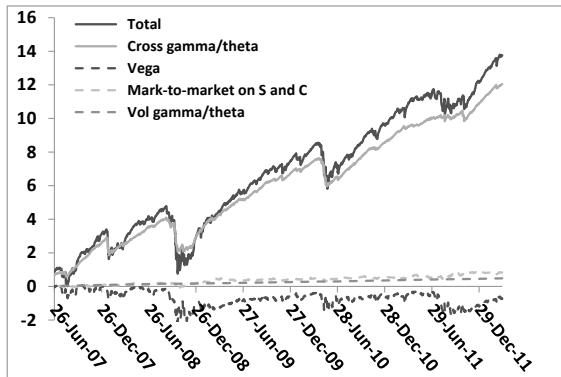


Figure 9.8: Cumulative P&Ls of the realized skew arbitrage strategy.

The volatility gamma/theta P&L, as well  $P\&L_{\text{MtM}}^{S,C}$  and  $P\&L_{\text{Vega}}$ , are both small, even though the latter is very noisy. As a rule, sharp market moves generate simultaneous downward moves of the spot and upward moves of implied volatilities. Our gamma-neutral position has negative vega. On these days, our option position loses money both because of our positive cross-gamma and negative vega positions: this is confirmed by the simultaneous drops in the cumulative cross gamma/theta and vega P&Ls in Figure 9.8. These drops indeed coincide with peaks of the realized skew seen in Figure 9.6.

Remember that in our historical sample the *realized* ATM skew is on average 20% lower than the *implied* ATM skew – or equivalently that the realized SSR is 20% lower than its model-independent value of 2. Our position has vanishing spot gamma; moreover Figure 9.8 shows that the volatility gamma/theta P&L is small. Thus, almost all of the theta is generated by the cross theta. The cumulative theta P&L of our strategy is 45.7€. 20% of this, that is about 9€ is a number that is indeed consistent with the cumulative cross gamma/theta P&L.

### 9.10.5 The “fair” ATMF skew

In conclusion, just as the difference between realized and implied volatility can be arbitrated – that is materialized as the P&L of an option strategy – for short maturities, the difference between the realized value of the SSR and its model-independent value of 2 – or equivalently the difference between the realized ATM skew and the implied ATM skew – can be materialized, to a good approximation, as the P&L of a dynamical option strategy.<sup>5</sup>

<sup>5</sup>In practice, we would presumably buy a little more 105% strike options in order to reduce our vega position. This would generate a positive spot gamma position – which helps in case of large market

There is then no inconsistency in the fact that  $\mathcal{R}^{\text{real}} < 2$ . The short ATM skew quantifies the *implied* value of the instantaneous spot/volatility covariance, which may differ from its *realized* value. In our sample, it is on average 20% lower than its *implied* counterpart. Our trading strategy approximately materializes this difference as a P&L.

We have shown that this is equivalent to materializing as a P&L the difference between implied skew and realized skew:

$$\mathcal{S}^r = \frac{1}{2\hat{\sigma}_0\delta t} \left\langle \frac{\delta S}{S} \frac{\delta\hat{\sigma}_0}{\hat{\sigma}_0} \right\rangle$$

Note that the fair level of the ATM skew is not determined by the covariance of the spot and *realized* volatility; rather it is given by the covariance of spot and ATM *implied* volatility, a circumstance that may surprise at first.

Imagine there is no options' market – our only hedge instrument is the underlying itself – and we are asked to quote “fair” vanilla option prices, “fair” meaning that we do not make or lose money on average. We then need to model the process of the instantaneous historical realized variance – or equivalently the process of expected future historical variances, that is forward variances. We use pricing equation (7.4), which expresses that, on average no money is made or lost as we delta-hedge our option. The only difference now is that  $\xi^\tau$  is no longer a market VS forward variance, but instead the expected future realized instantaneous variance.<sup>6</sup>

The results of Chapter 8 apply: at order one in volatility of volatility the “fair” ATM skew is given by the weighted integral of the *historical* covariance of spot and future (*expected*) *realized* variance.<sup>7</sup>

Now assume that there exists instead a market of ATM options – or variance swaps. Just as in the above backtest, we use these options to cancel the spot gamma, thus removing the sensitivity to *realized* volatility.

Since the vanilla hedge has to be dynamically readjusted, our position becomes sensitive to the joint dynamics of the spot and *implied* volatilities. Practically, this is

---

moves but costs some theta otherwise. Using the 100% strike along with the 95% and 105% strikes in order to cancel both gamma and vega does not work, as the resulting position is long the 95% and 105% strikes and short the 100% strike: this is mostly a volatility gamma/theta position. Also, unwinding and restarting a new position on a daily basis is not practical: in our backtest, factoring in a bid/offer spread of 0.2 points of volatility on each leg of our spread position wipes out the strategy's P&L.

<sup>6</sup>Interestingly the  $\xi^\tau$  are still driftless. The  $\xi^\tau$  in (7.4) are driftless because they can be delta-hedged by taking a position in market instruments (VSs) that require no financing. In the present context, the  $\xi^\tau$  are driftless in the historical probability measure, just because they are expectations (of future realized variances).

<sup>7</sup>One can also avoid modeling variances altogether by using the hedged Monte Carlo technique of Bouchaud, Potters and Sestovic – see [17], or equivalently, a method proposed by Bruno Dupire. It consists in (a) using consecutive sequences of historical returns as Monte Carlo paths, (b) simulating the daily delta-hedging – at a given implied volatility – of a vanilla option, (c) finding the implied volatility such that the average of the final payoff minus the P&L from the delta hedge equals the Black-Scholes price. Underlying this technique is the unstated – and strong – assumption of stationarity: we are averaging over different volatility regimes and the conditionality on spot level or past return history is lost.

materialized, at order one in volatility of volatility, as the cross gamma of spot and *implied* volatility.

### 9.10.6 Relevance of model-independent properties

It is not clear that other model-independent properties can be established, without resorting to more or less reasonable additional assumptions. For example, Peter Carr and Roger Lee show in [27] that if implied volatilities of power payoffs<sup>8</sup> are uncorrelated with the spot process, the density of the realized quadratic variation up to  $T$  can be extracted from the vanilla smile of maturity  $T$ , hence payoffs on realized variance can be replicated by dynamical trading in vanilla options.

Are these model-independent rules practically relevant? This is assessed by studying whether a violation of these rules can be arbitAGED, i.e. materialized as a P&L. It is not clear in particular that a violation of (9.13) could be practically arbitAGED.

This issue is connected to the general question of the practical relevance of calibration. Our study of the arbitrage of the short SSR is a sobering illustration of how difficult it can be to lock the value of a model parameter – in our case the covariance of spot and ATM implied volatility – by dynamically trading vanilla options.

Only when one is able to do so does it make sense to entrust a model with the task of backing the value of a dynamical parameter out of the vanilla smile.

## 9.11 Conclusion

Because the skew in stochastic volatility models is generated by the covariance of spot and forward variances, some features of the underlying spot/variance dynamics can be recovered from the resulting smile. Provided some reasonable assumptions hold – in particular time homogeneity – for a flat term structure of VS volatilities, at order one in the volatility of volatility, the Skew Stickiness ratio is bounded above and below:

$$\mathcal{R}_T \in [1, 2]$$

We also show that  $\mathcal{R}_T$  – a quantity that characterizes the model’s dynamics – is related to the scaling of the ATMF skew  $\mathcal{S}_T$  with maturity a static property of the smile. For short maturities,  $\mathcal{R}_T$  tends to the universal value of 2. For long maturities its scaling depends on the characteristic exponent  $\gamma$  of the underlying

<sup>8</sup>See Section 4.3 for the definition of power payoffs. In the derivation of [27] the assumption of no correlation between  $S$  and the instantaneous volatility is made. What matters however from a trading point of view is that  $S$  and implied volatilities of power payoffs be uncorrelated. Only then does the spot/implied volatility cross-gamma P&L vanish.

spot/variance covariance function. Stochastic volatility models fall into one of two classes depending on the value of  $\gamma$ .

- Type I models are such that  $\gamma > 1$ : for large  $T$ ,  $\mathcal{S}_T \propto \frac{1}{T}$  and  $\mathcal{R}_T \rightarrow 1$
- Type II models are such that  $\gamma < 1$ : for large  $T$ ,  $\mathcal{S}_T \propto \frac{1}{T^\gamma}$  and  $\mathcal{R}_T \rightarrow 2 - \gamma$

Thus, Type II models allow for a slower decay of the ATMF skew – which is consistent with market skews, for which  $\gamma \simeq \frac{1}{2}$  – and a non-trivial value of the long-maturity limit of the SSR, which, again, is compatible with the observed dynamics of market smiles. Moreover, the characteristic exponent of the spot/variance covariance function can be backed out of the scaling of  $\mathcal{S}_T$  and the long-maturity limit of  $\mathcal{R}_T$ .

This connection between  $\mathcal{S}_T$  and  $\mathcal{R}_T$  is a distinguishing feature of time-homogeneous stochastic volatility models – it is summarized by the following formula: for long maturities,

$$\mathcal{S}_T \propto \frac{1}{T^{2-\mathcal{R}_\infty}}$$

In time-homogeneous Jump-Lévy models, by contrast, while for long maturities  $\mathcal{S}_T \propto \frac{1}{T}$ ,  $\mathcal{R}_T$  vanishes for all  $T$ .

While putting together a genuine Type II model is difficult – because one needs to reconcile a non-trivial scaling of the spot/variance covariance function with the requirement of a low-dimensional Markov representation – Type II behavior can be achieved for a decent range of maturities in a model driven by simple Ornstein–Uhlenbeck processes.

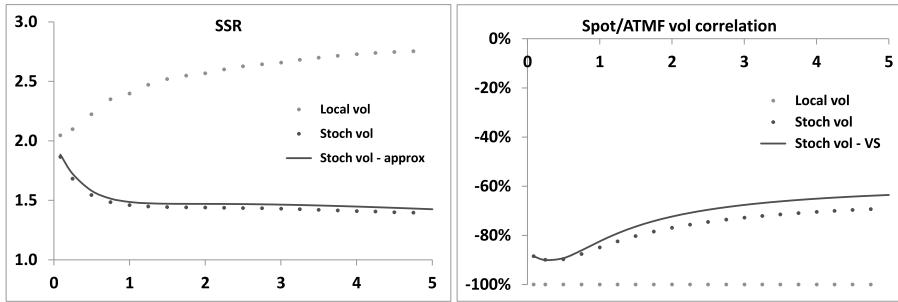
We have provided evidence that suitable parametrization of a simple two-factor model generates Type II scaling for a range of maturities that is practically relevant – while also enforcing the desired scaling of volatilities of volatilities.

Having analyzed how dynamical properties of spot and time-homogeneous stochastic volatility models are related to the smile they produce, the following questions naturally arise:

- How do they compare with dynamical properties of a local volatility model calibrated to the same smile?
- How do they compare with the *realized* behavior of spot and implied volatilities?

### 9.11.1 SSR in local and stochastic volatility models – and in reality

The behavior of  $\mathcal{R}_T$  as a function of  $T$  is structurally different in stochastic volatility models than in the local volatility model. In either Type I or Type II stochastic volatility models, the SSR starts from 2, then *decreases* either towards 1 or towards a non-trivial value in  $[1, 2]$ . In the local volatility model, instead, the SSR



**Figure 9.9:** Left: the SSR as a function of maturity (years) in the two-factor stochastic volatility model (dark dots) and in the local volatility model (light dots) calibrated on the smile generated by the stochastic volatility model. Parameters in Table 8.2, page 329, have been used. The solid line corresponds to the SSR of the stochastic volatility model calculated using expression (9.21), at order one in volatility of volatility. Right: correlation of spot and ATMF volatilities. The solid line corresponds to the correlation of spot and VS – rather than ATMF – volatilities in the two-factor model, which is easily evaluated exactly.

starts from 2 for short maturities, then *increases* for longer maturities, for typical equity index smiles – see Figure 2.4, page 59.

This is illustrated in Figure 9.9. We have used the parameters in Table 8.2 and a flat term structure of VS volatilities at 20% (same parameters as those used in Figure 9.1). The curves in Figure 9.9 are thus obtained with the same vanilla smile – shown in Figure 8.2, page 330, for select maturities.

In both models the SSR starts from the model-independent value of 2 for short maturities. Equivalently, the implied regression coefficient of the short ATMF volatility on the spot is model-independent and given by the ATMF skew.

Parameters in Table 8.2 are such that the ATMF skew approximately decays with the characteristic exponent  $\gamma = \frac{1}{2}$ . For the longer maturities in our graph, the SSR tends to 1.5 in the stochastic volatility model (approximately  $2 - \gamma$ ), while for the local volatility model it tends to a value close to 3.

For a smile such that the ATMF skew decays as a power law, the approximate expression (2.81), page 56, of  $\mathcal{R}_\infty$  in the local volatility model:

$$\mathcal{R}_\infty = \frac{2 - \gamma}{1 - \gamma}$$

indeed yields  $\mathcal{R}_\infty = 3$  for  $\gamma = \frac{1}{2}$ .

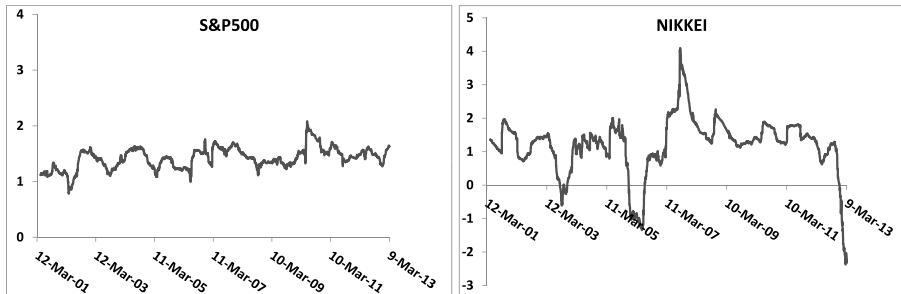
### The spot/ATMF volatility covariance

The SSR is a useful indicator as it measures the implied covariance of  $\ln S$  and the ATMF volatility  $\hat{\sigma}_{F_T T}$  – in units of the ATMF skew – generated by the pricing

model. Spot/volatility cross-gammas are one of the main risks of exotic equity payoffs.<sup>9</sup> Rather than directly comparing the implied spot/volatility covariance with its realized counterpart, we can convert the *realized* covariance of  $\ln S$  and  $\widehat{\sigma}_{F_T T}$  into a *realized* SSR.

Model-generated and realized SSR can then be compared to assess whether the pricing model is sufficiently conservative.

The realized SSR for the Euro Stoxx 50 and S&P 500 indexes, for  $T = 2$  years appear in Figure 9.3 – the estimator for the realized SSR is given by expression (9.22). For both indexes, the average realized SSR is around 1.5. For other indexes, the realized SSR can be very different. Figure 9.10 shows historical values of the realized 6-month value of  $\mathcal{R}_T$  for  $T = 1$  year, for the S&P 500 and Nikkei indexes. While the SSR for the S&P 500 is fairly stable around 1.5, for the Nikkei it reaches at times very negative values.



**Figure 9.10:** The sliding 6-month realized SSR for  $T = 1$  year, for the S&P 500 (left) and Nikkei (right) indexes.

The large negative dip of the SSR of the Nikkei at the end of 2012 can be traced to the impact of the vega hedging of autocalls on the liquidity of the Nikkei vanilla option market.

Autocalls<sup>10</sup> provide the buyer a negative volatility exposure which vanishes when the spot price goes above an upper threshold, whereupon the option expires. Upon selling an autocall the dealer hedges his/her vega position by selling vanilla

<sup>9</sup>What matters is the cross-gamma of the hedged position – see the discussion in Section 1.3.

<sup>10</sup>Consider an autocall of notional  $N$  on an underlying  $S$ . At inception the autocall buyer pays  $N$ . Periodically – say every 3 months – the buyer receives a fixed coupon proportional to  $N$  provided the spot value is below a given threshold, typically 105% of the initial spot value. If, on a coupon date, the spot value is above this threshold, the buyer receives his coupon, gets  $N$  back, and the autocall expires. At maturity  $T$  the buyer gets back  $N$  unless at some point during the autocall's life the spot crosses a lower barrier, typically 60% of the initial spot value, in which case the payoff at maturity is  $N \min(\frac{S_T}{S_0}, 1)$ .

Autocalls are typically written on an index or on the worst of 3/4 stocks:  $S_t \equiv \min_i (S_t^i)$ . The (large) coupon of an autocall compensates the buyer for the risk that he may not recover the invested notional, if the underlying drops significantly.

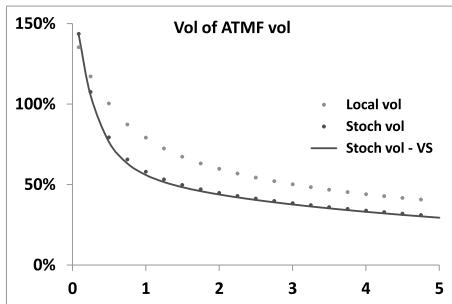
options. The presence of the upper threshold generates a large spot/volatility cross-gamma: as the spot moves up, the dealer needs to buy back vanilla options. In normal circumstances, as the spot moves up, implied volatilities decrease, and the unwinding of the vega hedge should generate on average positive cross-gamma P&L. This cross-gamma P&L is offset by a theta contribution which is quantified by the SSR of the pricing model:

$$\begin{aligned} P\&L &= -S\widehat{\sigma}_T \frac{d^2\Pi}{dSd\widehat{\sigma}_T} \left( \frac{\delta S}{S} \frac{\delta\widehat{\sigma}_T}{\widehat{\sigma}_T} - \mathcal{R}_T \mathcal{S}_T \frac{\sigma^2}{\widehat{\sigma}_T} \delta t \right) \\ &= -SS_T \frac{d^2\Pi}{dSd\widehat{\sigma}_T} (\mathcal{R}_T^r \sigma_r^2 - \mathcal{R}_T \sigma^2) \delta t \end{aligned} \quad (9.33)$$

Figure 9.9 illustrates how different the SSR of the local volatility model and the SSR of a stochastic volatility model are. When carrying a long spot/volatility cross-gamma position, it thus seems preferable to price with the local volatility model, while with a short spot/volatility cross-gamma position, a stochastic volatility model should be used.

We refer the reader to a similar discussion in Section 12.6 of Chapter 12, page 482, in the more general context of mixed local-stochastic volatility models.

### 9.11.2 Volatilities of volatilities



**Figure 9.11:** Volatility of ATMF volatilities as a function of maturity (years) in the two-factor stochastic volatility model (dark dots) and in the local volatility model (light dots) calibrated on the smile generated by the stochastic volatility two-factor model. Parameters in Table 8.2 have been used. The solid line corresponds to volatilities of VS – rather than ATMF – volatilities in the two-factor model calculated exactly using expression (7.39), page 227.

What about volatilities of volatilities? Volatilities of ATMF volatilities for both models appear in Figure 9.11. The term structure of volatilities of volatilities of the two-factor stochastic volatility model has been examined in Section 7.4. As for the

local volatility model, because  $\widehat{\sigma}_{F_T T}$  is a function of  $S$ , the volatility of  $\widehat{\sigma}_{F_T T}$  can be expressed using the SSR. The instantaneous lognormal volatility of  $\widehat{\sigma}_{F_T T}$  is given by (see Section 2.5.5):

$$\text{vol}(\widehat{\sigma}_{F_T T}) = \mathcal{R}_T \mathcal{S}_T \frac{\widehat{\sigma}_{F_0 0}}{\widehat{\sigma}_{F_T T}}$$

where the short ATMF volatility  $\widehat{\sigma}_{F_0 0}$  is also equal to the instantaneous volatility of  $S$ .

For  $T \rightarrow 0$   $\mathcal{R}_T$  tends to 2 and the volatility of the short ATMF volatility is twice the ATMF skew – we recover result (2.85) of Section 2.5.5.

For longer maturities, one can use the approximate expression (2.83):

$$\text{vol}(\widehat{\sigma}_{F_T T}) \simeq \left( \mathcal{S}_T + \frac{1}{T} \int_0^T \mathcal{S}_\tau d\tau \right) \frac{\widehat{\sigma}_{F_0 0}}{\widehat{\sigma}_{F_T T}}$$

which only involves the term structure of the ATMF skew.

Again, in contrast with the two-factor model, the local volatility model is not time-homogeneous; volatilities of volatilities depend not only on  $T - t$ , but also on  $t$ . Typically, volatilities of volatilities in the local volatility model are larger for small  $t$ /smaller for large  $t$ , compared with those of a time-homogeneous stochastic volatility model calibrated on the same smile.

### 9.11.3 Carry P&L of a partially vega-hedged position

Consider a delta- and vega-hedged option position. Assume that there is only one implied volatility  $\widehat{\sigma}$ , in addition to  $S$ . The gamma/theta P&L of the hedged position reads:<sup>11</sup>

$$P\&L = -\frac{1}{2} \frac{d^2 P}{dS^2} (\delta S^2 - \bullet \delta t) - \frac{1}{2} \frac{d^2 P}{d\widehat{\sigma}^2} (\delta \widehat{\sigma}^2 - \bullet \delta t) - \frac{d^2 P}{dS d\widehat{\sigma}} (\delta S \delta \widehat{\sigma} - \bullet \delta t)$$

The suitability of the model parametrization is assessed by comparing the model-generated values of the spot/volatility and volatility/volatility covariances – denoted by  $\bullet \delta t$  – with their realized counterparts, the latter evaluated using historical data for  $S$  and  $\widehat{\sigma}$ .

In practice delta-hedging is performed daily – *and delta is calculated using current market implied volatilities* – but vega-hedging may not, because of larger bid/offer costs. Vega-hedging and delta-hedging occur asynchronously. Which second order spot gamma, spot/volatility cross-gamma and volatility/volatility cross-gamma P&Ls does one then materialize?<sup>12</sup>

<sup>11</sup>Typically, either a term structure of implied volatilities  $\widehat{\sigma}_T$  – in the case of the two-factor model – or all vanilla implied volatilities  $\widehat{\sigma}_{KT}$  – when using the local volatility model or the local-stochastic volatility models of Chapter 12 – are taken as underliers in addition to  $S$ , and contribute to the P&L.

<sup>12</sup>This issue calls to mind that of spot/spot cross-gammas for underlyings trading in different time zones – see [12].

Imagine we vega-hedge our option position periodically every  $n$  days – say one week ( $n = 5$ ). We make the reasonable assumption that gammas and cross-gammas are constant during this period.

Rather than plodding through the calculation of this P&L, we choose a more enlightening route: start from the P&L generated by hedging once per period, then remove terms that can be offset by daily delta-hedging. We denote by  $\delta\hat{\sigma}_i$  and  $\delta S_i$  the respective daily increments over  $[t_{i-1}, t_i]$ .

### Spot gamma

Assuming we delta-hedge only once per period, the spot gamma P&L reads:

$$\begin{aligned} \frac{1}{2} \frac{d^2P}{dS^2} (\Sigma_i \delta S_i)^2 &= \frac{1}{2} \frac{d^2P}{dS^2} (\Sigma_i \delta S_i) (\Sigma_j \delta S_j) \\ &= \frac{1}{2} \frac{d^2P}{dS^2} (\Sigma_i \delta S_i^2) + \frac{d^2P}{dS^2} \Sigma_i (\Sigma_{j < i} \delta S_j) \delta S_i \end{aligned}$$

The prefactor in front of  $\delta S_i$  in the second piece involves spot increments that precede  $\delta S_i$ , thus are known at time  $t_{i-1}$ . This P&L linear in  $\delta S_i$  can then be cancelled by a delta strategy  $\Delta_{t_{i-1}} \delta S_i$  with  $\Delta_{t_{i-1}} = \frac{1}{2} \frac{d^2P}{dS^2} (\Sigma_{j < i} \delta S_j)$ . Only the first piece remains and we get the usual result for the spot gamma P&L over one period:

$$\frac{1}{2} \frac{d^2P}{dS^2} (\Sigma_i \delta S_i^2)$$

It involves the spot variance, measured using daily increments – as expected.

### Volatility gamma

Since no vega-hedging happens during the  $n$ -day period, the volatility gamma P&L is simply:

$$\frac{1}{2} \frac{d^2P}{d\hat{\sigma}^2} (\Sigma_i \delta\hat{\sigma}_i)^2$$

that is, the variance of volatility is sampled according to our weekly schedule – as expected.

### Spot/volatility cross-gamma

The spot/volatility cross-gamma P&L reads:

$$\frac{d^2P}{dSd\hat{\sigma}} (\Sigma_i \delta S_i) (\Sigma_j \delta\hat{\sigma}_j) = \frac{d^2P}{dSd\hat{\sigma}} \Sigma_i (\Sigma_{j < i} \delta\hat{\sigma}_j) \delta S_i + \frac{d^2P}{dSd\hat{\sigma}} \Sigma_i (\Sigma_{j \geq i} \delta\hat{\sigma}_j) \delta S_i$$

The first portion of the right-hand side can be offset by a delta position; our final P&L over one period reads:

$$\frac{d^2P}{dSd\hat{\sigma}} \Sigma_i (\Sigma_{j \geq i} \delta\hat{\sigma}_j) \delta S_i \tag{9.34}$$

which involves the product of spot increments and all subsequent volatility increments.

The spot/volatility estimator in (9.34) is what should be used to measure realized spot/volatility covariance, for the sake of comparing realized and model-implied levels. How does it differ from the usual estimator that applies to the situation of synchronous delta and vega-hedging?

Imagine that during the vega-hedging period  $S$  and  $\hat{\sigma}$  vary by  $\delta S$ ,  $\delta\hat{\sigma}$  and that both experience a trend:  $\delta S_i = \frac{\delta S}{n}$ ,  $\delta\hat{\sigma}_i = \frac{\delta\hat{\sigma}}{n}$ .

Then, using the above expressions, the spot gamma P&L is equal to  $\frac{1}{n} \left( \frac{1}{2} \frac{d^2 P}{dS^2} \delta S^2 \right)$  – that is delta-hedging has reduced it by a factor  $\frac{1}{n}$  – and the volatility gamma P&L is  $\frac{1}{2} \frac{d^2 P}{d\hat{\sigma}^2} \delta\hat{\sigma}^2$ .

As for the cross-gamma P&L, it is equal to  $\frac{n+1}{2n} \left( \frac{d^2 P}{dS d\hat{\sigma}} \delta S \delta\hat{\sigma} \right) \simeq \frac{1}{2} \frac{d^2 P}{dS d\hat{\sigma}} \delta S \delta\hat{\sigma}$ , for  $n$  large.

Thus, using current implied volatilities in the calculation of the delta – even without trading any intermediate vega hedges – has reduced the realized cross-gamma P&L by a factor  $\simeq \frac{1}{2}$ .

## Chapter's digest

### 9.1 The ATMF skew

- At order one in volatility of volatility, the ATMF skew is given, as a function of the spot/variance covariance function by:

$$\mathcal{S}_T = \frac{1}{2\sqrt{T}} \frac{1}{\left(\int_0^T \xi_0^\tau d\tau\right)^{3/2}} \int_0^T d\tau \int_\tau^T \mu(\tau, u, \xi_0) du$$



### 9.2 The Skew Stickiness Ratio (SSR)

- The SSR is defined as the instantaneous regression coefficient of the ATMF volatility on  $\ln S$ , in units of the ATMF skew:

$$\mathcal{R}_T = \frac{1}{\mathcal{S}_T} \frac{E[d \ln S d\hat{\sigma}_{F_T(S)T}]}{E[(d \ln S)^2]}$$

In jump-diffusion models  $\mathcal{R}_T = 0, \forall T$ . In the local volatility model, since the ATMF volatility is a function of  $S$ :  $\mathcal{R}_T = \frac{1}{\mathcal{S}_T} \frac{d\hat{\sigma}_{F_T(S)T}}{d \ln S}$ .

- At order one in volatility of volatility,  $\mathcal{R}_T$  is given, as a function of the spot/variance covariance function, by:

$$\mathcal{R}_T = \frac{\int_0^T \xi_0^\tau d\tau}{T \xi_0^0} \frac{T \int_0^T \mu(0, u, \xi_0) du}{\int_0^T d\tau \int_\tau^T \mu(\tau, u, \xi_0) du}$$



### 9.3 Short-maturity limit of the ATMF skew and the SSR

- For  $T \rightarrow 0$ ,  $\mathcal{S}_0 = \frac{\mu(0,0,\xi_0)}{4(\xi_0^0)^{3/2}}$  and  $\mathcal{R}_0 = 2$ . The short-maturity limit of the SSR is 2, as in the local volatility model.



### 9.4 Model-independent range of the SSR

- Assuming the VS term structure is flat and the spot/variance covariance function is time homogeneous –  $\mu(\tau, u, \xi_0) \equiv \mu(u - \tau)$  – one derives the following

expressions for ATMF skew and SSR:

$$\begin{aligned}\mathcal{S}_T &= \frac{1}{2\xi_0^{3/2}T^2} \int_0^T (T-t)\mu(t)dt \\ \mathcal{R}_T &= \frac{\int_0^T \mu(t)dt}{\int_0^T (1-\frac{t}{T})\mu(t)dt}\end{aligned}$$

Making the assumption that  $\mu(t)$  decays monotonically towards zero as  $t \rightarrow \infty$ , we get model-independent lower and upper bounds on  $\mathcal{R}_T$ :

$$\mathcal{R}_T \in [1, 2]$$

Because of its definition the SSR is inversely proportional to the short VS volatility. One should bear in mind this dependence when assessing SSRs of market smiles.



### 9.5 Scaling of ATMF skew and SSR – a classification of models

► Depending on the rate of decay of  $\mu(t)$  for  $t \rightarrow \infty$ , two long-maturity regimes for the ATMF skew and SSR can be defined, which lead to the division of stochastic volatility models into two classes. Assuming that  $\mu(t) \propto \frac{1}{t^\gamma}$  for  $t \rightarrow \infty$ , for long maturities:

- (Type I) If  $\gamma > 1$ :

$$\mathcal{S}_T \propto \frac{1}{T} \quad \text{and} \quad \lim_{T \rightarrow \infty} \mathcal{R}_T = 1 \quad (9.35)$$

- (Type II) If  $\gamma < 1$ :

$$\mathcal{S}_T \propto \frac{1}{T^\gamma} \quad \text{and} \quad \lim_{T \rightarrow \infty} \mathcal{R}_T = 2 - \gamma \quad (9.36)$$

Exponential decay of  $\mu(t)$  produces Type I behavior. In Type I models, in the long-maturity regime,  $\mathcal{S}_T$  and  $\mathcal{R}_T$  bear no signature of the rate of decay of  $\mu$ .

For both types of models, the long-maturity ATMF skew and SSR are related through:

$$\mathcal{S}_T \propto \frac{1}{T^{2-\mathcal{R}_\infty}}.$$



### 9.6 Type I models – the Heston model

► The Heston model produces Type I behavior. Its long-maturity SSR is 1, which can be checked using the order-one expansion of the ATMF skew in Chapter 6.



### 9.7 Type II models

► Even though the two-factor model is strictly of Type 1, since  $\mu(t)$  decays exponentially for large  $t$ , we can still achieve Type II scaling of the ATMF skew – and the corresponding value of the SSR – on a range of maturities that is practically relevant.

For a flat term-structure of VS volatilities, at order one in volatility of volatility, both quantities are given by:

$$\begin{aligned} S_T &= \nu \alpha_\theta \left[ (1-\theta) \rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta \rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2} \right] \\ R_T &= \frac{(1-\theta) \rho_{SX^1} \frac{1-e^{-k_1 T}}{k_1 T} + \theta \rho_{SX^2} \frac{1-e^{-k_2 T}}{k_2 T}}{(1-\theta) \rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta \rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2}} \end{aligned}$$

► The realized behavior of equity index smiles is consistent with Type II.



### 9.8 Numerical evaluation of the SSR

► The SSR of the two-factor model is easily evaluated numerically in a Monte Carlo simulation by simply shifting the initial values of processes  $X_t$  and  $Y_t$ .



### 9.9 The SSR for short maturities

► The value of the realized SSR of short-maturity equity index smiles is usually substantially lower than the model-independent value of stochastic volatility models. Can this difference be materialized as the P&L of a trading strategy?



### 9.10 Arbitraging the realized short SSR

► We use a lognormal model for the short ATM volatility. In our two asset-model – spot and short ATM volatility – the difference between the realized SSR and its model-independent value of 2 is materialized as a cross-gamma/theta P&L.

► We backtest a dynamical delta-hedged option strategy on the Euro Stoxx 50 index that consists in maintaining a short-skew position around the money. The resulting P&L approximately captures the cross-gamma/theta P&L corresponding to the difference between implied and realized skew, even though the residual vega of our position impacts the cumulative P&L negatively on large downward moves of the spot.



### 9.11 Conclusion

► The behavior of  $\mathcal{R}_T$  as a function of  $T$  is structurally different in stochastic volatility models than in the local volatility model. In the local volatility model, for typical equity index smiles, the SSR starts from 2 for short maturities, then *increases* for longer maturities.

For an ATMF skew that decays algebraically with exponent  $\gamma$ , in the local volatility model  $\mathcal{R}_\infty = \frac{2-\gamma}{1-\gamma}$ , while in a stochastic volatility model,  $\mathcal{R}_\infty = 2 - \gamma$ . For the typical value  $\gamma = \frac{1}{2}$ , the SSR of the local volatility model for long maturities is  $\mathcal{R}_\infty = 3$ , compared to  $\mathcal{R}_\infty = 1.5$  for a stochastic volatility model.

► Instantaneous volatilities of volatilities of long-dated vanilla options are also larger in the local volatility model than in a stochastic volatility model.

► In practice, vega hedging occurs less frequently than delta hedging, thus the spot/volatility cross/gamma-theta P&L is not materialized exactly. Assuming that delta-hedging is performed daily, using actual implied volatilities, while vega-hedging is performed less frequently, we obtain the approximate result that, in case of a trend in both spot and implied volatility in between two vega rehedges, half of the cross-gamma P&L is materialized.

**This page intentionally left blank**

# **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**

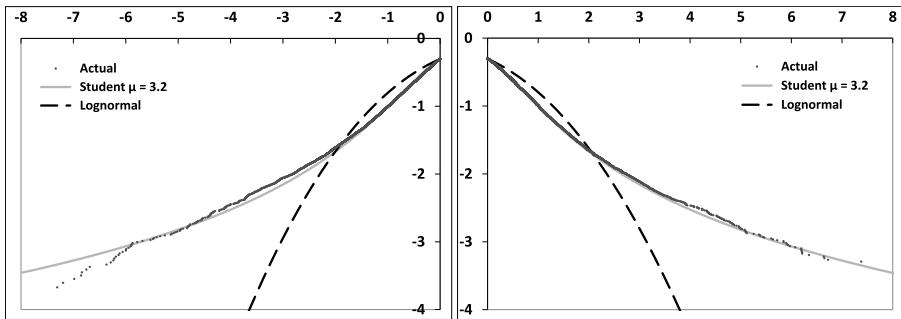
Figure 10.1 shows the cumulative distribution of normalized negative and positive returns of the Dow Jones index. We have used daily closing quotes from January 1, 1900 to July 20, 2014.<sup>1</sup>

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  $14.3$ .<sup>2</sup>

---

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

<sup>2</sup>With 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  $\mu$  is customarily called “number of degrees of freedom”.

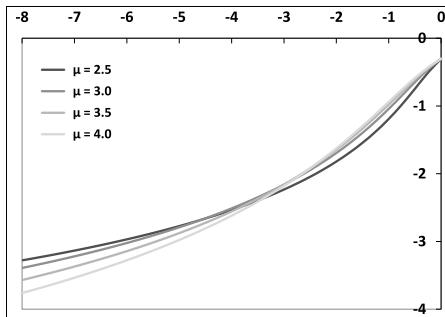
The smaller  $\mu$  the thicker the tails. For large values of  $x$  the density scales like  $\frac{1}{x^{1+\mu}}$ , thus the one-sided cumulative distribution function scales like  $\frac{1}{x^\mu}$ . Only moments of order smaller than  $\mu$  exist.

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  $\mu$  are shown in Figure 10.2. Any value in the interval [3, 4] provides an acceptable fit to empirical data.<sup>3</sup>



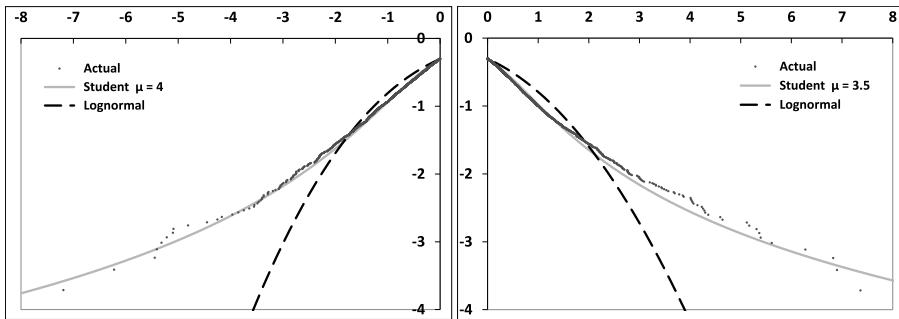
**Figure 10.2:** Logarithm ( $\log_{10}$ ) of the left tail of the distribution function of a Student random variable, normalized so that its variance is 1, for different values of  $\mu$ .

Thus our preliminary conclusions are:

- The empirical distribution of realized equity index returns is well approximated by a Student distribution with  $\mu$  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.

<sup>3</sup>For  $\mu < 4$ , the kurtosis  $\kappa$  of Student-distributed returns diverges. Obviously, the kurtosis estimator applied to any sample yields a finite number. However, the divergence of  $\kappa$  manifests itself in the fact that occasional large returns generate large jumps in the estimator for  $\kappa$ . Tails of the Student distribution for  $\mu < 4$  are thick enough that these events occur sufficiently frequently that defining an average finite value for  $\kappa$  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  $\kappa$  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  $\sigma_i$  is the instantaneous volatility and  $z_i$  are iid random variables with unit variance.

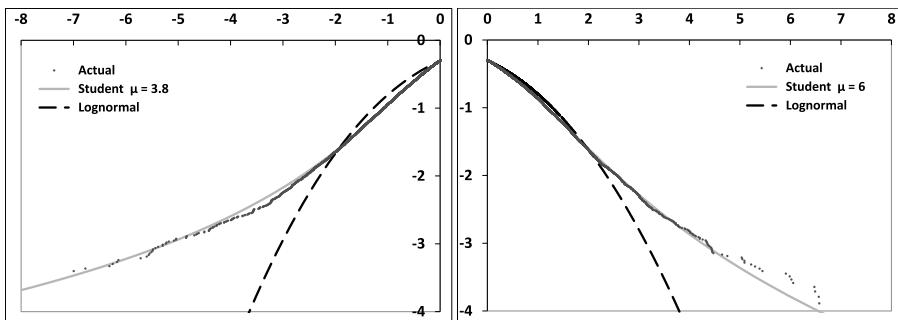
The rationale for this ansatz is that the variability of the probability distribution of  $r_i$  is condensed in that of the scale factor  $\sigma_i$  – the  $z_i$  are identically distributed.

When  $\sigma_i$  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  $\sigma_i$ , (b) the distribution of  $z_i$ ?

(10.2) is a natural ansatz but accessing  $\sigma_i$  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  $\sigma_i$  reduces the tail thickness, which is manifested in larger values for  $\mu$ . 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.<sup>4</sup>

<sup>4</sup>The fact that the conditional value of  $\mu$  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  $\xi_t^\tau$  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  $\sigma_t$  is given by:

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

and the dynamics of forward variances  $\xi_t^\tau$  in the two-factor model is given by equation (7.28), page 226:<sup>5</sup>

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

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)  $\delta Z$  is a random variable with mean 0 and variance  $\Delta$ . In the standard version of the two-factor model  $\delta Z$  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  $\delta W^S$  and  $\delta X^i$  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  $\delta Z$  from a two-sided Student distribution with different values of  $\mu$  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  $\mu_+, \mu_-$  the parameters of the corresponding Student distributions. We define  $\delta Z$  as:

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

---

<sup>5</sup>No need for volatility-of-volatility smile here.

where  $X_\mu$  denotes a Student random variable with  $\mu$  degrees of freedom. The density of  $\delta Z$  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  $\sigma_+, \sigma_-$  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  $\sigma_+, \sigma_-$  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  $\alpha_-$ . 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  $\delta Z$  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  $\delta Z$  with  $\delta X^1, \delta X^2$  and (b) to be able to degenerate  $\delta Z$  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  $\delta W^S$  supplied by the Monte Carlo engine of our standard two-factor model and map them into  $\delta Z$  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  $\mathcal{N}_{\mu}^{-1}$  is the inverse cumulative distribution function of a Student random variable with  $\mu$  degrees of freedom.<sup>6</sup>

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  $\mu_+, \mu_-$  to assess the influence of the one-day smile we must ensure that other features of the model are unchanged. As we replace  $\delta W^S$  with  $\delta Z$  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  $\rho_{1S}^*, \rho_{2S}^*$  so that the instantaneous correlations – or covariances – of  $\delta Z$  and  $\delta X^1$  and of  $\delta Z$  and  $\delta X^2$  using  $\rho_{1S}^*, \rho_{2S}^*$  are identical to the covariances in the standard two-factor model parametrized with  $\rho_{1S}, \rho_{2S}$ , 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  $\phi$  is the probability density of the standard normal variable.

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

---

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

<sup>7</sup> 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  $\mu$ ,  $\rho_{iS}^*/\rho_{iS}$  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  $\delta S$  is exact for short time intervals – see the discussion below.

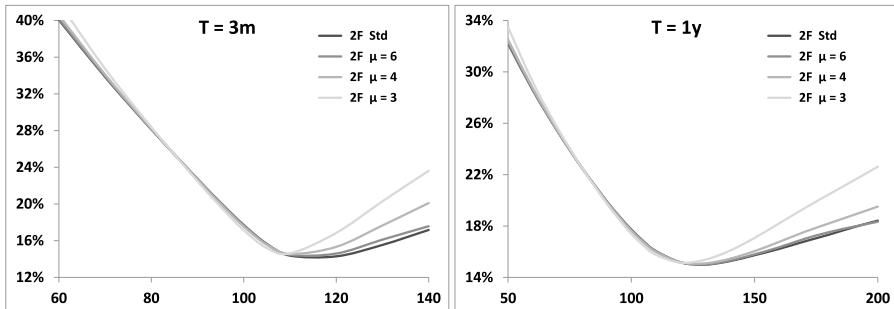
We use throughout a flat variance curve:  $\xi_0^\tau = \xi_0 = 20\%^2$  and parameters in Table 10.1.<sup>8</sup>

$\nu$	$\theta$	$k_1$	$k_2$	$\rho$	$\rho_{SX^1}$	$\rho_{SX^2}$
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:  $\mu_+ = \mu_- = \mu$ .

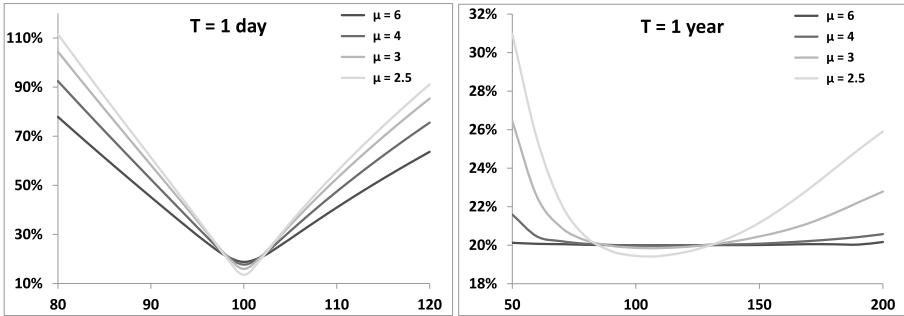
Figure 10.5 displays three-month and one-year smiles in the standard and fat-tailed version of the two-factor model.



**Figure 10.5:** Smiles in the two-factor model with Student-distributed returns for different values of  $\mu$ ,  $p_+ = p_- = 1/2$  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  $\mu$  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.

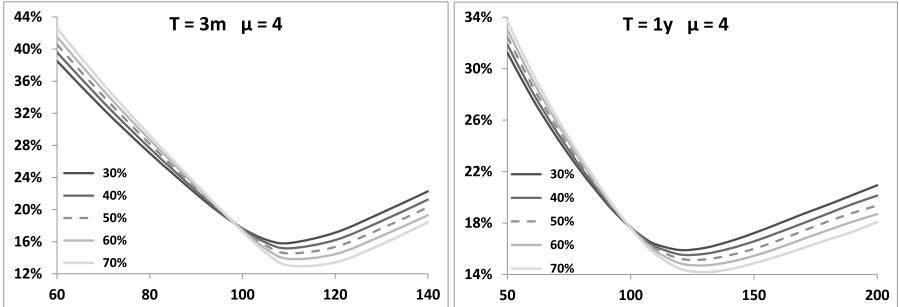
<sup>8</sup>These were typical of the Euro Stoxx 50 index as of July 2014.



**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  $\mu_+$  and  $\mu_-$  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  $p_+ = 30\%, 40\%, 50\%, 60\%, 70\%$ .

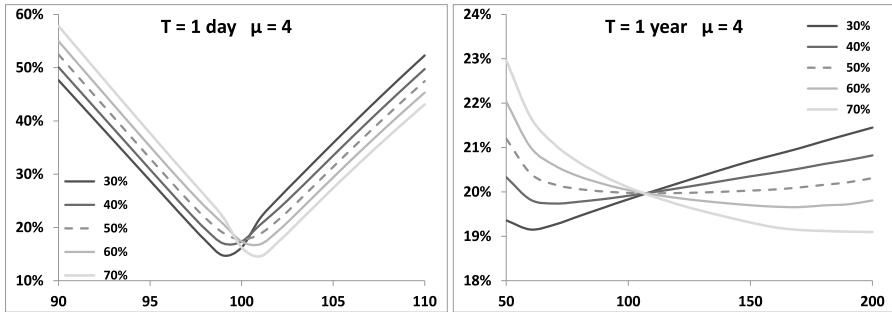
Again let us turn off stochastic volatility; the 1-day and 1-year smiles are shown in Figure 10.8.



**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  $p_+$ .<sup>9</sup>

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  $\Delta$  is the time scale of the discrete model.

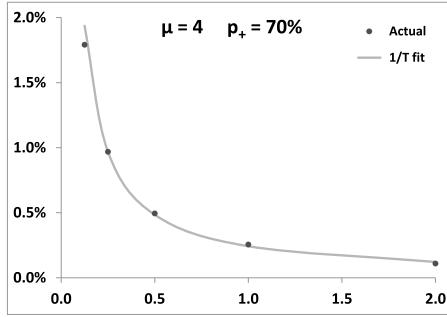
These models generate the ATMF skew through two mechanisms: (a) the ATMF skew for time scale  $\Delta$ , (b) the covariance of spot and forward variances – not unlike our fat-tailed stochastic volatility model.

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

<sup>9</sup>See 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 :

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

$\Delta$  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  $\delta Z$  is a random variable with vanishing mean and variance  $\Delta$ , and  $\sigma$  is the instantaneous volatility. Inserting expression (10.9) into (10.8) and expanding in powers of  $\delta Z$  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  $\delta S$ , 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  $\Delta$ , 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  $\Delta$  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).<sup>10</sup>
- 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  $\delta Z$ . 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  $\sigma$ , and in figures 10.5 and 10.7 with stochastic volatility.<sup>11</sup>

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

---

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

<sup>11</sup>When 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  $\mu$  and  $\nu$  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  $\delta Z$ . Only higher-order contributions to the P&L will depend on the distribution of  $\delta Z$ .

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  $\delta S^k$  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  $\delta S$  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.

$\mu$	$\infty$	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  $\mu$  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  $\mu$ ), and as we vary  $p^+$ , thus increasing the ATMF skew (bottom section of table).<sup>12</sup>

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

<sup>12</sup>Because 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.

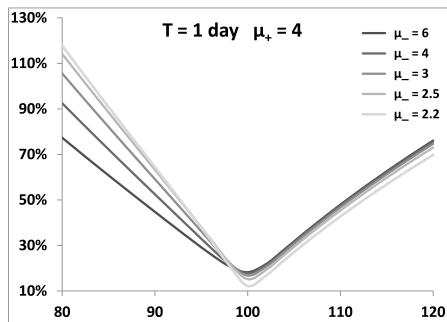
Typical strikes for these daily puts are in the range of 75% to 90%. Daily cliques are also typically of the knock-out type – the option expires once one coupon has paid off and the premium is paid on a quarterly basis – making them very similar to CDS contracts.

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  $\mu_-$  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.

$\mu_-$	$\infty$	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  $\mu_-$  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  $\mu_-$  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  $\mu_-$  this low are typically needed to match market prices.

This implied value of  $\mu_-$  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  $\mu_-$ . For example, setting  $\nu = 257\%$  increases the rightmost price in Table 10.3 by about 5 bps.<sup>13</sup>

---

### **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.  $\mu_+$ ,  $\mu_-$  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.

---

<sup>13</sup>Stochastic 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  $\delta S$  is cancelled by the delta, so the P&L starts with a  $\delta S^2$  term. The first step in model building focuses on terms of order  $\delta S^2$  – 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  $\sigma$ :

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

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).<sup>14</sup>

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  $\delta S$ , 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  $\delta S$  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  $\sigma$  – the scale of the return – is modeled with a two-factor forward variance model, and the distribution of  $\delta Z$  – or equivalently the conditional distribution of  $\delta S$  – 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.

---

<sup>14</sup>The 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  $\delta S^2$  portion, which acts as an extra contribution to the volatility of  $S$ . Let us now assign an annualized frequency  $\lambda$  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  $\lambda$ .
- 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  $\beta$ , and that the rate of return required by the bank on its capital is  $\mu$ , the amount charged to the desk is then of the form (10.13) with  $\lambda = \beta \mu$ ;  $\lambda$  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  $\lambda$  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  $\Delta P$  to the option price.  $\Delta P$  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:

$$P + \Delta P$$

$\Delta P$  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  $\delta S$ . 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  $\Delta P$ .

$\Delta P$  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  $\sigma$  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  $\rho_z$  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  $\rho_Z$ ,  $L(\tau, q)$ , which is the logarithm of the characteristic function of  $\rho_z$ , 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  $\ln S$ :<sup>15</sup>

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

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  $\kappa_3$  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.<sup>16</sup>

---

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

<sup>16</sup>For  $T \rightarrow 0$ , 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  $\kappa_3$

## 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  $\sigma$  replaced with the local volatility  $\sigma(t, S)$ , chosen so that market prices of vanilla options are recovered.<sup>17</sup>

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  $\rho_z$  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.

<sup>17</sup>Space 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  $\Sigma_{KT}^{\text{LV}}$  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.<sup>18</sup>

---

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].

<sup>18</sup>How 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  $\lambda_i$  and jump density  $\rho_i^*$ . 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  $\delta W^S$  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**

# Chapter 11

---

## Multi-asset stochastic volatility

Multi-asset stochastic volatility models are not mere juxtapositions of single-asset models. Obviously we need to specify an asset/asset correlation matrix, just as in a multi-asset Black-Scholes model.

In addition to the individual spot/volatility and volatility/volatility covariance functions whose role we have examined in Chapter 8, multi-asset stochastic volatility models involve cross spot/volatility and volatility/volatility correlations. What is the effect of these new parameters?

We propose a parametrization of the global correlation matrix based on observable physical quantities, study the ATMF skew of a basket and the correlation swap.

In the course of our discussion we use the example of the multi-asset local volatility model, the most popular multi-asset stochastic volatility model.

---

### 11.1 The short ATMF basket skew

This section focuses on the short-maturity ATMF basket skew. Consider a basket  $B$  consisting of  $n$  assets, with weights  $\alpha_i$ :

$$B = \sum_i \alpha_i S_i$$

and let us define relative weights  $w_i = \frac{\alpha_i S_i}{B}$ :  $\sum_i w_i = 1$ . The variation of  $B$  is:

$$\frac{dB}{B} = \sum_i w_i \frac{dS_i}{S_i} \quad (11.1)$$

While the  $\alpha_i$  are fixed, the  $w_i$  are not; as we will be studying the short-maturity case we will make the approximation that the  $w_i$  are constant. From the results of Section 8.5, the short ATMF implied volatility  $\widehat{\sigma}_B$  and skew  $S_B$  are given by – see formula (8.36), page 322:

$$\widehat{\sigma}_B = \sigma_B \quad (11.2a)$$

$$S_B = \frac{1}{2\widehat{\sigma}_B^2} \frac{\langle d \ln B d\widehat{\sigma}_B \rangle}{dt} \quad (11.2b)$$

where  $\sigma_B$  is the instantaneous volatility of the basket which, from (11.1) is given by:

$$\sigma_B^2 = \sum_{ij} w_i w_j \rho_{ij} \sigma_i \sigma_j$$

$\sigma_i$  is the instantaneous volatility of asset  $i$  – equal to the short ATM volatility  $\widehat{\sigma}_i$  – and  $\rho_{ij}$  is the correlation of  $S_i$  and  $S_j$ . The relationship between the short ATM basket volatility and the short ATM volatilities of the constituents reads:

$$\widehat{\sigma}_B = \sqrt{\sum_{ij} w_i w_j \rho_{ij} \sigma_i \sigma_j} = \sqrt{\sum_{ij} w_i w_j \rho_{ij} \widehat{\sigma}_i \widehat{\sigma}_j} \quad (11.3)$$

$d\widehat{\sigma}_B$  is given by:

$$d\widehat{\sigma}_B = \frac{1}{\widehat{\sigma}_B} \sum_{ij} w_i w_j \rho_{ij} \widehat{\sigma}_i d\widehat{\sigma}_j$$

and we get:

$$\begin{aligned} \frac{1}{2\widehat{\sigma}_B^2} \frac{\langle d \ln B d\widehat{\sigma}_B \rangle}{dt} &= \frac{1}{2\widehat{\sigma}_B^3} \sum_{ijk} w_i w_j w_k \rho_{ij} \widehat{\sigma}_i \frac{\langle d \ln S_k d\widehat{\sigma}_j \rangle}{dt} \\ &= \frac{1}{2\widehat{\sigma}_B^3} \sum_{ij} w_i w_j^2 \rho_{ij} \widehat{\sigma}_i \frac{\langle d \ln S_j d\widehat{\sigma}_j \rangle}{dt} + \frac{1}{2\widehat{\sigma}_B^3} \sum_{i j \neq k} w_i w_j w_k \rho_{ij} \widehat{\sigma}_i \frac{\langle d \ln S_k d\widehat{\sigma}_j \rangle}{dt} \end{aligned}$$

The basket skew is thus given by:

$$\mathcal{S}_B = \frac{1}{\widehat{\sigma}_B^3} \sum_{ij} w_i w_j^2 \rho_{ij} \widehat{\sigma}_i \widehat{\sigma}_j^2 \mathcal{S}_j + \frac{1}{2\widehat{\sigma}_B^3} \sum_{i j \neq k} w_i w_j w_k \rho_{ij} \widehat{\sigma}_i \frac{\langle d \ln S_k d\widehat{\sigma}_j \rangle}{dt} \quad (11.4)$$

where we have separated terms involving the covariance on an asset with its own volatility and have used (11.2b) to relate  $\langle d \ln S_j d\widehat{\sigma}_j \rangle$  to  $\mathcal{S}_j$ , the ATM skew of basket component  $S_j$ . Expression (11.4) shows that the basket ATM skew is generated partly by the ATM skew of the basket components and partly by the covariance of each component with other components' ATM volatilities.

### 11.1.1 The case of a large homogeneous basket

Assume that all volatilities and correlations are equal and that the  $n$  components of the basket are equally weighted:  $\widehat{\sigma}_i \equiv \widehat{\sigma}$ ,  $\rho_{ij} \equiv \rho_{SS}$ ,  $\mathcal{S}_i = \mathcal{S}$ ,  $w_i = \frac{1}{n}$ . Expressions (11.3) and (11.4) for the basket ATM volatility and skew simplify to:

$$\widehat{\sigma}_B = \sqrt{\frac{1 + (n - 1)\rho_{SS}}{n}} \widehat{\sigma} \quad (11.5a)$$

$$\mathcal{S}_B = \frac{1 + (n - 1)\rho_{SS}}{n} \frac{\widehat{\sigma}^3}{\widehat{\sigma}_B^3} \left[ \frac{\mathcal{S}}{n} + \frac{n - 1}{n} \frac{1}{2\widehat{\sigma}^2} \frac{\langle d \ln S d\widehat{\sigma} \rangle_{\text{cross}}}{dt} \right] \quad (11.5b)$$

where  $\langle d \ln S d\hat{\sigma} \rangle_{\text{cross}}$  denotes the instantaneous covariation of a basket component with another component's ATMF volatility. Observe that the portion contributed to  $\mathcal{S}_B$  by each component's ATMF skew scales like  $\frac{1}{n}$ , hence goes to zero for a large basket.

For a large basket ( $n \gg 1$ ) – which is the case of most equity indexes – we get the following simpler formulas:

$$\hat{\sigma}_B \simeq \sqrt{\rho_{SS}} \hat{\sigma} \quad (11.6a)$$

$$\mathcal{S}_B \simeq \frac{1}{\sqrt{\rho_{SS}}} \frac{1}{2\hat{\sigma}^2} \frac{\langle d \ln S d\hat{\sigma} \rangle_{\text{cross}}}{dt} \quad (11.6b)$$

Equation (11.6b) makes it clear that the short ATMF skew of a large basket only depends on cross-asset spot-volatility covariances. Consider the following opposite cases:

- $\langle d \ln S d\hat{\sigma} \rangle_{\text{cross}} = 0$ : a component's volatility is uncorrelated with other components. In this case  $\mathcal{S}_B = 0$ .
- $\langle d \ln S d\hat{\sigma} \rangle_{\text{cross}} = \langle d \ln S d\hat{\sigma} \rangle_{\text{diag}}$  where  $\langle d \ln S d\hat{\sigma} \rangle_{\text{diag}}$  is the instantaneous co-variation of a basket component with its own implied volatility. This situation is obtained for example by assuming that volatilities of all assets are driven by a single Brownian motion: the  $\hat{\sigma}_i$  are 100% correlated. Using:

$$\mathcal{S} = \frac{1}{2\hat{\sigma}^2} \frac{\langle d \ln S d\hat{\sigma} \rangle_{\text{diag}}}{dt} \quad (11.7)$$

we get:

$$\mathcal{S}_B \simeq \frac{1}{\sqrt{\rho_{SS}}} \mathcal{S} \quad (11.8)$$

The basket skew is larger than the component's skew.

Before turning to the real case, we examine the prediction of the local volatility model.

### 11.1.2 The local volatility model

A multi-asset local volatility model is a peculiar stochastic volatility model in that it only takes as inputs asset-asset correlations. Implied volatilities are *functions* of spot and time:  $\hat{\sigma} = \hat{\sigma}(t, S)$ , thus are driven by the same Brownian motion as the spot process. For a homogeneous basket, such that the local volatility functions of all components are identical, this entails that:

$$\langle d \ln S d\hat{\sigma} \rangle_{\text{cross}} = \rho_{SS} \langle d \ln S d\hat{\sigma} \rangle_{\text{diag}}$$

Using again (11.7), expressions (11.5) become:

$$\hat{\sigma}_B = \sqrt{\frac{1 + (n - 1)\rho_{SS}}{n}} \hat{\sigma} \simeq \sqrt{\rho_{SS}} \hat{\sigma} \quad (11.9a)$$

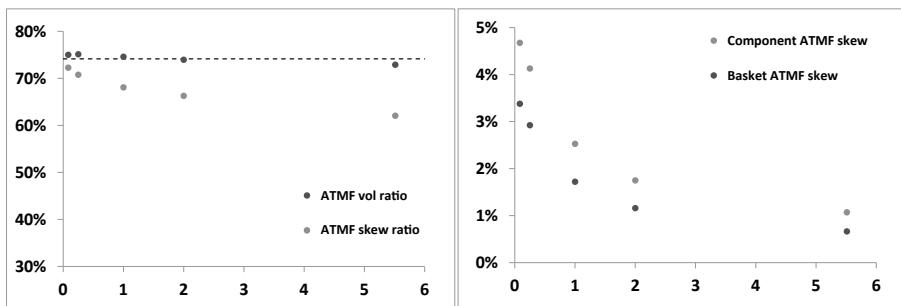
$$\mathcal{S}_B = \sqrt{\frac{1 + (n - 1)\rho_{SS}}{n}} \mathcal{S} \simeq \sqrt{\rho_{SS}} \mathcal{S} \quad (11.9b)$$

One consequence of (11.9) is that, in the local volatility model, for short maturities, the ratios of ATMF basket skew to ATMF component skew and of ATMF basket volatility to ATMF component volatility are identical and equal to  $\sqrt{\rho_{SS}}$ .

$$\frac{S_B}{S} \simeq \frac{\hat{\sigma}_B}{\hat{\sigma}} \simeq \sqrt{\rho_{SS}} \quad (11.10)$$

Figure 11.1 provides an illustration of how well relationship (11.10) is actually obeyed, for a range of maturities. We have taken  $n = 10$ ,  $\rho_{SS} = 50\%$ , and have used for the 10 components the same smile.<sup>1</sup>

Observe how  $\frac{\hat{\sigma}_B}{\hat{\sigma}}$  stays remarkably close to its theoretical value  $\sqrt{\frac{1+(n-1)\rho_{SS}}{n}} = 74.2\%$ , even for long maturities.  $\frac{S_B}{S}$ , on the other hand, is in fact lower than its short-maturity value (11.9b).



**Figure 11.1:** Left: ratios  $\frac{\hat{\sigma}_B}{\hat{\sigma}}$  and  $\frac{S_B}{S}$  as a function of maturity (years) in the local volatility model for an equally weighted basket of 10 assets with the same smile. The dashed line indicates the theoretical short-maturity value  $\sqrt{\frac{1+(n-1)\rho_{SS}}{n}}$ . Right: the ATMF skew expressed as the difference of the implied volatilities of the 95% and 105% strikes.

### 11.1.3 The basket skew in reality

Typically, implied correlations levels for equity indexes – derived from (11.6a) – are  $\rho_{SS} \simeq 60\%$ : ATMF index implied volatilities are about 25% lower than stock implied volatilities. Market ATMF skews of indexes are stronger than stock skews. They are typically about 25% larger – this is in fact roughly what formula (11.8) for the case of 100% correlated volatilities yields.

This stands in stark contrast with the prediction of the local volatility model: the short-maturity approximate formula (11.10) implies that the index skew should be about 20% *lower* than the stock skew; values of  $\frac{S_B}{S}$  in Figure 11.1 – obtained with  $\rho_{SS} = 50\%$  – show that the actual ratio in the local volatility model is lower still.

<sup>1</sup>We have used for the components the smile of the Euro Stoxx 50 index of June 12th, 2012, which is steeper than typical stock skews – using weaker skews wouldn't alter our conclusions.

### 11.1.4 Digression – how many stocks are there in an index?

Formulas (11.5a) and (11.5b) for  $\hat{\sigma}_B$  and  $\mathcal{S}_B$  apply to the case of an equally weighted basket; the large-basket regime is reached when  $n \gg 1$ . Equity indexes are not equally weighted baskets. Given a particular index, which value of  $n$  should be used in (11.5a) and (11.5b) for assessing whether the asymptotic regime is applicable?

Let us assume that the volatilities of the index components are all equal to  $\sigma$ .<sup>2</sup>  $\sigma_B^2$  is given by:

$$\sigma_B^2 = \sigma^2 \sum_{ij} w_i w_j \rho_{ij} = \sigma^2 \left( \sum_i w_i^2 + \rho_{SS} \sum_{i \neq j} w_i w_j \right) \quad (11.11)$$

Comparing this formula with (11.5a), we see that the effective number of components of an index,  $n^*$ , should be defined as:<sup>3</sup>

$$n^* = \frac{1}{\sum_i w_i^2} \quad (11.12)$$

Table 11.1 lists the value of  $n$  and  $n^*$  for a few equity indexes;  $n^*$  can be sizeably smaller than  $n$ .

	S&P 500	Euro Stoxx 50	NIKKEI	KOSPI	FTSE	SMI	CAC	Russell 2000
$n$	500	50	225	200	101	20	40	2000
$n^*$	143	37	47	17	36	8	20	1056

**Table 11.1:** Actual ( $n$ ) and effective ( $n^*$ ) number of components of some equity indexes. In (11.12) values of  $w_i$  as of August 6, 2013 have been used.

## 11.2 Parametrizing multi-asset stochastic volatility models

Given a set of single-asset models, we need to define (a) correlations for the spot processes, (b) a parametrization of cross spot/volatility and volatility/volatility correlations.

Rather than directly parametrizing cross factor/factor and spot/factor correlations, whose significance depends on the specifics of the factor structure, we aim for a parsimonious parametrization of “physical” quantities.

Such parametrization can be used regardless of the particular factor structure of the model at hand.

<sup>2</sup>Or less crudely that there is no systematic correlation between  $w_i$  and  $\sigma_i$ .

<sup>3</sup>The same value of  $n^*$  can be extracted from the second component of  $\sigma_B^2$  since  $\sum_i w_i = 1$ . Interestingly,  $n^*$  is used by economists as a measure of market concentration under the name of Herfindahl-Hirschman Index.

We know from Chapter 8 that, at order two in volatilities of volatilities, the smile is determined by the spot/variance and variance/variance covariance functions  $\mu(t, u, \xi) = \frac{\langle d \ln S_t \, d\xi_t^u \rangle}{dt}$  and  $\nu(t, u, u', \xi) = \frac{\langle d\xi_t^u \, d\xi_t^{u'} \rangle}{dt}$ . We then focus on these objects and define cross spot/factor and factor/factor correlations so that cross-covariance functions are related simply to their diagonal counterparts.

### 11.2.1 A homogeneous basket

Consider the case a basket of  $n$  assets, each driven by the two-factor model of Section 7.4 with identical parameters  $\nu, \theta, k_1, k_2, \rho_{X_1 X_2}$  (previously noted  $\rho_{12}, \rho_{SX_1}, \rho_{SX_2}$ ). For the sake of readability, we decide to carry the factor indices as subscripts rather than superscripts, in the present section.

The instantaneous spot/variance and variance/variance covariance functions of an asset  $S_t$  and its forward variances  $\xi_t^u$  are given by expressions (8.50) and (8.51), page 327:

$$\mu^{\text{diag}}(t, u, \xi) = 2\nu\alpha_\theta\sqrt{\xi^u}\xi^u \left( (1-\theta)\rho_{SX_1}^{\text{diag}} e^{-k_1(u-t)} + \theta\rho_{SX_2}^{\text{diag}} e^{-k_2(u-t)} \right) \quad (11.13)$$

$$\begin{aligned} \nu^{\text{diag}}(t, u, u', \xi) &= 4\nu^2\xi^u\xi^{u'}\alpha_\theta^2 \left[ (1-\theta)^2\rho_{X_1 X_1}^{\text{diag}} e^{-k_1(u+u'-2t)} \right. \\ &\quad + \theta^2\rho_{X_2 X_2}^{\text{diag}} e^{-k_2(u+u'-2t)} \\ &\quad \left. + \rho_{X_1 X_2}^{\text{diag}}\theta(1-\theta) \left( e^{-k_1(u-t)}e^{-k_2(u'-t)} + e^{-k_2(u-t)}e^{-k_1(u'-t)} \right) \right] \end{aligned} \quad (11.14)$$

where the *diag* subscript indicates that these covariance functions apply to spot and variances of the same asset. For the diagonal case, we have:

$$\rho_{SX_1}^{\text{diag}} = \rho_{SX_1}, \quad \rho_{SX_2}^{\text{diag}} = \rho_{SX_2}, \quad \rho_{X_1 X_1}^{\text{diag}} = 1, \quad \rho_{X_2 X_2}^{\text{diag}} = 1, \quad \rho_{X_1 X_2}^{\text{diag}} = \rho_{X_1 X_2}$$

Upon replacing *diag* with *cross* superscripts,  $\mu_{\text{cross}}, \nu_{\text{cross}}$  are given by expressions (11.13) and (11.14) as well, except the *diag* superscript is replaced with *cross*.

Correlations  $\rho_{X_1 X_2}, \rho_{SX_1}, \rho_{SX_2}$  are set. We need to define  $\rho_{SX_1}^{\text{cross}}, \rho_{SX_2}^{\text{cross}}, \rho_{X_1 X_1}^{\text{cross}}$ ,  $\rho_{X_2 X_2}^{\text{cross}}, \rho_{X_1 X_2}^{\text{cross}}$  as well as the asset-asset correlation  $\rho_{SS}$ .

Our aim is to parametrize the multi-asset model so that  $\mu^{\text{cross}}$  and  $\nu^{\text{cross}}$  are defined in terms of their diagonal counterparts. Specifically, we require that the term structure of  $\mu^{\text{cross}}$  and  $\nu^{\text{cross}}$  – that is their dependency on  $u - t$  and  $u' - t$  – match that of  $\mu^{\text{diag}}$  and  $\nu^{\text{diag}}$ , once corrected for the effect of the term structure of forward variances. Note that prefactors  $\sqrt{\xi^u}\xi^u$  and  $\xi^u\xi^{u'}$  in (11.13) and (11.14) would no longer appear if we worked with instantaneous spot/variance and variance/variance correlations rather than covariances.

We introduce parameter  $\chi_{S\sigma}$  and define cross-correlations  $\rho_{SX_1}^{\text{cross}}, \rho_{SX_2}^{\text{cross}}$  through:

$$\begin{cases} \rho_{SX_1}^{\text{cross}} &= \chi_{S\sigma} \rho_{SX_1}^{\text{diag}} \\ \rho_{SX_2}^{\text{cross}} &= \chi_{S\sigma} \rho_{SX_2}^{\text{diag}} \end{cases}$$

$\chi_{S\sigma}$  simply measures how much larger or smaller cross spot/volatility correlations are with respect to their diagonal counterparts. Using (11.13) we can check that, for a flat term structure of VS volatilities:  $\mu^{\text{cross}}(t, u, \xi) = \chi_{S\sigma} \mu^{\text{diag}}(t, u, \xi)$  or, equivalently:

$$\rho^{\text{cross}}(S, \hat{\sigma}_T) = \chi_{S\sigma} \rho^{\text{diag}}(S, \hat{\sigma}_T)$$

where  $\rho^{\text{diag}}(S, \hat{\sigma}_T)$  is the instantaneous correlation between  $S$  and its implied VS volatility of maturity  $T$  and  $\rho^{\text{cross}}(S, \hat{\sigma}_T)$  is its cross-counterpart.

Turning now to  $\nu$ , we similarly introduce coefficient  $\chi_{\sigma\sigma}$  and define  $\rho_{X_1 X_1}^{\text{cross}}, \rho_{X_2 X_2}^{\text{cross}}, \rho_{X_1 X_2}^{\text{cross}}$  as:

$$\begin{cases} \rho_{X_1 X_1}^{\text{cross}} &= \chi_{\sigma\sigma} \rho_{X_1 X_1}^{\text{diag}} \\ \rho_{X_2 X_2}^{\text{cross}} &= \chi_{\sigma\sigma} \rho_{X_2 X_2}^{\text{diag}} \\ \rho_{X_1 X_2}^{\text{cross}} &= \chi_{\sigma\sigma} \rho_{X_1 X_2}^{\text{diag}} \end{cases}$$

Thus,

$$\rho_{X_1 X_1}^{\text{cross}} = \chi_{\sigma\sigma}, \rho_{X_2 X_2}^{\text{cross}} = \chi_{\sigma\sigma}, \rho_{X_1 X_2}^{\text{cross}} = \chi_{\sigma\sigma} \rho_{X_1 X_2}$$

One can check on (11.14) that with this parametrization, the instantaneous correlation between  $\xi_i^u$  and  $\xi_j^{u'}$  where  $i, j$  denote different assets is equal to  $\chi_{\sigma\sigma}$  times the correlation of  $\xi_i^u$  and  $\xi_i^{u'}$ . For a flat term structure of VS volatilities  $\nu^{\text{cross}}(t, u, u', \xi) = \chi_{\sigma\sigma} \nu^{\text{diag}}(t, u, u', \xi)$  or, equivalently:

$$\rho^{\text{cross}}(\hat{\sigma}_T, \hat{\sigma}_{T'}) = \chi_{\sigma\sigma} \rho^{\text{diag}}(\hat{\sigma}_T, \hat{\sigma}_{T'})$$

where  $\rho^{\text{diag}}(\hat{\sigma}_T, \hat{\sigma}_{T'})$  is the instantaneous correlation between VS volatilities of maturities  $T, T'$  of the same underlying and  $\rho^{\text{cross}}(\hat{\sigma}_T, \hat{\sigma}_{T'})$  is a similarly-defined quantity for correlations of VS volatilities of different underlyings.

Finally we introduce the (uniform) correlation between spot processes,  $\rho_{SS}$ . Each asset is associated to 3 Brownian motions:  $W_i^S, W_i^1, W_i^2$ . The global correlation matrix is thus of dimension  $3n \times 3n$ . What are the conditions on  $\rho_{SS}$ ,  $\chi_{\sigma\sigma}$ ,  $\chi_{S\sigma}$  so that it is positive?

### Conditions on $\rho_{SS}$ , $\chi_{\sigma\sigma}$ , $\chi_{S\sigma}$

Let us compute the eigenvalues of matrix  $\Omega$ , defined by:  $\Omega = \frac{1}{dt} \langle dU dU^\top \rangle$  where  $U^\top = (W_1^S \cdots W_n^S, W_1^1 \cdots W_n^1, W_1^2 \cdots W_n^2)$ .

Let  $\lambda$  be an eigenvalue of  $\Omega$ , associated to the eigenvector  $T$ , whose components we write  $T^\top = (s_1 \cdots s_n, x_1 \cdots x_n, y_1 \cdots y_n)$ . Expressing that  $\Omega T = \lambda T$  we get, for  $1 \leq i \leq n$ :

$$\begin{cases} (s_i + \rho_{SS} \bar{\Sigma}_j s_j) + \rho_{SX_1}(x_i + \chi_{S\sigma} \bar{\Sigma}_j x_j) + \rho_{SX_2}(y_i + \chi_{S\sigma} \bar{\Sigma}_j y_j) &= \lambda s_i \\ \rho_{SX_1}(s_i + \chi_{S\sigma} \bar{\Sigma}_j s_j) + (x_i + \chi_{\sigma\sigma} \bar{\Sigma}_j x_j) + \rho_{X_1 X_2}(y_i + \chi_{\sigma\sigma} \bar{\Sigma}_j y_j) &= \lambda x_i \\ \rho_{SX_2}(s_i + \chi_{S\sigma} \bar{\Sigma}_j s_j) + \rho_{X_1 X_2}(x_i + \chi_{\sigma\sigma} \bar{\Sigma}_j x_j) + (y_i + \chi_{\sigma\sigma} \bar{\Sigma}_j y_j) &= \lambda y_i \end{cases}$$

where  $\bar{\Sigma}_j$  is a shorthand notation for  $\Sigma_{j \neq i}$  and we have removed the *diag* superscripts in  $\rho_{X_1 X_2}^{\text{diag}}$ ,  $\rho_{S X_1}^{\text{diag}}$ ,  $\rho_{S X_2}^{\text{diag}}$  for notational economy. The correlation structure is invariant under permutation of the  $i, j$  indices and in particular under translation:  $\rho(W_i^\circ, W_j^\bullet) = \rho(W_{i+k \bmod n}^\circ, W_{j+k \bmod n}^\bullet) \forall k$  where  $\circ, \bullet$  stand for  $S, 1, 2$ .

We thus diagonalize  $\Omega$  in the basis of eigenvectors of the discrete translation operator.  $T$  is parametrized with the 4 numbers  $\theta_k, s, x, y$ :

$$T^\top = (s e^{i\theta_k} \cdots s e^{in\theta_k}, x e^{i\theta_k} \cdots x e^{in\theta_k}, y e^{i\theta_k} \cdots y e^{in\theta_k})$$

with  $\theta_k = \frac{2k\pi}{n}$  where  $k = 0 \cdots n - 1$  and  $i = \sqrt{-1}$ . We have:

$$\sum_{j=0, j \neq i}^{n-1} e^{ij\theta_k} = \sum_{j=i+1}^{n-1+i} e^{ij\theta_k} = e^{ii\theta_k} \sum_{j=1}^{n-1} e^{ij\theta_k} = \begin{cases} (n-1)e^{i\theta_k} & k=0 \\ -e^{i\theta_k} & k \neq 0 \end{cases}$$

We thus define  $A_{SS}(k)$  by:

$$A_{SS}(k) = \left( 1 + \rho_{SS} \sum_{j=1}^{n-1} e^{ij\theta_k} \right) = \begin{cases} 1 + (n-1)\rho_{SS} & k=0 \\ 1 - \rho_{SS} & k \neq 0 \end{cases} \quad (11.15)$$

$A_{\sigma\sigma}(k), A_{S\sigma}(k)$  are defined similarly in terms of  $\chi_{\sigma\sigma}, \chi_{S\sigma}$ . We have:

$$\begin{cases} A_{\sigma\sigma}(k=0) = 1 + (n-1)\chi_{\sigma\sigma} \\ A_{\sigma\sigma}(k \neq 0) = 1 - \chi_{\sigma\sigma} \end{cases} \quad \text{and} \quad \begin{cases} A_{S\sigma}(k=0) = 1 + (n-1)\chi_{S\sigma} \\ A_{S\sigma}(k \neq 0) = 1 - \chi_{S\sigma} \end{cases}$$

The resulting system for  $s, x, y$  is:

$$\begin{cases} A_{SS}s + \rho_{SX_1} A_{S\sigma}x + \rho_{SX_2} A_{S\sigma}y = \lambda s \\ \rho_{SX_1} A_{S\sigma}s + A_{\sigma\sigma}x + \rho_{X_1 X_2} A_{\sigma\sigma}y = \lambda x \\ \rho_{SX_2} A_{S\sigma}s + \rho_{X_1 X_2} A_{\sigma\sigma}x + A_{\sigma\sigma}y = \lambda y \end{cases}$$

The condition is thus that the following symmetric matrix

$$\omega(k) = \begin{pmatrix} A_{SS} & \rho_{SX_1} A_{S\sigma} & \rho_{SX_2} A_{S\sigma} \\ \rho_{SX_1} A_{S\sigma} & A_{\sigma\sigma} & \rho_{X_1 X_2} A_{\sigma\sigma} \\ \rho_{SX_2} A_{S\sigma} & \rho_{X_1 X_2} A_{\sigma\sigma} & A_{\sigma\sigma} \end{pmatrix}$$

be positive. This implies in particular that  $A_{SS}(k) \geq 0, A_{\sigma\sigma}(k) \geq 0 \forall k$ , which places the following bounds on  $\rho_{SS}, \chi_{\sigma\sigma}$ :

$$-\frac{1}{n-1} \leq \rho_{SS}, \chi_{\sigma\sigma} \leq 1$$

These conditions, in the equity context, are not very restrictive.

We now define a “correlation matrix”  $\rho(k)$  obtained by rescaling the symmetric matrix  $\omega(k)$ :  $\rho_{ij}(k) = \frac{\omega_{ij}(k)}{\sqrt{\omega_{ii}(k)\omega_{jj}(k)}}$ :

$$\rho(k) = \begin{pmatrix} 1 & \zeta\rho_{SX_1} & \zeta\rho_{SX_2} \\ \zeta\rho_{SX_1} & 1 & \rho_{X_1 X_2} \\ \zeta\rho_{SX_2} & \rho_{X_1 X_2} & 1 \end{pmatrix} \quad (11.16)$$

where  $\zeta(k)$  is given by:

$$\zeta(k) = \frac{A_{S\sigma}(k)}{\sqrt{A_{SS}(k)A_{\sigma\sigma}(k)}} = \begin{cases} \frac{1+(n-1)\chi_{S\sigma}}{\sqrt{(1+(n-1)\rho_{SS})(1+(n-1)\chi_{\sigma\sigma})}} & k=0 \\ \frac{1-\chi_{S\sigma}}{\sqrt{(1-\rho_{SS})(1-\chi_{\sigma\sigma})}} & k \neq 0 \end{cases}$$

$\rho(k)$  in (11.16) is in fact the correlation matrix of the single-asset case, but with spot/volatility correlations rescaled by the factor  $\zeta(k)$ , which may be smaller or larger than 1 depending on the values of  $\rho_{SS}, \chi_{\sigma\sigma}, \chi_{S\sigma}$ .

$\chi_{S\sigma}$  is allowed to be larger than 1, unlike  $\rho_{SS}$  and  $\chi_{\sigma\sigma}$ . This will be needed in situations when the index skew is much steeper than the component skew – see the discussion in Section 11.1.1.

The case  $\zeta(k) = 1$  is obtained by taking  $\chi_{S\sigma} = \chi_{\sigma\sigma} = \rho_{SS}$ .  $\rho(k)$  is then equal to the single-asset correlation matrix thus the positivity of the  $3n \times 3n$  global covariance matrix is ensured. In this model, cross-correlations (spot/spot, spot/volatility, volatility/volatility) are all equal to  $\rho_{SS}$  times their diagonal counterparts. This is the case of the multi-asset local volatility model.<sup>4</sup>

### Conclusion

Given correlations  $\rho_{SX_1}, \rho_{SX_2}, \rho_{X_1 X_2}$  for the single-asset case, we parametrize the  $3n \times 3n$  global correlation matrix of the multi-asset case by introducing three additional parameters:  $\rho_{SS}, \chi_{\sigma\sigma}, \chi_{S\sigma}$ .

$\rho_{SS}$  is the spot/spot correlation.  $\chi_{\sigma\sigma}, \chi_{S\sigma}$  define cross spot/volatility and volatility/volatility covariance functions in terms of their diagonal counterparts.

$\rho_{SS}$  and  $\chi_{\sigma\sigma}$  are restricted to the interval  $[-\frac{1}{n-1}, 1]$ . The necessary and sufficient condition on  $\rho_{SS}, \chi_{\sigma\sigma}, \chi_{S\sigma}$  for the global correlation matrix to be positive is that the following effective correlation matrix for the single-asset case:

$$\begin{pmatrix} 1 & \zeta\rho_{SX_1} & \zeta\rho_{SX_2} \\ \zeta\rho_{SX_1} & 1 & \rho_{X_1 X_2} \\ \zeta\rho_{SX_2} & \rho_{X_1 X_2} & 1 \end{pmatrix}$$

be positive for the two following values of  $\zeta$ :

$$\zeta = \frac{1 + (n - 1)\chi_{S\sigma}}{\sqrt{(1 + (n - 1)\rho_{SS})(1 + (n - 1)\chi_{\sigma\sigma})}} \quad (11.17a)$$

$$\zeta = \frac{1 - \chi_{S\sigma}}{\sqrt{(1 - \rho_{SS})(1 - \chi_{\sigma\sigma})}} \quad (11.17b)$$

in addition to the  $\zeta = 1$  case. This is the case if  $|\zeta\rho_{SX_1}| \leq 1$ ,  $|\zeta\rho_{SX_2}| \leq 1$  and:

$$-1 \leq \frac{\rho_{X_1 X_2} - \zeta^2 \rho_{SX_1} \rho_{SX_2}}{\sqrt{1 - \zeta^2 \rho_{SX_1}^2} \sqrt{1 - \zeta^2 \rho_{SX_2}^2}} \leq 1$$

---

<sup>4</sup>The local volatility model's assumption of setting volatility/volatility correlations identical to spot/spot correlations is not very realistic. Typically, correlations of implied volatilities of underlyings belonging to the same index are usually larger than the correlations of the underlyings themselves.

It is easy to check, going backwards through the derivation in Section 11.2.1, that these necessary conditions are also sufficient: we have effectively built the  $3n$  eigenvectors of  $\Omega$ .<sup>5</sup>

While in what follows we will be using a homogeneous model, in practice, parameters  $\nu, \theta, k_1, k_2, \rho_{X_1 X_2}, \rho_{SX_1}, \rho_{SX_2}$  are different for different components. In particular, even though we typically use identical values for  $\nu, \theta, k_1, k_2, \rho_{X_1 X_2}$  so that volatilities of volatilities are identical, we may use different values for  $\rho_{SX_1}^{\text{diag}}, \rho_{SX_2}^{\text{diag}}$ . Consider two components  $S_i, S_j$ . Should we set  $\rho_{S^i X_1^j}^{\text{cross}} = \chi_{S\sigma} \rho_{S^j X_1^j}^{\text{diag}}$  or  $\rho_{S^i X_1^j}^{\text{cross}} = \chi_{S\sigma} \rho_{S^i X_1^j}^{\text{diag}}$ ?

When setting the cross spot/volatility correlations of  $S_i$  and  $\widehat{\sigma}_j^T$ , it is more reasonable to derive the  $T$ -dependence of  $\rho(S_i, \widehat{\sigma}_j^T)$  from the term-structure of  $\rho(S_j, \widehat{\sigma}_j^T)$  rather than that of  $\rho(S_i, \widehat{\sigma}_i^T)$ . We would thus set:

$$\rho_{S^i X_1^j}^{\text{cross}} = \chi_{S\sigma} \rho_{S^j X_1^j}^{\text{diag}}$$

and likewise for  $\rho_{S^i X_2^j}^{\text{cross}}$ . Going back to implied volatilities, this is equivalent to setting:

$$\rho(S^i, \widehat{\sigma}_T^j) = \chi_{S\sigma} \rho(S^j, \widehat{\sigma}_T^j)$$

Likewise, with respect to volatility/volatility correlations, we could set

$$\begin{aligned} \text{either } \rho(\widehat{\sigma}_T^i, \widehat{\sigma}_{T'}^j) &= \chi_{\sigma\sigma} \rho(\widehat{\sigma}_T^i, \widehat{\sigma}_{T'}^i) \\ \text{or } \rho(\widehat{\sigma}_T^i, \widehat{\sigma}_{T'}^j) &= \chi_{\sigma\sigma} \rho(\widehat{\sigma}_T^j, \widehat{\sigma}_{T'}^j) \end{aligned}$$

In practice, volatility/volatility correlations are sufficiently similar across underlyings that either choice results in very similar values of  $\chi_{\sigma\sigma}$  – see the examples in Table 11.2 below.

The global correlation matrix we thus build for a non-homogeneous basket is not guaranteed to be positive – but is not expected to be very negative either; in case it is we use the algorithm proposed in [61] to generate the closest positive correlation matrix.<sup>6</sup>

### 11.2.2 Realized values of $\chi_{S\sigma}$ and $\chi_{\sigma\sigma}$

Typical realized values of  $\chi_{S\sigma}$  and  $\chi_{\sigma\sigma}$  for indexes appear in Table 11.2. For each pair of underlyings  $(S^1, S^2)$ , two values of  $\chi_{S\sigma}$  and  $\chi_{\sigma\sigma}$  have been computed, obtained by averaging respectively either  $\rho(\widehat{\sigma}_T^1, S^2)/\rho(\widehat{\sigma}_T^1, S^1)$  or

<sup>5</sup> $\Omega$  has 3 non-degenerate eigenvalues (obtained with  $k = 0$ ) and 3 degenerate eigenvalues (obtained with  $k \neq 0$ , thus each with  $(n - 1)$  degeneracy).

<sup>6</sup>The algorithm in [61] generates the correlation matrix that is closest to a candidate input matrix. In the formula for the matrix distance, appropriate weights can be used so as to enforce, for example, that diagonal correlations  $\rho_{S^i X_1^i}^{\text{diag}}, \rho_{S^i X_2^i}^{\text{diag}}, \rho_{X_1^i X_2^i}^{\text{diag}}$ , as well as spot/spot correlations are least altered.

$\rho(\hat{\sigma}_T^2, S^1)/\rho(\hat{\sigma}_T^2, S^2)$  over several maturities. Likewise, for the determination of  $\chi_{\sigma\sigma}$ , two values are obtained by averaging either  $\rho(\hat{\sigma}_T^1, \hat{\sigma}_{T'}^2)/\rho(\hat{\sigma}_T^1, \hat{\sigma}_{T'}^1)$  or  $\rho(\hat{\sigma}_T^1, \hat{\sigma}_{T'}^2)/\rho(\hat{\sigma}_T^2, \hat{\sigma}_{T'}^2)$ , for all  $(T, T')$  couples.

Note that it is important to use asynchronous estimators – lest we underestimate correlations; see [12].

June 2008 - June 2013					June 2003 - June 2008					
	S&P500	Stoxx50	S&P500	Stoxx50	NIKKEI	S&P500	Stoxx50	S&P500	Stoxx50	NIKKEI
$\rho$	Stoxx50	FTSE	NIKKEI	NIKKEI	KOSPI	Stoxx50	FTSE	NIKKEI	NIKKEI	KOSPI
$\rho$	83%	89%	55%	60%	67%	71%	86%	45%	59%	66%
$\chi_{S\sigma}$	87%	95%	59%	63%	85%	78%	89%	56%	61%	108%
$\chi_{S\sigma}$	86%	93%	71%	78%	78%	75%	91%	53%	77%	65%
$\chi_{\sigma\sigma}$	84%	92%	54%	60%	72%	72%	85%	31%	40%	50%
$\chi_{\sigma\sigma}$	84%	92%	55%	61%	73%	71%	85%	31%	41%	50%

**Table 11.2:** Historical spot/spot correlations and values of  $\chi_{S\sigma}$  and  $\chi_{\sigma\sigma}$  for different pairs of indexes, measured on two 5-year samples.  $\chi_{S\sigma}$  and  $\chi_{\sigma\sigma}$  are evaluated by averaging pairwise ratios of spot/volatility and volatility/volatility correlations using implied ATMF volatilities with maturities 3 months, 6 months, 1 year, 2 years. The asynchronous estimator in [12] has been used.

Empirically, diagonal volatility/volatility correlations across indexes are very similar, thus the two ways of estimating  $\chi_{\sigma\sigma}$  yield very similar value; we can thus use an average of the two estimates.

This is less the case for  $\chi_{S\sigma}$ . Indeed, historical regimes of spot/volatility correlations for the Nikkei index, for example, can be quite different from those of the S&P 500 and Euro Stoxx 50 indexes. This is also reflected in the different behavior of their SSRs – see Figure 9.10, page 381.

### 11.3 The ATMF basket skew

After focusing in Section 11.1 on short maturities, we derive now an approximation of the basket ATMF skew at order one in volatility of volatility by using expression (8.22) which involves the spot/variance covariance function evaluated at order one in volatility of volatility. We use the notation  $\hat{\sigma}_T^B$  and  $\mathcal{S}_T^B$  for the basket VS volatility and ATM skew and denote *basket* forward variances by  $\zeta_t^u$ .

We will assume that the underlying model is time-homogeneous and that the VS term structures of all components are flat and identical.

From equation (8.22), page 314,  $\mathcal{S}_T^B$  is given by:

$$\mathcal{S}_T^B = \hat{\sigma}_T^B \frac{C_B^{x\xi}(T)}{2((\hat{\sigma}_T^B)^2 T)^2} \quad (11.18)$$

where  $C_B^{x\xi}(T)$  is the doubly-integrated spot/variance covariance function:

$$C_B^{x\xi}(T) = \int_0^T dt \int_t^T du \mu_B(t, u) \quad (11.19)$$

$$\mu_B(t, u) = \lim_{dt \rightarrow 0} \frac{1}{dt} E_t [d \ln B_t d \zeta_t^u] \quad (11.20)$$

where  $B_t = \Sigma_i w_i S_{it}$ .  $\mu_B(t, u)$  is evaluated at order one in volatility of volatility on the initial basket variance curve.

Let us assume that the basket is homogeneous and equally weighted; spot correlations  $\rho_{ij}, i \neq j$  are all identical, equal to  $\rho$  and we use the correlation parametrization of Section 11.2.1.

We also make the approximation that weights are frozen, equal to their initial values, which we denote by  $w_{i0}$ . Basket forward variances  $\zeta_t^T$  are given by:

$$\zeta_t^u = E_t [\sum_{ij} w_i w_j \rho_{ij} \sqrt{\xi_{iu}^u} \sqrt{\xi_{ju}^u}] \quad (11.21)$$

$$= \sum_{ij} \rho_{ij} w_{i0} w_{j0} E_t [\sqrt{\xi_{iu}^u} \sqrt{\xi_{ju}^u}] \quad (11.22)$$

Thus

$$d\zeta_t^u = \sum_{ij} w_{i0} w_{j0} \rho_{ij} E_t \left[ \frac{\sqrt{\xi_{ju}^u}}{\sqrt{\xi_{iu}^u}} d\xi_{iu}^u \right]$$

At order one in volatility of volatility, it is sufficient to replace the prefactor inside the expectation with its value in the unperturbed state, that is using forward variances values at  $t = 0$ . At this order:

$$d\zeta_t^u = \sum_{ij} w_{i0} w_{j0} \rho_{ij} \frac{\sqrt{\xi_{j0}^u}}{\sqrt{\xi_{i0}^u}} E_t [d\xi_{iu}^u] = \sum_{ij} w_{i0} w_{j0} \rho_{ij} \frac{\sqrt{\xi_{j0}^u}}{\sqrt{\xi_{i0}^u}} d\xi_{iu}^u$$

Let us now compute  $\mu_B(t, u)$ . We have:

$$\frac{dB_t}{B_t} = \sum_i w_{k0} \frac{dS_{k,t}}{S_{k,t}}$$

We then get the expression of  $\mu_B(t, u)$  as a function of the diagonal spot/variance covariance function, which we denote compactly by  $\mu(u - t) = \frac{1}{dt} \langle d \ln S_{it} d \zeta_{it}^u \rangle$ , as

we have assumed a time-homogeneous model:

$$\begin{aligned}\mu_B(t, u) &= \frac{1}{dt} \langle d \ln B_t d\xi_t^u \rangle = \sum_{ijk} w_{i0} w_{j0} w_{k0} \rho_{ij} \frac{\sqrt{\xi_{j0}^u}}{\sqrt{\xi_{i0}^u}} \frac{\langle d \ln S_{kt} d\xi_{it}^u \rangle}{dt} \\ &= \sum_{ij} w_{i0}^2 w_{j0} \rho_{ij} \frac{\sqrt{\xi_{j0}^u}}{\sqrt{\xi_{i0}^u}} \mu(u-t) + \chi_{S\sigma} \sum_{i \neq k, j} w_{i0} w_{j0} w_{k0} \rho_{ij} \frac{\sqrt{\xi_{j0}^u}}{\sqrt{\xi_{i0}^u}} \mu(u-t)\end{aligned}$$

For a homogeneous basket with flat and identical term structures of VS volatilities, this simplifies to:

$$\mu_B(t, u) = \frac{1 + (n-1)\chi_{S\sigma}}{n} \frac{1 + (n-1)\rho_{SS}}{n} \mu(u-t)$$

Using expression (11.19) for  $C_B^{x\xi}(T)$ :

$$C_B^{x\xi}(T) = \left[ \frac{1 + (n-1)\chi_{S\sigma}}{n} \right] \left[ \frac{1 + (n-1)\rho_{SS}}{n} \right] \int_0^T dt \int_0^T du \mu(u-t)$$

Finally, the ATMF skew of a homogeneous basket is given, at order one in volatility of volatility, by:

$$\mathcal{S}_T^B = \frac{\hat{\sigma}_{0T}^B}{2((\hat{\sigma}_{0T}^B)^2 T)^2} \left[ \frac{1 + (n-1)\chi_{S\sigma}}{n} \right] \left[ \frac{1 + (n-1)\rho_{SS}}{n} \right] \int_0^T dt \int_t^T du \mu(u-t) \quad (11.23)$$

where  $\hat{\sigma}_{0T}^B$  is the basket VS volatility at order zero in volatility of volatility, that is with zero volatility of volatility. Assuming that the VS term structures of all components are flat, and equal to  $\hat{\sigma}^2$ , and that the basket is equally weighted,  $\hat{\sigma}_{0T}^B$  is given by:

$$(\hat{\sigma}_{0T}^B)^2 = \frac{1 + (n-1)\rho_{SS}}{n} \hat{\sigma}^2$$

Using this expression of  $\hat{\sigma}_{0T}^B$  in (11.23) and remembering that the single-asset skew is given by expression (8.22), page 314, with  $C^{x\xi}$  given by (8.13):

$$\mathcal{S}_T = \frac{\hat{\sigma}}{2(\hat{\sigma}^2 T)^2} \int_0^T dt \int_t^T du \mu(u-t)$$

we get our final expression for the basket skew, as a function of the components' skew:

$$\mathcal{S}_T^B = \mathcal{S}_T \sqrt{\frac{n}{1 + (n-1)\rho_{SS}}} \left[ \frac{1 + (n-1)\chi_{S\sigma}}{n} \right] \quad (11.24)$$

Thus an implied value for  $\chi_{S\sigma}$  can be backed out of the ratio of basket skew to component skew.

Let us check that in the limit  $T \rightarrow 0$ , (11.24) yields back the short-maturity result (11.5b) for  $\mathcal{S}_T^B$ . Using the fact that  $\int_0^T dt \int_t^T du = \frac{T^2}{2}$  and that  $\mathcal{S}_{T=0} = \frac{\mu(0)}{4\hat{\sigma}_{T=0}^3}$ , expression (11.23) for  $\mathcal{S}_T^B$  yields:

$$\begin{aligned}\mathcal{S}_{T=0}^B &= \frac{1}{4(\hat{\sigma}_{T=0}^B)^3} \left[ \frac{1 + (n-1)\chi_{S\sigma}}{n} \right] \left[ \frac{1 + (n-1)\rho_{SS}}{n} \right] \mu(0) \\ &= \frac{1}{4(\hat{\sigma}_{T=0}^B)^3} \left[ \frac{1 + (n-1)\chi_{S\sigma}}{n} \right] \left[ \frac{1 + (n-1)\rho_{SS}}{n} \right] 4\hat{\sigma}_{T=0}^3 \mathcal{S}_{T=0} \\ &= \mathcal{S}_{T=0} \left( \frac{\hat{\sigma}_{T=0}}{\hat{\sigma}_{T=0}^B} \right)^3 \left[ \frac{1 + (n-1)\chi_{S\sigma}}{n} \right] \left[ \frac{1 + (n-1)\rho_{SS}}{n} \right]\end{aligned}\quad (11.25)$$

where we have omitted the 0 subscript in  $\hat{\sigma}_{0T}^B$ , as, for  $T = 0$ ,  $\hat{\sigma}_{0T}^B = \hat{\sigma}_T^B$ . (11.25) is identical to (11.5b) for  $T = 0$ .

Again, expression (11.24) for  $\mathcal{S}_T^B$  involves the factor  $\frac{1+(n-1)\chi_{S\sigma}}{n}$ : the portion contributed to the basket skew by the components' skew scales like  $\frac{1}{n}$  and the bulk of the basket skew is generated by cross spot/volatility correlations.

### 11.3.1 Application to the two-factor model

In the two-factor model the dynamics of each component reads:

$$\begin{cases} dS_t &= \sqrt{\xi_t^S} S_t dW_t^S \\ d\xi_t^T &= 2\nu \xi_t^T \alpha_\theta ((1-\theta)e^{-k_1(T-t)} dW_t^1 + \theta e^{-k_2(T-t)} dW_t^2) \end{cases}$$

where  $\nu$  is the lognormal volatility of a very short-dated implied VS volatility, and  $\alpha_\theta = ((1-\theta)^2 + \theta^2 + 2\rho_{X_1 X_2} \theta(1-\theta))^{-1/2}$ . The cross spot/variance and variance/variance correlations are parametrized using  $\chi_{S\sigma}$  and  $\chi_{\sigma\sigma}$  defined in Section 11.2.1 and the correlations between the  $S_i$  are equal to  $\rho$ .

For a flat term structure of VS volatilities equal to  $\hat{\sigma}$ , the spot/variance covariance function  $\mu(\tau)$  is given by:

$$\mu(\tau) = (2\nu) \hat{\sigma}^3 \alpha_\theta ((1-\theta)\rho_{SX_1} e^{-k_1\tau} + \theta\rho_{SX_2} e^{-k_2\tau}) \quad (11.26)$$

and the component's ATMF skew at order one in volatility of volatility is given by expression (8.55):

$$\mathcal{S}_T = \nu \alpha_\theta \left[ (1-\theta) \rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta \rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2} \right] \quad (11.27)$$

Using (11.24) the basket skew – at order one in volatility of volatility – is given by:

$$\begin{aligned}\mathcal{S}_T^B &= \nu \alpha_\theta \sqrt{\frac{n}{1 + (n-1)\rho_{SS}}} \left[ \frac{1 + (n-1)\chi_{S\sigma}}{n} \right] \\ &\times \left[ (1-\theta) \rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta \rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2} \right]\end{aligned}\quad (11.28)$$

While, for the sake of calculating  $S_T^B$  we have needed  $\hat{\sigma}_T^B$  at order zero in volatility of volatility only, it can be calculated exactly – with the assumption of frozen weights:

$$\begin{aligned} (\hat{\sigma}_T^B)^2 &= \frac{1}{T} \int_0^T \sum_{ij} w_{i0} w_{j0} \rho_{ij} E \left[ \sqrt{\xi_{it}^t} \sqrt{\xi_{jt}^t} \right] dt \\ &= \frac{1}{T} \int_0^T \sum_{ij} w_{i0} w_{j0} \rho_{ij} \sqrt{\xi_{i0}^t} \sqrt{\xi_{j0}^t} h_{ij}(t) dt \end{aligned}$$

with  $h_{ij}(\tau)$  given by:  $h_{ii}(\tau) = 1$ ,  $h_{i \neq j}(\tau) = h(\tau)$ , where  $h(\tau)$  reads:

$$h(\tau) = e^{-\nu^2 \alpha_\theta^2 (1-\chi_{\sigma\sigma}) \left( (1-\theta)^2 \frac{1-e^{-2k_1\tau}}{2k_1} + \theta^2 \frac{1-e^{-2k_2\tau}}{2k_2} + 2\theta(1-\theta)\rho_{X_1 X_2} \frac{1-e^{-(k_1+k_2)\tau}}{k_1+k_2} \right)} \quad (11.29)$$

For flat and identical term-structures of VS volatilities:

$$(\hat{\sigma}_T^B)^2 = \hat{\sigma}^2 \frac{1 + (n-1)\rho_{SS} \frac{1}{T} \int_0^T h(\tau) d\tau}{n} \quad (11.30)$$

### 11.3.2 Numerical examples

We now test the accuracy of formulas (11.24), (11.30) for  $S_T^B$ ,  $\hat{\sigma}_T^B$  with the two-factor model.

To generate stock-like parameters, we start from the index-like parameters in Table 8.2, page 329, and reduce  $\rho_{SX^1}$ ,  $\rho_{SX^2}$  by about 30% so as to reduce the ATM skew by the same relative amount. The other parameters are left unchanged. The basket component's parameters we obtain appear in Table 11.3.

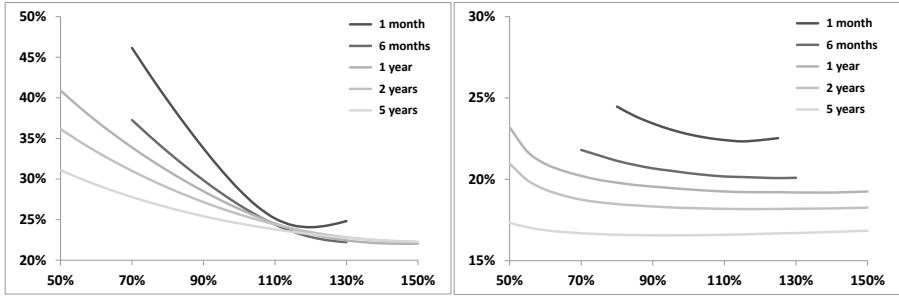
$\nu$	$\theta$	$k_1$	$k_2$	$\rho_{X^1 X^2}$	$\rho_{SX^1}$	$\rho_{SX^2}$
174%	0.245	5.35	0.28	0%	-53.0%	-33.9%

Table 11.3: Basket component's parameters in the two-factor model.

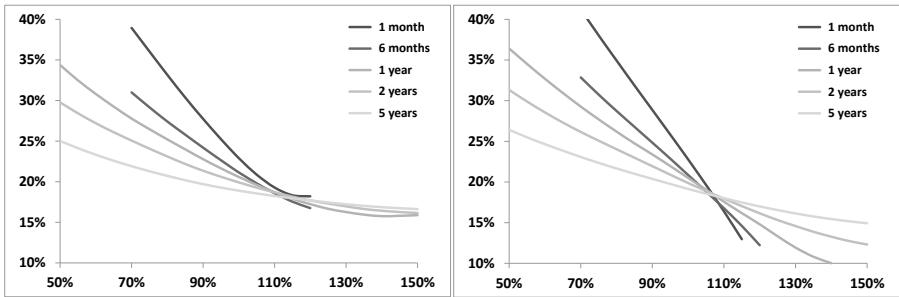
The component's term-structure of volatilities of volatilities is that corresponding to Set II in Figure 7.2, page 229. The VS term structure of volatilities is flat at 30%. We take  $n = 10$  components and will use  $\rho_{SS} = 60\%$  for the component's correlations. We use vanishing rate and repo.

Let us first set  $\chi_{S\sigma} = 0$ . The basket smile – with  $\chi_{\sigma\sigma} = 0$  – is shown in Figure 11.2, along with the component's smile for comparison. The basket ATM skew almost vanishes – a reflection of the fact that the basket ATM skew is mostly generated by cross spot/volatility correlations.

We now use more reasonable values for  $\chi_{S\sigma}$ . Figure 11.3 shows basket smiles generated with  $(\chi_{S\sigma} = 80\%, \chi_{\sigma\sigma} = 80\%)$  and  $(\chi_{S\sigma} = 115\%, \chi_{\sigma\sigma} = 80\%)$ .



**Figure 11.2:** Left: component's smile with parameters of Table 11.3. Right: basket smile with  $\rho_{SS} = 60\%$ ,  $\chi_{S\sigma} = 0$  and  $\chi_{\sigma\sigma} = 0$ .



**Figure 11.3:** Basket smile with:  $\rho_{SS} = 60\%$ ,  $\chi_{\sigma\sigma} = 80\%$ . Left:  $\chi_{S\sigma} = 80\%$ . Right:  $\chi_{S\sigma} = 115\%$ .

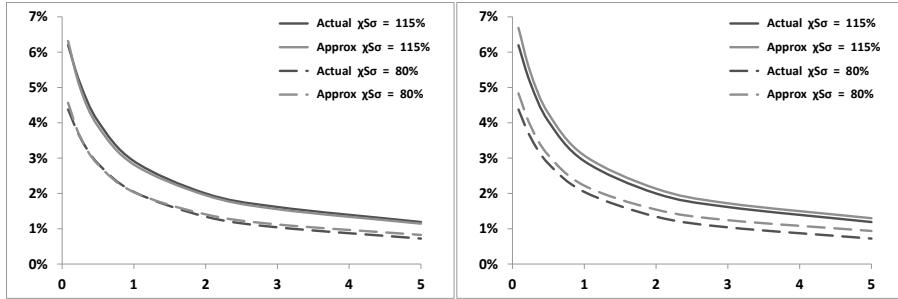
Comparing with the left-hand graph in Figure 11.2 it is apparent that the basket ATM skew with  $(\chi_{S\sigma} = 115\%, \chi_{\sigma\sigma} = 80\%)$  is now steeper than the component's ATM skew.

How accurate are formulas for (11.24), (11.30) for  $S_T^B, \hat{\sigma}_T^B$ ?

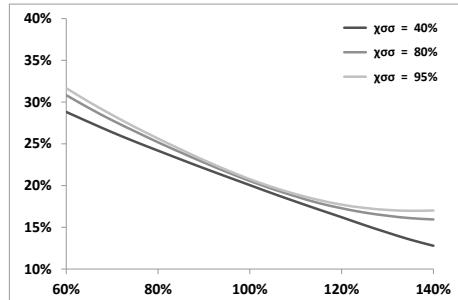
### Basket ATM skew

Figure (11.4) shows the 95/105 ATM basket skew ( $\hat{\sigma}_{K=0.95S_0,T}^B - \hat{\sigma}_{K=1.05S_0,T}^B$ ), either calculated in a Monte Carlo simulation, or evaluated using (11.24), with  $S_T$  given either by the actual component's ATM skew, or given by (11.27), for two different values of  $\chi_{S\sigma}$ . The left-hand graph confirms that expression (11.24) is very accurate. The small overestimation of  $S_T^B$  in the right-hand graph is due to the slight overestimation of the component's skew in (11.27), evidenced in Figure 8.3, page 331.

Figure (11.5) provides another confirmation that the basket ATM skew is controlled by  $\chi_{S\sigma}$ . Here we have varied  $\chi_{\sigma\sigma}$  while keeping  $\rho_{SS} = 60\%$  and  $\chi_{S\sigma} = 80\%$  fixed. The ATM skew hardly changes when  $\chi_{\sigma\sigma}$  is varied. For the three values



**Figure 11.4:** Left: basket 95/105 skew in volatility points (Actual) compared to formula (11.24) (Approx) where the actual component's 95/105 skew has been used. Right: basket 95/105 skew (Actual) compared to formula (11.28) (Approx). Maturities are in years. The two values of  $\chi_{S\sigma} = 80\%$ ,  $115\%$  have been used. All other parameters are kept constant, including  $\chi_{\sigma\sigma} = 80\%$ .



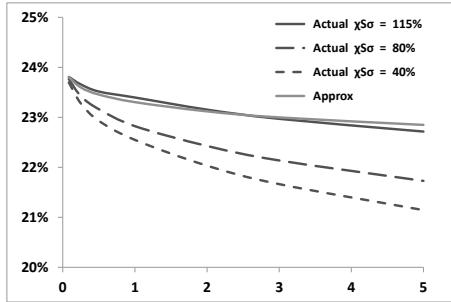
**Figure 11.5:** One-year basket smiles for different values of  $\chi_{\sigma\sigma}$ . All other parameters are kept constant:  $\rho_{SS} = 60\%$  and  $\chi_{S\sigma} = 80\%$ .

of  $\chi_{\sigma\sigma}$  used: 40%, 80%, 95%, the 95/105 one-year skew values are respectively: 2.01%, 2.05%, 2.06%.

Using identical and flat term structures of VS volatilities for the basket components, as well as identical sets of parameters, leads to the particularly simple formulas (11.24) and (11.28) but the derivation of  $\mathcal{S}_T^B$  in the general case presents no particular difficulty. One only needs to express  $\mu_B(t, u)$  as a function of the component's spot/volatility covariance functions  $\mu(t, u, \xi)$ , which are given by expression (8.50) in the two-factor model.

### Basket VS volatility

$\hat{\sigma}_T^B$ , either evaluated in a Monte Carlo simulation or given by (11.30), is graphed in Figure 11.6 for different values of  $\chi_{S\sigma}$ . In expression (11.30), with the approximation



**Figure 11.6:**  $\hat{\sigma}_T^B$ , either evaluated in a Monte Carlo simulation of the two-factor model, for three values of  $\chi_{S\sigma}$ : 40%, 80%, 115%, or given by expression (11.30). Maturities are in years. The component's parameters are listed in Table 11.3 and  $\rho_{SS} = 60\%$ ,  $\chi_{\sigma\sigma} = 80\%$ .

of frozen weights, only spot/spot ( $\rho_{SS}$ ) and volatility/volatility ( $\chi_{\sigma\sigma}$ ) correlations appear.

Figure 11.6 makes it clear that this assumption is not adequate:  $\hat{\sigma}_T^B$  does depend on cross spot/volatility correlations.

### 11.3.3 Mimicking the local volatility model

Can our two-factor stochastic volatility model mimic a multi-asset local volatility model? In the local volatility model,  $\rho^{\text{cross}}(S, \hat{\sigma}_T) = -\rho_{SS}$  and  $\rho^{\text{cross}}(\hat{\sigma}_T, \hat{\sigma}_{T'}) = \rho_{SS}$ . We have:  $\rho^{\text{diag}}(S, \hat{\sigma}_T) = -1$  and  $\rho^{\text{diag}}(\hat{\sigma}_T, \hat{\sigma}_{T'}) = 1$ .<sup>7</sup> Thus, in the local volatility model:

$$\begin{aligned}\rho^{\text{cross}}(S, \hat{\sigma}_T) &= \rho_{SS} \rho^{\text{diag}}(S, \hat{\sigma}_T) \\ \rho^{\text{cross}}(\hat{\sigma}_T, \hat{\sigma}_{T'}) &= \rho_{SS} \rho^{\text{diag}}(\hat{\sigma}_T, \hat{\sigma}_{T'})\end{aligned}$$

Thus, with regard to spot/volatility and volatility/volatility correlations, the local volatility model can be viewed as a particular breed of multi-asset stochastic volatility with  $\chi_{S\sigma}$ ,  $\chi_{\sigma\sigma}$  given by:

$$\chi_{S\sigma} = \rho_{SS} \quad (11.31a)$$

$$\chi_{\sigma\sigma} = \rho_{SS} \quad (11.31b)$$

With this choice of cross-parametrization the values of  $\zeta$  in (11.17) are both equal to 1, thus positivity conditions are satisfied.

---

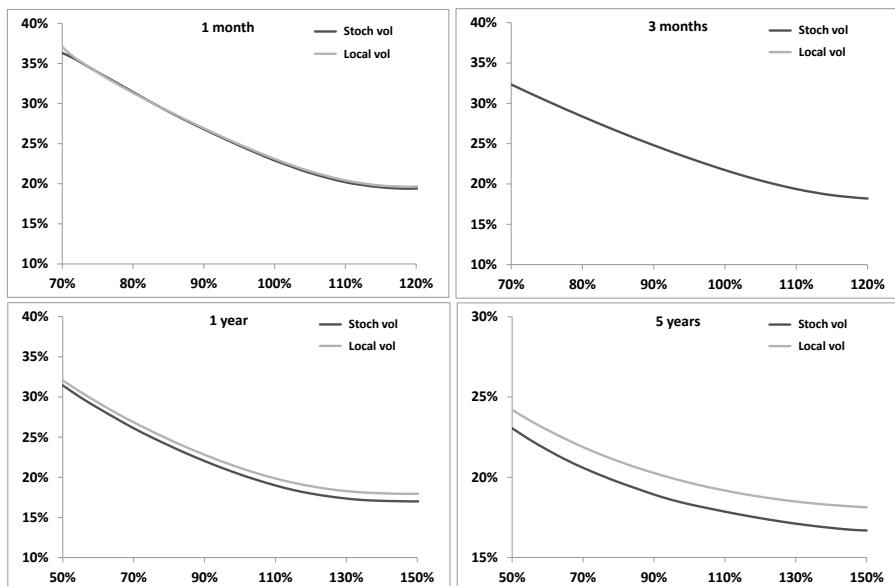
<sup>7</sup>We use here  $\rho^{\text{diag}}(S, \hat{\sigma}_T) = -1$  as we are considering the typical case of negatively sloping equity smiles. We could equivalently have written  $\rho^{\text{diag}}(S, \hat{\sigma}_T) = 1$ .

It is easy to check that in case spot/spot correlations are not all equal, the correlation structure of the multi-asset local volatility model is still given by (11.31) with  $\rho_{SS}$ ,  $\chi_{S\sigma}$ ,  $\chi_{\sigma\sigma}$  adorned with  $ij$  superscripts, as  $\rho^{\text{cross}}(S^i, \hat{\sigma}_T^j) = -\rho_{SS}^{ij}$  and  $\rho^{\text{cross}}(\hat{\sigma}_T^i, \hat{\sigma}_{T'}^j) = \rho_{SS}^{ij}$ .

We now test this mapping with the same parameter values used in Section 11.3.2;  $\rho_{SS} = 60\%$ . We use two-factor-model parameters in Table 11.3 to generate a vanilla smile. Using this as the component's vanilla smile, we calibrate the component's local volatility function and price the basket smile using  $\rho_{SS} = 60\%$ .

We then compare the resulting basket smile with that produced by our multi-asset two-factor model with  $\chi_{\sigma\sigma} = \chi_{S\sigma} = \rho_{SS} = 60\%$ . Results appear in Figure 11.7.

It is apparent that the mapping (11.31) is accurate for short maturities; for longer maturities implied volatilities in the stochastic volatility model are lower than in the local volatility model. This is not surprising: because volatilities – especially forward volatilities – are more volatile in the stochastic volatility model than in the local volatility model: the level of “effective” spot/spot correlation is lower than  $\rho_{SS}$ . This affects all multi-asset products, including the correlation swap which we now study.



**Figure 11.7:** Comparison of basket smiles in local volatility and stochastic volatility models, for several maturities. The mapping of local to stochastic volatility is obtained with  $\chi_{\sigma\sigma} = \chi_{S\sigma} = \rho_{SS} = 60\%$ . The basket consists of  $n = 10$  components, with identical smiles given by parameters in Table 11.3.

## 11.4 The correlation swap

Consider a basket of stocks or indexes. A correlation swap of maturity  $T$  pays at  $T$  the average pairwise realized correlation of the basket components minus a fixed strike  $\hat{\rho}$

$$\frac{1}{n(n-1)} \sum_{i \neq j} \rho_{ij} - \hat{\rho} \quad (11.32)$$

where  $\hat{\rho}$  is set so that the swap's initial value vanishes. Correlation  $\rho_{ij}$  of  $S_i, S_j$  is defined with the standard realized correlation estimator, using daily log-returns:

$$\rho_{ij} = \frac{\sum r_k^i r_k^j}{\sqrt{\sum r_k^i} \sqrt{\sum r_k^j}} \quad (11.33)$$

where  $r_k^i = \ln(S_k^i / S_{k-1}^i)$  and the sums runs from  $k = 1$  to  $k = N$ , where  $N$  is the number of returns used for estimating covariances and variances in (11.33).

$n$  is the number of securities: 2 or 3 when the components are indexes and up to 50 for a correlation swap on the constituents of the Euro Stoxx 50 index. We have used in (11.32) the typical equal weighting.

Strike  $\hat{\rho}$  is also called the implied correlation of the swap. Indeed, in a constant volatility model, for daily returns, with all spot/spot correlations  $\rho_{SS}$  equal and a large number of returns – thus a long maturity –  $\hat{\rho} = \rho_{SS}$ .

For shorter maturities, (11.33) is biased. Let us make the assumption of centered log-returns:  $r_k^i = \sigma^i \sqrt{\Delta t} Z_k^i$ , where  $\sigma^i$  is the (constant) volatility of asset  $i$ ,  $\Delta t$  the interval between two spot observations – here one day, and the  $Z_k^i$  are iid standard normal random variables.

The bias and standard deviation of the correlation estimator are derived in Appendix A, at lowest order in  $\frac{1}{N}$ :

$$E[\rho_{ij}] = \rho_{SS} \left( 1 - \frac{1 - \rho_{SS}^2}{2N} \right) \quad (11.34)$$

$$\text{Stdev}(\rho_{ij}) = \frac{1}{\sqrt{N}} (1 - \rho_{SS}^2) \quad (11.35)$$

where  $N$  is the number of returns in the historical sample. Typically the bias in  $E[\rho_{ij}]$  is about one point of correlation for  $\rho_{SS} = 60\%$  and  $T = 1$  month.

Correlation swaps were introduced as a means of trading correlation and making implied correlation an observable parameter.

As a measure of correlation, however, averaging pairwise correlations results in a poorly defined estimator.

One would expect of an adequately defined estimator that its standard deviation vanishes either in the limit of a large sample size ( $N \rightarrow \infty$ ) or as the number of

basket constituents is increased ( $n \rightarrow \infty$ ). This is not the case for the correlation swap estimator.

Making the same assumption of constant volatility centered log-returns as above, the bias and standard deviation of the correlation swap estimator in the limit  $n \rightarrow \infty$  are given, at lowest order in  $\frac{1}{N}$ , by:

$$\begin{aligned} \lim_{n \rightarrow \infty} E\left[\frac{1}{n(n-1)} \sum_{i \neq j} \rho_{ij}\right] &= \rho_{SS} \left(1 - \frac{1 - \rho_{SS}^2}{2N}\right) \\ \lim_{n \rightarrow \infty} \text{Stdev}\left(\frac{1}{n(n-1)} \sum_{i \neq j} \rho_{ij}\right) &= \frac{1}{\sqrt{N}} (1 - \rho_{SS}^2) \frac{\sqrt{2} |\rho_{SS}|}{1 + \rho_{SS}} \end{aligned} \quad (11.36)$$

Observe that the right-hand side of (11.36) ( $n = \infty$ ) is only marginally smaller than that of (11.35) ( $n = 2$ ). For the sake of creating an estimator of average realized correlation, it would have been more judicious to average covariances and divide them by the average of variances – the standard deviation of the resulting correlation estimator tends to zero as the number of constituents is increased.

One would hope that the correlation swap is the perfect instrument for calibrating spot/spot correlation levels, but it is not so. Let us assume that our basket is homogeneous and let  $\sigma_t^i$  be the instantaneous (lognormal) volatility of  $S_t^i$ . In the limit of short returns,  $\hat{\rho}$  is given by:

$$\hat{\rho}(T) = E[\rho_{ij}] = \rho_{SS} E\left[\frac{\int_0^T \sigma_t^i \sigma_t^j dt}{\sqrt{\int_0^T \sigma_t^i{}^2 dt} \sqrt{\int_0^T \sigma_t^j{}^2 dt}}\right] \quad (11.37)$$

where we explicitly keep track of the maturity dependence of  $\hat{\rho}$ .

In case instantaneous volatilities  $\sigma_t^i$  and  $\sigma_t^j$  are collinear –  $\sigma_t^j = \lambda \sigma_t^i$  – then  $\hat{\rho}$  is indeed equal to  $\rho_{SS}$ . In all other cases,  $\hat{\rho} \leq \rho_{SS}$ , and the amount by which  $\hat{\rho}$  differs from  $\rho_{SS}$  depends on volatilities and correlations of instantaneous volatilities. Instantaneous volatilities are unobservable quantities – how does this manifest itself practically?

### Risk-managing the correlation swap in the Black-Scholes model

Consider risk-managing a correlation swap on two underlyings  $S^1, S^2$  with the Black-Scholes model. Since the dollar gammas and vegas of the correlation swap do not depend on  $S^1, S^2$ , it is natural to use variance swaps as hedge instruments.

Denote by  $\sigma_{1t}, \sigma_{2t}, \rho_t$  the realized volatilities and correlation of log returns over  $[0, t]$  and  $\hat{\sigma}_{1t}, \hat{\sigma}_{2t}$  implied VS volatilities at time  $t$  for maturity  $T$ , which we are using as implied volatilities in our Black-Scholes model. Denote by  $\hat{\rho}_0$  the (implied) correlation level at which we are risk-managing the correlation swap, which we keep constant.

The value at  $t$  of the correlation swap in the Black-Scholes model is given by:<sup>8</sup>

$$P = \frac{t\rho_t\sigma_{1t}\sigma_{2t} + (T-t)\hat{\rho}_0\hat{\sigma}_{1t}\hat{\sigma}_{2t}}{\sqrt{t\sigma_{1t}^2 + (T-t)\hat{\sigma}_{1t}^2}\sqrt{t\sigma_{2t}^2 + (T-t)\hat{\sigma}_{2t}^2}}$$

At  $t = 0$ ,  $P = \hat{\rho}_0$  and  $\frac{dP}{d(\hat{\sigma}_{1t}^2)} = \frac{dP}{d(\hat{\sigma}_{2t}^2)} = 0$ : the correlation swap has no sensitivity to implied volatilities at inception. We use variances rather than volatilities as state variables, since variance swaps provide perfect delta hedges with respect to variances. Let us now examine the correlation swap's sensitivities at later times.

Consider a short position in a correlation swap vega-hedged with variance swaps of maturity  $T$ . Let us calculate the P&L over  $[t, t + \delta t]$  generated by variations  $\delta(\hat{\sigma}_{1t}^2)$ ,  $\delta(\hat{\sigma}_{2t}^2)$  of implied VS variances, at order two. Computing second derivatives of  $P$  with respect to  $\hat{\sigma}_{1t}^2$ ,  $\hat{\sigma}_{2t}^2$  is straightforward.

Consider the particular situation when realized volatilities and correlation over  $[0, t]$  match the implied values at time  $t$ :  $\sigma_{1t} = \hat{\sigma}_{1t}$ ,  $\sigma_{2t} = \hat{\sigma}_{2t}$ ,  $\rho_t = \hat{\rho}_0$  – this allows for a more compact expression or the P&L. The P&L at order two in  $\hat{\sigma}_{1t}^2$ ,  $\hat{\sigma}_{2t}^2$  – for a short position – reads:

$$\begin{aligned} P\&L &= \frac{\hat{\rho}_0}{8} \frac{t(T-t)}{T^2} \left( \frac{(\delta(\hat{\sigma}_{1t}^2))^2}{(\hat{\sigma}_{1t}^2)^2} + \frac{(\delta(\hat{\sigma}_{2t}^2))^2}{(\hat{\sigma}_{2t}^2)^2} - 2 \frac{\delta(\hat{\sigma}_{1t}^2)\delta(\hat{\sigma}_{2t}^2)}{\hat{\sigma}_{1t}^2\hat{\sigma}_{2t}^2} \right) \\ &= \frac{\hat{\rho}_0}{8} \frac{t(T-t)}{T^2} \left( \frac{\delta(\hat{\sigma}_{1t}^2)}{\hat{\sigma}_{1t}^2} - \frac{\delta(\hat{\sigma}_{2t}^2)}{\hat{\sigma}_{2t}^2} \right)^2 \end{aligned} \quad (11.38)$$

Had we used different values for  $\sigma_{1t}$ ,  $\sigma_{2t}$ ,  $\rho_t$ , (11.38) would have been replaced with a more complicated quadratic form of  $\delta(\hat{\sigma}_{1t}^2)$ ,  $\delta(\hat{\sigma}_{2t}^2)$ , but the prefactor  $\frac{t(T-t)}{T^2}$  would have remained.

(11.38) confirms – for the special case of realized volatilities and correlations at time  $t$  matching their implied values – that if  $\frac{\delta\hat{\sigma}_{1t}}{\hat{\sigma}_{1t}} = \frac{\delta\hat{\sigma}_{2t}}{\hat{\sigma}_{2t}}$  no P&L is generated by the variation of VS volatilities.<sup>9</sup>

Correlation swaps are thus exotic volatility instruments that depend on the dynamics of forward variances. Practically, this manifests itself through volatility/volatility cross-gamma P&Ls.

### 11.4.1 Approximate formula in the two-factor model

We now compute  $\hat{\rho}(T)$  in (11.37) at order two in volatility of volatility in the two-factor model, for two underlyings with the same parameters, with a flat VS term

<sup>8</sup>We have made the assumption of very short returns, so that the contribution from the risk-neutral drift of  $\ln S_t$  is negligible – this is adequate for daily returns of equities.

<sup>9</sup>In contrast with options on realized variance, there is no way that the vega hedge of the correlation swap may also function as gamma hedge since VSs on  $S^1$  and  $S^2$  generate gamma P&Ls proportional to  $(\delta S^1)^2$  and  $(\delta S^2)^2$  while the gamma P&L of the correlation swap also includes a  $\delta S^1 \delta S^2$  term. In addition to the P&L in (11.38), our P&L thus comprises a spreaded position of diagonal spot gammas against spot cross-gammas, accompanied by their respective thetas.

structure, in the limit of short returns. The instantaneous variance  $\xi_t^t$  is given by:

$$\xi_t^t = \xi_0^t e^{2\nu x_t^t - 2\nu^2 E[(x_t^t)^2]}$$

where  $x_t^T$  has been defined in (7.30), page 226. For  $T = t$ :

$$x_t^t = \alpha_\theta [(1 - \theta) X_t^1 + \theta X_t^2] \quad (11.39)$$

and  $\alpha_\theta = ((1 - \theta)^2 + \theta^2 + 2\rho_{X_1 X_2} \theta(1 - \theta))^{-1/2}$ . Remember  $\nu$  is the volatility of a very short volatility. Expanding at order two in  $\nu$ :

$$\xi_t^t = \xi_0^t (1 + 2\nu x_t^t + 2\nu^2 [(x_t^t)^2 - E[(x_t^t)^2]]) \quad (11.40)$$

In an expansion of  $\hat{\rho}(T)$  at order two in  $\nu$ , the  $\nu^2$  term in (11.40) is multiplied by a constant: when evaluating its expectation its contribution vanishes. For the sake of calculating  $\hat{\rho}(T)$  at order two in  $\nu$ , the expansion of  $\xi_t^t$  at order one is thus sufficient:

$$\xi_t^t = \xi_0^t (1 + 2\nu x_t^t)$$

We consider two underlyings each with a flat VS term structure and denote by  $x_t, y_t$  the  $x_t^t$  processes for, respectively, the first and second underlying. At second order in  $\nu$ ,  $\hat{\rho}(T)$  reads:

$$\hat{\rho}(T) = \rho_{SS} E \left[ \frac{\int_0^T \sqrt{1 + 2\nu x_t} \sqrt{1 + 2\nu y_t} dt}{\sqrt{\int_0^T (1 + 2\nu x_t) dt} \sqrt{\int_0^T (1 + 2\nu y_t) dt}} \right]$$

Working out the expansion at order two in  $\nu$  we get:

$$\begin{aligned} \hat{\rho}(T) &= \rho_{SS} E \left[ \frac{1 + \nu (\bar{x} + \bar{y}) - \frac{\nu^2}{2} \overline{(x - y)^2}}{1 + \nu (\bar{x} + \bar{y}) - \frac{\nu^2}{2} \overline{(\bar{x} - \bar{y})^2}} \right] \\ &= \rho_{SS} E \left[ 1 - \frac{\nu^2}{2} \left( \overline{(x - y)^2} - \overline{x - y}^2 \right) \right] \end{aligned}$$

which at order two in  $\nu$  can be rewritten as:

$$\hat{\rho}(T) = \rho_{SS} e^{-\frac{\nu^2}{2} E[(\bar{x} - \bar{y})^2]} \quad (11.41)$$

where:

$$\begin{aligned} \bar{x} &= \frac{1}{T} \int_0^T x_t dt \\ \overline{(x - y)^2} &= \frac{1}{T} \int_0^T (x_t - y_t)^2 dt \\ \overline{x - y}^2 &= \left( \frac{1}{T} \int_0^T (x_t - y_t) dt \right)^2 \end{aligned}$$

Using the exponential in (11.41) ensures that for large values of  $\nu$ ,  $\hat{\rho}$  at most vanishes, but does not become negative. From the definitions above we have the pathwise inequality  $(x - y)^2 \geq \overline{x - y}^2$  thus  $E[\overline{(x - y)^2}] \geq E[\overline{x - y}^2]$ .

(11.41) thus also ensures that the property  $\hat{\rho} \leq \rho_{SS}$  that expression (11.37) for  $\hat{\rho}(T)$  implies, is obeyed in our expansion at second order in  $\nu$ . If forward variances for the two underlyings are collinear processes –  $x_t = y_t$  – we recover that  $\hat{\rho} = \rho_{SS}$ .

Carrying out the computation for the two-factor model with  $x_t^t$  given by (11.39) yields:

$$\hat{\rho}(T) = \rho_{SS} e^{-(1-\chi_{\sigma\sigma})(\nu^2 T) \frac{1}{T} \int_0^T f(t) dt} \quad (11.42)$$

with the dimensionless  $f(t)$  given by:

$$\begin{aligned} f(t) = & \alpha_\theta^2 \left[ (1-\theta)^2 \frac{1-e^{-2k_1 t}}{2k_1 T} \left( 1 - 2 \frac{1-e^{-k_1(T-t)}}{k_1 T} \right) \right. \\ & + \theta^2 \frac{1-e^{-2k_2 t}}{2k_2 T} \left( 1 - 2 \frac{1-e^{-k_2(T-t)}}{k_2 T} \right) \\ & \left. + 2\rho_{X_1 X_2} \theta (1-\theta) \frac{1-e^{-(k_1+k_2)t}}{(k_1+k_2)T} \left( 1 - \frac{1-e^{-k_1(T-t)}}{k_1 T} - \frac{1-e^{-k_2(T-t)}}{k_2 T} \right) \right] \end{aligned}$$

### 11.4.2 Examples

We use parameter values in Table 11.3, page 435, for each of our two underlyings, which correspond to the smiles in Figure 11.2. We take  $\rho_{SS} = 60\%$  and use three different values for  $\chi_{\sigma\sigma}$ : 40%, 60%, 80%.

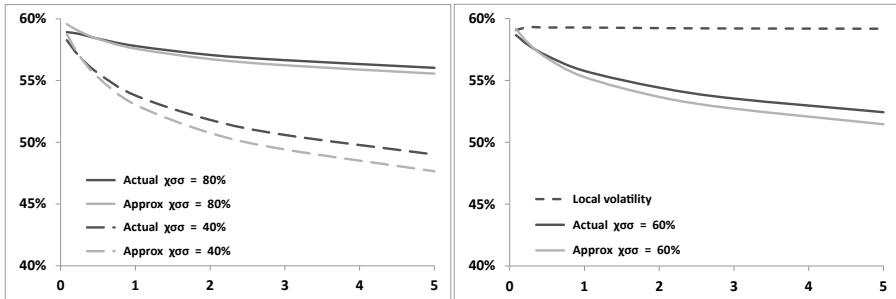
$\hat{\rho}(T)$  appears in Figure 11.8, both evaluated numerically in a Monte Carlo simulation, and as given by formula (11.42). The right-hand side of Figure 11.8 shows the same curves for  $\chi_{\sigma\sigma} = 60\%$ , together with the values generated by a local volatility model parametrized as in the previous section, that is calibrated on the smile generated by the two-factor model, and with  $\rho_{SS} = 60\%$ .

Observe first that the accuracy of the second-order expansion in volatility of volatility is satisfactory, even though it is systematically biased low, except for short maturities. For short maturities  $\hat{\rho}(T)$  in (11.42) surges above the exact value: this is due to the fact that we have carried out our expansion in the continuous-time version of the correlation swap.

Thus  $\hat{\rho}(T)$  in (11.42) is not subject to the bias in (11.34), which instead affects the actual Monte Carlo estimation – as it uses discrete daily returns – and is material for small values of  $N$ , that is short maturities.

The right-hand graph is a compelling illustration of how different a local volatility and a stochastic volatility model can be, even though they are calibrated on the same smile.

In the local volatility model,  $\hat{\rho}(T)$  hardly depends on  $T$ : this is likely due to the fact that forward variances – especially long-dated ones – have very little volatility.



**Figure 11.8:** Left:  $\hat{\rho}(T)$  for maturities  $T = 1$  month to 5 years, evaluated in a Monte Carlo simulation (Actual) and with formula (11.42) at second order in volatility of volatility for  $\chi_{\sigma\sigma} = 80\%$  and  $\chi_{\sigma\sigma} = 40\%$ . Right: same curves for  $\chi_{\sigma\sigma} = 60\%$ , along with  $\hat{\rho}(T)$  given by the local volatility model, calibrated on the smile generated by the two-factor model with parameters in Table 11.3.

Contrast this with the graphs in Figure 11.7. While for the basket smile the mapping  $\chi_{\sigma\sigma} = \chi_{S\sigma} = \rho_{SS} = 60\%$  makes both models equivalent, they are not with respect to the correlation swap.

Figure 11.8 shows that the longer the maturity, the larger the impact of volatility of volatility, thus the lower  $\hat{\rho}(T)$ . This deviates from market practice: typically the term structure of implied correlation swap correlations rises, rather than decreasing.

Finally, we have taken  $\chi_{S\sigma} = 80\%$  in our numerical tests; using different values results, as expected, in no change of the actual and approximate values of  $\hat{\rho}(T)$ .

## 11.5 Conclusion

- Life in the multi-asset local volatility framework used to be simple. Say we chose  $\rho_{SS} = 60\%$  as correlation for the driving Brownian motions, priced multi-asset options and backed out the implied correlation in a multi-asset Black-Scholes model. Depending on the payoff at hand – say, an ATM basket call option, a VS on the basket, a correlation swap, a forward on the worst-performing asset – we would get different numbers, but clustered within a small range below 60%.
- This is not so with multi-asset stochastic volatility: exotic payoffs are sensitive to spot/volatility and volatility/volatility correlation levels, which, unlike in the local volatility model, can now be set separately. The richness of correlation risks manifests itself in the fact that implied spot/spot correlations backed

out in a multi-asset Black-Scholes model for different payoffs will now vary widely.

- In our simple parametrization, in addition to  $\rho_{SS}$ , we introduce two quantities:  $\chi_{S\sigma}$  and  $\chi_{\sigma\sigma}$ . By choosing  $\chi_{\sigma\sigma} = \chi_{S\sigma} = \rho_{SS}$  we are able to mimic the local volatility model, for basket European payoffs, for short maturities.
- Do  $\chi_{S\sigma}$  and  $\chi_{\sigma\sigma}$  have implied counterparts? We have shown that only a small fraction of the basket skew is contributed by the components' skew; the bulk of it is determined by  $\chi_{S\sigma}$ . Thus, the ratio of basket-to-component skew is a measure of the implied value of  $\chi_{S\sigma}$ , once  $\rho_{SS}$  has been set. Similarly  $\chi_{\sigma\sigma}$  determines the strike of the correlation swap, again once  $\rho_{SS}$  has been set.

It would be useful to have good approximations for the basket VS and ATMF volatilities. It doesn't help that baskets are typically of the arithmetic – rather than geometric – type: the weights of each asset's return in the basket's return are not constant. Thus even simple objects such as basket VS and ATMF volatilities depend in a complicated way on  $\rho_{SS}$ ,  $\chi_{S\sigma}$ ,  $\chi_{\sigma\sigma}$ , let alone payoffs on worst-ofs.

## Appendix A – bias/standard deviation of the correlation estimator

We present here the derivation of formulas (11.34) and (11.35), for the case of Gaussian-distributed returns with constant volatility. Since volatility is constant, we can normalize returns by their standard deviation and evaluate correlations with normalized returns. Consider the  $i$ -th and  $j$ -th underlyings and denote by  $x_i^\tau$ ,  $x_j^\tau$  their  $\tau$ -th daily return.  $x_i^\tau$ ,  $x_j^\tau$  are assumed to be iid standard normal random variables with  $\langle x_i^\tau x_j^\tau \rangle = \rho$  if  $i \neq j$  and  $\langle x_i^\tau x_i^\tau \rangle = 1$  where  $\rho$  is the correlation between the two underlyings.

The correlation estimator for the  $(i, j)$  couple is:

$$\hat{\rho}_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}$$

with

$$C_{ij} = x_i \cdot x_j \equiv \frac{1}{N} \sum_{\tau} x_i^\tau x_j^\tau$$

where  $N$  is the number of returns in our historical sample.  $C_{ij}$  is the average of independent random variables  $x_i^\tau x_j^\tau$ . As  $N \rightarrow \infty$ , fluctuations around its mean  $\langle C_{ij} \rangle = \rho$  tend to zero, so let us write

$$\begin{aligned} C_{i \neq j} &= \rho + \varepsilon_{ij} & \varepsilon_{ij} &= x_i \cdot x_j - \rho \\ C_{ii} &= 1 + \varepsilon_{ii} & \varepsilon_{ii} &= x_i \cdot x_i - 1 \end{aligned}$$

We have:

$$\hat{\rho}_{ij} = \frac{\rho + \varepsilon_{ij}}{\sqrt{1 + \varepsilon_{ii}} \sqrt{1 + \varepsilon_{jj}}}$$

We now expand  $\hat{\rho}_{ij}$  in powers of fluctuations  $\varepsilon_{ij}$ . By construction the  $\varepsilon_{ij}$  are centered:  $\langle \varepsilon_{ij} \rangle = 0$ . The first non-trivial contribution to the bias of  $\hat{\rho}_{ij}$  is thus generated by second-order terms. At second order in the  $\varepsilon_{ij}$  we have:

$$\hat{\rho}_{ij} = \rho - \left( \rho \frac{\varepsilon_{ii} + \varepsilon_{jj}}{2} - \varepsilon_{ij} \right) - \left( \varepsilon_{ij} \frac{\varepsilon_{ii} + \varepsilon_{jj}}{2} - \rho \left( \frac{3}{8} (\varepsilon_{ii}^2 + \varepsilon_{jj}^2) + \frac{1}{4} \varepsilon_{ii} \varepsilon_{jj} \right) \right) \quad (11.43)$$

At this order, the expectation of the correlation estimator is given by:

$$\langle \hat{\rho}_{ij} \rangle = \rho - \left\langle \varepsilon_{ij} \frac{\varepsilon_{ii} + \varepsilon_{jj}}{2} - \rho \left( \frac{3}{8} (\varepsilon_{ii}^2 + \varepsilon_{jj}^2) + \frac{1}{4} \varepsilon_{ii} \varepsilon_{jj} \right) \right\rangle$$

This formula involves moments of Gaussian variables  $x_i^\tau, x_j^\tau$ . The Wick theorem yields the following expressions for 4th order moments:

$$\begin{aligned} \langle \varepsilon_{ii}^2 \rangle &= \frac{2}{N} & \langle \varepsilon_{ii} \varepsilon_{jj} \rangle &= \frac{2\rho^2}{N} & \langle \varepsilon_{ii} \varepsilon_{ij} \rangle &= \frac{2\rho}{N} & \langle \varepsilon_{ij}^2 \rangle &= \frac{1+\rho^2}{N} \\ \langle \varepsilon_{ij} \varepsilon_{kk} \rangle &= \frac{2\rho^2}{N} & \langle \varepsilon_{ij} \varepsilon_{kl} \rangle &= \frac{2\rho^2}{N} & \langle \varepsilon_{ij} \varepsilon_{il} \rangle &= \frac{\rho+\rho^2}{N} \end{aligned}$$

where indices  $i, j, k, l$  are all different.

We then get, from (11.43):

$$\begin{aligned} \langle \hat{\rho}_{ij} \rangle &= \rho - \left( \frac{2\rho}{N} - \rho \left( \frac{3}{8} \frac{4}{N} + \frac{1}{4} \frac{2\rho^2}{N} \right) \right) \\ &= \rho \left( 1 - \frac{1-\rho^2}{2N} \right) \end{aligned}$$

which is the result in (11.34).

We now turn to the variance of  $\hat{\rho}_{ij}$ . The first non-trivial contribution is generated by the square of the order-one correction to  $\rho$  in (11.43), and is of order  $\frac{1}{N}$ . For the sake of calculating the variance of  $\hat{\rho}_{ij}$  we can then ignore the bias just derived, which generates a contribution of order  $\frac{1}{N^2}$ . Using the expressions of 4th order moments above, we have:

$$\left\langle \left( \rho \frac{\varepsilon_{ii} + \varepsilon_{jj}}{2} - \varepsilon_{ij} \right)^2 \right\rangle = \frac{1}{N} (1 - \rho^2)^2$$

from which (11.35) follows.

Consider now the correlation estimator used in the definition of the correlation swap, namely the average of all pairwise correlation estimators:

$$\hat{\rho} = \frac{1}{n(n-1)} \sum_{i \neq j} \hat{\rho}_{ij}$$

Assuming all pairwise correlations are equal to  $\rho$ , at order one in the fluctuations  $\varepsilon_{ij}$ :

$$\hat{\rho} = \rho - \frac{1}{n(n-1)} \sum_{i \neq j} \left( \rho \frac{\varepsilon_{ii} + \varepsilon_{jj}}{2} - \varepsilon_{ij} \right) \quad (11.44)$$

Calculating now the expectation of the square of the right-hand side of (11.44) and taking the limit  $n \rightarrow \infty$  yields expression (11.36) for the standard deviation of  $\hat{\rho}$ .

## Chapter's digest

### 11.1 The short ATMF basket skew

► We consider the case of an equally weighted basket of  $n$  components. Assuming identical correlations  $\rho_{SS}$  among components, the ATM implied volatility and skew of the basket are given, for short maturities, by:

$$\widehat{\sigma}_B = \sqrt{\frac{1 + (n - 1)\rho_{SS}}{n}} \widehat{\sigma}$$

$$\mathcal{S}_B = \frac{1 + (n - 1)\rho_{SS}}{n} \frac{\widehat{\sigma}^3}{\widehat{\sigma}_B^3} \left[ \frac{\mathcal{S}}{n} + \frac{n - 1}{n} \frac{1}{2\widehat{\sigma}^2} \frac{\langle d \ln S d\widehat{\sigma} \rangle_{\text{cross}}}{dt} \right]$$

where  $\widehat{\sigma}$  and  $\mathcal{S}$  are the components' ATMF volatility and skew, assumed to be identical for all components, and  $\langle d \ln S d\widehat{\sigma} \rangle_{\text{cross}}$  is the cross spot/volatility covariance.

Only a fraction  $\frac{1}{n}$  of the basket ATMF skew is contributed by the ATM skew of the components. The bulk of the basket skew is generated by cross spot/volatility correlations.

► Specializing to the case of a large homogeneous basket:

$$\widehat{\sigma}_B \simeq \sqrt{\rho_{SS}} \widehat{\sigma}$$

$$\mathcal{S}_B \simeq \frac{1}{\sqrt{\rho_{SS}}} \frac{1}{2\widehat{\sigma}^2} \frac{\langle d \ln S d\widehat{\sigma} \rangle_{\text{cross}}}{dt}$$

► For vanishing cross spot/volatility correlations, the basket skew vanishes.

► If cross spot/volatility covariances are identical to their diagonal counterparts,  $\mathcal{S}_B \simeq \frac{1}{\sqrt{\rho_{SS}}} \mathcal{S}$ : the basket skew is larger than the component's.

► In the local volatility model:

$$\widehat{\sigma}_B \simeq \sqrt{\rho_{SS}} \widehat{\sigma}$$

$$\mathcal{S}_B \simeq \sqrt{\rho_{SS}} \mathcal{S}$$

The basket skew is smaller than the component's skew.



### 11.2 Parametrizing multi-asset stochastic volatility models

► We wish to parametrize a multi-asset stochastic volatility model (a) with few additional parameters, (b) in a manner that does not depend on the model's factor structure. We introduce dimensionless numbers  $\chi_{S\sigma}$  and  $\chi_{\sigma\sigma}$  to specify the

cross spot/volatility and volatility/volatility covariance functions, in terms of their diagonal counterparts:

$$\begin{aligned}\rho^{\text{cross}}(S, \hat{\sigma}_T) &= \chi_{S\sigma} \rho^{\text{diag}}(S, \hat{\sigma}_T) \\ \rho^{\text{cross}}(\hat{\sigma}_T, \hat{\sigma}_{T'}) &= \chi_{\sigma\sigma} \rho^{\text{diag}}(\hat{\sigma}_T, \hat{\sigma}_{T'})\end{aligned}$$

For a homogeneous basket, the conditions on  $\chi_{S\sigma}, \chi_{\sigma\sigma}$  so that the global correlation matrix is positive are easily expressed. While  $\chi_{\sigma\sigma} \leq 1$ ,  $\chi_{S\sigma}$  can go above 1.

- Realized values of  $\chi_{S\sigma}$  and  $\chi_{\sigma\sigma}$  can be measured using historical data for spot and implied volatilities.



### 11.3 The ATMF basket skew

- At order one in volatility of volatility, for a homogeneous basket, the ATMF skew is given by:

$$S_T^B = S_T \sqrt{\frac{n}{1 + (n - 1)\rho_{SS}}} \left[ \frac{1 + (n - 1)\chi_{S\sigma}}{n} \right]$$

where  $\rho_{SS}$  is the correlation among basket components and  $S_T$  their ATMF skew.

- Numerical experiment show that this formula for the basket skew is accurate, and that the latter hardly depends on  $\chi_{\sigma\sigma}$ .

- The approximate expression for the basket VS volatility derived, as is the case for  $S_T^B$ , with the assumption of frozen weights, is, on the other hand, not accurate. In reality it does depend on  $\chi_{S\sigma}$ .

- The multi-asset local volatility model is mimicked by setting  $\chi_{S\sigma}$  and  $\chi_{\sigma\sigma}$  equal to the spot/spot correlation.

In practice, while the shape of the resulting smile indeed matches that of the multi-asset local volatility model, the level of basket ATMF volatility is shifted downwards, especially for long maturities.



### 11.4 The correlation swap

- The correlation swap is an exotic volatility instrument. It is very sensitive to the correlation of the volatility processes of the basket's components, thus to parameter  $\chi_{\sigma\sigma}$ .

At lowest order in volatility of volatility, the fair strike of the correlation swap, for the case of a homogeneous pair of underlyings, is given by:

$$\widehat{\rho}(T) = \rho_{SS} e^{-(1-\chi_{\sigma\sigma})(\nu^2 T)^{\frac{1}{T}} \int_0^T f(t) dt}$$

where  $f(t)$  is a simple function that involves model parameters.

- Numerical experiments show that the accuracy of our approximate formula is acceptable. For long maturities, the correlation swap's fair strike lies much lower than the correlation of spot processes. This contrasts with the behavior in the local volatility model, which yields a fair strike almost equal to the correlation of spot processes.

**This page intentionally left blank**

# **Chapter 12**

---

## ***Local-stochastic volatility models***

This chapter should be considered as a natural sequel to Chapter 2 on the local volatility model – which we urge the reader to read if she or he has not done so – and Chapter 7 on forward variance models.

Local-stochastic volatility models are market models that possess a Markovian representation in terms of  $t, S$  plus a few additional state variables.

We begin by motivating their study, then cover their calibration to market smiles before we get to practical modeling issues: what does the carry P&L look like in these models? Which models can be used in trading applications? How can we adjust spot/volatility and volatility/volatility break-even levels?

---

### **12.1 Introduction**

The local volatility model is a market model for the spot and vanilla options, or, equivalently, implied volatilities  $\widehat{\sigma}_{KT}$  – it is covered in detail in Chapter 2. A market model takes as inputs any non-arbitrageable configuration of all hedging instruments – spot and vanilla options – generates a delta on each of them, and is characterized by the covariance structure it generates for the assets it models. The latter translates into a genuine gamma/theta analysis of the carry P&L.

The local volatility model is the simplest of all market models for vanilla options, as all instruments have a one-dimensional Markovian representation in terms of  $t, S$ . With this frugality comes, however, a total lack of control on the dynamics of implied volatilities generated by the model.

Forward variance models are surveyed in Chapter 7. They are market models for  $S$  together with a (one-dimensional) term structure of implied volatilities, for example VS or ATMF implied volatilities. Unlike local volatility, forward variance models afford a great deal of flexibility as to the dynamics of implied volatilities they are able to generate. However, while their parameters – volatilities of volatilities and spot/volatility correlations – can be chosen so as to best match a given market smile, they typically cannot be calibrated exactly to the full set of implied volatilities  $\widehat{\sigma}_{KT}$ .

What we are really aiming for is a market model that lets us specify – at least partially and possibly indirectly – the joint dynamics of the spot and implied volatilities. Local-stochastic volatility models are a modest step in that direction.

There is a natural reason for considering them. In practice, only models having a low-dimensional Markovian representation can be employed. Starting with the simplest, the local volatility model, which is Markovian in  $t, S$ , which model is next in the hierarchy of market models? The answer is local-stochastic volatility models. They can be defined as market models that possess a Markovian representation in terms of  $t, S$  plus a few additional state variables, for example the factors  $X_t^1, X_t^2$  of the two-factor model of Chapter 7.

We only consider diffusive models, which, practically, means that their carry P&L is characterized by their break-even levels for gammas and cross-gammas.

---

## 12.2 Pricing equation and calibration

### 12.2.1 Pricing

In local-stochastic volatility models – which we also call mixed models – we choose the following ansatz for the instantaneous volatility of  $S_t$ :

$$\sigma_t = \sqrt{\zeta_t^t} \sigma(t, S_t) \quad (12.1)$$

$\zeta_t^t$  is a positive process that has a Markovian representation in terms of a small number of factors.<sup>1</sup>

Let us assume that  $\zeta_t^t$  is a process driven by the two-factor model of Section 7.4. From equation (7.28), page 226, the SDEs for  $S_t$  and  $\zeta_t^T$  read:

$$\begin{cases} dS_t = (r - q)S_t dt + \sigma(t, S_t) \sqrt{\zeta_t^t} S_t dW_t^S \\ d\zeta_t^T = 2\nu \zeta_t^T \alpha_\theta ((1 - \theta) e^{-k_1(T-t)} dW_t^1 + \theta e^{-k_2(T-t)} dW_t^2) \end{cases} \quad (12.2)$$

where  $\alpha_\theta = 1/\sqrt{(1 - \theta)^2 + \theta^2 + 2\rho_{12}\theta(1 - \theta)}$ . We use the notation  $\zeta_t^T$  rather than  $\xi_t^T$ , as  $\zeta_t^T$  is no longer a forward variance. In the mixed model forward variances are given by:

$$\xi_t^T = E_t [\sigma_T^2] = E_t [\zeta_T^T \sigma(T, S_T)^2]$$

They are not known analytically.

---

<sup>1</sup>Though typical, there is nothing mandatory about making the instantaneous volatility the *product* of stochastic and local volatility components.

The pricing equation in the mixed model is almost identical to that of the underlying stochastic volatility model, but for the local volatility component. Starting from equation (7.4), page 219, for the  $n$ -factor model, we get the pricing equation for the mixed model:

$$\begin{aligned} \frac{dP}{dt} + (r - q) S \frac{dP}{dS} + \frac{\zeta^t \sigma(t, S)^2}{2} S^2 \frac{d^2 P}{dS^2} \\ + \frac{1}{2} \int_t^T du \int_t^T du' \nu(t, u, u', \zeta) \frac{d^2 P}{\delta \zeta^u \delta \zeta^{u'}} + S \sigma(t, S) \int_t^T du \mu(t, u, \zeta) \frac{d^2 P}{dS \delta \zeta^u} = rP \end{aligned} \quad (12.3)$$

with  $\nu(t, u, u', \zeta)$  and  $\mu(t, u, \zeta)$  given by expressions (8.50) and (8.51), page 327, for the special case of the two-factor model:

$$\mu(t, u, \xi) = 2\nu\xi^u \sqrt{\xi^t} \alpha_\theta \left[ \rho_{SX^1} (1 - \theta) e^{-k_1(u-t)} + \rho_{SX^2} \theta e^{-k_2(u-t)} \right] \quad (12.4)$$

$$\begin{aligned} \nu(t, u, u', \xi) = 4\nu^2 \xi^u \xi^{u'} \alpha_\theta^2 \left[ (1 - \theta)^2 e^{-k_1(u+u'-2t)} + \theta^2 e^{-k_2(u+u'-2t)} \right. \\ \left. + \rho_{12} \theta (1 - \theta) \left( e^{-k_1(u-t)} e^{-k_2(u'-t)} + e^{-k_2(u-t)} e^{-k_1(u'-t)} \right) \right] \end{aligned}$$

$\nu$  – which we term “volatility of volatility” – is the volatility of a VS volatility with vanishing maturity. Practically, it functions as a scale factor of volatilities of volatilities.

### 12.2.2 Is it a price?

Equation (12.3) is derived from the corresponding pricing equation (7.4) of forward variance models through the ansatz:

$$\sqrt{\zeta_t^t} \rightarrow \sqrt{\zeta_t^t} \sigma(t, S_t) \quad (12.5)$$

The pricing equation of forward variance models arises from a replication analysis – see Section 7.1, page 217 – hence the P&L of a delta-hedged position has the usual gamma/theta expression, with well-defined break-even levels.

In local-stochastic volatility models, we use the seemingly innocuous ansatz (12.5). However, equation (12.3) does not arise from a replication-based argument, thus there is no reason that solving it produces a price – that is that the P&L of a hedged position is of the usual gamma/theta form.

If it is not, what will it be, may the reader ask? The carry P&L could include, in addition to gamma/theta terms, additional contributions that cause P&L leakage. This happens for example if we use an arbitrary pricing function  $P(t, S, \hat{\sigma}_{KT})$ ; in this case the model is not usable.

This will need to be assessed a posteriori – in Section 12.3. As it turns out, most local-stochastic volatility models are not usable models.

### Simulation

In the original version of the underlying stochastic volatility model, initial values of forward variances are analytically calibrated to the term structure of log-contract volatilities:  $\xi_{t=0}^T = \frac{d}{dT}(T\hat{\sigma}_T^2)$ , or numerically to the term structure of ATMF volatilities, or the term structure of implied volatilities for a given moneyness.

Because the underlying stochastic volatility model has a Markovian representation in terms of two factors, solving SDE (12.2) or PDE (12.3) boils down to the simulation of 3 processes:  $S_t, X_t^1, X_t^2$ .

The SDE for  $S_t, X_t^1, X_t^2$  are:

$$\begin{cases} dS_t = (r - q)S_t dt + \sqrt{\zeta_t^t} \sigma(t, S_t) S_t dW_t^S \\ dX_t^1 = -k_1 X_t^1 dt + dW_t^1 \\ dX_t^2 = -k_2 X_t^2 dt + dW_t^2 \end{cases} \quad (12.6)$$

with  $X_{t=0}^1 = X_{t=0}^2 = 0$ .  $X_t^1, X_t^2$  are Ornstein–Ühlenbeck processes that are easily simulated exactly – see Section 7.3.1, page 222.

$\zeta_t^t$  in (12.6) is given by:

$$\begin{cases} \zeta_t^t = \zeta_0^t f(t, X_t^1, X_t^2) \\ f(t, x_1, x_2) = e^{2\nu\alpha_\theta[(1-\theta)x_1+\theta x_2]-\frac{(2\nu)^2}{2}\chi(t,t)} \end{cases} \quad (12.7)$$

where  $\chi(t, T \geq t)$  is given by expression (7.35), page 227.

The only remaining task left is calibration of the local volatility function  $\sigma(t, S)$ .

#### 12.2.3 Calibration to the vanilla smile

Consider a diffusive model and  $\sigma_t$  the instantaneous volatility in that model; the condition that vanilla option prices be matched at  $t = 0$  is given in (2.6), page 28:

$$E[\sigma_t^2 | S_t = S] = 2 \left. \frac{\frac{dC}{dT} + qC + (r - q) K \frac{dC}{dK}}{K^2 \frac{d^2 C}{dK^2}} \right|_{\substack{K=S \\ T=t}} = \sigma_{\text{Mkt}}(t, S)^2 \quad (12.8)$$

where  $C(K, T)$  is the market price for a call option of strike  $K$ , maturity  $T$ , and  $\sigma_{\text{Mkt}}(t, S)$  is the local volatility function associated to the market smile.

The simplest way of complying with (12.8) is to choose  $\sigma_t \equiv \sigma_{\text{Mkt}}(t, S)$  – this is the local volatility model.

In mixed models, (12.8) translates into:

$$E[\zeta_t^t \sigma(t, S)^2 | S_t = S] = \sigma_{\text{Mkt}}(t, S)^2$$

thus  $\sigma(t, S)$  is given by:

$$\sigma(t, S)^2 = \frac{\sigma_{\text{Mkt}}(t, S)^2}{E[\zeta_t^t | S_t = S]} \quad (12.9)$$

This is a self-consistent equation for  $\sigma(t, S)$ : the unknown local volatility function appears both explicitly in the left-hand side, and implicitly in the right-hand side, in the expectation in the denominator. It is not clear how the solution (12.9) should be approached.

Algorithms for solving (12.9) start with the discretization of time and proceed forward, starting from  $t = 0$ . Imagine the density  $\varphi(t, S, X)$  is known at time  $t$ , where  $S$  is the underlying and  $X$  represents the state variables of the underlying stochastic volatility model. We use  $\varphi(t, S, X)$  to calculate the conditional expectation in the denominator of (12.9). (12.9) then yields the time- $t$  slice of the local volatility function, which we use to build the density  $\varphi(t + \delta t, S, X)$ , and so on.

How should  $\varphi$  be calculated practically? It can be done in two ways, depending on the dimensionality of the underlying stochastic volatility model.

- For a (very) small number of factors one can use the PDE technique described below – it is really usable for a one-factor model and consists in solving the (two-dimensional) forward equation for the joint density  $\varphi(t, S, X)$ .
- For a larger number of factors, the method of choice is the particle method, which we outline next, first introduced by Pierre Henry-Labordère and Julien Guyon in [52]. This algorithm is much more straightforward than the PDE technique and is immune to the curse of dimensionality.

#### 12.2.4 PDE method

Let us assume that  $\zeta_t$  is driven by a one-factor model – typically, through a Markovian representation as a function of an Ornstein–Uhlenbeck (OU) process  $X_t$ :

$$\begin{cases} \zeta_t^t = \zeta_0^t f(t, X_t) \\ dX_t = -k X_t dt + dZ_t, \quad X_0 = 0 \end{cases}$$

The SDE for  $S_t$  is:

$$dS_t = (r - q) S_t dt + \sqrt{\zeta_0^t f(t, X_t)} \sigma(t, S_t) S_t dW_t \quad (12.10)$$

and we denote by  $\rho$  the correlation between  $Z_t$  and  $W_t$ .

Consider the density  $\varphi(t, S, X) = E[\delta(S - S_t) \delta(X - X_t)]$ . Equation (12.9) for  $\sigma(t, S)$  can be rewritten as:

$$\sigma(t, S)^2 = \frac{\sigma_{\text{Mkt}}(t, S)^2}{\zeta_0^t} \frac{\int_{-\infty}^{+\infty} \varphi(t, S, X) dX}{\int_{-\infty}^{+\infty} \varphi(t, S, X) f(t, X) dX} \quad (12.11)$$

$\varphi(t \geq 0, S, X)$  is obtained by solving the forward Kolmogorov equation:

$$\frac{d\varphi}{dt} = \mathcal{L}\varphi \quad (12.12a)$$

$$\mathcal{L} = \mathcal{L}_S + \mathcal{L}_X + \mathcal{L}_{SX} \quad (12.12b)$$

with the initial condition:

$$\varphi(t=0, S, X) = \delta(S - S_0) \delta(X - X_0)$$

Linear operators  $\mathcal{L}_S$ ,  $\mathcal{L}_X$ ,  $\mathcal{L}_{SX}$  are defined by their action on a function  $\psi$ :

$$\mathcal{L}_S\psi = -(r - q) \frac{d}{dS}(S\psi) + \frac{1}{2} \frac{d^2}{dS^2}(f(t, X)\sigma(t, S)^2 S^2 \psi) \quad (12.13a)$$

$$\mathcal{L}_X\psi = k \frac{d}{dX}(X\psi) + \frac{1}{2} \frac{d^2\psi}{dX^2} \quad (12.13b)$$

$$\mathcal{L}_{SX}\psi = \frac{d^2}{dSdX}(\rho\sqrt{f(t, X)}\sigma(t, S)S\psi) \quad (12.13c)$$

$\mathcal{L}_S$ ,  $\mathcal{L}_{SX}$  involve the local volatility function  $\sigma(t, S)$ , thus (12.12) has to be solved self-consistently with (12.11). The idea of calibrating  $\sigma(t, S)$  via a forward PDE-based algorithm was first proposed by Alex Lipton in [70].

### Finite-difference algorithm

(12.12) is usually solved with a finite-difference algorithm. We assume that the reader has some familiarity with the numerical solution of parabolic equations – see [83] for an introduction.  $X$  and  $S$  – or more typically  $\ln S$  – are discretized on a two-dimensional grid  $(S_i, X_j)$ ;  $i = 0 \dots n_S - 1$ ,  $j = 0 \dots n_X - 1$ , with uniform spacings  $\delta S$  and  $\delta X$ :  $S_{i+1} - S_i = \delta S$  and  $X_{i+1} - X_i = \delta X$ . Density  $\varphi$  is replaced with a vector of dimension  $n_S n_X$ :  $\varphi_{i+n_S j} = \varphi(S_i, X_j)$ .

Derivatives are replaced by their centered finite-difference approximations:

$$\frac{df}{dx} \simeq \frac{f_{i+1} - f_{i-1}}{2\delta x}, \quad \frac{d^2 f}{dx^2} \simeq \frac{f_{i+1} + f_{i-1} - 2f_i}{\delta x^2} \quad (12.14a)$$

$$\frac{d^2 f}{dxdy} \simeq \frac{f_{i+1,j+1} - f_{i-1,j+1} - f_{i+1,j-1} + f_{i-1,j-1}}{4\delta x\delta y} \quad (12.14b)$$

whose errors are of order two in  $\delta x, \delta y$ . Action of operators  $\mathcal{L}_S$ ,  $\mathcal{L}_X$ ,  $\mathcal{L}_{SX}$  on  $\varphi$  in the right-hand side of (12.12) thus becomes a matrix/vector multiplication.

In what follows, we use the same notation for operators  $\mathcal{L}_S$ ,  $\mathcal{L}_X$ ,  $\mathcal{L}_{SX}$  or their discretized version, and likewise for  $\varphi$ .<sup>2</sup> Let us first assume for simplicity that

<sup>2</sup> Matrices  $\mathcal{L}_S$ ,  $\mathcal{L}_X$  are easy to invert as they are block-diagonal. The inversion of  $\mathcal{L}_S$ , for example, consists in the independent inversion of  $n_X$  sub-matrices, each of dimension  $n_S$ . Moreover, because derivatives at point  $S_i$  are approximated using only values for  $S_{i-1}, S_i, S_{i+1}$ , these sub-matrices are tridiagonal, thus the computational cost of each inversion is linear in  $n_S$ .

The total cost of inverting  $\mathcal{L}_S$  is thus of order  $n_S n_X$  – and likewise for  $\mathcal{L}_X$ . Compare this to the cost of inverting  $\mathcal{L}_{SX}$ , which is proportional to  $(n_S n_X)^3$ . Multiplication by  $\mathcal{L}_{SX}$ , on the other hand, is achieved at a cost proportional to  $n_S n_X$ .

$\mathcal{L}_S$ ,  $\mathcal{L}_X$ ,  $\mathcal{L}_{SX}$  do not depend explicitly on time. The formal solution of (12.12) over  $[t, t + \delta t]$  reads:

$$\varphi(t + \delta t) = e^{(\mathcal{L}_S + \mathcal{L}_X + \mathcal{L}_{SX})\delta t} \varphi(t)$$

where the exponential of operator  $\mathcal{O}$  is defined by:  $e^{\mathcal{O}} \varphi = \Sigma_0^\infty \frac{\mathcal{O}^n}{n!} \varphi$ .

### Vanishing correlation

Assume that  $\rho = 0$  so that  $\mathcal{L}_{SX}$  vanishes.

Several numerical schemes exist that approximate  $e^{(\mathcal{L}_S + \mathcal{L}_X + \mathcal{L}_{SX})\delta t}$  up to second order in  $\delta t$ . The most popular is the Peaceman-Rachford (PR) algorithm, which consists in the following sequence and makes use of the intermediate vector  $\varphi^*$ :

$$\left(1 - \frac{\delta t}{2} \mathcal{L}_X\right) \varphi^* = \left(1 + \frac{\delta t}{2} \mathcal{L}_S\right) \varphi(t) \quad (12.15a)$$

$$\left(1 - \frac{\delta t}{2} \mathcal{L}_S\right) \varphi(t + \delta t) = \left(1 + \frac{\delta t}{2} \mathcal{L}_X\right) \varphi^* \quad (12.15b)$$

Each step involves the inversion of a block-diagonal matrix (this is the so-called implicit sub-step), which as mentioned above, is computationally economical and a multiplication of a block-diagonal matrix on a vector (the so called explicit sub-step), which is computationally frugal as well. Expressing  $\varphi(t + \delta t)$  directly in terms of  $\varphi(t)$ :

$$\begin{aligned} \varphi(t + \delta t) &= \mathcal{U}_{t,t+\delta t} \varphi(t) \\ \mathcal{U}_{t,t+\delta t} &= \left(1 - \frac{\delta t}{2} \mathcal{L}_S\right)^{-1} \left(1 + \frac{\delta t}{2} \mathcal{L}_X\right) \left(1 - \frac{\delta t}{2} \mathcal{L}_X\right)^{-1} \left(1 + \frac{\delta t}{2} \mathcal{L}_S\right) \end{aligned}$$

One can check by hand that  $\mathcal{U}_{t,t+\delta t}$  approximates  $e^{(\mathcal{L}_S + \mathcal{L}_X)\delta t}$  up to order two in  $\delta t$ .

$$\mathcal{U}_{t,t+\delta t} = 1 + \delta t (\mathcal{L}_S + \mathcal{L}_X) + \frac{\delta t^2}{2} (\mathcal{L}_S + \mathcal{L}_X)^2 + o(\delta t^2)$$

Note that no assumption is made regarding the commutation of  $\mathcal{L}_S$ ,  $\mathcal{L}_X$ .

Let us introduce operators/matrices  $\mathcal{E}_S$  (standing for explicit) and  $\mathcal{I}_S$  (standing for implicit):

$$\mathcal{E}_S = \left(1 + \frac{\delta t}{2} \mathcal{L}_S\right) \quad \mathcal{I}_S = \left(1 - \frac{\delta t}{2} \mathcal{L}_S\right)^{-1}$$

and likewise for  $\mathcal{E}_X$  and  $\mathcal{I}_X$ . With these notations the PR algorithm simply reads:

$$\varphi(t + \delta t) = \mathcal{I}_S \mathcal{E}_X \mathcal{I}_X \mathcal{E}_S \varphi(t) \quad (12.16)$$

$\mathcal{E}$  and  $\mathcal{I}$  commute; moreover:

$$\mathcal{E}\mathcal{I} = \mathcal{I}\mathcal{E} = 2\mathcal{I} - 1 \quad (12.17)$$

thus there exist different equivalent implementations of the PR algorithm, for example:

$$\varphi(t + \delta t) = \mathcal{I}_S (2\mathcal{I}_X - 1) \mathcal{E}_S \varphi(t)$$

which is implemented through the following sequence:

$$\begin{aligned}\varphi^* &= \mathcal{E}_S \varphi(t) \\ \mathcal{I}_X^{-1} \varphi^{**} &= \varphi^* \\ \mathcal{I}_S^{-1} \varphi(t + \delta t) &= (2\varphi^{**} - \varphi^*)\end{aligned}$$

### Non-vanishing correlation

If  $\rho \neq 0$ ,  $\mathcal{L}_{SX}$  does not vanish. An algorithm that is correct up to order two in  $\delta t$  is the so-called predictor-corrector algorithm, which consists in two successive iterations of the PR algorithm, where  $\mathcal{L}_{SX}$  is always treated explicitly – see footnote 2 on page 458. This is the well-known Craig-Sneyd algorithm.

The predictor step reads:

$$\mathcal{I}_X^{-1} \varphi^{**} = \mathcal{E}_S \varphi(t) + \frac{\delta t}{2} \mathcal{L}_{SX} \varphi(t) \quad (12.18a)$$

$$\mathcal{I}_S^{-1} \varphi^*(t + \delta t) = \mathcal{E}_X \varphi^{**} + \frac{\delta t}{2} \mathcal{L}_{SX} \varphi(t) \quad (12.18b)$$

and generates  $\varphi^*(t + \delta t)$ . The corrector step is similar, except  $\mathcal{L}_{SX}$  is applied to the average of  $\varphi(t)$  and  $\varphi^*(t + \delta t)$ . Define  $\bar{\varphi}$  as:

$$\bar{\varphi} = \frac{1}{2}(\varphi(t) + \varphi^*(t + \delta t))$$

The corrector step reads:

$$\mathcal{I}_X^{-1} \varphi^{**} = \mathcal{E}_S \varphi(t) + \frac{\delta t}{2} \mathcal{L}_{SX} \bar{\varphi} \quad (12.19a)$$

$$\mathcal{I}_S^{-1} \varphi(t + \delta t) = \mathcal{E}_X \varphi^{**} + \frac{\delta t}{2} \mathcal{L}_{SX} \bar{\varphi} \quad (12.19b)$$

We use the same notation  $\varphi^{**}$  for the intermediate results in (12.18) and (12.19) – they are different vectors. The predictor step (12.18) reads:

$$\varphi^*(t + \delta t) = \mathcal{I}_S \mathcal{I}_X [\mathcal{E}_X \mathcal{E}_S + \delta t \mathcal{L}_{SX}] \varphi(t) \quad (12.20)$$

where we have used (12.17). The full scheme is compactly expressed through:

$$\begin{aligned}\varphi(t + \delta t) &= \mathcal{U}_{t,t+\delta t} \varphi(t) \\ \mathcal{U}_{t,t+\delta t} &= \mathcal{I}_S \mathcal{I}_X \left( \mathcal{E}_X \mathcal{E}_S + \delta t \mathcal{L}_{SX} \left( \frac{1}{2} + \frac{1}{2} \mathcal{I}_S \mathcal{I}_X (\mathcal{E}_X \mathcal{E}_S + \delta t \mathcal{L}_{SX}) \right) \right)\end{aligned} \quad (12.21)$$

While the predictor step is of order one in  $\delta t$ ,  $\mathcal{U}_{t,t+\delta t}$  is correct up to second order in  $\delta t$ . The reader is invited to check that indeed:

$$\mathcal{U}_{t,t+\delta t} = 1 + \delta t (\mathcal{L}_S + \mathcal{L}_X + \mathcal{L}_{SX}) + \frac{\delta t^2}{2} (\mathcal{L}_S + \mathcal{L}_X + \mathcal{L}_{SX})^2 + o(\delta t^2)$$

Again, there exist different corrector/predictor sequences implementing (12.21).

When  $\mathcal{L}_S$ ,  $\mathcal{L}_X$ ,  $\mathcal{L}_{SX}$  explicitly depend on  $t$  – which is the case in practice, if only because  $\sigma(t, S)$  enters  $\mathcal{L}_S$  and  $\mathcal{L}_{SX}$ , and also because of the presence of  $f(t, X)$  – then in (12.18) and (12.19)  $\mathcal{L}_S$  and  $\mathcal{L}_{SX}$  have to be evaluated at time  $t + \frac{\delta t}{2}$  to preserve the order-two accuracy of our numerical scheme. Thus, over each interval  $[t, t + \delta t]$  the local volatility is needed – and is determined – in  $t + \frac{\delta t}{2}$ .

### Implementation

In practice,  $\ln S$ , rather than  $S$ , is used – we still use  $S$  in the discussion as there is no difference in implementation. The local volatility  $\sigma(t, S)$  is discretized on the same spot grid as  $\varphi$ . The algorithm is started with the initial condition  $\varphi(t = 0, S, X) = \delta(S - S_0)\delta(X - X_0)$ , where  $S_0$  is the initial spot value and we take  $X_{t=0} = 0$ . In our discretized grid this translates into the following initial condition  $\varphi_{i \neq i_0, j \neq j_0} = 0$ ,  $\varphi_{i_0, j_0} = \frac{1}{\delta S} \frac{1}{\delta X}$  where  $i_0, j_0$  are the indexes for the initial values of  $S, X$ :  $S_{i_0} = S_{t=0}$  and  $X_{j_0} = 0$ .  $\sigma(\frac{\delta t}{2}, S)$  is simply initialized as:  $\sigma(\frac{\delta t}{2}, S) \equiv \hat{\sigma}(K = S, T = \frac{\delta t}{2}) / (\zeta_0^0 f(0, 0))$ .

Time is discretized with a step  $\delta t$ . Application of the finite difference algorithm generates the density  $\varphi$  at times  $t_k = k\delta t$ .

Assume we have the density  $\varphi$  at time  $t$  and the local volatility at time  $t - \frac{\delta t}{2}$ . Generation of  $\varphi(t + \delta t)$  and  $\sigma(t + \frac{\delta t}{2}, S)$  involves the following steps:

- Run the predictor-corrector scheme, where the local volatility  $\sigma(t + \frac{\delta t}{2}, S)$  in  $\mathcal{L}_S$ ,  $\mathcal{L}_{SX}$  is taken equal to that determined in the previous step:  $\sigma(t - \frac{\delta t}{2}, S)$ . This generates  $\varphi(t + \delta t)$ .
- Compute  $\sigma(t + \delta t, S)$  using (12.11) applied at time  $t + \delta t$ , using  $\varphi(t + \delta t)$ , then average the values at  $t$  and  $t + \delta t$ :

$$\sigma\left(t + \frac{\delta t}{2}, S\right)^2 = \frac{1}{2} \left( \sigma(t, S)^2 + \sigma(t + \delta t, S)^2 \right) \quad (12.22)$$

- Run the predictor-corrector scheme over  $[t, t + \delta t]$  again, this time using this final value for  $\sigma(t + \frac{\delta t}{2}, S)$  in  $\mathcal{L}_S$ ,  $\mathcal{L}_{SX}$ . This generates our final estimate for  $\varphi(t + \delta t)$ . This step guarantees that our scheme is overall of order two in time – in practice, though, this does not seem necessary.

Some additional points are worthy of note:

- The width of the grids in  $S$  (or  $\ln S$ ) and  $X$  is defined by choosing a percentile  $\varepsilon$  and setting  $S_{\min}, S_{\max}$  and  $X_{\min}, X_{\max}$  so that  $p(S_t \leq S_{\min}) \leq \varepsilon$ ,  $p(S_t \geq S_{\max}) \leq \varepsilon$  and  $p(X_t \leq X_{\min}) \leq \varepsilon$ ,  $p(X_t \geq X_{\max}) \leq \varepsilon$  for all  $t \in [0, T]$ . Typically, setting  $S_{\min}, S_{\max}$  and  $X_{\min}, X_{\max}$  so that  $p(S_T \leq S_{\min}) = p(S_T \geq S_{\max}) = \varepsilon$  and  $p(X_T \leq X_{\min}) = p(X_T \geq X_{\max}) = \varepsilon$  for the furthest maturity of interest  $T$  is adequate. Finding  $X_{\min}, X_{\max}$  is easy as  $X_T$  is Gaussian, thus its cumulative density is known in closed form.  $S_{\min}, S_{\max}$  are also easily found, since  $p(S_T \leq S_{\min}), p(S_T \geq S_{\max})$  are undiscounted prices of European digital options of maturity  $T$ , thus can be read off the market smile directly – see equation 1.24, page 21.

- Boundary conditions for  $\varphi$  need to be specified for  $S = S_0, S_{n_S}$  and  $X = X_0, X_{n_X}$ . Non-trivial boundary conditions for the density are not easy to derive. Typically, one takes wide grids in  $S$  (or  $\ln S$ ) and  $X$  and imposes that  $\varphi$  vanishes on the edges of the grid.  $\varphi(t)$  is a density so should integrate to one for all  $t$ ; this is not guaranteed by the algorithm above. Once  $\varphi(t + \delta t)$  is determined, one typically rescales it so that it integrates (numerically) to one. We refer the reader to Appendix A for an implementation that ensures that  $\varphi$  integrates to one and such that boundary conditions are automatically taken care of.
- For very small/large values of  $S$ , the density is small, thus the denominator in the right-hand side of (12.11) is small and subject to numerical noise. It is preferable to extrapolate  $\sigma(t, S)$  starting from values of  $S$  for which the denominator in (12.11) is still appreciable.

### 12.2.5 Particle method

The PDE technique outlined in the previous section can in practice only be used for one-factor stochastic volatility models. For models with more than one stochastic volatility factor, the PDE is of dimension three or higher. It is then best to calibrate  $\sigma(t, S)$  using the particle method, a Monte Carlo algorithm first published by Pierre Henry-Labordère and Julien Guyon in [52].

The particle algorithm is general, does not depend on the dimensionality of the process driving  $\zeta_t$  and can also be used to calibrate the local volatility function in a hybrid local-stochastic volatility/stochastic interest rate model, or to calibrate the local correlation of a cross-FX rate or of a basket of equity underlyings to an index smile.

As the particle method is documented in Pierre Henry-Labordère and Julien Guyon's book [53], we only sketch it. The particle method is a Monte Carlo algorithm based on simultaneous simulation of interacting paths.

Time is discretized and we set up a grid of spot values  $S^*$  for which the local volatility function will be determined.

- Draw  $N$  paths for the pair  $(S_t, \zeta_t)$  – we denote them by  $(S_t^k, \zeta_t^k)$ ,  $k = 1 \dots N$ , starting from  $S_{t=0}^k = S_0, \zeta_{t=0}^k = \zeta_0$ . Each pair  $(S_t^k, \zeta_t^k)$  obeys the model's SDE and the Brownian motions driving different pairs are all independent.
- Assume that the local volatility function is known at  $t_i$ . Use it to propagate the particles until  $t_{i+1}$ . At  $t_{i+1}$  use the empirical density defined by:

$$\varphi_{em}(t_{i+1}, S, \zeta) = \frac{1}{N} \sum_k \delta(S - S_{t_{i+1}}^k) \delta(\zeta - \zeta_{t_{i+1}}^k) \quad (12.23)$$

to evaluate the conditional expectation  $E[\zeta_t | S_t = S]$  for spot values  $S^*$ . Rather than straight Dirac peaks on  $S$  one uses in (12.23) a smoother kernel  $\phi$ :

$$\varphi_{em}(t_{i+1}, S, \zeta) = \frac{1}{N} \sum_k \phi(S - S_{t_{i+1}}^k) \delta(\zeta - \zeta_{t_{i+1}}^k)$$

Efficient operation of the particle algorithm depends in fact on a proper choice of  $\phi$ . Using expression (12.9) the local volatility at time  $t_{i+1}$  is thus given by:

$$\sigma(t_{i+1}, S^*)^2 = \sigma_{\text{Mkt}}(t_{i+1}, S^*)^2 \frac{\Sigma_k \phi(S^* - S_{t_{i+1}}^k)}{\Sigma_k \zeta_{t_{i+1}}^k \phi(S^* - S_{t_{i+1}}^k)}$$

Interpolate/extrapolate local volatilities calculated on spot values  $S^*$  of the grid to obtain  $\sigma(t_{i+1}, S)$ . Store this  $t_{i+1}$  slice of the local volatility function and simulate the particles until  $t_{i+2}$ .

Calibration of the local volatility function and pricing can be performed in one single simulation, using the same paths.

---

## 12.3 Usable models

We're now fully equipped for calibrating a mixed model and pricing exotic options. Before we do this, we need to address a concern expressed in Section 12.2.2: are the resulting numbers prices?

The reason for using local-stochastic volatility models is that they are calibrated, by construction, to the vanilla smile. When we build them what we are really trying to do is build a market model for spot and vanilla options.

In such a market model, the hedge instruments are the spot and vanilla options. The price of a derivative is a function of the values of these hedge instruments, and the P&L of a delta-hedged, vega-hedged position is the sum of cross-gamma contributions involving second-order moments of variations of the spot and implied volatilities, accompanied by matching thetas defined by a break-even covariance matrix.

Do mixed models indeed supply a genuine gamma/theta breakdown of the carry P&L? Provided they do, what are the implied break-even volatilities of volatilities and the break-even spot/volatility and volatility/volatility correlations?

We have already carried out such an analysis for the local volatility model, in Section 2.7, page 66. We ask the reader to read that portion of Chapter 2.

As will be made clear shortly, unlike the local volatility model, most mixed models are *not* market models for spot and implied volatilities. Only particular types of mixed models, which we now characterize, give rise to a genuine theta/gamma breakdown of the carry P&L.

### 12.3.1 Carry P&L

Consider the general case of a mixed model and denote by  $P^M(t, x)$  the price of a derivative.  $x$  is the vector of inputs:  $x_1$  is the spot price,  $x_2$  the local volatility

function,  $x_3 \cdots x_n$  are state variables of the underlying stochastic volatility model. For example, if the underlying stochastic volatility model is the two-factor forward variance model, the dynamics of the mixed model is given by (12.2) and  $P^M$  reads:

$$P^M(t, x) \equiv P^M(t, S, \sigma, \zeta^u)$$

where the  $\zeta^u$  make up a curve.

Consider instead a local-stochastic volatility model built on the Heston model:

$$\begin{cases} dS_t = (r - q)S_t dt + \sigma(t, S_t)\sqrt{V_t}S_t dW_t^S \\ dV_t = -k(V_t - V^0)dt + \nu\sqrt{V_t}dW_t^V \end{cases}$$

Here:

$$P^M(t, x) \equiv P^M(t, S, \sigma, V)$$

$P^M(t, x)$  takes as input the local volatility function. Consider now the pricing function  $P(t, \hat{x})$  which takes the set of implied volatilities as an input, rather than the local volatility function;  $\hat{x}_1$  is the spot price,  $\hat{x}_2$  the set of implied volatilities  $\hat{\sigma}_{KT}, \hat{x}_3 \cdots \hat{x}_n$  are again the state variables of the underlying stochastic volatility model.

$P(t, \hat{x})$  is the pricing function we use in trading applications as it takes as inputs market observables, in addition to the state variables of the underlying stochastic volatility model. For the two-factor model:

$$\begin{aligned} P^M(t, x) &\equiv P^M(t, S, \sigma, \zeta^u) \\ P(t, \hat{x}) &\equiv P(t, S, \hat{\sigma}_{KT}, \zeta^u) \end{aligned}$$

The reason why we explicitly include the local volatility function  $\sigma$  as an argument of  $P^M$  is that  $\sigma$  is not frozen. Indeed,  $P(t, S, \hat{\sigma}_{KT}, \zeta^u)$  implicitly involves recalibration of  $\sigma$  whenever the arguments of  $P$  change; as we risk-manage our exotic option using the pricing function  $P$ ,  $\sigma$  will change and our carry P&L accounts for these changes as well.

In the mixed model, for a set local volatility function, implied volatilities are a *function* of time, spot value, local volatility function and other state variables:  $\hat{x} \equiv \hat{x}(t, x)$  and we have:

$$P^M(t, x) = P(t, \hat{x}(t, x))$$

For example, if we use the two-factor model as the underlying stochastic volatility model, we have:

$$\hat{\sigma}_{KT} \equiv \Sigma_{KT}^M(t, S, \sigma, \zeta^u)$$

and the following relationship between  $P^M$  and  $P$ :

$$P^M(t, S, \sigma, \zeta^u) = P(t, S, \Sigma_{KT}^M(t, S, \sigma, \zeta^u), \zeta^u)$$

In a trading context, P&L accounting is done with  $P$  and involves derivatives of  $P$  with respect to  $t, S, \hat{\sigma}_{KT}$ . The pricing equation (12.3), however, involves  $P^M$  and its derivatives.

Let us thus change variables from  $(t, x)$  to  $(t, \hat{x})$ :

$$(t, x) \rightarrow (t, \hat{x}(t, x))$$

The pricing equation (12.3), page 455, of the mixed model – with a set local volatility function – reads:

$$\frac{dP^M}{dt} + \left( \Sigma_k \mu_k \frac{d}{dx_k} + \frac{1}{2} \Sigma_{kl} a_{kl} \frac{d^2}{dx_k dx_l} \right) P^M = 0 \quad (12.24)$$

where we assume zero interest rates without loss of generality – otherwise consider that  $P$  is the undiscounted price.

Switching now to variables  $\hat{x}$ , the pricing equation reads:

$$\frac{dP}{dt} + \left( \Sigma_i \hat{\mu}_i \frac{d}{d\hat{x}_i} + \frac{1}{2} \Sigma_{ij} \hat{a}_{ij} \frac{d^2}{d\hat{x}_i d\hat{x}_j} \right) P = 0 \quad (12.25)$$

with:

$$\begin{cases} \hat{\mu}_i = \frac{d\hat{x}_i}{dt} + \Sigma_k \mu_k \frac{d\hat{x}_i}{dx_k} + \frac{1}{2} \Sigma_{kl} a_{kl} \frac{d^2\hat{x}_i}{dx_k dx_l} \\ \hat{a}_{ij} = \Sigma_{kl} a_{kl} \frac{d\hat{x}_i}{dx_k} \frac{d\hat{x}_j}{dx_l} \end{cases} \quad (12.26)$$

$\hat{\mu}_i$  is the drift of  $\hat{x}_i$  and  $\hat{a}_{ij}$  is the covariance matrix of  $\hat{x}_i$  and  $\hat{x}_j$  – as generated by the mixed model with a fixed local volatility function.

Derivatives  $\frac{d\hat{x}_i}{dx_k}$  are calculated keeping the local volatility function constant; in the two-factor model they involve derivatives  $\frac{d\Sigma_{KT}^M}{dt}, \frac{d\Sigma_{KT}^M}{dS}, \frac{d\Sigma_{KT}^M}{d\zeta^u}$ .

While the differential operator in (12.24) does not involve derivatives with respect to the local volatility function, the operator in (12.25) does involve derivatives with respect to implied volatilities.

### 12.3.2 P&L of a hedged position

Consider the P&L of a short option position – unhedged for now – during  $\delta t$ :

$$P&L = -P(t + \delta t, \hat{x} + \delta\hat{x}) + P(t, \hat{x})$$

Remember that as  $t, S, \hat{\sigma}_{KT}$  move by  $\delta t, \delta S$  and  $\delta\hat{\sigma}_{KT}$ , the local volatility function of our mixed model is recalibrated.

Expand at order two in  $\delta\hat{x}$  and one in  $\delta t$ , and use (12.25) to express  $\frac{dP}{dt}$  in terms of derivatives with respect to the  $\hat{x}_i$ :

$$\begin{aligned} P\&L &= -\frac{dP}{dt}\delta t - \Sigma_i \frac{dP}{d\hat{x}_i} \delta\hat{x}_i - \frac{1}{2} \Sigma_{ij} \frac{d^2P}{d\hat{x}_i d\hat{x}_j} \delta\hat{x}_i \delta\hat{x}_j \\ &= -\Sigma_i \frac{dP}{d\hat{x}_i} (\delta\hat{x}_i - \hat{\mu}_i \delta t) - \frac{1}{2} \Sigma_{ij} \frac{d^2P}{d\hat{x}_i d\hat{x}_j} (\delta\hat{x}_i \delta\hat{x}_j - \hat{a}_{ij} \delta t) \end{aligned}$$

Among components of  $\hat{x}$  we now make a distinction between those that correspond to market observables –  $S, \hat{\sigma}_{KT}$  – which we denote by  $O_i$ , and those corresponding to state variables of the underlying stochastic volatility model, which we denote by  $\lambda_k$ :  $\hat{x} \equiv (O, \lambda)$ .

$$\begin{aligned} P\&L &= -\Sigma_i \frac{dP}{dO_i} (\delta O_i - \hat{\mu}_i \delta t) - \frac{1}{2} \Sigma_{ij} \frac{d^2P}{dO_i dO_j} (\delta O_i \delta O_j - \hat{a}_{ij} \delta t) \\ &\quad - \Sigma_k \frac{dP}{d\lambda_k} (\delta \lambda_k - \hat{\mu}_k \delta t) \\ &\quad - \frac{1}{2} \Sigma_{kl} \frac{d^2P}{d\lambda_k d\lambda_l} (\delta \lambda_k \delta \lambda_l - \hat{a}_{kl} \delta t) - \Sigma_{ik} \frac{d^2P}{dO_i d\lambda_k} (\delta O_i \delta \lambda_k - \hat{a}_{ik} \delta t) \end{aligned} \tag{12.27}$$

Consider now a delta and vega-hedged position, so that order-one contributions in  $\delta O_i$  vanish. Because equation (12.27) for the P&L during  $\delta t$  obviously also holds for (a) the underlying itself, (b) vanilla options, cancelling the  $\delta O_i$  terms also cancels the  $\hat{\mu}_i \delta t$  contributions. We denote by  $P_H$  the value of the delta-hedged, vega-hedged, position.

### Using implied volatilities or option prices

Parameters  $O_i$  reflect the values of hedge instruments. For the spot we use the spot value itself, but for vanilla options we may use straight option prices, or their implied volatilities, or yet a different parametrization. For the sake of the present discussion, we treat the spot separately from vanilla options. The value of the hedged position is:

$$P_H = P - \Delta_S S - \sum_i \Delta_i f_i(t, S, O_i)$$

where the sum runs over vanilla options used as hedges and  $f_i$  is the value of a vanilla option as a function of parameter  $O_i$ . The delta is  $\Delta_S = \frac{dP}{dS} \Big|_{\lambda, O}$ . The vega hedge ratios  $\Delta_i$  are given by:  $\Delta_i = \frac{dP}{dO_i} \left( \frac{df_i}{dO_i} \right)^{-1}$ .

- If  $O_i$  is an implied volatility  $\hat{\sigma}_{KT}$  then  $f_i$  is the value of a delta-hedged vanilla option in the Black-Scholes model and  $P_H$  reads:

$$P_H = P - \frac{dP}{dS} \Big|_{\lambda, \hat{\sigma}_{KT}} S - \sum_{KT} \Delta_{KT} \left( P_{KT} - \frac{dP_{KT}^{BS}}{dS} S \right) \tag{12.28}$$

$$\text{with } \Delta_{KT} = \frac{dP}{d\hat{\sigma}_{KT}} \Big|_{\lambda, S} \left( \frac{dP_{KT}^{BS}}{d\hat{\sigma}_{KT}} \right)^{-1}.$$

- If instead  $O_i$  is the vanilla option price  $P_{KT}$ ,  $\Delta_{KT} = \frac{dP}{dP_{KT}} \Big|_{\lambda, S}$  and  $P_H$  reads:

$$P_H = P - \frac{dP}{dS} \Big|_{\lambda, P_{KT}} S - \sum_{KT} \Delta_{KT} P_{KT} \quad (12.29)$$

There is no inconsistency in these two expressions of  $P_H$ . Obviously, the composition of the hedge portfolio cannot depend on how we decide to represent option prices, either using straight option prices or Black-Scholes implied volatilities: canceling at order one (a) the sensitivity to  $S$  and vanilla option prices, or (b) the sensitivity to  $S$  and Black-Scholes implied volatilities, is equivalent.

Thus,  $\Delta_{KT}$  in (12.28) and (12.29) are identical, and so are the deltas in both portfolios:

$$\frac{dP}{dS} \Big|_{\lambda, \hat{\sigma}_{KT}} - \sum_{KT} \Delta_{KT} \frac{dP_{KT}^{BS}}{dS} = \frac{dP}{dS} \Big|_{\lambda, P_{KT}}$$

We refer the reader to a similar discussion of market-model and sticky-strike deltas in the local volatility model in Section 2.7, page 66.

### Splitting the P&L of a hedged position

The P&L of the hedged position reads:

$$P\&L_H = -\frac{1}{2} \sum_{ij} \frac{d^2 P_H}{dO_i dO_j} (\delta O_i \delta O_j - \hat{a}_{ij} \delta t) \quad (12.30a)$$

$$- \sum_k \frac{dP_H}{d\lambda_k} (\delta \lambda_k - \hat{\mu}_k \delta t) \quad (12.30b)$$

$$- \frac{1}{2} \sum_{kl} \frac{d^2 P_H}{d\lambda_k d\lambda_l} (\delta \lambda_k \delta \lambda_l - \hat{a}_{kl} \delta t) - \sum_{ik} \frac{d^2 P_H}{dO_i d\lambda_k} (\delta O_i \delta \lambda_k - \hat{a}_{ik} \delta t) \quad (12.30c)$$

- By construction, the mixed model is calibrated to the market values of hedge instruments  $\forall t$ , thus we have:  $\frac{dO_i}{d\lambda_k} = 0$ ,  $\forall i, \forall k$ . This implies that  $\frac{dP_H}{d\lambda_k} = \frac{dP}{d\lambda_k}$ ,  $\forall k$  and  $\frac{d^2 P_H}{d\lambda_k d\lambda_l} = \frac{d^2 P}{d\lambda_k d\lambda_l}$  as well as  $\frac{d^2 P_H}{dO_i d\lambda_k} = \frac{d^2 P}{dO_i d\lambda_k}$ : all sensitivities of the hedged position involving  $\lambda_k$  are those of the unhedged position.  $P\&L_H$  can be rewritten as:

$$P\&L_H = -\frac{1}{2} \sum_{ij} \frac{d^2 P_H}{dO_i dO_j} (\delta O_i \delta O_j - \hat{a}_{ij} \delta t) \quad (12.31a)$$

$$- \sum_k \frac{dP}{d\lambda_k} (\delta \lambda_k - \hat{\mu}_k \delta t) \quad (12.31b)$$

$$- \frac{1}{2} \sum_{kl} \frac{d^2 P}{d\lambda_k d\lambda_l} (\delta \lambda_k \delta \lambda_l - \hat{a}_{kl} \delta t) - \sum_{ik} \frac{d^2 P}{dO_i d\lambda_k} (\delta O_i \delta \lambda_k - \hat{a}_{ik} \delta t) \quad (12.31c)$$

- Contribution (12.31a) to  $P\&L_H$  is the regular theta/gamma P&L involving second-order moments of the variations of market instruments. The matching deterministic terms  $\widehat{a}_{ij}\delta t$  are genuine thetas, as  $\widehat{a}$  is a valid (positive) covariance matrix – this is obvious from its definition in (12.26). The break-even covariances  $\widehat{a}_{ij}^*$  are those generated by the model with a fixed local volatility function.
- (12.31b) and (12.31c) are unwanted contributions to the P&L, generated by variations of the  $\lambda_k$  state variables, that have no financial significance:  $V$  for the Heston model, the  $\zeta^u$  for the two-factor model. These terms were absent from P&L expression (2.105), page 69, in the local volatility model.
- Note that the  $\delta\lambda_k$  are fully in our control. Imagine setting  $\delta\lambda_k = \widehat{\mu}_k\delta t$  so that (12.31b) cancels out. There remain the theta and gamma terms in (12.30c).

The conclusion is that, generally, when using a local-stochastic volatility model, a hedged position will generate spurious P&L leakage that does not correspond to any regular gamma/theta P&L.<sup>3,4</sup>

The contribution from  $\delta\lambda_k$  obviously vanishes if (a)  $\delta\lambda_k = 0$ , (b)  $\widehat{\mu}_k = 0$  and  $\widehat{a}_{kl} = \widehat{a}_{ik} = 0, \forall i, \forall l$ . This is the case if  $\lambda_k$  is frozen – in other words if  $\lambda_k$  is a constant parameter, for example the long-run variance  $V^0$  in the Heston model, or constants  $k_1, k_2, \theta, \rho_{12}$  in the two-factor model.

### 12.3.3 Characterizing usable models

Are there instances of mixed models that lead to regular theta/gamma P&L accounting without P&L leakage? The answer is yes.

All we need is for  $P$  to not depend on  $\lambda_k$ :<sup>5</sup>

$$\frac{dP}{d\lambda_k} \Big|_{S, \widehat{\sigma}_{KT}} = 0, \forall k \quad (12.32)$$

---

<sup>3</sup> Could it be that – for some configurations of  $t, S, O_i, \lambda$  and some payoffs – there exists a positive matrix  $\widehat{a}_{ij}^*$  such that all  $\delta t$  contributions to the P&L are absorbed in the theta portion of the theta/gamma P&L (12.31a)?  $\widehat{a}_{ij}^*$  would then be such that:

$$\frac{1}{2} \sum_{ij} \frac{d^2 P_H}{dO_i dO_j} \widehat{a}_{ij}^* = \frac{1}{2} \sum_{ij} \frac{d^2 P_H}{dO_i dO_j} \widehat{a}_{ij} + \Sigma_k \frac{dP}{d\lambda_k} \widehat{\mu}_k + \frac{1}{2} \sum_{kl} \frac{d^2 P}{d\lambda_k d\lambda_l} \widehat{a}_{kl} + \Sigma_{ik} \frac{d^2 P}{dO_i d\lambda_k} \widehat{a}_{ik}$$

By setting  $\delta\lambda_k = 0$ , our P&L would simply consist of contribution (12.31a) with effective break-even covariances  $\widehat{a}_{ij}^*$ . Assuming this could hold for (a) all values of  $t, S, O, \lambda$ , (b) all payoffs, is a very unrealistic assumption. It amounts to hoping that a theta that was engineered to offset one cross-gamma offsets a different cross-gamma.

<sup>4</sup> Given a leaky model, can we size up the P&L leakage? Imagine for example that we use the Heston model as the underlying model. Then  $V\sigma^2 T \frac{d^2 P}{dV^2} \Big|_{S, \widehat{\sigma}_{KT}}$  is a (very) rough estimate of the leakage generated by the first piece in (12.31c). To assess the magnitude of the leakage from the second piece, one needs second-order derivatives  $\frac{d^2 P}{dV d\widehat{\sigma}_{KT}} \Big|_{S, \widehat{\sigma}_{KT}}$  and  $\frac{d^2 P}{dV dS} \Big|_{S, \widehat{\sigma}_{KT}}$ .

<sup>5</sup> Ex-physicists will be tempted to call this a condition of gauge-invariance.

Our P&L then simply reads:

$$P\&L_H = -\frac{1}{2}\Sigma_{ij}\frac{d^2P_H}{dO_idO_j}(\delta O_i\delta O_j - \hat{a}_{ij}\delta t) \quad (12.33)$$

and we have indeed a market model for spot and vanilla options – see our discussion in Section 1.1 of Chapter 1.

Condition (12.32) happens to be fulfilled for the two-factor model. Indeed, consider pricing equations (12.6) together with (12.7), page 456, for the mixed model. Perform the following transformation on the  $\zeta^u$  and the local volatility function:

$$\begin{aligned} \zeta_0^u &\rightarrow \varphi^u \zeta_0^u \\ \sigma(u, S) &\rightarrow \sqrt{\frac{1}{\varphi^u}}\sigma(u, S) \end{aligned}$$

where  $\varphi^u$  are arbitrary constants. One can see from (12.6) and (12.7) that this leaves the process for  $S_t$  and its instantaneous volatility unchanged, thus:

$$\frac{\delta}{\delta\zeta^u}P(t, S, \hat{\sigma}_{KT}, \zeta^u) = 0, \forall u$$

This also holds in the version of the two-factor model that generates volatility-of-volatility smile:  $f$  in (12.7) is then the sum of two exponentials – see Section 7.7.1, page 263.

- Condition (12.32) does not hold if we use the Heston model as the underlying model since:

$$\frac{d}{dV}P(t, S, \hat{\sigma}_{KT}, V) \neq 0$$

- It does not hold in the Bloomberg model either, defined by the following SDEs – see [45]:

$$\begin{cases} dS_t = (r - q)S_t dt + \sigma(t, S_t)\lambda_t S_t dW_t^S \\ d\lambda_t = -k(\lambda_t - \theta)dt + \xi(t)\lambda_t dW_t^\lambda \end{cases} \quad (12.34)$$

where  $\xi(t)$  is a deterministic function of  $t$ .

- Condition (12.32) does hold if the instantaneous volatility of the underlying stochastic volatility model is lognormal, as in the SABR model:

$$\begin{cases} dS_t = (r - q)S_t dt + \sigma(t, S_t)\lambda_t S_t dW_t^S \\ d\lambda_t = \nu\lambda_t dW_t^\lambda \end{cases} \quad (12.35)$$

### Discussion

We are aiming for a market model that takes as inputs the spot and implied volatilities. In case we make option prices dependent on additional non-financial state variables,<sup>6</sup> it is not surprising that there is P&L leakage, as the model allocates part of its theta as compensation – on average – for second-order contributions from these extra state variables as well.

$\frac{dP}{dV} \neq 0$  is a signal that prices depend on more state variables than hedge instruments, hence the inconsistency in P&L accounting and the fact that the model is unsuitable for trading purposes.

This dependence is of a different nature than the dependence on model parameters, such as  $V^0$  in the Heston model, or  $k_1, k_2, \theta, \rho_{12}$  in the two-factor model. Model parameters do not generate any P&L leakage, and only give rise to discrete P&Ls when they are changed and the option position is remarked with new parameter values.

In the case of the Heston model used as underlying stochastic volatility model, the “solution” to the P&L leakage – if it can be called a solution – is to include one additional hedge instrument, to which  $V$  is then calibrated – say a barrier or forward-start option. We then have as many hedge instruments as there are state variables in our model, and the carry P&L of a hedged option position is again of the genuine gamma/theta form, provided we dynamically trade this additional instrument.

The interesting aspect of models in the admissible class is that the underlying stochastic volatility model *does* affect the dynamics of spot and option prices; yet these additional degrees of freedom do not require any additional hedge instruments, and only impact the covariance structure of hedge instruments.<sup>7</sup>

The carry P&L is of the regular gamma/theta form and involves hedge instruments only.

## 12.4 Dynamics of implied volatilities

Having derived expression (12.33) of the P&L in an admissible model, we now need to size up covariances  $\hat{a}_{ij}^*$  to assess the suitability of our model’s break-even levels.

<sup>6</sup>Consider that  $V$  is, for example, the temperature in the Luxembourg garden.

<sup>7</sup>It is of course always possible to express a model using more processes than necessary, but this is just cosmetic. Take the Black-Scholes model and write the driving Brownian motion as the sum of two different Brownian motions. The resulting model is still Black-Scholes with exactly the same dynamics and option prices as the original model.

Option prices in the admissible class *are* impacted by the dynamics of the underlying stochastic volatility model.

The  $\hat{a}_{ij}$  are model-generated covariances for spot and implied volatilities for *fixed* strikes and maturities. With regard to the task of calculating realized covariances, however, working with *floating* strikes corresponding to a set moneyness – for example ATMF – and *relative* maturities, is much more convenient.

In what follows, we derive approximations of model-generated variances and covariances of spot and ATMF volatilities. Spot/volatility covariances can equivalently be quantified using the SSR.

In practice, bid/offer costs on options are such that vega hedging is performed less frequently than delta hedging. The resulting carry P&L is different than (12.33) – see the discussion in Section 9.11.3, page 383.

We use the two-factor model as underlying stochastic volatility model and derive expressions of the SSR and the volatility of the ATMF volatility at lowest order, both in:

- the local volatility function – as in Section 2.5
- the volatility of volatility of the underlying two-factor model – as in the expansion of Section 8.2

#### 12.4.1 Components of the ATMF skew

Consider SDE (12.6) for  $S_t$  in the mixed two-factor model. Since the two-factor model is in the admissible class, we have  $\frac{dP}{d\zeta^\tau} = 0 \forall \tau$ , thus we can take  $\zeta^\tau = 1, \forall \tau$ . The SDE for  $S_t$  is:

$$dS_t = (r - q)S_t dt + \sigma(t, S_t) \sqrt{f(t, X_t^1, X_t^2)} S_t dW_t^S \quad (12.36)$$

where  $f$  is defined in (12.7).

We now perform an expansion by writing the local volatility function as:

$$\sigma(t, S) = \bar{\sigma}(t) + \delta\sigma(t, S)$$

with  $\delta\sigma(t, S)$  given by:

$$\delta\sigma(t, S) = \alpha(t) x, \quad x = \ln \frac{S}{F_t}$$

This is the form we use in (2.44), page 46, where we derive the approximate expression of the ATMF skew in the local volatility model. With respect to Section 2.4.5, here we perform the expansion in  $\delta\sigma$  around  $\bar{\sigma}(t)$  rather than a constant volatility  $\sigma_0$ .

We now turn to  $f$  and expand it at order one in  $\nu$ . For  $\nu = 0$ ,  $f = 1$ ; at order one in volatility of volatility  $\nu$ ,

$$\sqrt{f(t, X_t^1, X_t^2)} = 1 + \frac{\nu}{2} g(t, X_t^1, X_t^2)$$

SDE (12.36) now reads:

$$dS_t = (r - q)S_t dt + \left( \bar{\sigma}(t) + \delta\sigma(t, S_t) \right) \left( 1 + \frac{\nu}{2} g(t, X_t^1, X_t^2) \right) S_t dW_t^S \quad (12.37)$$

Write  $S_t$  as:

$$S_t = S_t^0 + \delta S_t^{\text{LV}} + \delta S_t^{\text{SV}} \quad (12.38)$$

where  $\delta S_t^{\text{LV}}$  and  $\delta S_t^{\text{SV}}$  are, respectively, the order-one corrections in  $\delta\sigma$  and in  $\nu$ . Inserting this expression in SDE (12.37) and equating terms linear in  $\delta\sigma$  and  $\nu$  yields the following SDEs for  $S_t^0, \delta S_t^{\text{LV}}, \delta S_t^{\text{SV}}$ :

$$\begin{cases} dS_t^0 = (r - q)S_t^0 dt + \bar{\sigma}(t) S_t^0 dW_t^S & S_{t=0}^0 = S_0 \\ d\delta S_t^{\text{LV}} = (r - q)\delta S_t^{\text{LV}} dt + \delta\sigma(t, S_t^0) S_t^0 dW_t^S & \delta S_{t=0}^{\text{LV}} = 0 \\ d\delta S_t^{\text{SV}} = (r - q)\delta S_t^{\text{SV}} dt + \bar{\sigma}(t) \frac{\nu}{2} g S_t^0 dW_t^S & \delta S_{t=0}^{\text{SV}} = 0 \end{cases}$$

Expansion (12.38) translates into an expansion of option prices:

$$P = P_0 + \delta P_{\text{LV}} + \delta P_{\text{SV}} \quad (12.39)$$

where  $P_0$  is the price generated by  $S_t^0$ , that is the Black-Scholes price with time-deterministic volatility  $\bar{\sigma}(t)$ , that is with implied volatilities  $\hat{\sigma}_\tau$  defined by:

$$\hat{\sigma}_\tau = \sqrt{\frac{1}{\tau} \int_0^\tau \bar{\sigma}_u^2 du}$$

(12.39) in turn translates into the following expansion of implied volatilities and of the ATM skew  $\mathcal{S}_T = \left. \frac{d\hat{\sigma}_{KT}}{d \ln K} \right|_{F_T}$ :

$$\hat{\sigma}_{KT} = \hat{\sigma}_T + \delta\hat{\sigma}_{KT}^{\text{LV}} + \delta\hat{\sigma}_{KT}^{\text{SV}} \quad (12.40a)$$

$$\mathcal{S}_T = \mathcal{S}_T^{\text{LV}} + \mathcal{S}_T^{\text{SV}} \quad (12.40b)$$

where the two order-one contributions to  $\hat{\sigma}_{KT}$  have been calculated already:  $\delta\hat{\sigma}_{KT}^{\text{LV}}$  in Chapter 2 – see equation (2.40), page 44;  $\delta\hat{\sigma}_{KT}^{\text{SV}}$  in Chapter 8 – see equation (8.20), page 314.

- $\mathcal{S}_T^{\text{LV}}$  is generated by  $\delta S_t^{\text{LV}}$ , which, according to its SDE, corresponds to the perturbation at order one of the Black-Scholes model with deterministic volatility  $\bar{\sigma}(t)$  by a local volatility function  $\delta\sigma(t, S)$ . We can use the skew-averaging expression (2.60), page 51, and get:

$$\mathcal{S}_T^{\text{LV}} = \frac{1}{T} \int_0^T \frac{\hat{\sigma}_t^2 t}{\hat{\sigma}_T^2 T} \frac{\bar{\sigma}(t)}{\hat{\sigma}_T} \alpha(t) dt \quad (12.41)$$

If we expand around a constant volatility  $\sigma_0$ , rather than a deterministic volatility  $\bar{\sigma}(t)$  we get – replacing  $\sigma(t)$ ,  $\hat{\sigma}_t^2$  and  $\hat{\sigma}_T$  with  $\sigma_0$  – expression (2.48), page 46:

$$\mathcal{S}_T^{\text{LV}} = \frac{1}{T} \int_0^T \frac{t}{T} \alpha(t) dt \quad (12.42)$$

- Likewise,  $\delta S_t^{\text{SV}}$  is generated by the perturbation at order one in volatility of volatility in a two-factor forward variance model with an initial variance curve given by:

$$\xi_0^\tau = \bar{\sigma}^2(\tau)$$

We can readily recycle expression (8.54), page 329, of the ATMF skew in the two-factor model:

$$\mathcal{S}_T^{\text{SV}} = \frac{\nu\alpha_\theta}{\hat{\sigma}_T^3 T^2} \int_0^T dt \bar{\sigma}(t) \int_t^T du \bar{\sigma}^2(u) \left[ (1-\theta)\rho_{SX^1} e^{-k_1(u-t)} + \theta\rho_{SX^2} e^{-k_2(u-t)} \right] \quad (12.43)$$

where  $\hat{\sigma}_T = \sqrt{\frac{1}{T} \int_0^T \xi_0^t dt} = \sqrt{\frac{1}{T} \int_0^T \bar{\sigma}^2(t) dt}$ . Expanding around a constant volatility  $\sigma_0$  leads to the simpler expression (8.55), page 330:

$$\mathcal{S}_T^{\text{SV}} = \nu\alpha_\theta \left[ (1-\theta)\rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta\rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2} \right] \quad (12.44)$$

Summing both contributions, the ATMF skew in the mixed model at order one both in the local volatility component and in volatility of volatility is thus:

$$\begin{aligned} \mathcal{S}_T &= \frac{1}{T} \int_0^T \frac{\hat{\sigma}_t^2 t}{\hat{\sigma}_T^2 T} \frac{\bar{\sigma}(t)}{\hat{\sigma}_T} \alpha(t) dt \\ &\quad + \frac{\nu\alpha_\theta}{\hat{\sigma}_T^3 T^2} \int_0^T dt \bar{\sigma}(t) \int_t^T du \bar{\sigma}^2(u) \left[ (1-\theta)\rho_{SX^1} e^{-k_1(u-t)} + \theta\rho_{SX^2} e^{-k_2(u-t)} \right] \end{aligned} \quad (12.45)$$

which, when taking  $\bar{\sigma}(t)$  constant, equal to  $\sigma_0$ , simplifies to:

$$\begin{aligned} \mathcal{S}_T &= \frac{1}{T} \int_0^T \frac{t}{T} \alpha(t) dt \\ &\quad + \nu\alpha_\theta \left[ (1-\theta)\rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta\rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2} \right] \end{aligned}$$

Note that the latter expression of  $\mathcal{S}_T$  does not depend on the value of  $\sigma_0$ , the constant volatility around which the order-one expansion is performed.

Expression (12.45) of  $\mathcal{S}_T$  is, in itself, useless, as our model is calibrated to market, thus  $\mathcal{S}_T$  can be read off the market smile. The expression of  $\mathcal{S}_T^{\text{SV}}$ , though, will come in handy in what follows.

How should  $\bar{\sigma}(t)$  be chosen? We can simply take it constant, equal to the ATMF volatility of the maturity  $T$  of interest.

For smiles with a marked term structure, it is preferable to calibrate  $\bar{\sigma}(t)$  to the term structure of ATMF volatilities, which are readily read off the market smile, or VS volatilities.

In what follows, we make the former choice.  $\hat{\sigma}_T$  denotes the ATMF volatility of maturity  $T$ :  $\hat{\sigma}_T = \hat{\sigma}_{F_T T}$ .

### 12.4.2 Dynamics of ATMF volatilities

We now focus on the ATMF volatility  $\hat{\sigma}_T$ .

From (12.40) the variation during  $dt$  of  $\hat{\sigma}_T$ , at order one in  $\alpha(t)$  and  $\nu$  consists of two pieces:

$$d\hat{\sigma}_T = d\delta\hat{\sigma}_T^{\text{LV}} + d\delta\hat{\sigma}_T^{\text{SV}}$$

For the sake of calculating covariances, we are only interested in the diffusive portion of  $d\hat{\sigma}_T$  – not in its drift. Thus, in what follows, only the diffusive contributions appear in the expressions of  $d\delta\hat{\sigma}_T^{\text{LV}}$  and  $d\delta\hat{\sigma}_T^{\text{SV}}$ .

- $d\delta\hat{\sigma}_T^{\text{LV}}$  is generated by the local volatility component of the mixed model. From (2.82), page 57, we have:

$$d\hat{\sigma}_T^{\text{LV}} = \mathcal{R}_T^{\text{LV}} \mathcal{S}_T^{\text{LV}} d \ln S$$

where  $\mathcal{R}_T^{\text{LV}}$  is the SSR of the local volatility component, given by expression (2.65), page 52:

$$\mathcal{R}_T^{\text{LV}} = 1 + \frac{1}{T} \int_0^T \frac{\bar{\sigma}^2(t)}{\hat{\sigma}_t \hat{\sigma}_T} \frac{\mathcal{S}_t^{\text{LV}}}{\mathcal{S}_T^{\text{LV}}} dt \quad (12.46)$$

Thus:

$$d\hat{\sigma}_T^{\text{LV}} = \left( \mathcal{S}_T^{\text{LV}} + \frac{1}{T} \int_0^T \frac{\bar{\sigma}^2(t)}{\hat{\sigma}_t \hat{\sigma}_T} \mathcal{S}_t^{\text{LV}} dt \right) d \ln S$$

- $d\delta\hat{\sigma}_{KT}^{\text{SV}}$  is given by expression (7.36), page 227:

$$d\delta\hat{\sigma}_T^{\text{SV}} = \nu \alpha_\theta \hat{\sigma}_T \left( (1 - \theta) A_1 dW^1 + \theta A_2 dW^2 \right)$$

with  $A_i$  given by:

$$A_i = \frac{\int_0^T \xi_0^\tau e^{-k_i \tau} d\tau}{\int_0^T \xi_0^\tau d\tau} = \frac{\int_0^T \bar{\sigma}^2(\tau) e^{-k_i \tau} d\tau}{\int_0^T \bar{\sigma}^2(\tau) d\tau} \quad (12.47)$$

$\delta\hat{\sigma}_{KT}^{\text{SV}}$  can be calculated from knowledge of the term-structure of ATMF volatilities – to which  $\bar{\sigma}(t)$  is calibrated – and the parameters of the two-factor model.

$d\delta\hat{\sigma}_T^{\text{LV}}$  on the other hand depends on  $\mathcal{S}_t^{\text{LV}}$ , that is the ATMF skew generated by the local volatility component of our model, which we do not know explicitly. This is readily taken care of by using (12.40):  $\mathcal{S}_t^{\text{LV}} = \mathcal{S}_t - \mathcal{S}_t^{\text{SV}}$ .

- Bringing now everything together:

$$\begin{aligned} d\hat{\sigma}_T &= \left( (\mathcal{S}_T - \mathcal{S}_T^{\text{SV}}) + \frac{1}{T} \int_0^T \frac{\bar{\sigma}^2(t)}{\hat{\sigma}_t \hat{\sigma}_T} (\mathcal{S}_t - \mathcal{S}_t^{\text{SV}}) dt \right) d \ln S \\ &\quad + \nu \alpha_\theta \hat{\sigma}_T \left( (1 - \theta) A_1 dW^1 + \theta A_2 dW^2 \right) \end{aligned} \quad (12.48)$$

where  $\mathcal{S}_t^{\text{SV}}$  is given by (12.43).

- If instead we expand around a constant volatility, equal to the ATMF volatility of the maturity  $T$  of interest:  $\sigma_0 = \hat{\sigma}_T$ , (12.48) simplifies to:

$$\begin{aligned} d\hat{\sigma}_T &= \left( (\mathcal{S}_T - \mathcal{S}_T^{\text{SV}}) + \frac{1}{T} \int_0^T (\mathcal{S}_t - \mathcal{S}_t^{\text{SV}}) dt \right) d \ln S \\ &\quad + \nu \alpha_\theta \hat{\sigma}_T \left( (1 - \theta) \frac{1 - e^{-k_1 T}}{k_1 T} dW^1 + \theta \frac{1 - e^{-k_2 T}}{k_2 T} dW^2 \right) \end{aligned} \quad (12.49)$$

where  $\mathcal{S}_T^{\text{SV}}$  is given by (12.44).

#### 12.4.2.1 SSR

Recall the definition of the SSR:

$$\mathcal{R}_T = \frac{1}{\mathcal{S}_T} \frac{\langle d\hat{\sigma}_T d \ln S \rangle}{\langle (d \ln S)^2 \rangle} \quad (12.50)$$

Writing  $d\hat{\sigma}_T$  as  $d\delta\hat{\sigma}_T^{\text{LV}} + d\delta\hat{\sigma}_T^{\text{SV}}$  we have:

$$\mathcal{R}_T = \frac{\mathcal{R}_T^{\text{LV}} \mathcal{S}_T^{\text{LV}} + \mathcal{R}_T^{\text{SV}} \mathcal{S}_T^{\text{SV}}}{\mathcal{S}_T^{\text{LV}} + \mathcal{S}_T^{\text{SV}}} \quad (12.51)$$

where  $\mathcal{R}_T^{\text{LV}}$  is the SSR generated by the local volatility component, at order one in  $\delta\sigma$  and  $\mathcal{R}_T^{\text{SV}}$  the SSR generated by the stochastic volatility component, at order one in  $\nu$ .

Using formula (12.46) for the SSR in the local volatility model and the fact that  $\mathcal{S}_t^{\text{LV}} = \mathcal{S}_t - \mathcal{S}_t^{\text{SV}}$ :

$$\mathcal{R}_T^{\text{LV}} \mathcal{S}_T^{\text{LV}} = \mathcal{R}_T^{\text{LV}}(\text{Mkt}) \mathcal{S}_T - \mathcal{R}_T^{\text{LV}}(\text{SV}) \mathcal{S}_T^{\text{SV}}$$

where  $\mathcal{R}_T^{\text{LV}}(\text{Mkt})$  (resp.  $\mathcal{R}_T^{\text{LV}}(\text{SV})$ ) are the SSRs of the local volatility model calibrated to the market smile (resp. to the smile generated by the stochastic volatility component), that is given by (12.46) with  $\mathcal{S}_t^{\text{LV}}$  replaced with  $\mathcal{S}_t$  (resp. with  $\mathcal{S}_t^{\text{SV}}$ ).

Inserting this in (12.51) yields:

$$\mathcal{R}_T = \frac{\mathcal{R}_T^{\text{LV}}(\text{Mkt}) \mathcal{S}_T - \mathcal{R}_T^{\text{LV}}(\text{SV}) \mathcal{S}_T^{\text{SV}} + \mathcal{R}_T^{\text{SV}} \mathcal{S}_T^{\text{SV}}}{\mathcal{S}_T}$$

which supplies our final expression for the SSR of the mixed model:

$$\mathcal{R}_T = \mathcal{R}_T^{\text{LV}}(\text{Mkt}) + \frac{\mathcal{S}_T^{\text{SV}}}{\mathcal{S}_T} [\mathcal{R}_T^{\text{SV}} - \mathcal{R}_T^{\text{LV}}(\text{SV})] \quad (12.52)$$

Let us recall the expressions of  $\mathcal{R}_T^{\text{LV}}(\text{Mkt})$ ,  $\mathcal{R}_T^{\text{LV}}(\text{SV})$ ,  $\mathcal{R}_T^{\text{SV}}$  for the case of the two-factor model.

### No term structure

$$\mathcal{R}_T^{\text{LV}}(\text{Mkt}) = 1 + \frac{1}{T} \int_0^T \frac{\mathcal{S}_t}{\mathcal{S}_T} dt \quad (12.53)$$

$$\mathcal{R}_T^{\text{LV}}(\text{SV}) = 1 + \frac{1}{T} \int_0^T \frac{\mathcal{S}_t^{\text{SV}}}{\mathcal{S}_T^{\text{SV}}} dt \quad (12.54)$$

$$\mathcal{S}_T^{\text{SV}} = \nu \alpha_\theta \left[ (1 - \theta) \rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta \rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2} \right]$$

$\mathcal{R}_T^{\text{SV}}$  is given by expression (9.21), page 364:

$$\mathcal{R}_T^{\text{SV}} = \frac{(1 - \theta) \rho_{SX^1} \frac{1 - e^{-k_1 T}}{k_1 T} + \theta \rho_{SX^2} \frac{1 - e^{-k_2 T}}{k_2 T}}{(1 - \theta) \rho_{SX^1} \frac{k_1 T - (1 - e^{-k_1 T})}{(k_1 T)^2} + \theta \rho_{SX^2} \frac{k_2 T - (1 - e^{-k_2 T})}{(k_2 T)^2}} \quad (12.55)$$

### Using the term structure of ATMF volatilities

$$\begin{aligned} \mathcal{R}_T^{\text{LV}}(\text{Mkt}) &= 1 + \frac{1}{T} \int_0^T \frac{\bar{\sigma}^2(t)}{\hat{\sigma}_t \hat{\sigma}_T} \frac{\mathcal{S}_t}{\mathcal{S}_T} dt \\ \mathcal{R}_T^{\text{LV}}(\text{SV}) &= 1 + \frac{1}{T} \int_0^T \frac{\bar{\sigma}^2(t)}{\hat{\sigma}_t \hat{\sigma}_T} \frac{\mathcal{S}_t^{\text{SV}}}{\mathcal{S}_T^{\text{SV}}} dt \end{aligned}$$

$$\mathcal{S}_T^{\text{SV}} = \frac{\nu \alpha_\theta}{\hat{\sigma}_T^3 T^2} \int_0^T dt \bar{\sigma}(t) \int_t^T du \bar{\sigma}^2(u) \left[ (1 - \theta) \rho_{SX^1} e^{-k_1(u-t)} + \theta \rho_{SX^2} e^{-k_2(u-t)} \right]$$

$\mathcal{R}_T^{\text{SV}}$  is given by expression (9.19), page 364, with  $\xi_0^t = \bar{\sigma}^2(t)$ :

$$\mathcal{R}_T^{\text{SV}} = \frac{1}{\bar{\sigma}(0)} \frac{\hat{\sigma}_T^2 T \int_0^T \bar{\sigma}^2(t) \left[ (1 - \theta) \rho_{SX^1} e^{-k_1 t} + \theta \rho_{SX^2} e^{-k_2 t} \right] dt}{\int_0^T dt \bar{\sigma}(t) \left( \int_t^T \bar{\sigma}^2(u) \left[ (1 - \theta) \rho_{SX^1} e^{-k_1(u-t)} + \theta \rho_{SX^2} e^{-k_2(u-t)} \right] du \right)}$$

As observed in Section 9.4, page 359, in the discussion on the bounds of the SSR for a stochastic volatility model,  $\mathcal{R}_T$  is very sensitive to the short end of the term structure of ATMF volatilities, that is  $\bar{\sigma}(0)$ .

## A sanity check

- If the market smile happens to be that generated by the underlying stochastic volatility model of our mixed model, so that the local volatility component is a constant, then  $\mathcal{S}_T = \mathcal{S}_T^{\text{SV}}$ ,  $\mathcal{R}_T^{\text{LV}}(\text{SV}) = \mathcal{R}_T^{\text{LV}}(\text{Mkt})$ . We recover  $\mathcal{R}_T = \mathcal{R}_T^{\text{SV}}$ .
- If volatility of volatility is switched off so that we are really using a local volatility model,  $\mathcal{S}_T^{\text{SV}} = 0$ . We recover  $\mathcal{R}_T = \mathcal{R}_T^{\text{LV}}(\text{Mkt})$ .

To gain accuracy on  $\mathcal{R}_T$ , as given by (12.52), it is preferable to calculate  $\mathcal{S}_T^{\text{SV}}$  numerically. This ensures that, were the input smile generated by the underlying stochastic volatility model, we would have  $\frac{\mathcal{S}_T^{\text{SV}}}{\mathcal{S}_T} = 1$  in (12.52) thus would exactly get back the SSR of the stochastic volatility model. Computing  $\mathcal{S}_T^{\text{SV}}$  numerically can be done very efficiently in a Monte Carlo simulation; see the examples below.

### 12.4.2.2 Volatilities of volatilities

Using (12.49) or (12.48), the instantaneous volatility of  $\hat{\sigma}_T$  is readily evaluated, as well as its covariance with  $S_t$ , hence all volatility/volatility correlations and spot/volatility correlations, at order one in  $\alpha(t)$  and in  $\nu$ .

### 12.4.3 Numerical evaluation of the SSR and volatilities of volatilities

How do we calculate numerically the exact values of SSR and volatilities of volatilities?

They can be evaluated in a Monte Carlo simulation of the mixed model, without any recalibration, as the gamma/theta break-even levels are those generated by the model *with a fixed local volatility function* – see Section 12.3.2.

Our derivation is similar to that in the pure two-factor model, in Section 9.8, page 368.

In the mixed two-factor model, with a fixed local volatility function, the ATMF volatility  $\hat{\sigma}_T$  is a function of  $S, X^1, X^2$ :

$$\hat{\sigma}_T \equiv \hat{\sigma}_{FTT}(\ln S, X^1, X^2)$$

Expanding at first order in  $d \ln S, dX^1, dX^2$ :

$$d\hat{\sigma}_T = \frac{d\hat{\sigma}_T}{d \ln S} d \ln S + \frac{d\hat{\sigma}_T}{dX^1} dX^1 + \frac{d\hat{\sigma}_T}{dX^2} dX^2 \quad (12.56)$$

From the definition of the SSR in (12.50) and using that

$$E[(d \ln S_t)^2] = \sigma_0^2 dt, \quad E[d \ln S_t dX_t^1] = \rho_{SX^1} \sigma_0 dt, \quad E[d \ln S_t dX_t^2] = \rho_{SX^2} \sigma_0 dt$$

we get:

$$\mathcal{R}_T = \frac{1}{\mathcal{S}_T} \frac{E[d \ln S_t d\hat{\sigma}_T]}{E[(d \ln S_t)^2]} = \frac{1}{\mathcal{S}_T} \frac{1}{\sigma_0} \left( \frac{d\hat{\sigma}_T}{d \ln S} \sigma_0 + \frac{d\hat{\sigma}_T}{dX^1} \rho_{SX^1} + \frac{d\hat{\sigma}_T}{dX^2} \rho_{SX^2} \right)$$

where  $\sigma_0$  is the instantaneous volatility:

$$\sigma_0 = \sigma(0, S_0)$$

Thus  $\mathcal{R}_T$  can be simply evaluated numerically by computing  $\widehat{\sigma}_T$  with one repricing:

$$\mathcal{R}_T \simeq \frac{1}{\mathcal{S}_T} \frac{1}{\sigma_0} \frac{\widehat{\sigma}_T(\ln S_0 + \varepsilon \sigma_0, X_0^1 + \varepsilon \rho_{SX^1}, X_0^2 + \varepsilon \rho_{SX^2}) - \widehat{\sigma}_T(\ln S_0, X_0^1, X_0^2)}{\varepsilon}$$

where  $\varepsilon$  is a small offset. Typically we take  $X_0^2 = X_0^1 = 0$ .

As for volatilities of volatilities,  $\text{vol}(\widehat{\sigma}_{FTT})$  is obtained by squaring (12.56) and taking its expectation. We have:

$$\begin{aligned} E[d \ln S^2] &= \sigma_0^2 dt & E[(dX^1)^2] &= E[(dX^2)^2] = dt \\ E[d \ln S dX^1] &= \rho_{SX^1} \sigma_0 dt & E[d \ln S dX^2] &= \rho_{SX^2} \sigma_0 dt \end{aligned}$$

We need  $\frac{d\widehat{\sigma}_{FTT}}{d \ln S}$ ,  $\frac{d\widehat{\sigma}_{FTT}}{d X^1}$ ,  $\frac{d\widehat{\sigma}_{FTT}}{d Y}$ . Each derivative is obtained with one Monte Carlo simulation (or two when using centered differences). Numerical evaluation of a volatility of volatility thus requires three repricings (or six).

While the order-one approximations in the previous section only apply to ATMF (or VS) volatilities, one can of course numerically compute volatilities of volatilities of arbitrary strikes.

## 12.5 Numerical examples

We now test our approximations for SSR, volatility of volatility and spot/volatility correlation in a mixed model whose underlying stochastic volatility model is the two-factor model.

We use as market smile a smile generated by the two-factor model, rather than a real market smile, so that the case of pure stochastic volatility is attainable in the mixed model. The parameters we use generate a typical index smile – say, of the Euro Stoxx 50 index; our conclusions hold for general market smiles as well.

The “market smile” is generated with flat VS volatilities equal to 20% and the parameters in Table 12.1, using the mixing-solution technique of Section A.1, Chapter 8.

$\rho_{12}$  is taken equal to zero. Parameters  $\nu, \theta, k_1, k_2$  are chosen so that volatilities of VS volatilities best match the benchmark form (7.40):<sup>8</sup>

$$\nu_T^B(t) = \sigma_0 \left( \frac{\tau_0}{T-t} \right)^\alpha$$

---

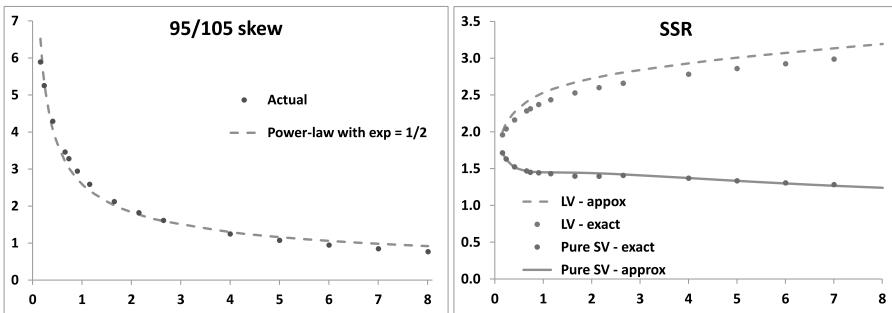
<sup>8</sup>See page 228 for a discussion of the parametrization of the two-factor model.

$\nu$	$\theta$	$k_1$	$k_2$	$\rho_{12}$	$\rho_{SX^1}$	$\rho_{SX^2}$
310%	0.139	8.59	0.47	0%	-54.0%	-62.3%

**Table 12.1:** Parameters of the two-factor model used for generating the “market smile”.

with  $\alpha = 0.6$ ,  $\tau_0 = 3$  months and the (lognormal) volatility of the VS volatility for a 3-month VS volatility is  $\sigma_0 = 125\%$ . Correlations  $\rho_{SX^1}$  and  $\rho_{SX^2}$  are chosen so as to generate an ATMF skew that approximately decays like  $\frac{1}{\sqrt{T}}$ , a typical scaling of index smiles. We use zero rate and repo for simplicity.

The ATMF skew of the “market smile” as well as the SSR of the pure two-factor stochastic volatility model are shown in Figure 12.1. In the expansion that produces the approximate formulas of Section 12.4.1 and 12.4.2,  $\hat{\sigma}_T$  has been taken equal to the VS volatility for maturity  $T$ , here 20%.



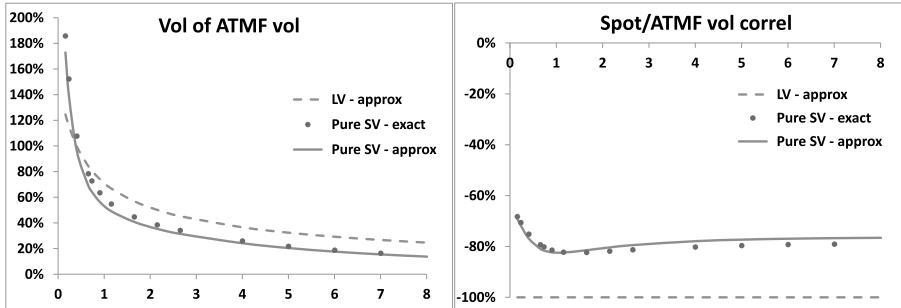
**Figure 12.1:** Left: term structure of the ATMF skew of the “market smile” expressed as the difference of implied volatilities of the 95% and 105% strikes in volatility points as a function of maturity (years) together with a power-law fit  $\frac{1}{T^\gamma}$  with  $\gamma = \frac{1}{2}$ . Right: SSR of (a) the two-factor model, (b) the local volatility model calibrated on the smile of the former, as a function of maturity (years), either calculated in a Monte Carlo simulation (exact) or using, respectively, approximate formulas (12.55) and (12.53) (approx).

Notice how approximate values for (a) the SSR of the two-factor model and (b) the SSR of the local volatility model agree with actual values.

Both values of the SSR start from the value of 2 for very short maturities. Since the ATMF skew decays approximately like  $\frac{1}{\sqrt{T}}$ , that is with an exponent  $\frac{1}{2}$ , we expect that, for long maturities, approximately:

- the SSR of the stochastic volatility model tends to  $2 - \frac{1}{2} = 1.5$
- the SSR of the local volatility model tends to the value of  $\frac{2 - \frac{1}{2}}{1 - \frac{1}{2}} = 3$

This is indeed observed in Figure 12.1.<sup>9</sup> Volatilities of ATMF volatilities and spot/ATMF volatility correlations in both models appear in Figure 12.2.



**Figure 12.2:** Volatilities of ATMF volatilities in the two-factor model and in the local volatility model calibrated on the same smile (left), and spot/ATMF volatility correlation (right) as a function of maturity (years).

Unsurprisingly, spot/volatility correlations in the local volatility model are equal to  $-1$ . The approximate values derived from expression (12.48) for  $d\hat{\sigma}_T$  agree well with “exact” values calculated in a Monte Carlo simulation.

### Halving spot/volatility correlations – Figure 12.3

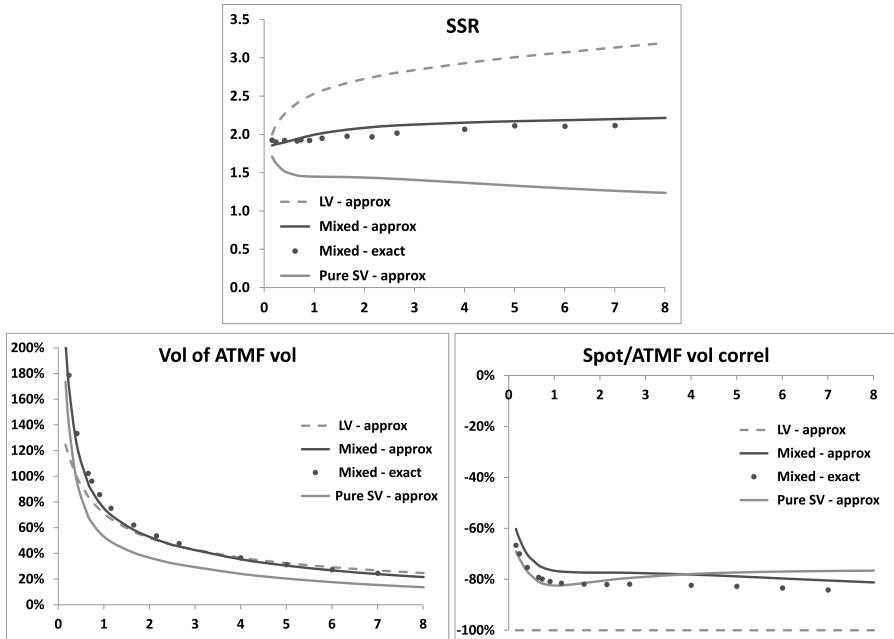
Using the same “market smile”, we still use the two-factor model as underlying stochastic volatility model, but halve the values of  $\rho_{SX^1}$  and  $\rho_{SX^2}$ . Roughly half of the skew now needs to be generated by the local volatility component and we expect the SSR of the mixed model to lie in between the two curves in the right-hand graph of Figure 12.1. Figure 12.3 – where those two curves are shown for reference – shows that it is indeed the case, and that formula (12.52) works well.

The “exact” values in figure 12.3 is obtained in a Monte Carlo simulation, with the local volatility component calibrated using the particle method.

### Halving volatilities of volatilities – Figure 12.4

Rather than halving  $\rho_{SX^1}$  and  $\rho_{SX^2}$  we now halve  $\nu$ . Again, about half of the ATMF skew now needs to be generated by the local volatility component. The SSR, volatilities of ATMF volatilities and spot/ATMF volatility correlation appear in Figure 12.4. Again, the curves for the case of a pure stochastic volatility model and local volatility model, graphed in Figure 12.2, are shown for reference.

<sup>9</sup>We refer the reader to Section 9.5, page 361, for a discussion of the relationship of the long-maturity limit of the SSR to the decay of the ATMF skew in time-homogeneous stochastic volatility models, and to Section 2.5.4, page 56, for a discussion of the corresponding relationship in the context of the local volatility model.



**Figure 12.3:** Top: SSR of the mixed model, compared to that generated by (a) the local volatility model, (b) the two-factor model used to generate the “market smile”, as a function of maturity (years). In the mixed model,  $\rho_{SX^1}$  and  $\rho_{SX^2}$  are halved. Bottom: volatilities of ATM volatilities (left) and spot/ATM vol correlations (right) in (a) the two-factor model, (b) the local volatility model calibrated to the smile of the former, (c) the mixed model.

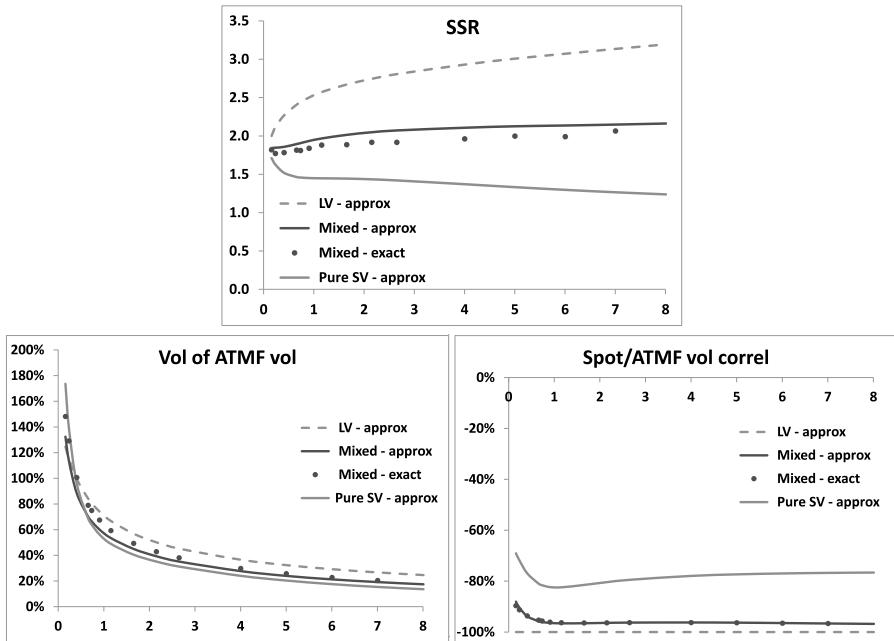
### Raising volatilities of volatilities – Figure 12.5

What if we increase  $\nu$  so that the underlying stochastic volatility model generates a steeper smile than the input smile, with the effect that the local volatility component now generates positive skew?

We could use as market smile that produced by the two-factor model with parameters in Table 12.1 and raise  $\nu$  – say by 50%. With such high level of volatility of volatility, however, calibration by the particle method does not function well anymore.<sup>10</sup>

To circumvent this difficulty, we generate a different “market smile” using parameters in Table 12.1, but with  $\nu$  halved –  $\nu = 155\%$  – and then use  $\nu = 232.5\%$  in the mixed model. Numerical results are reported in Figure 12.5.

<sup>10</sup>There is indeed no mathematical guarantee that, given a non-arbitrageable market smile and an underlying stochastic volatility model, there exists a local volatility function such that the mixed model recovers the market smile. Deterioration of the quality of calibration is typically observed for (very) large levels of volatility of volatility.



**Figure 12.4:** Top: SSR of the mixed model, compared to that generated by (a) the local volatility model, (b) the two-factor model used to generate the “market smile”, as a function of maturity (years). In the mixed model  $\nu$  is halved.

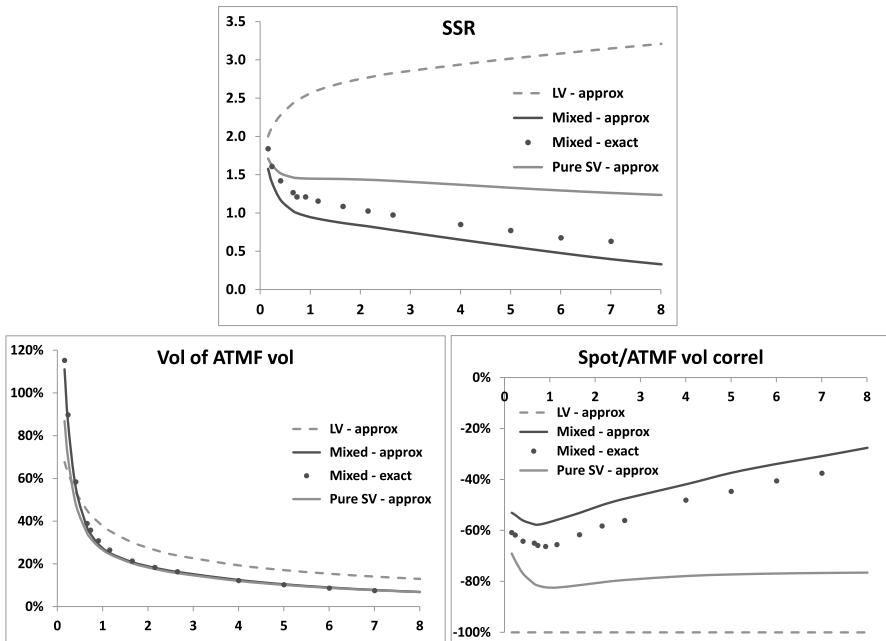
Bottom: volatilities of ATM volatilities (left) and spot/ATMF volatility correlations (right) in (a) the two-factor model, (b) the local volatility model calibrated to the smile of the former, (c) the mixed model.

While qualitatively correct, our expansion at order one in (a) volatility of volatility, (b) local volatility is, in this situation, less accurate, especially for the SSR and the spot/ATMF correlation.

Observe that, by parametrizing the underlying stochastic volatility model so that the ATM skew it generates is stronger than in the “market smile”, the SSR is lower than that of the stochastic volatility model used for generating the “market smile”. For sufficiently long maturities,  $\mathcal{R}_T \leq 1$ .

## 12.6 Discussion

Graphs in figures 12.3, 12.4, 12.5 confirm that our objective in developing local-stochastic volatility models was reached . Mixed models – in the admissible class



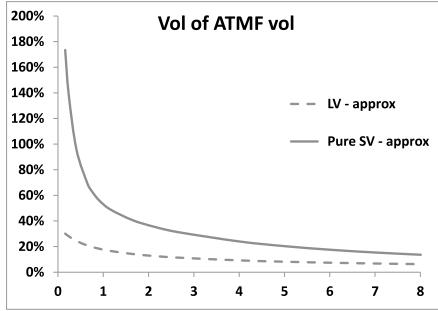
**Figure 12.5:** Top: SSR of the mixed model, compared to that generated by (a) the local volatility model, (b) the two-factor model used to generate the “market smile”, as a function of maturity (years). In the mixed model  $\nu$  is multiplied by 1.5.  
 Bottom: volatilities of ATM volatilities (left) and spot/ATMF volatility correlations (right) in (a) the two-factor model, (b) the local volatility model calibrated to the smile of the former, (c) the mixed model.

– do afford some flexibility as to the SSR, volatility of volatility and spot/volatility correlation.

While our examples seem to show that levels of volatility of volatility in the mixed model vary minimally in figures 12.3 and 12.4, this is due to our choice of parameters: volatilities of volatilities in the two-factor model used to generate the “market smile” and the local volatility model are very similar, in the first place.

This is apparent in the left-hand graph of Figure 12.2. Had we chosen vanishing values of  $\rho_{SX^1}$  and  $\rho_{SX^2}$  to generate the “market smile”, the latter would have been U-shaped with the effect that volatilities of volatilities in the local volatility model would have been almost vanishing.

For example, start from parameters in Table 12.1 and reduce correlations by a factor of 4:  $\rho_{SX^1} = -13.5\%$  and  $\rho_{SX^2} = -15.6\%$ . Figure 12.6 shows volatilities of ATM volatilities in the two-factor model and in the local volatility model – it is apparent that the difference between volatility-of-volatility levels is larger than in Figure 12.2.



**Figure 12.6:** Volatilities of ATM volatilities in the two-factor model and in the local volatility model calibrated on the same smile of the latter, as a function of maturity (years). The parameters of the two-factor model are those of Table 12.1, page 479, except  $\rho_{SX^1} = -13.5\%$  and  $\rho_{SX^2} = -15.6\%$ .

How can the different SSRs in figures 12.1, 12.3, 12.4 be reconciled with the fact that the corresponding vanilla smiles are all identical? The same can be asked of the SSRs in Figure 9.9, page 380.

Recall expression (8.32), page 319, relating the ATM skew to the weighted average of instantaneous covariances of  $S$  and the square of the ATM volatility for the residual maturity  $\hat{\sigma}_{FTT}$ :

$$\mathcal{S}_T = \frac{1}{2\hat{\sigma}_T^3(0)T} \int_0^T \frac{T-t}{T} \frac{\langle d\ln S_t d\hat{\sigma}_{FTT}^2(t) \rangle}{dt} dt \quad (12.57)$$

This general expression is correct at order one in volatility of volatility and holds as long as  $\langle d\ln S_t d\hat{\sigma}_{FTT}^2(t) \rangle$  does not depend on  $S_t$ .

In our analysis above,  $\hat{\sigma}_{FTT}^2$  is the sum of three pieces. From (12.40):

$$\hat{\sigma}_{FTT} = \hat{\sigma}_T + \delta\hat{\sigma}_{FTT}^{LV} + \delta\hat{\sigma}_{FTT}^{SV}$$

where  $\hat{\sigma}_T$  is the order-zero contribution, which is static, and  $\delta\hat{\sigma}_{FTT}^{LV}$  and  $\delta\hat{\sigma}_{FTT}^{SV}$  are respectively the contribution of the local volatility and stochastic volatility components, at order one in  $\alpha(t)$  and  $\nu$ . At order one in  $\alpha(t)$  and  $\nu$ , we can rewrite (12.57) as:

$$\mathcal{S}_T = \frac{1}{\hat{\sigma}_T^3(0)T} \int_0^T \frac{T-t}{T} \hat{\sigma}_T(t) \frac{\langle d\ln S_t d\hat{\sigma}_{FTT}(t) \rangle}{dt} dt \quad (12.58)$$

where  $\hat{\sigma}_T(t) = \sqrt{\frac{1}{T-t} \int_t^T \bar{\sigma}^2(u) du}$ . The local volatility function is given by:

$$\sigma(t, S) = \bar{\sigma}(t) + \alpha(t) \ln \frac{S}{F_t}$$

Its contribution to the covariance in (12.58) is:

$$\begin{aligned}\langle d\ln S_t d\delta\hat{\sigma}_{F_T T}^{\text{LV}}(t) \rangle &= \left\langle \frac{d\delta\hat{\sigma}_{F_T T}^{\text{LV}}}{d\ln S}(d\ln S_t)^2 \right\rangle \\ &= \left( \frac{1}{T-t} \int_t^T \frac{\bar{\sigma}(u)}{\hat{\sigma}_T(u)} \alpha(u) du \right) \sigma^2(t) dt\end{aligned}$$

where we have made use of expression (2.60c), page 51, for  $\frac{d\delta\hat{\sigma}_{F_T T}^{\text{LV}}}{d\ln S}$ . Thus  $\langle d\ln S_t d\delta\hat{\sigma}_{F_T T}^{\text{LV}}(t) \rangle$  does not depend on  $S_t$ . From (12.49)  $d\delta\hat{\sigma}_{F_T T}^{\text{SV}}$  is given by:

$$d\delta\hat{\sigma}_{F_T T}^{\text{SV}} = \nu \alpha_\theta \hat{\sigma}_T \left( (1-\theta) A_1 dW^1 + \theta A_2 dW^2 \right)$$

thus  $\langle d\ln S_t d\delta\hat{\sigma}_{F_T T}^{\text{SV}}(t) \rangle$  does not depend on  $S_t$  either.

The conclusion is that expression (12.57) for  $\mathcal{S}_T$  holds in our mixed model, at order one in  $\alpha(t)$  and  $\nu$ .

Consider two models generating the same smile – say the two-factor stochastic volatility model and a local volatility model calibrated on the smile of the former. From (12.57), the ATMF skew of the smile used for calibration sets the value of the integrated spot/volatility covariance, for all  $T$ .

### Time-homogeneous models

If the model were time-homogeneous, that is if  $\langle d\ln S_t d\hat{\sigma}_T^2(t) \rangle$  were a function of  $T-t$ , knowledge of  $\mathcal{S}_T$  for all  $T$  would determine  $\langle d\ln S_t d\hat{\sigma}_T^2(t) \rangle$  and in particular its value for  $t=0$ : the instantaneous covariance of  $\ln S$  with the ATMF volatility of maturity  $T$ .

Two time-homogeneous models calibrated to the same smile would generate identical SSRs – provided the assumptions needed for (12.57) hold – and also the same future ATMF skews.

The reason why the local volatility model generates a different SSR than a time-homogeneous stochastic volatility model is due to the non-homogeneity of the former. The larger SSRs generated by the local volatility model in Figure 9.9 point to the fact that, in the integral in (12.57), the contribution from covariances for  $t$  near 0 to the integral in (12.57) is larger in the local volatility model than in the two-factor model.

#### 12.6.1 Future smiles in mixed models

The implication of the larger SSRs in the local volatility model is that future spot/volatility covariances are lower in the local volatility model; thus accounting for weaker future skews. Mixed models parametrized such that their SSR is lower than that of the local volatility model will thus produce larger future skews. Because, as seen in Section 2.6, page 63, future skews generated by the local volatility component

quickly die off, future skews for far-away forward dates are predominantly generated by the stochastic volatility component of the mixed model.

We illustrate these properties using the example of foward-start call spreads struck at  $K_{\text{lo}} = 95\%$  and  $K_{\text{hi}} = 105\%$  with maturity 3 months. The payoff of each forward-start option is

$$\left( \frac{S_{T+\Delta}}{S_T} - K_{\text{lo}} \right)^+ - \left( \frac{S_{T+\Delta}}{S_T} - K_{\text{hi}} \right)^+$$

where  $\Delta = 3$  months and the forward-start dates  $T$  are quarterly dates, from  $T = 0$  to  $T = 4$  years and 9 months. We thus have 20 such options.

We take vanishing rate and repo, use as “market smile” the smile generated by the two-factor model, parametrized as in Table 12.1, page 479, and a term structure of VS volatilities flat at 20%.

We price our 20 forward-start options in four different models.<sup>11</sup>

**I** the two-factor model thus parametrized.

**II** a mixed model calibrated to the “market smile”, with the underlying two-factor model having a value of  $\nu$  halved – this is the model that was used to produce Figure 12.4.

**III** the local volatility model calibrated to the “market smile”.

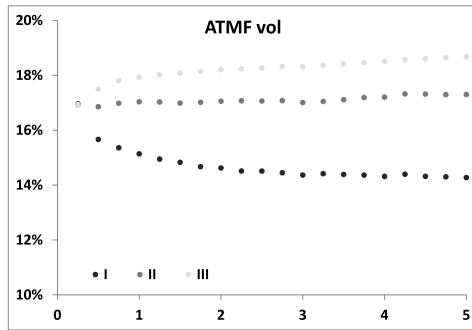
**IV** the two-factor model parametrized as in I, but with  $\nu$  halved. This is the stochastic volatility portion of model II.

For each foward-start date  $T$ , we also price the forward-start ATM option that pays  $\left( \frac{S_{T+\Delta}}{S_T} - 1 \right)^+$  and imply a Black-Scholes volatility. These implied volatilities appear in Figure 12.7.

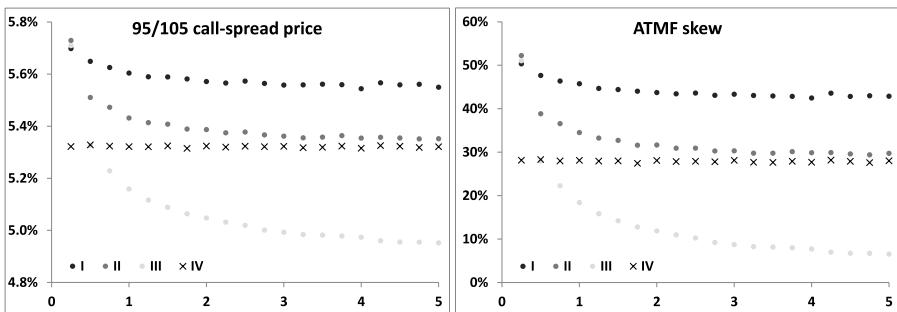
Observe that for  $T = 0$ , our forward option is simply a vanilla option of maturity 3 months, thus has identical prices in models I, II, III as they generate the same vanilla smile. Thus the three curves collapse onto the same value for the first option.

For further-away forward-start dates, implied volatilities – hence prices – are lower in model I. This can be traced mainly to the higher volatility of volatility in model I. We refer the reader to the discussion of forward-start options in Section 3.1.6 of Chapter 3, page 111. For a forward-start ATM call option, adjustment  $\delta P_1$  is negative and more so when volatility of volatility is higher.

We now turn to the foward call spreads. From the price of each call spread we imply a forward skew  $\mathcal{S}$  by equating the model price to a Black-Scholes price calculated using implied volatilities  $\hat{\sigma}_{ATM} + \mathcal{S} \ln(K_{\text{lo}})$  for strike  $K_{\text{lo}}$ , and  $\hat{\sigma}_{ATM} + \mathcal{S} \ln(K_{\text{hi}})$  for strike  $K_{\text{hi}}$ .  $\hat{\sigma}_{ATM}$  is the ATM implied volatility of the corresponding



**Figure 12.7:** Implied volatilities of forward-start ATM options of maturity as a function of their maturities  $T + \Delta$ .



**Figure 12.8:** Prices (left) and implied skews (right, multiplied by  $-1$ ) of forward-start 95/105 call spreads as a function of their maturities  $T + \Delta$ .

ATM call option. Prices and forward skews (multiplied by  $-1$ ) are shown in Figure 12.8.

- First note the similarity of both graphs in Figure 12.8: the price of a narrow call spread centered on the money directly reflect the ATM skew.
- Then observe how the forward skew in mixed model II is lower than that of the pure stochastic volatility model I, while the forward skew in local volatility model IV is lower still. This confirms what we mentioned above, based on formula (12.58) for the SSR: higher SSRs translate into lower forward skews.
- Finally observe how prices – and implied skews – in model II tend towards those of model IV, for far-away forward dates. Model IV is in fact the stochastic volatility component of model II. This confirms that, indeed, the local volatility

<sup>11</sup>I thank Pierre Henry-Labordère for generating these results.

component of mixed model II hardly generates any forward skew far into the future.

Thus parameters of the underlying two-factor model can be chosen to control the forward skew of the mixed model. Long-dated cliques hardly have any sensitivity to the local volatility component.

---

## 12.7 Conclusion

- Provided condition (12.32) holds, local-stochastic volatility models are legitimate models that can be used in trading applications. They are market models for the spot and vanilla options that possess a Markovian representation in terms of the spot price and the state variables of the underlying stochastic volatility model.

Unlike the local volatility model – also a market model for the same instruments – local-stochastic volatility models afford some control on spot/volatility and volatility/volatility break-even levels, while maintaining calibration to the vanilla smile. This is achieved by appropriately choosing the parameters of the underlying stochastic volatility model.

- The approximate expressions for  $d\hat{\sigma}_T$  in Section 12.4.1 and the resulting approximate formulas for SSR, volatilities of volatilities and spot/volatility correlations allow for a quick assessment of the model's break-even levels for a given market smile, and how the latter vary if the market smile changes – which of course is bound to happen as we risk-manage a derivative position through time.
- In practice, unlike delta hedging, vega hedging is typically not performed on a daily basis, because of larger bid/offer costs. See the discussion in Section 9.11.3, page 383, regarding the carry P&L of a delta-hedged, vega-hedged position when vega and delta rehedging frequencies differ.
- A lower SSR than in the local volatility model also translates into stronger future skews. Future smiles in mixed models are mostly generated by the stochastic volatility component. Thus, mixed models can be parametrized so as to achieve given levels of future skews – which are related to *future* break-even levels for the spot/volatility covariance – rather than choosing parameters to achieve desired *present* break-even levels.
- Local-stochastic volatility models can be used to price and risk-manage path-dependent payoffs that involve spot observations.

Unlike the forward variances models of Chapter 7, they are not well-suited to options involving VIX observations, as forward variances – let alone VIX futures – are not directly accessible in the model.

Whenever one needs to (a) have a handle on future skews, (b) have explicit access to VIX futures or forward variances, the discrete forward variance models of Chapter 7 should be employed.

---

## Appendix A – alternative schemes for the PDE method

We present here a technique for deriving schemes for the forward equation (12.12) for the density from schemes for the backward equation for prices – this idea was first proposed by Jesper Andreasen and Brian Huge – see [1]. We use here the notations of Section 12.2.4. While the density  $\varphi(t, S, X) = E[\delta(S - S_t)\delta(X - X_t)]$  obeys the forward equation (12.12), the undiscounted price  $P(t, S, X)$  of a European option satisfies the usual backward equation:

$$\frac{dP}{dt} = -LP \quad (12.59a)$$

$$L = L_S + L_X + L_{SX} \quad (12.59b)$$

with terminal condition  $P(t = T, S, X) = g(S)$  where  $T$  is the option's maturity and  $g$  is the option's payoff. Operators  $L_S$ ,  $L_X$ ,  $L_{SX}$  read:

$$\begin{aligned} L_S &= (r - q)S \frac{d}{dS} + \frac{1}{2}f(t, X)\sigma(t, S)^2 S^2 \frac{d^2}{dS^2} \\ L_X &= -kX \frac{d}{dX} + \frac{1}{2} \frac{d^2}{dX^2} \\ L_{SX} &= \rho\sqrt{f(t, X)}\sigma(t, S)S \frac{d^2}{dSdX} \end{aligned}$$

Assume we have  $\varphi$  and  $P$  at time  $t$ . Then the price at  $t = 0$ ,  $p_0 = P(t = 0, S_0, X_0)$  can be written as:

$$p_0 = \iint \varphi(t, S, X) P(t, S, X) dSdX \quad (12.60)$$

Since this holds for any  $t \in [0, T]$ , the derivative with respect to  $t$  of the right-hand side vanishes:

$$\iint \left( \frac{d\varphi}{dt} P + \varphi \frac{dP}{dt} \right) dSdX = 0 \quad (12.61)$$

(12.61) is a consistency condition relating the forward PDE (12.12) for  $\varphi$  and the backward PDE (12.59) for  $P$ .

Given operator  $L$  for the backward equation, the forward operator  $\mathcal{L}$  is such that:

$$\iint (\mathcal{L}\varphi) P \, dSdX = \iint \varphi (LP) \, dSdX \quad (12.62)$$

(12.62) can be considered a definition of  $\mathcal{L}$ .<sup>12</sup>

Consider now a discretization of  $\varphi$  and  $P$  on the same  $(S, X)$  grid with spacings  $\delta S, \delta X$ . As we did in Section 12.2.4 we now use the notation  $\varphi$  and  $P$  to also denote their discretized version –  $\varphi$  and  $P$  are then vectors of size  $n_S n_X$ ; similarly,  $L$ ,  $L$  will also designate matrices, generated by replacing differential operators with their discretized versions in (12.14).

Consider a discrete time step  $\delta t$  and assume we have a scheme – i.e. a matrix  $U_{t,t+\delta t}$  – for the backward evolution of  $P$  over  $[t, t + \delta t]$ :

$$P(t) = U_{t,t+\delta t} P(t + \delta t) \quad (12.63)$$

In the discretized version of our problem, integrals are converted into sums, and equation (12.60) giving the price at  $t = 0$  – a scalar – translates into:

$$p_0 = N \varphi(t)^\top P(t)$$

where  $\varphi(t)^\top$  denotes the transpose of vector  $\varphi(t)$  and  $N$  is a normalization constant:  $N = \delta S \delta X$ . Now use (12.63):

$$\begin{aligned} p_0 &= N \varphi(t)^\top U_{t,t+\delta t} P(t + \delta t) \\ &= N \left( U_{t,t+\delta t}^\top \varphi(t) \right)^\top P(t + \delta t) \end{aligned} \quad (12.64)$$

Identifying the right-hand side of (12.64) with  $N \varphi(t + \delta t)^\top P(t + \delta t)$  yields the following relationship between  $\varphi(t)$  and  $\varphi(t + \delta t)$ :

$$\varphi(t + \delta t) = U_{t,t+\delta t}^\top \varphi(t) \quad (12.65)$$

The upshot is that we can obtain a numerical scheme for the forward equation for  $\varphi$  by simply taking the transpose of a scheme for the backward equation for prices – the boundary conditions are automatically taken care of.

Imagine the terminal payoff  $g(S)$  is a constant  $g$  – then  $P(t, S, X) = g \forall t, S, X$ . The condition that our numerical scheme for  $P$  complies with this requirement reads:

$$U_{t,t+\delta t} \mathbf{1} = \mathbf{1} \quad (12.66)$$

where vector  $\mathbf{1}$  has its  $n_S n_X$  components all equal to 1.  $U_{t,t+\delta t} = e^{\delta t(L_S + L_X + L_{SX})}$ . The matrices representing  $L_S$ ,  $L_X$  and  $L_{SX}$ , must thus be such that when acting on vector  $\mathbf{1}$ , whose  $n_S n_X$  components are all equal to 1, the resulting vector vanishes – equivalently the components of each line of  $L_S$ ,  $L_X$ ,  $L_{SX}$  add

---

<sup>12</sup>Starting from the left-hand side of (12.62) and integrating by parts to generate the right-hand equivalent, one can check that one indeed obtains  $\mathcal{L}_S$ ,  $\mathcal{L}_X$ ,  $\mathcal{L}_{SX}$  in (12.13).

up to zero. Consider for example the operator  $\frac{d^2}{dS^2}$ . Owing to (12.14) the non-zero matrix elements of line  $i$  of its discretized version are given for  $i > 0$  by:

$$\frac{d^2}{dS^2}_{i,i-1} = \frac{1}{\delta S^2}, \quad \frac{d^2}{dS^2}_{i,i} = -\frac{2}{\delta S^2}, \quad \frac{d^2}{dS^2}_{i,i+1} = \frac{1}{\delta S^2}$$

and sum up to zero. For  $i = 0$  and  $i = n_S - 1$  we can enforce the typical boundary condition  $\frac{d^2 P}{dS^2} = 0$  – which is natural for vanilla payoffs, whose asymptotic profiles are affine, which translates into all elements of the first line ( $i = 0$ ) and last line ( $i = n_S - 1$ ) vanishing. The same conditions we impose in the  $X$  directions. With regard to  $L_{SX}$  conditions on edges and corners have to be such that each line sums up to zero.

Multiplying each side of (12.65) on the left by  $\mathbf{1}^\top$  yields:

$$\mathbf{1}^\top \varphi(t + \delta t) = \mathbf{1}^\top U_{t,t+\delta t}^\top \varphi(t) = \mathbf{1}^\top \varphi(t) \quad (12.67)$$

$\mathbf{1}^\top \varphi(t)$  is the sum of the elements of  $\varphi(t)$  – up to the normalizing constant  $N$ , the integral  $\iint \varphi(t, S, X) dS dX$  evaluated numerically. Thus, using  $U_{t,t+\delta t}^\top$  as a scheme for  $\varphi(t)$  ensures that the numerical integral of the probability density  $\varphi(t, S, X)$  is conserved in the forward scheme: starting with an initial density that integrates to one, further densities do so as well.

Besides taking care of boundary conditions, this is another attractive feature of using the transpose of the backward scheme. Note though, that generally it will not be possible to ensure that elements of  $\varphi(t)$  are all positive.

The final recipe is simple: choose a scheme for the backward equation and transpose it.

### Vanishing correlation

The backward equation is identical to the forward equation, except the initial vector is  $P(t + \delta t)$  and the final one is  $P(t)$ ; we can recycle the schemes used in Section 12.2.4.

If  $\rho$  vanishes the Peaceman-Rachford algorithm, expressed by (12.16) for the forward equation, reads – for  $P$ :

$$P(t) = I_S E_X I_X E_S P(t + \delta t) \quad (12.68)$$

where  $E_S, I_S$  are given by

$$E_S = 1 + \frac{\delta t}{2} L_S \quad I_S = \left(1 - \frac{\delta t}{2} L_S\right)^{-1}$$

and likewise for  $E_X, I_X$ . To get the numerical algorithm for  $\varphi$ , simply transpose (12.68):

$$\varphi(t + \delta t) = E_S^\top I_X^\top E_X^\top I_S^\top \varphi(t)$$

which is implemented, for example, through the following sequence:

$$\begin{aligned} \left(1 - \frac{\delta t}{2} L_S^\top\right) \varphi^* &= \varphi(t) \\ \left(1 - \frac{\delta t}{2} L_X^\top\right) \varphi^{**} &= \left(1 + \frac{\delta t}{2} L_X^\top\right) \varphi^* \\ \varphi(t + \delta t) &= \left(1 + \frac{\delta t}{2} L_S^\top\right) \varphi^{**} \end{aligned}$$

It entails the same number of multiplications/inversions of tridiagonal matrices as in (12.15).

### Non-vanishing correlation

Take the predictor-corrector scheme (12.21), express it using backward, rather than forward, operators, and transpose it:

$$\varphi(t + \delta t) = \left( E_S^\top E_X^\top + \frac{\delta t}{2} L_{SX}^\top + \frac{\delta t}{2} (E_S^\top E_X^\top + \delta t L_{SX}^\top) I_X^\top I_S^\top L_{SX}^\top \right) I_X^\top I_S^\top \varphi(t) \quad (12.69)$$

This can be implemented through the following sequence consisting of a “predictor” step:

$$\begin{aligned} \left(1 - \frac{\delta t}{2} L_S^\top\right) \varphi^0 &= \varphi(t) \\ \left(1 - \frac{\delta t}{2} L_X^\top\right) \varphi^* &= \varphi^0 \\ \tilde{\varphi} &= \delta t L_{SX}^\top \varphi^* \end{aligned}$$

a “corrector” step:

$$\begin{aligned} \left(1 - \frac{\delta t}{2} L_S^\top\right) \varphi^1 &= \tilde{\varphi} \\ \left(1 - \frac{\delta t}{2} L_X^\top\right) \varphi^2 &= \varphi^1 \\ \bar{\varphi} &= \left(1 + \frac{\delta t}{2} L_S^\top\right) \left(1 + \frac{\delta t}{2} L_X^\top\right) \varphi^2 + \delta t L_{SX}^\top \varphi^2 \end{aligned}$$

and a final step:

$$\varphi(t + \delta t) = \left(1 + \frac{\delta t}{2} L_S^\top\right) \left(1 + \frac{\delta t}{2} L_X^\top\right) \varphi^* + \frac{1}{2} (\tilde{\varphi} + \bar{\varphi})$$

Again, (12.69) can be implemented in multiple manners.

## Chapter's digest

### 12.2 Pricing equation and calibration

► In local-stochastic volatility models, the instantaneous volatility of the underlying is written as the product of a local volatility component and a stochastic volatility component, generated by an underlying stochastic volatility model:  $\sigma_t = \sqrt{\zeta_t^t} \sigma(t, S_t)$ .

► The local volatility component is calibrated so that the vanilla smile is recovered. Two methods can be used: (a) a PDE-based method based on the solution of the forward equation for the density, practically applicable to the one-factor case only, (b) the particle method, a general Monte Carlo technique that can be used regardless of the dimensionality of the underlying stochastic volatility model.



### 12.3 Usable models

► Local-stochastic volatility models are calibrated to the spot value and the vanilla smile. Are they market models for these assets – that is can they be used for trading purposes? Can the P&L of a delta-hedged, vega-hedged position be written as the sum of gamma/theta contributions involving all hedging instruments, with well-defined and payoff-independent break-even levels?

► The condition for a local-stochastic volatility model to be usable is that prices should not depend on the state variables of the underlying stochastic volatility model. Otherwise, spurious contributions to the P&L appear, that have no financial meaningfulness, causing P&L leakage. This condition is not met for most models proposed in the literature.

► The two-factor model satisfies the admissibility condition. The Heston model does not; neither does the Bloomberg model. A model whose underlying stochastic volatility model has a lognormal instantaneous volatility meets the admissibility criterion.



### 12.4 Dynamics of implied volatilities

► Models belonging to the admissible class possess well-defined break-even gamma/theta levels for the spot and vanilla implied volatilities. The latter can be computed in a Monte Carlo simulation of the model, but it is useful to have approximate values for volatilities of volatilities, spot/volatility correlations, and SSR, for the sake of sizing up model-generated break-even levels, and choosing parameters of the underlying stochastic volatility model.

- We derive approximate levels of volatilities of ATMF volatilities, spot/ATMF volatility correlations, and SSR in an expansion at order one in the volatility of volatility of the underlying stochastic volatility model, and in the local volatility component. We obtain expressions that only involve the market smile and parameters of the underlying stochastic volatility model. The (calibrated) local volatility does not appear explicitly.



## 12.6 Discussion

- We run numerical tests using as market smile that generated by the underlying stochastic volatility model, so that the case of pure stochastic volatility can be spanned.
- The accuracies of volatilities of ATMF volatilities, spot/ATMF volatility correlations, and SSR obtained in the order-one expansion are sufficient for choosing the parameters of the underlying two-factor model so as to generate the desired level of SSR.
- By selecting the level of volatility of volatility of the underlying stochastic volatility model, it is possible to cover the range of SSRs from that of the pure stochastic volatility to that of the pure local volatility, and also to explore SSRs outside of this range.
- A local volatility model calibrated to the smile of a stochastic volatility model produces higher SSRs. This is consistent with the fact that future smiles are weaker in the local volatility model. Future smiles in local-stochastic volatility models are mostly generated by the stochastic volatility component of the model.

---

## *Epilogue*

Bornons ici cette carrière.  
Les longs ouvrages me font peur.  
Loin d'épuiser une matière,  
On n'en doit prendre que la fleur.

La Fontaine, *Fables*, VI

Et pour ceux qui joignent le bon sens avec l'étude, lesquels seuls je souhaite  
pour mes juges [...]

Descartes, *Discours de la méthode*

Quant au mouvement en lui-même, je vous le déclare avec humilité, nous  
sommes impuissants à le définir.

Balzac, *La Peau de chagrin*

– This excursion, Reader, ends here. I hope the journey has been pleasant and instructive.

We attempt to build market models for hedging instruments – the underlying along with vanilla options, or a subset of them, or variance swaps, or yet other convex payoffs – such that (a) the P&L of a hedged option position is of the typical gamma/theta form, with payoff-independent break-even levels, (b) these break-even levels are – at least partially – in our control.

We could consider higher-order contributions but, in practice, managing the risk of all second-order greeks at a book level is a formidable task already.

Vanilla options are not quite independent instruments, however; they are related to one another and to the underlying itself through their terminal condition. At order one in volatility of volatility, for example, the SSR of homogeneous stochastic volatility models is related to the decay of the AMTF skew.

Characterizing the restrictions that the initial configuration of vanilla implied volatilities places on their future evolution is an unsettled issue, that may be more easily addressed by modeling other convex payoffs, for example the power payoffs of Chapter 4.

Even achieving objective (a) is not as straightforward, as highlighted by our analysis of mixed models in Chapter 12. Expressing the instantaneous volatility as the product of local and stochastic volatility components, a seemingly reasonable and innocuous ansatz, generally leads to non-functional models, for lack of a proper breakdown of the carry P&L.

This serves as a reminder that derivatives modeling does not start with the assumption of a process for  $S_t$ , but with clearly articulated modeling objectives

relating to observable quantities, regardless of how these objectives are achieved mathematically.

A commonplace statement one has been hearing at recent quant conferences is that the age of modeling is now over, that quantitative finance has become but a tedious form of accounting, the preserve of a new order of adjusters – xxA-quants: CVA-quants, FVA-quants, KVA-quants, etc.

Focusing on derivatives' risks, on their proper modeling and the meaningfulness of model-generated prices – the things being adjusted – is the symptom of an old-fashioned, outdated, mindset.

I respectfully dissent. Product risks and modeling choices are still begging for a proper understanding; forty years after Black-Scholes, work on the next generation of models has just started.

What about calibration, a deceptive notion we should strive to abolish? No one says that they “calibrate” the spot value in the Black-Scholes model: a model should naturally take as inputs the market values of instruments used as hedges. Then, *parameters* should be chosen so as to generate the desired break-even levels for the carry P&L.

Should they be calibrated to market prices? This is meaningful only if the difference between their calibrated and realized values can be materialized as the P&L of an actual trading strategy, a rare occurrence.

Therefore, Reader, you will do well to resist the compulsion of calibration and the addictive psychological reward that comes with it.

With these last words of encouragement, Reader, I bid you farewell.

---

## Bibliography

- [1] Andreasen, J., Huge B.: *Random grids*, Risk Magazine, July, pp. 66–71, 2011.
- [2] Andersen, L., Andreasen, J.: *Jump-Diffusion Processes: Volatility Smile Fitting and Numerical Methods for Pricing*, available at SSRN: <http://ssrn.com/abstract=171438>.
- [3] Avellaneda, M., Levy, A., Paras, A.: *Pricing and hedging derivative securities in markets with uncertain volatilities*, Applied Mathematical Finance 2(2), pp. 73–88, 1995.
- [4] Avellaneda, M., Paras, A.: *Pricing and hedging derivative securities in markets with uncertain volatilities: the Lagrangian uncertain volatility model*, Applied Mathematical Finance, 3, pp. 21–52, 1996.
- [5] Backus, D., Foresi, S., Wu, L.: *Accounting for biases in Black-Scholes*, available at SSRN: <http://ssrn.com/abstract=585623>, 1997.
- [6] Berestycki, H., Busca, J., Florent, I.: *Asymptotics and calibration of local volatility models*, Quantitative Finance, 2(1), pp. 61–69, 2002.
- [7] Berestycki, H., Busca, J., Florent, I.: *Computing the implied volatility in stochastic volatility models*, Communications on Pure and Applied Mathematics, Vol. LVII, pp. 1352–1373, 2004.
- [8] Bergomi, L.: *Smile dynamics*, Risk Magazine, September, pp. 117–123, 2004. Also available at SSRN: <http://ssrn.com/abstract=1493294>.
- [9] Bergomi, L.: *Smile dynamics II*, Risk Magazine, October, pp. 67–73, 2005. Also available at SSRN: <http://ssrn.com/abstract=1493302>.
- [10] Bergomi, L.: *Smile dynamics III*, Risk Magazine, October, pp. 90–96, 2008. Also available at SSRN: <http://ssrn.com/abstract=1493308>.
- [11] Bergomi, L.: *Smile dynamics IV*, Risk Magazine, December, pp. 94–100, 2009. Also available at SSRN: <http://ssrn.com/abstract=1520443>.
- [12] Bergomi, L.: *Correlations in asynchronous markets*, Risk Magazine, November, pp. 76–82, 2010. Also available at SSRN: <http://ssrn.com/abstract=1635866>.
- [13] Bergomi, L., Guyon, J.: *Stochastic volatility's orderly smiles*, Risk Magazine, May, pp. 60–66, 2012. Also available at SSRN: <http://ssrn.com/abstract=1967470>.

- [14] Bick, A.: *Quadratic-variation-based dynamic strategies*, Management Science, 41(4), pp. 722–732, 1995.
- [15] Björk, T., Blix, M., Landén, C.: *On finite dimensional realizations for the term structure of futures prices*, International Journal of Theoretical and Applied Finance, 9(03), pp. 281–314, 2006.
- [16] Bos, M., Vandermark, S.: *Finessing fixed dividends*, Risk Magazine, September, pp. 157–158, 2002.
- [17] Bouchaud, J. Ph., Potters, M., Sestovic, D.: *Hedge your Monte Carlo*, Risk Magazine, March, pp. 133–136, 2001. Also available at SSRN: <http://ssrn.com/abstract=238868>.
- [18] Breeden, D., Litzenberger, R.: *Prices of state contingent claims implicit in option prices*, Journal of Business, 51, pp. 621–651, 1978.
- [19] Buehler, H.: *Volatility and dividends – volatility modeling with cash dividends and simple credit risk*, available at SSRN: <http://ssrn.com/abstract=1141877>.
- [20] Buehler, H.: *Consistent variance curve models*, Finance and Stochastics, 10(2), pp. 178–203, 2006. Also available at SSRN: <http://ssrn.com/abstract=687258>.
- [21] Carmona, R., Nadtochiy, S.: *Local volatility dynamic models*, Finance and Stochastics (13), pp. 1–48, 2009.
- [22] Carr, P., Chou, A.: *Breaking barriers*, Risk Magazine, September, pp. 139–145, 1997.
- [23] Carr, P., Madan, D.: *Determining volatility surfaces and option values from an implied volatility smile*, Quantitative Analysis in Financial Markets, Vol II, M. Avellaneda, ed, pp. 163–191, 1998.
- [24] Carr, P., Lewis, K.: *Corridor variance swaps*, Risk Magazine, February, pp. 67–72, 2004.
- [25] Carr, P., Madan, D.: *Towards a theory of volatility trading*, Volatility, Risk Publications, Robert Jarrow, ed., pp. 417–427, 2002.
- [26] Carr, P., Geman, H., Madan, D., Yor, M.: *Stochastic volatility for Lévy processes*, Mathematical Finance 13(3), pp. 345–382, 2003.
- [27] Carr, P., Lee, R.: *Robust replication of volatility derivatives*, available at: <http://www.math.uchicago.edu/~rl/rrvd.pdf>, 2009.
- [28] Carr, P., Lee, R. *Hedging variance options on continuous semimartingales*, Finance and Stochastics, 14 (2), pp. 179–207, 2010.
- [29] Castagna, A., Mercurio, F.: *The vanna-volga method for implied volatilities*, Risk Magazine, January, pp. 106–111, 2007.

- [30] Cheyette, O.: *Markov representation of the Heath-Jarrow-Morton model*, available at SSRN: <http://ssrn.com/abstract=6073>.
- [31] Cherny, A., Dupire, B.: *On certain distributions associated with the range of martingales*, Optimality and Risk – Modern Trends in Mathematical Finance, pp. 29–38, 2010.
- [32] Chriss, N., Morokoff, W.: *Market risk of variance swaps*, Risk Magazine, October, pp. 55–59, 1999.
- [33] Cont, R., Tankov, P.: *Financial modeling with jump processes*, Chapman & Hall / CRC Press, Financial Mathematics Series, 2003.
- [34] Cox, A., Wang, J.: *Optimal robust bounds for variance options*, <http://arxiv.org/abs/1308.4363>, 2013.
- [35] Davis, M. H., Panas, V. G., Zariphopoulou, T.: *European option pricing with transaction costs*, SIAM Journal on Control and Optimization, 31(2), pp. 470–493, 1993.
- [36] De Marco, S., Henry-Labordère, P.: *Linking vanillas and VIX options: a constrained martingale optimal transport problem*, available at SSRN: <http://ssrn.com/abstract=2354898>.
- [37] Derman, E., Kani, I.: *Riding on a smile*, Risk Magazine, February, pp. 32–39, 1994.
- [38] Duanmu, Z.: *Rational pricing of options on realized volatility – the Black Scholes way*, Global Derivatives conference, Madrid, 2004.
- [39] Derman, E., Kani, I.: *Stochastic implied trees: arbitrage pricing with stochastic term and strike structure of volatility*, International Journal of Theoretical and Applied Finance, 1(01), pp. 61–110, 1998.
- [40] Dupire, B.: *Pricing with a smile*, Risk Magazine, January, pp. 18–20, 1994.
- [41] Dupire, B.: *Functional Ito calculus and volatility hedge*, Global Derivatives conference, Paris, 2009.
- [42] Dupire, B.: *Arbitrage bounds for volatility derivatives as free boundary problem*, available at [http://www.math.kth.se/pde\\_finance/presentations/Bruno.pdf](http://www.math.kth.se/pde_finance/presentations/Bruno.pdf), 2005.
- [43] Durrleman, V.: *From implied to spot volatilities*, Finance and Stochastics, 14 (4), pp. 157–177, 2010.
- [44] Durrleman, V., El Karoui, N.: *Coupling smiles*, Quantitative Finance, 8(6), pp. 573–590, 2008. Also available at SSRN: <http://ssrn.com/abstract=1005332>.
- [45] Fisher, T., Tataru, G.: *Non-parametric stochastic local volatility modeling*, Global Derivatives conference, Paris, 2010.

- [46] Fukasawa, M.: *The normalizing transformation of the implied volatility smile*, Mathematical Finance, 22(4), pp. 753–762, 2012. Preprint version available at <http://arxiv.org/abs/1008.5055>, 2010.
- [47] Galichon, A., Henry-Labordère, P., Touzi, N.: *A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options*, Annals of Applied Probability, 24(1), pp. 312–336, 2014.
- [48] Gatheral, J.: *The volatility surface: a practitioner’s guide*, Wiley Finance, 2006.
- [49] Gatheral, J., Jacquier, A.: *Arbitrage-free SVI volatility surfaces*, Quantitative Finance, 14(1), pp. 59–71, 2014. Also available at SSRN: <http://ssrn.com/abstract=2033323>.
- [50] Green, R.C., Jarrow, R.A.: *Spanning and completeness in markets with contingent claims*, Journal of Economic Theory, 41, pp. 202–210, 1987.
- [51] Guyon, J., Henry-Labordère, P.: *From spot volatilities to implied volatilities*, Risk Magazine, June, pp. 79–84, 2011. Also available at SSRN: <http://ssrn.com/abstract=1663878>.
- [52] Guyon, J., Henry-Labordère, P.: *Being particular about calibration*, Risk Magazine, January, pp. 92–97, 2012.
- [53] Guyon, J., Henry-Labordère, P.: *Nonlinear option pricing*, Chapman & Hall/CRC Press, Financial Mathematics Series, 2013.
- [54] Gyöngi, I.: *Mimicking the one-dimensional marginal distributions of processes having an Ito differential*, Probability Theory and Related Fields, 71, pp. 501–516, 1986.
- [55] Hagan, P.S., Kumar, D., Lesniewski, A.S., Woodward, D.E.: *Managing smile risk*, Wilmott Magazine, September, pp. 84–108, 2002.
- [56] Henry-Labordère, P.: *Analysis, geometry, and modeling in finance: advanced methods in option pricing*, Chapman & Hall/CRC Press, Financial Mathematics Series, 2008.
- [57] Henry-Labordère, P.: *Automated option pricing: numerical methods*, available at SSRN: <http://ssrn.com/abstract=1968344>, 2011.
- [58] Henry-Labordère, P.: *Calibration of local stochastic volatility models to market smiles: a Monte-Carlo approach*, Risk Magazine, September, pp. 113–117, 2009. Also available at SSRN: <http://ssrn.com/abstract=1493306>.
- [59] Henry-Labordère, P.: *Vega decomposition of exotics on vanillas: a Monte-Carlo approach*, available at SSRN: <http://ssrn.com/abstract=2229990>.
- [60] Heston, S.: *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, Review of Financial Studies, 6(2), pp. 327–343, 1993.

- [61] Higham, N. J.: *Computing the nearest correlation matrix – a problem from finance*, IMA Journal of Numerical Analysis, 22(3), pp. 329–343, 2002.
- [62] Hobson, D.: *The Skorokhod embedding problem and model-independent bounds for option prices*, Paris-Princeton Lectures on Mathematical Finance, pp. 267–318, 2010.
- [63] Hobson, D., Klimmek, M.: *Model-independent hedging strategies for variance swaps*, Finance and Stochastics, 16(4), pp. 611–649, 2012.
- [64] Hunt, P., Kennedy, J., Pelsser, A.: *Markov-functional interest rate models*, Finance and Stochastics, 4(4), pp. 391–408, 2000.
- [65] Lee, R.: *Weighted variance swap*, Encyclopedia of Quantitative Finance, 2010.  
Also available at [http://math.uchicago.edu/~rl/EQF\\_weightedvarianceswap.pdf](http://math.uchicago.edu/~rl/EQF_weightedvarianceswap.pdf).
- [66] Lee, R.: *Implied and local volatilities under stochastic volatility*, International Journal of Theoretical and Applied Finance, 4(01), pp. 45–89, 2001.
- [67] Lee, R.: *The moment formula for implied volatility at extreme strikes*, Mathematical Finance, 14(3), pp. 469–480, 2004.
- [68] Leland, H. E.: *Option pricing and replication with transaction costs*, The Journal of Finance, 40(5), pp. 1283–1301, 1985.
- [69] Lewis, A.: *Option valuation under stochastic volatility*, Finance Press, 2000.
- [70] Lipton, A.: *The vol smile problem*, Risk Magazine, February, pp. 61–65, 2002.
- [71] Lyons, T.: *Uncertain volatility and the risk-free synthesis of derivatives*, Applied Mathematical Finance, 2(2), pp. 117–133, 1995.
- [72] Marcinkiewicz, J.: *Sur une propriété de la loi de Gauss*, Mathematische Zeitschrift, 44, pp. 612–618, 1939.
- [73] Matytsin, A.: *Perturbative analysis of volatility smiles*, Columbia Practitioners Conference on the Mathematics of Finance, New York, 2000.
- [74] Nachman, D.: *Spanning and completeness with options*, Review of Financial Studies, 3, 31, pp. 311–328, 1988.
- [75] Neuberger, A.: *The log contract*, Journal of Portfolio Management 20, 1994.
- [76] Nicolay, D.: *Asymptotic Chaos Expansions in Finance: Theory and Practice*, Springer Finance Lecture Notes, 2014.
- [77] Papanicolaou, A.: *Extreme-strike comparisons and structural bounds for SPX and VIX Options*, available at SSRN: <http://ssrn.com/abstract=2532020>.
- [78] Piterbarg, V.: *Time to smile*, Risk Magazine, May, pp. 71–75, 2005.

- [79] Romano, M., Touzi, N.: *Contingent claims and market completeness in a stochastic volatility model*, Mathematical Finance, 7(4), pp. 399–412, 1997.
- [80] Schönbucher, P. J.: *A market model for stochastic implied volatility*, Phil. Trans. Roy. Soc., Ser. A 357, pp. 2071–2092, 1999.
- [81] Schweizer, M., Wissel, J.: *Term structures of implied volatilities: absence of arbitrage and existence results*, Mathematical Finance, 18(1), pp. 77–114, 2008.
- [82] Sepp, A.: *Efficient Numerical PDE Methods to Solve Calibration and Pricing Problems in Local Stochastic Volatility Models*, available at <http://kodu.ut.ee/~spartak/>.
- [83] Strikwerda, J.: *Finite Difference Schemes and Partial Differential Equations*, Society for Industrial and Applied Mathematics, 2nd edition, 2007.
- [84] Whalley, A. E., Wilmott P.: *An asymptotic analysis of an optimal hedging model for option pricing with transaction costs*. Mathematical Finance 7(3), pp. 307–324, 1997.
- [85] Willard, G.A.: *Calculating prices and sensitivities for path-independent derivative securities in multifactor models*. Journal of Derivatives, 5(1), pp. 45–61, 1997.
- [86] Wissel, J.: *Arbitrage-free market models for option prices*, Working Paper Series, Working Paper No. 248, FINRISK, 2007.

# Stochastic Volatility Modeling

Packed with insights, Lorenzo Bergomi's **Stochastic Volatility Modeling** explains how stochastic volatility is used to address issues arising in the modeling of derivatives, including:

- *Which trading issues do we tackle with stochastic volatility?*
- *How do we design models and assess their relevance?*
- *How do we tell which models are usable and when does calibration make sense?*

This manual covers the practicalities of modeling local volatility, stochastic volatility, local-stochastic volatility, and multi-asset stochastic volatility. In the course of this exploration, the author, *Risk's* 2009 Quant of the Year and a leading contributor to volatility modeling, draws on his experience as head quant in Société Générale's equity derivatives division. Clear and straightforward, the book takes readers through various modeling challenges, all originating in actual trading/hedging issues, with a focus on the practical consequences of modeling choices.

## Features

- Covers forward-start options, variance swaps, options on realized variance, timer options, VIX futures and options, and daily cliques
- Includes an in-depth study of the dynamics of the local volatility model, its carry P&L, and its delta
- Surveys the uncertain volatility model and its usage
- Discusses the parametrization of local-stochastic volatility and multi-asset stochastic volatility models
- Characterizes the links between static and dynamics features of stochastic volatility models
- Contains a wealth of unpublished results and insights



**CRC Press**  
Taylor & Francis Group  
an **informa** business  
[www.crcpress.com](http://www.crcpress.com)

6000 Broken Sound Parkway, NW  
Suite 300, Boca Raton, FL 33487  
711 Third Avenue  
New York, NY 10017  
2 Park Square, Milton Park  
Abingdon, Oxon OX14 4RN, UK

K23465  
ISBN: 978-1-4822-4406-9  
9 0000  
  
 9 781482 244069  
[WWW.CRCPRESS.COM](http://WWW.CRCPRESS.COM)