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 ξ^T.

The ξ^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 ξ is the variance curve.

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

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

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 ξ^u is a function of u. Our total P&L during δt, including now our delta- and vega-hedge, is:

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

We remind the reader of the fact that (forward) VSs – which provide exact delta hedges for the ξ^u – can be entered into at zero cost. The ξ^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 δt and two in δS and $\delta\xi^u$:[1]

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

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 $\nu\left(t,u,u'\right)$ and $\mu\left(t,u\right)$ the instantaneous break-even covariances for, respectively, $\delta\xi^u, \delta\xi^{u'}$ and $\delta\xi^u$, $\frac{\delta S}{S}$:

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

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

Obviously μ is only defined for $u \geq t$ and ν for $u \geq t,\ u' \geq t$.

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

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

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

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

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

---

[1]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: $\sigma(t,S,\xi)^2 \;=\; (\widehat{\sigma}_t^t)^2 \;=\; \xi^t$.

Identifying the δt terms in (7.1) and (7.3) supplies us with the pricing equation in our model:

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

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 ξ_t^T if the payoff involves observations of VS volatilities.[2]

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

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

with the following SDEs for S_t and ξ_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 λ_t^u and correlations between W_t^S and W_t^u such that:

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

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

7.2 A Markov representation

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

[2] Consider for example VS swaptions.

In addition to this technical condition, we also require that the dynamics of forward variances be financially motivated.[3]

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

If there existed a market of options on ξ^T with maturities ranging from t to T, the volatility risk of ξ^T could be hedged away and the volatility of ξ^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 cliquets. We have already considered cliquets 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 ξ^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 ξ^T only depends on $T-t$, with a dependence that is adjustable so that, for example, volatilities of spot-starting VS volatilities $\widehat{\sigma}_T$ in the model can be made to match their historical counterparts. Let us then write:

$$d\xi_t^T = \omega\left(T-t\right)\xi_t^T dW_t^T \tag{7.7}$$

The solution of this SDE is

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

Imagine the same Brownian motion W_t drives the dynamics of all ξ^T. Equation (7.8) expresses $\ln\xi^T$ as a weighted average of increments of W_t, with a weight $\omega\left(T-t\right)$ that depends on T, hence is specific to forward variance ξ^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 ξ^T have to be simulated individually.

However, if ω is of the form:

$$\omega\left(u\right) \ = \ \omega e^{-ku} \tag{7.9}$$

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

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.[4]

[3] 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 ξ^T are indeed akin to a futures curve, as they are driftless.

[4] In [20] Hans Buehler studies Markov representations of the variance curve of the type: $\xi_t^T = G\left(\mathbf{X}_t, T-t\right)$ 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-Ühlenbeck (OU) process X_t:

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

X_t and its variance are given by:

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

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

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

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

7.3 N-factor models

Let us use N Brownian motions and write the SDE of ξ_t^T as:

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

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

$$\omega \;=\; 2\nu \tag{7.12a}$$

$$\alpha_w \;=\; \frac{1}{\sqrt{\sum_{ij} w_i w_j \rho_{ij}}} \tag{7.12b}$$

The solution of (7.11) is given by:

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

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

$$dX_t^i \;=\; -k_i X_t^i dt + dW_t^i,\ X_{t=0}^i \;=\; 0 \tag{7.14}$$

---

that ensure that ξ_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-Ühlenbeck 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:

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

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

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

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

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

where ρ_{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.[5]

[5] 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 δW^S is a Gaussian random variable of variance δt. The covariance of δW^S and δX^i can be computed at once using their expressions:

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

We get:

$$E\left[\delta W^S \delta X^i\right] \;=\; \rho_{iS}\frac{1 - e^{-k_i \delta\tau}}{k_i} \tag{7.18}$$

where ρ_{iS} is the correlation between W^i and W^S.

Using expressions (7.17) and (7.18) for the various covariances, Gaussian random variables δW^S and δX^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 ξ_t^T is, from SDE (7.11):

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

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

$$\rho_t(\xi^T, \xi^{T'}) = \frac{\Sigma_{ij} w_i w_j \rho_{ij} e^{-\left(k_i(T-t)+k_j(T'-t)\right)}}{\sqrt{\Sigma_{ij} w_i w_j \rho_{ij} e^{-(k_i+k_j)(T-t)}}\sqrt{\Sigma_{ij} w_i w_j \rho_{ij} e^{-(k_i+k_j)(T'-t)}}} \tag{7.20}$$

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

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

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

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

The instantaneous volatility of a very short VS volatility is ν:

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

As is clear from (7.22) ν is a global scale factor for volatilities of volatilities.

What about volatilities of forward VS volatilities? Consider two dates T_1, T_2 with $t \le T_1 \le T_2$ and define the foward VS volatility $\widehat{\sigma}_{T_1T_2}$ as:

$$\widehat{\sigma}_{T_1T_2}(t) = \sqrt{\frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \xi_t^\tau d\tau}$$

The instantaneous volatility $\nu_{T_1T_2}(t)$ of $\widehat{\sigma}_{T_1T_2}$ is given by:

$$\begin{cases} \nu_{T_1T_2}(t) = \nu\alpha_w \sqrt{\sum_{ij} w_i w_j \rho_{ij}\, f_i(t,T_1,T_2)\, f_j(t,T_1,T_2)} \\ f_i(t,T_1,T_2) = \dfrac{\int_{T_1}^{T_2} \xi_t^\tau e^{-k_i(\tau-t)} d\tau}{\int_{T_1}^{T_2} \xi_t^\tau d\tau} \end{cases} \tag{7.23}$$

Flat term structure of VS volatilities

In the case of a flat term structure of VS volatilities, ξ_t^τ does not depend on τ. 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 $\widehat{\sigma}_T$ at time t:

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

where

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

Likewise, the instantaneous volatility $\nu_{T_1T_2}(t)$ of the *forward* VS volatility $\widehat{\sigma}_{T_1T_2}$ is given by:

$$\nu_{T_1T_2}(t) = \nu\alpha_w \sqrt{\sum_{ij} w_i w_j \rho_{ij} I\big(k_i(T_2-T_1)\big) I\big(k_j(T_2-T_1)\big) e^{-(k_i+k_j)(T_1-t)}} \tag{7.26}$$

As is clear from (7.24), whenever the VS term structure is flat at time t, $\nu_T(t)$ and $\nu_{T_1T_2}(t)$ 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 δt the usual gamma/theta $P\&L$ with break-even levels specified by the covariance functions μ and ν in the pricing equation (7.4).

From SDE (7.11) for ξ_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:

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

$$\nu(t,u,u',\xi) \;=\; \omega^2\alpha_w^2\xi_t^u\xi_t^{u'}\sum_{ij}\rho_{ij}w_iw_j e^{-k_i(u-t)}e^{-k_j\left(u'-t\right)} \tag{7.27b}$$

where ρ_{ij} is the correlation of W^i and W^j and ρ_{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 $\widehat{\sigma}_T$ in the case of a flat term structure of VS volatilities at time t is:

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

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 ρ_{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 α_θ the normalization constant previously noted α_w in (7.12) – recall that the instantaneous lognormal volatility of the instantaneous *variance* ξ_t^t is equal to 2ν:

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

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

We introduce processes x_t^T defined by:

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

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:

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

whose quadratic variation is given by:

$$\left\langle (dx_t^T)^2 \right\rangle = \eta^2 (T-t)\, dt \tag{7.31a}$$

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

By definition of α_θ, $\eta(0) = 1$.

SDE (7.28) now simply reads:

$$d\xi_t^T \;=\; (2\nu)\xi_t^T \, dx_t^T \tag{7.32}$$

Its solution is:

$$\xi_t^T \;=\; \xi_0^T f^T\left(t, x_t^T\right) \tag{7.33}$$

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

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

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

(7.33) expresses the property that ξ_t^T has a Markov representation as a function of x_t^T – a Gaussian process. We have a Markov-functional model for ξ_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):

$$\frac{d\widehat{\sigma}_T}{\widehat{\sigma}_T} = \nu\alpha_\theta \left((1-\theta) \frac{\int_t^T \xi_t^\tau e^{-k_1(\tau-t)} d\tau}{\int_t^T \xi_t^\tau d\tau} dW_t^1 + \theta \frac{\int_t^T \xi_t^\tau e^{-k_2(\tau-t)} d\tau}{\int_t^T \xi_t^\tau d\tau} dW_t^2 \right) + \bullet\, dt \tag{7.36}$$

$$= \nu\alpha_\theta \Big((1-\theta) A_1 dW_t^1 + \theta A_2 dW_t^2 \Big) + \bullet\, dt \tag{7.37}$$

with A_i given by:

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

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

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

For a flat term-structure of VS volatilities:

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

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 $\nu_T(t)$:

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

where τ_0 is a reference maturity and σ_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 σ_0 derived from prices of options on realized variance are about twice as large – hence our choice for σ_0.

Figure 7.1 shows $\nu_T^{\mathrm{B}}(t)$, 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 ν, θ, k_1, k_2 so as to best match our benchmark (7.40) for maturities from one month to 5 years.

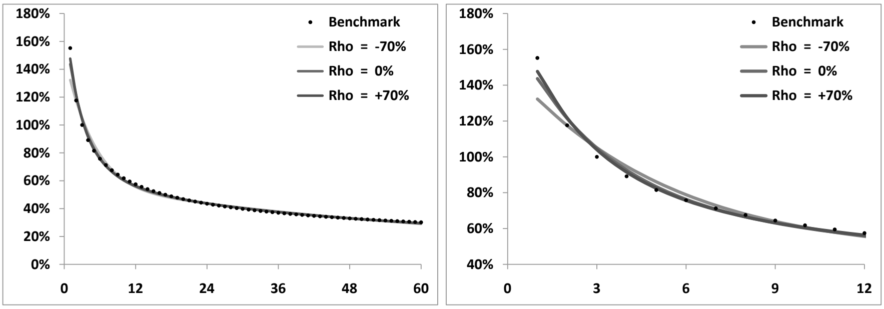


Figure 7.1: The left-hand graph displays the term structure of instantaneous volatilities at $t = 0$ of VS volatilities $\nu_T(t)$ (y axis) as a function of T (x axis, in months) generated by the benchmark form (7.40) as well as the two-factor model, with the different sets of parameters listed in Table 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 α. Moreover, for a given α, many different sets of parameters exist that provide an equally acceptable fit to our benchmark $\nu_T^{\mathrm{B}}(t)$. Table 7.1 displays the parameters used.

	ν	θ	k_1	k_2	ρ_{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 $\nu_T^{\mathrm{B}}(t)$ 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. ν 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 parameters sets corresponding to the same value of α? VS variances are equally weighted baskets of forward variances: $\widehat{\sigma}_T^2 = \frac{1}{T}\int_0^T \xi^\tau d\tau$. 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.

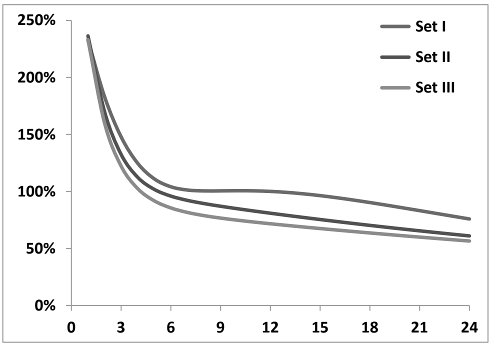


Figure 7.2: Instantaneous volatilities of forward variances ξ_t^T at $t = 0$ as a function of T (in months), using parameter sets in Table 7.1.

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 –

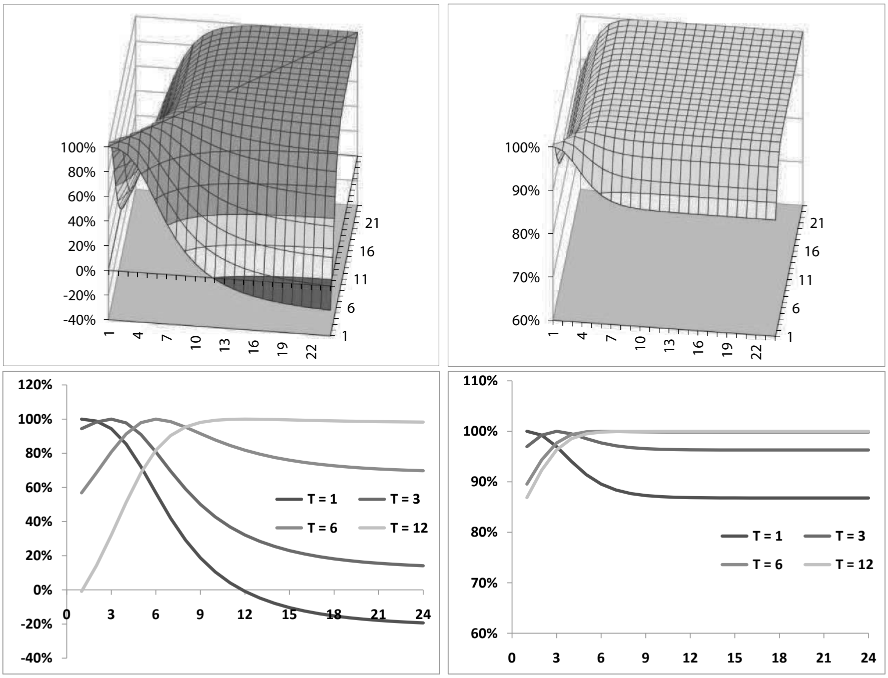


Figure 7.3: Top: $\rho(\xi_t^T, \xi_t^{T'})$ at $t = 0$ as a function of T, T' (in months) in Set I (left) and Set III (right). Bottom: slices of $\rho(\xi_t^T, \xi_t^{T'})$ for $T' = 1$ month, 3 months, 6 months, 12 months, in Set I (left) and Set III (right).

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 ξ_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 ξ^T as lognormal processes, based on historical evidence that VS volatilities are lognormal rather than normal – see Figure 6.2. VS *variances* $\widehat{\sigma}_T^2$, which are baskets of the ξ^T, will not be exactly lognormal and neither will VS *volatilities* $\widehat{\sigma}_T$. 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 $\left(\sigma_r^2 - K\right)$, where σ_r is the realized volatility over $[T_1, T_2]$.

The option is exercised at T_1 only if the forward VS volatility $\widehat{\sigma}_{T_1T_2}$ observed at T_1 is larger than $\sqrt{K}$: we exercise the swaption and sell a VS struck at the market implied VS volatility $\widehat{\sigma}_{T_1T_2}(T_1)$. The payout of this strategy at T_2 is:[6]

$$\left(\sigma_r^2 - K\right) - \left(\sigma_r^2 - \widehat{\sigma}_{T_1T_2}^2(T_1)\right) \;=\; \widehat{\sigma}_{T_1T_2}^2(T_1) - K$$

The underlying of the VS swaption is thus the *forward* VS volatility $\widehat{\sigma}_{T_1T_2}$ and the VS swaption is a call option of maturity T_1 on its square.[7] Expressing $\widehat{\sigma}_{T_1T_2}^2(T_1)$ 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 $\widehat{\sigma}_{T_1T_2}^2(T_1)$ have been computed by simply inverting the Black-Scholes formula as $\widehat{\sigma}_{T_1T_2}^2$ is driftless. In the two-factor model $\widehat{\sigma}_{T_1T_2}(T_1)$ is a function of two Gaussian variables $X_{T_1}^1, X_{T_1}^2$: variance swaptions are simply priced by two-dimensional quadrature.[8]

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 $\widehat{\sigma}_{T_1T_2}$ at time t, let us make the approximation that the VS term structure at time t is flat. The instantaneous volatility of $\widehat{\sigma}_{T_1T_2}$ at t is then given by $\nu_{T_1T_2}(t)$ in (7.26). The instantaneous volatility of $\widehat{\sigma}_{T_1T_2}^2(t)$ is twice as large. We then get the following strike-independent approximation of the implied volatility $2\widehat{\nu}_{T_1T_2}(T_1)$ by integrating the square of $\nu_{T_1T_2}(t)$ in

[6] See the footnote on page 151 for the normalization of VS payoffs in actual VS term sheets.

[7] Note the similarity with cliquets – see Section 1.3.2.

[8] 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.

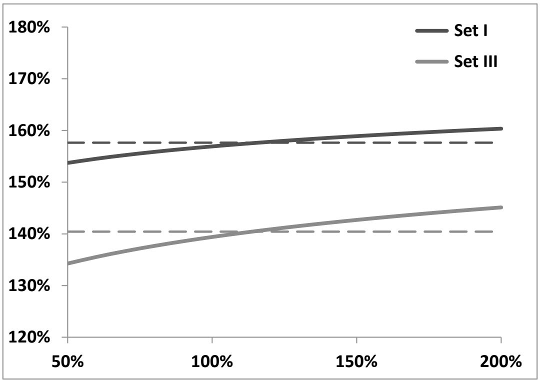


Figure 7.4: Implied volatilities of variance swaptions – that is of $\widehat{\sigma}^2_{T_1 T_2}(T_1)$ – as a function of *volatility* moneyness: $\sqrt{K}/\widehat{\sigma}_{T_1 T_2}(t=0)$ with $T_1 = 3$ months, $T_2 = 6$ months in sets I and III. Dotted lines correspond to the strike-independent level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

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

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

$2\widehat{\nu}_{T_1 T_2}(T_1)$ 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 ξ^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 $\widehat{\sigma}^2_T$, thus increasing ν_T – see expression (7.22) for $\nu_T(t)$.

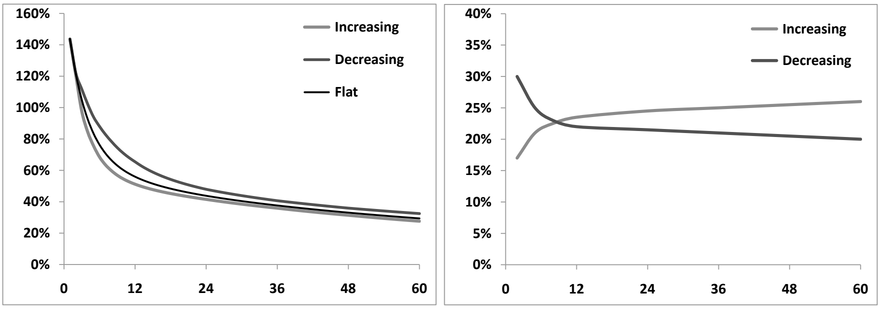


Figure 7.5: Left: instantaneous volatilities at $t = 0$ of VS volatilities ν_T as a function of T (months) computed in Set II for three different term structures of VS volatilities. Right: increasing and decreasing term structures of VS volatilities used in left-hand graph.

Note that simply multiplying all $\widehat{\sigma}_T$ by the same constant leaves volatilities of volatilities unchanged, as the ξ^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 $\nu_T^{\mathrm{B}}(t)$ in (7.40) can be matched over a wide range of maturities. Furthermore, very similar term structures of volatilities of spot-starting VS volatilities $\widehat{\sigma}_T$ 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 α 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 ξ_0^t so that the term structure of ATMF or ATM volatilities – or implied volatilities for an arbitrary moneyness – is recovered.[9] 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.[10]

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

[9]Efficient techniques for generating vanilla smiles are surveyed in Appendix A of Chapter 8.

[10]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 inputing 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.[11]

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:

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

where $\widehat{\boldsymbol{\sigma}}$ is the (volatility) strike, $\sigma_r\left(T\right)$ 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, $\widehat{\boldsymbol{\sigma}} = \widehat{\sigma}_T$ and at order one in $\sigma_r\left(T\right) - \widehat{\sigma}_T$, the payoff of the ATM call on variance is simply $\left(\sigma_r\left(T\right) - \widehat{\sigma}_T\right)^+$.

Volatility swaps, whose payoff is:

$$\sigma_r\left(T\right) - \widehat{\boldsymbol{\sigma}}$$

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

[11]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.[12]

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 $(\sigma_r^2 - \widehat{\boldsymbol{\sigma}}^2)^+$ – 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 $[0, t]$:

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

where $\sigma_r\left(t\right)$ is the realized volatility over $[0, t]$. The option's payoff is $\left(\frac{1}{T}Q_T - \widehat{\boldsymbol{\sigma}}^2\right)^+$. 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 $\widehat{\sigma}_T\left(t\right)$, whose payoff at T is:

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

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 $(T-t)\widehat{\sigma}_T^2\left(t\right)$, 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 $(T-t')\widehat{\sigma}_T^2\left(t'\right)$. The P&L over $[t, t']$ of our VS position is thus:

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

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:

$$U_t \ = \ \frac{Q_t + \left(T-t\right)\widehat{\sigma}_T^2\left(t\right)}{T} \tag{7.43}$$

[12] 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 $Q_t + (T-t)\,\widehat{\sigma}_T^2(t)$ 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:

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

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

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

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

$$\big(U_T - \widehat{\boldsymbol{\sigma}}^2\big)^+$$

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

A dynamics for U_t

From (7.43) we have:

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

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

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

which we rewrite as:

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

with R_t given by:

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

SDE (7.46) is problematic as the dynamics of U_t is not autonomous – it involves $\widehat{\sigma}_T^2(t)$ 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:

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

R_τ is the ratio of two integrated variances, thus is a number of order one.

We now make an approximation: let us replace R_τ with a constant – we approximate R_τ 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 σ_{eff}:

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

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

The number of VS contracts of maturity T needed at time t to hedge our option is:[13]

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

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 $U_{t=0} \ = \ \widehat{\sigma}_T^2(0)$, so that $R_{t=0} \ = \ 1$. The option premium is then given by:[14]

---

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

[14] 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 σ 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 $\sigma_{\text{eff}} = \frac{\sigma}{\sqrt{3}}$.

$$P(t=0) = P_{\mathrm{BS}}\left(t=0, \widehat{\sigma}_T^2(0), \sigma_{\mathrm{eff}}, T\right) \tag{7.50a}$$

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

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 $\widehat{\sigma}_T(t)$, 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 σ_{eff}. We have replaced R_τ 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 σ_{eff} is the instantaneous volatility of the VS volatility for maturity T observed at t, $\widehat{\sigma}_T(t)$. In our benchmark (7.40), $\nu_T(\tau)$ is specified directly:

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

In the two-factor model, instead, the model generates its own dynamics for $\widehat{\sigma}_T(t)$. 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 σ_{eff}?

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

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

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 $\widehat{\sigma}_T(t)$ 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 ξ_t^τ are equal to their initial values ξ_0^τ. 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 $\widehat{\sigma}_T(t)$ 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 σ_{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:

$$\frac{1}{2\widehat{\sigma}_T}\left(\sigma_r^2 - \widehat{\boldsymbol{\sigma}}^2\right)^+$$

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

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

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 $\nu_T(t) \simeq \nu_T^{\mathrm{B}}(t)$. 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 α in our benchmark, 0.4, is close to 0.5. For $\alpha = 0.5$, σ_{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_τ will be different than 1 whenever the realized volatility over $[0, \tau]$, $\sigma_r(\tau)$, is different than $\widehat{\sigma}_T(\tau)$.

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 $\widehat{\sigma}_T(\tau)$ are random, the expectation of their squares is equal to $20\%^2$, thus the expectations of the numerator and denominator of R_τ 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_τ 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_τ 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: $Q_t/t < \widehat{\sigma}_T^2(t)$. 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 τ of $\widehat{\sigma}_T$ matches our pricing level $\nu_T(\tau)$. Can we hedge this exposure of R_τ 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_τ 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_τ in (7.47) are:

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

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

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

Formula (7.49b) for σ_{eff} now becomes:

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

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

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

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

In what follows we use this new expression for σ_{eff} instead of (7.50b).

σ_{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 $\widehat{\sigma}_\tau^2$. Writing $\widehat{\sigma}_{\tau T}^2(t)$ as:

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

(7.55) can be rewritten as:

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

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

$$\lambda(\tau) = \left.\frac{\delta P}{\delta\widehat{\sigma}_\tau^2}\right|_t$$
$$= -8\frac{dP}{d\sigma_{\text{eff}}^2}\,\frac{(\tau-t)(T-\tau)}{(T-t)T}\,\frac{T\widehat{\sigma}_{\tau T}^2(t)}{\left(Q_t+(T-t)\,\widehat{\sigma}_T^2(t)\right)^2}\,\nu_T^2(\tau) \tag{7.58}$$

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

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

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_τ – hence of σ_{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 σ_{eff}^2 as:

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

Recall that a position in a VS of maturity T replicates the package $Q_t + (T-t)\,\widehat{\sigma}_T^2(t)$ exactly – see equation (7.42). As is manifest in equation (7.59) $\widehat{\sigma}_\tau^2(t)$ does not appear alone, but associated with Q_t in such a way that the VS position that hedges the exposure to $\widehat{\sigma}_\tau^2(t)$ actually hedges $Q_t + (\tau-t)\,\widehat{\sigma}_\tau^2(t)$ – the same goes for the exposure to $\widehat{\sigma}_T^2(t)$. In other words, the extra VS position that hedges the sensitivity of σ_{eff}^2 to the VS term structure also hedges the sensitivity of R_τ 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 σ_{eff}^2 to the term structure of VS volatilities exactly offsets the gamma/theta P&L generated by the dependence of σ_{eff}^2 on the quadratic variation Q_t and VS volatilities $\widehat{\sigma}_\tau$ 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 μ_T of hedging VSs are related by a simple equation. Expression (7.56a) shows that σ_{eff} is unchanged if all VS volatilities are rescaled uniformly – thus for a small change $\delta\widehat{\sigma}_T^2 = \varepsilon\widehat{\sigma}_T^2$, at order one in ε, $\delta\sigma_{\text{eff}}^2 = 0$:

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

which implies that

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

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

$$\int_0^T \lambda(\tau)\,d\tau + \mu_T \;=\; 0 \tag{7.60}$$

The aggregate vega of the VS position that hedges σ_{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 σ_{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 τ of a VS of maturity $\tau \in [0, T]$ are, respectively:

$$\frac{1}{2\widehat{\sigma}_T}\left(\frac{Q_T}{T} - \widehat{\boldsymbol{\sigma}}^2\right)^+, \qquad \frac{1}{2\widehat{\sigma}_\tau}\left(\frac{Q_\tau}{\tau} - \widehat{\sigma}_\tau^2\right) \tag{7.61}$$

The vega of a VS contract is thus 1.

With the benchmark

We first use the SM with our benchmark form for $\nu_T^{\text{B}}(t)$. We use a flat term structure of VS volatilities: while the *price* computed using either expression (7.56a) or expression (7.50b) for σ_{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 τ_i in Table 7.3 we have used a simple piecewise affine interpolation of $\tau\widehat{\sigma}_\tau^2$ over each interval $[\tau_i,\ \tau_{i+1}]$. This ensures that as $\widehat{\sigma}_{\tau_i}$ is shifted, only forward variances ξ_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.[15]

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 $\frac{1}{2\widehat{\sigma}_\tau}$ 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 τ is given by:

$$S^2 \frac{d^2\text{VS}}{dS^2} \;=\; \frac{1}{\widehat{\sigma}_\tau \tau} \tag{7.62}$$

[15] 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%
Aggregate	65.1%	182%

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).[16]

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 σ_{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 σ_{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 σ_{eff} vanishes.

[16] 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[17] – 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 δS: the quadratic variation is shifted from Q to $Q + \left(\frac{\delta S}{S}\right)^2$. 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 δQ small where $Q_0 = 0$ is the quadratic variation at $t = 0$, and δQ 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.[18] Using convention (7.61) for the variance option payoff and denoting by $\widehat{\boldsymbol{\sigma}}$ its (volatility) strike, at first order in δQ:

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

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

$$S^2 \frac{d^2 P}{dS^2} \; = \; -\frac{1}{\widehat{\boldsymbol{\sigma}} T} \frac{dP}{d\widehat{\boldsymbol{\sigma}}} \tag{7.63}$$

[17] 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.

[18] 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 δQ.

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_τ 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.

	VS hedge		VS dollar gamma	
Maturities	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 σ^2_{eff} makes use of $\nu^{\text{B}}_T(t)$, while in the latter case $\nu^0_T(t)$ in (7.52) is used.

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

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

This sensitivity of the instantaneous volatility of $\widehat{\sigma}_T(t)$ 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 $\nu^0_T(t)$ 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, $\nu^0_T(t)$ is unchanged, which implies that the sum of the additional VS hedges contributed by the sensitivity of $\nu^0_T(t)$ 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.[19] 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

[19] One only needs to: (a) make it a one-factor model ($\theta = 0$) and set $k_1 = 0$, (b) make ν in (7.28) time-dependent: $\nu \to \nu(t) = \sigma_0 \left(\frac{\tau_0}{T-t}\right)^\alpha$. Instantaneous volatilities of $\widehat{\sigma}_T(t)$ are then exactly equal to $\nu^{\text{B}}_T(t)$ 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 σ_0, τ_0, α, 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 ν, θ, k_1, k_2 of the two-factor model so that, for each term structure at hand, $\nu_T^0(t) \simeq \nu_T^{\mathrm{B}}(t)$, where $\nu_T^0(t)$ is given by (7.52). For $\nu_T^{\mathrm{B}}(t)$, 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	ν	θ	k_1	k_2	ρ_{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 $\widehat{\sigma}_T$ is 25% whereas in the 6-month case $\widehat{\sigma}_T = 31\%$ (resp. $\widehat{\sigma}_T = 19\%$) 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 $\widehat{\sigma}_T = 25\%$, 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 $\widehat{\sigma}_T\,(t)$, which is the only source of randomness.

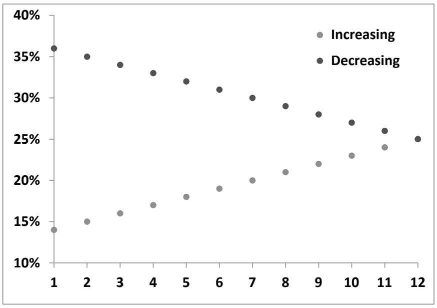


Figure 7.6: Term structures of VS volatilities at $t = 0$ used for generating prices in Table 7.5. Maturities are expressed in months.

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:

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

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:

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

$$\simeq \sigma\sqrt{\Delta t}Z \tag{7.64b}$$

where Z is a gaussian random variable with unit variance and Δt 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 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 Δt reads:

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

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

Let us assume that the volatility of $\widehat{\sigma}_T^2$ 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 Δt as:

$$r = \sigma_t \sqrt{\Delta t} Z \tag{7.66}$$

where Z is a (possibly non-gaussian) random variable with unit variance, and σ_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: $\sigma_t = \widehat{\sigma}_T(t)$. The second piece in (7.65) then reads:

$$\frac{\widehat{\sigma}_T^2\,(t)\,\Delta t}{T}\left(Z^2 - 1\right)$$

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:

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

where κ 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 $E[U_t] = \widehat{\sigma}_T^2$ and let us replace U_t with its expectation. The prefactor in (7.67) is then simply equal to 1. This yields:

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

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

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

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 σ_{eff}^2 which supersedes (7.56a):

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

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 σ_{eff}^2 is largest for short maturities.

Which value should we pick for κ? κ 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 σ_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 κ is attributable to fluctuations of σ_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 σ_{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 σ_{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.[20]

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 σ_{eff} given by (7.70) are in good agreement with prices computed in a Monte Carlo simulation of the full-blown forward variance model –

[20]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 ν 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.

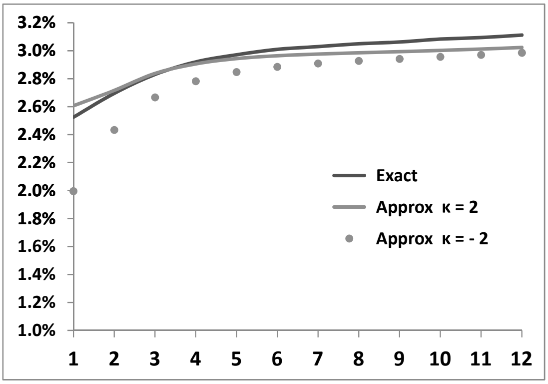


Figure 7.7: Price of an ATM call on realized variance as a function of maturity, from 1 month to 1 year, computed either in a Monte Carlo simulation of the forward variance model (exact), or in the approximate model with σ_{eff} given by (7.70) with $\kappa = 2$ and $\kappa = -2$.

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 $\overline{\mathcal{P}}(F)$ is obtained as the highest price generated by processes for the instantaneous volatility σ_t such that vanilla option prices are recovered:

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

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 $\underline{\mathcal{P}}(F)$, $\overline{\mathcal{P}}(F)$ 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 ($\sigma = \sigma_{\min} = 0$).[21]

[21]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[22] and a heuristic upper bound, proposed by Peter Carr and Roger Lee and published in [28].[23]

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 $\mathcal{Q} = \widehat{\boldsymbol{\sigma}}^2 T$, where $\widehat{\boldsymbol{\sigma}}$ is the (volatility) strike of the option on realized variance.[24]. The option's value is given by: $\mathcal{P}^f_{BS}(S_t, Q_t; \mathcal{Q})$, where $\mathcal{P}^f_{BS}$ is given by expression (5.68), page 183, and Q_t is the realized quadratic variation since $t = 0$.

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

$$P\&L^{Q_T<\mathcal{Q}}_{\text{Final}} \;=\; \mathcal{P}^f_{BS}(S_T, Q_T; \mathcal{Q}) - \mathcal{P}^f_{BS}(S_T, Q_T; Q_T) \tag{7.71}$$

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:

$$P\&L^{Q_T>\mathcal{Q}}_{\text{Final}} \;=\; -\frac{1}{2}\int_\tau^T S_t^2 \left.\frac{d^2 f}{dS^2}\right|_{S_t} \overline{\sigma}_t^2 dt \tag{7.72}$$

where $\overline{\sigma}_t$ 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, Γ: $\frac{S^2}{2}\frac{d^2 f}{dS^2} = \theta(S)\Gamma$ and $\Gamma \geq 0$, with $\theta(S) = 0$ or 1. Then:

$$\begin{aligned} P\&L^{Q_T>\mathcal{Q}}_{\text{Final}} \;&\geq\; -\Gamma \int_\tau^T \overline{\sigma}_t^2 dt &(7.73)\\ &\geq\; -\Gamma(Q_T - Q_\tau) \;=\; -\Gamma(Q_T - \mathcal{Q}) &(7.74)\end{aligned}$$

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

with $\sigma_{\min} = 0$, $\sigma_{\max} = +\infty$. In this equation P is the price of payoff $F - \Sigma\lambda_i O_i$ 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 $\frac{S^2}{2}\frac{d^2P}{dS^2} + \frac{dP}{dQ} = 0$. These are exactly conditions (5.70), page 184, that characterize timer options. I thank Pierre Henry-Labordère for pointing this out to me.

[22]Presented at the 2005 Global Derivatives conference.

[23]In what follows all $P\&Ls$ are computed at order one in δt and order two in δS; 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 δS to our daily $P\&L$ can be ignored.

[24]See Section 5.9 on timer options.

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

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

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

In case $Q_T > \mathcal{Q}$, this option's payout more than offsets the negative P&L $P\&L^{Q_T>\mathcal{Q}}_{\text{Final}}$: 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 $P\&L^{Q_T<\mathcal{Q}}_{\text{Final}}$ is positive as well.

Thus our short position in the European payoff $f(S)$ combined with a long position in $2\widehat{\sigma}_T T\Gamma$ options on realized variance of strike $\widehat{\boldsymbol{\sigma}}$ 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:

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

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

$$(2\widehat{\sigma}_T T\Gamma)C^{\widehat{\boldsymbol{\sigma}}}_{\text{Mkt}} \;-\; \Big(P^f_{\text{Mkt}} - \mathcal{P}^f_{BS}(S_0, 0; \mathcal{Q})\Big) \;\geq\; 0$$

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

$$C^{\widehat{\boldsymbol{\sigma}}}_{\text{Mkt}} \;\geq\; \frac{1}{2\widehat{\sigma}_T T}\frac{1}{\Gamma}\Big(P^f_{\text{Mkt}} - \mathcal{P}^f_{BS}(S_0, 0; \widehat{\boldsymbol{\sigma}}^2 T)\Big) \tag{7.75}$$

where we have replaced $\mathcal{Q}$ with $\widehat{\boldsymbol{\sigma}}^2 T$. 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 Γ. Which payoff should we pick, so that the lower bound for $C^{\widehat{\boldsymbol{\sigma}}}_{\text{Mkt}}$ 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:

$$\frac{1}{2\widehat{\sigma}_T T}\frac{1}{\Gamma}\int_0^\infty 2\theta(K)\frac{\Gamma}{K^2}\Big(P^K_{\text{Mkt}} - \mathcal{P}^K_{BS}(S_0, 0; \widehat{\boldsymbol{\sigma}}^2 T)\Big)dK \tag{7.76}$$

where P^K_{Mkt} is the market price for a vanilla option of strike K and $\mathcal{P}^K_{BS}$ is its Black-Scholes price. There is no need to distinguish between calls and puts, since by call-put parity $P^K_{\text{Mkt}} - \mathcal{P}^K_{BS}$ 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 $\widehat{\sigma}_K$: $P^K_{\text{Mkt}} = \mathcal{P}^K_{BS}(S_0, 0; \widehat{\sigma}^2_K T)$. (7.76) is thus equal to:

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

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

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

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: $\mathbf{1}_{\widehat{\sigma}_K > \widehat{\boldsymbol{\sigma}}} = 1 \; \forall K$. This implies also that $\widehat{\sigma}_T \geq \widehat{\boldsymbol{\sigma}}$: 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 $\widehat{\boldsymbol{\sigma}}$. This yields:

$$C^{\widehat{\boldsymbol{\sigma}}}_{\text{Mkt}} \;\geq\; \frac{1}{2\widehat{\sigma}_T}\left(\widehat{\sigma}^2_T - \widehat{\boldsymbol{\sigma}}^2\right)$$

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

Likewise, if $\mathbf{1}_{\widehat{\sigma}_K > \widehat{\boldsymbol{\sigma}}} = 0 \; \forall K$, $\widehat{\sigma}_T \leq \widehat{\boldsymbol{\sigma}}$: the call on realized variance is out of the money, and (7.77) again expresses that $C^{\widehat{\boldsymbol{\sigma}}}_{\text{Mkt}}$ 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 $\Sigma = \widehat{\boldsymbol{\sigma}}^2 T$ where $\widehat{\boldsymbol{\sigma}}$ 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 τ is defined as the time when S_τ 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:

$$\frac{S^2}{2}\frac{d^2\mathcal{B}_{LH}}{dS^2} + \frac{d\mathcal{B}_{LH}}{dQ} \; = \; 0 \tag{7.78}$$

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:

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

  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 δS, delta-hedging the profile $-2\ln S$ at zero implied volatility generates the realized quadratic variation:

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

  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 τ, 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:

$$Q_T - Q_\tau \; = \; (f_{LH}(S_T) - f_{LH}(S_\tau)) - \int_\tau^T \left.\frac{df_{LH}}{dS}\right|_{S_t} dS_t \tag{7.80a}$$

$$= \; f_{LH}(S_T) - \int_\tau^T \left.\frac{df_{LH}}{dS}\right|_{S_t} dS_t \tag{7.80b}$$

  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:

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

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: $\frac{1}{2\widehat{\sigma}_T T}(Q_T - \widehat{\boldsymbol{\sigma}}^2 T)$, whose price is $C^{\widehat{\boldsymbol{\sigma}}}_{\text{Mkt}}$, we have:

$$C^{\widehat{\boldsymbol{\sigma}}}_{\text{Mkt}} \ \leq \ \frac{1}{2\widehat{\sigma}_T T}\Big(\mathcal{B}_{LH}(S_0, 0; \widehat{\boldsymbol{\sigma}}^2 T) + \mathcal{G}_{LH}(S_0, T)\Big) \tag{7.81}$$

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.

$\mathcal{B}_{LH}(S_0, 0; \widehat{\boldsymbol{\sigma}}^2 T)$ 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 $\mathcal{B}_{LH}(S_0, 0; \widehat{\boldsymbol{\sigma}}^2 T) = 0$ 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 $-2\ln\left(\frac{S_T}{S_0}\right)$ 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.[25]

[25] 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 δS 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 $\widehat{\sigma}_T(T_1), T \in [T_1, T_2]$ given, for example in the SM, by expressions (7.56), but where we sit at $t = T_1$ rather than at $t = 0$:

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

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 σ_{eff} with 0 in (7.82); the value at T_1 of our option is then simply the call – or put – payoff applied to $\widehat{\sigma}_{T_2}^2 (T_1)$. 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.[26] An option on forward realized variance is thus more expensive than the corresponding VS swaption.[27]

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.

[26] Variance swaption smiles in the two-factor model are shown in Section 7.4.3.

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

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 ξ_t^T 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.[28] 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):

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

where $Q_{\mathrm{mkt},t}^T$ 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

[28]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 ith future, and $T - t = \Delta = 30$ days, and use the notation $\widehat{\sigma}_{\mathrm{VIX},T_i}$ 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:

$$\widehat{\sigma}^2_{\mathrm{VIX},T_i}(T_i) \;=\; \frac{2e^{r\Delta}}{\Delta}\left(\int_0^{F_{T_i}^{T_i+\Delta}} P^{K,T_i+\Delta}_{\mathrm{mkt},T_i}\frac{dK}{K^2} + \int_{F_{T_i}^{T_i+\Delta}}^{\infty} C^{K,T_i+\Delta}_{\mathrm{mkt},T_i}\frac{dK}{K^2}\right) \tag{7.83}$$

where $P^{KT}_{\mathrm{mkt},t}$ (resp. $C^{KT}_{\mathrm{mkt},t}$) 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 ∞ but is cut off for small and large strikes wherever a zero bid price is encountered for two consecutive strikes.[29] 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 $\widehat{\sigma}_{\mathrm{VIX},T_i}$ 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.[30]

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 $\widehat{\sigma}_{\mathrm{VS},T}$. $\widehat{\sigma}_T = \widehat{\sigma}_{\mathrm{VS},T}$ and forward log-contract and VS variances are identical, equal to ξ_t^T.

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

[29] We refer the reader to *www.cboe.com/micro/vix/vixwhite.pdf* for the exact procedure for computing $\widehat{\sigma}_{\mathrm{VIX},T}$.

[30] The VIX itself is a number published in real time: $\widehat{\sigma}_{\mathrm{VIX},t}$ 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.

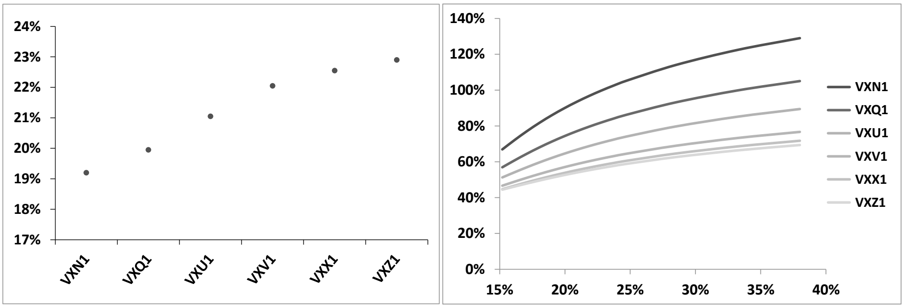


Figure 7.8: Left: VIX futures as of June 14, 2011 for expiries ranging from July (VXN1) to December (VXZ1). Right: Smiles of VIX futures.

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:

$$F_{t=T_i}^i = \widehat{\sigma}_{\mathrm{VIX},T_i}(T_i) \tag{7.84}$$

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

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

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

$$F_t^i = E[F_{T_i}^i] \tag{7.86}$$

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 ξ_t^T are given by (7.33) and (7.34):

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

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

A Markov-functional model

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

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

where $\eta^2(T-t)$ is the instantaneous variance of x_t^T. η 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 ξ_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 ξ^T reads:

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

The instantaneous volatility of ξ^T is thus given by:

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

$f^{T^{-1}}$ 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 ξ_t^T and t, thus what we have is a local volatility model for ξ_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 ξ^T are positively sloping, as for large values of x the linear combination is dominated by terms with large values of ω. 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 γ_T, β_T, ω_T and set:

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

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

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

Let us introduce the dimensionless parameter ζ^T and write ω_T as:

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

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

$$\mathrm{vol}(\xi_t^T)\big|_{t=0} \;=\; 2\nu\zeta_T\eta\,(T)$$

ζ_T is thus simply a scale factor for the instantaneous volatility of ξ_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 ξ_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.[31] Mathematically our model needs to ensure that:

$$\begin{aligned}
E_t\left[\widehat{\sigma}_{\mathrm{VIX},T_i}(T_i)\right] \;&=\; F_t^i \\
E_t\left[\,(\widehat{\sigma}_{\mathrm{VIX},T_i}(T_i)-K)^+\,\right] \;&=\; \mathcal{C}_t^{Ki},\;\; E_t\left[\,(K-\widehat{\sigma}_{\mathrm{VIX},T_i}(T_i))^+\,\right] \;=\; \mathcal{P}_t^{Ki}
\end{aligned}$$

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 $\widehat{\sigma}_{\mathrm{VIX},T_i}$ 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 ξ_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:

$$\widehat{\sigma}^2_{\mathrm{VIX},T_i}\left(t, X_t^1, X_t^2\right) \;=\; \xi_0^i\frac{1}{\Delta}\int_{T_i}^{T_{i+1}} f^\tau(t, x_t^\tau)\,d\tau \tag{7.91}$$

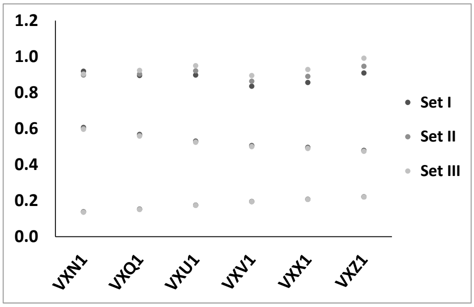
and both the VIX future and prices of options on $\widehat{\sigma}_{\mathrm{VIX},T_i}(T_i)$ are obtained through a two-dimensional Gaussian quadrature on $(X_{T_i}^1,\; X_{T_i}^2)$.

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

[31] 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.[32] Calibration is performed by least-squares minimization with a sufficiently large weight on the future itself that it is exactly calibrated.[33]

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.



VXN1	20-Jul-11
VXQ1	17-Aug-11
VXU1	21-Sep-11
VXV1	19-Oct-11
VXX1	16-Nov-11
VXZ1	21-Dec-11

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: β^i, middle: γ^i, top: ζ^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.[34] 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 ζ is appreciably larger in Set III than in Sets I and II. This is expected. Sets I, II, III generate almost identical term

[32] 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.

[33] The reader may wonder why, rather than calibrating ξ^i along with $\gamma^i, \beta^i, \zeta^i$, we do not determine ξ^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 ξ^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 ξ^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.

[34] 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. γ, β, ζ 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 ζ.

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 σ_t as the square root of the expected instantaneous variance of F_t^i:

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

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). σ_t is thus simply given by:

$$\sigma_t \;=\; \sqrt{E\left[\alpha_{X^1}^2 + \alpha_{X^2}^2 + 2\rho_{12}\alpha_{X^1}\alpha_{X^2}\right]}$$

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 σ_t for the VIX future expiring on December 21, 2011, for dates t ranging from June 16, 2011 to the future's expiry.

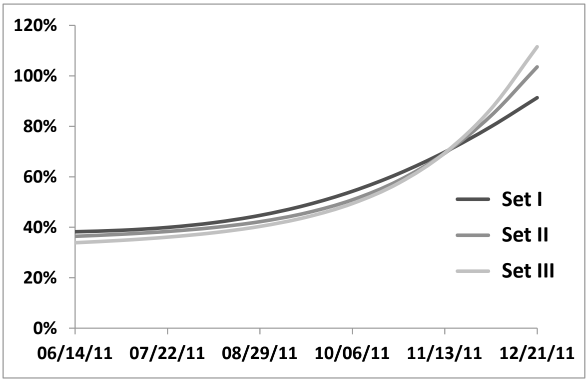


Figure 7.10: σ_t – as defined in (7.92) – for VIX future VXZ1, for dates ranging from June 14 to its expiry date, in Sets I, II, III.

While the integrated value of σ_t^2 is identical in Sets I, II, III, the distribution is different.[35]

[35]By definition, σ_t^2 is the expected instantaneous realized variance of F_t^i. The integral of σ_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 $\left(X_t^1, X_t^2\right)$, however each instantaneous forward variance ξ^T is a *function* of x_t^T. We have a one-factor Markov-functional model for each *instantaneous* forward variance ξ^T – i.e. a local volatility model[36] – 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 $\left(X_t^1, X_t^2\right)$; 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 ξ_t^T are modeled as a function of x_t^T. Forward variances $\widehat{\sigma}^2_{\mathrm{VIX},T_i}(t)$ are explicitly known for all t.

Consider VIX future F^i. At $t = T_i$, $F^i_{T_i} = \sqrt{\widehat{\sigma}^2_{\mathrm{VIX},T_i}(T_i)}$: 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), $F^i_{t=T_i}$ is a function of $X^1_{T_i}$ and $X^2_{T_i}$ given by (7.84). F_t^i is given by:

$$F_t^i \;=\; E_t\!\left[F^i_{T_i}\left(X^1_{T_i}, X^2_{T_i}\right)\right] \tag{7.93}$$

[36] See Section 2.10.

While $\widehat{\sigma}^2_{\text{VIX},T_i}(t)$ is readily available, F^i_t for $t < T_i$ has to be computed by two-dimensional quadrature on $X^1_{T_i}, X^2_{T_i}$.[37]

In case F^i_t 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.[38] Denoting by X_t the value of the ETF:

$$\frac{dX_t}{X_t} = rdt + \sum_{i,\ T_i > t} w^i_t \frac{dF^i_t}{F^i_t} \tag{7.94}$$

where the ETF's allocation strategy is expressed in the weight w^i_t. 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 $w^i_t \equiv w(T_i - t)$ with $w(\tau)$ given by:

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

where Δ 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:

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

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

The VXX prospectus states that it is the *number* of futures that is proportional to τ and $\Delta - \tau$, rather than the *notional* invested in both futures:[39]

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

---

[37] This is achieved efficiently by first finding the linear combination of $X^1_{T_i}$ and $X^2_{T_i}$ that accounts for the bulk of the variance of $F^i_{T_i}$. The resulting quadrature is then almost one-dimensional, especially since Δ is small. The chosen algorithm should ensure that F^i_t is a martingale so that self-financing strategies that invest in VIX futures – such as VIX ETNs – have the correct forward.

[38] 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.

[39] See the prospectus of the VXX ETN at *www.ipathetn.com/static/pdf/vix-prospectus.pdf*

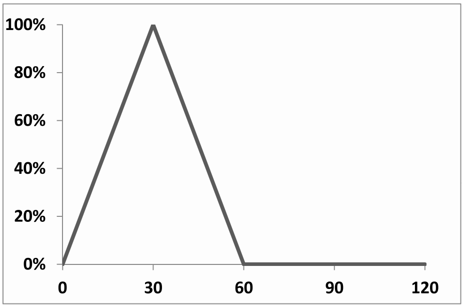


Figure 7.11: $w(\tau)$ for the VXX ETN, as a function of τ (days). We have made the simple assumption that VIX expiries are spaced 30 days apart.

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.[40]

[40]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 ζ_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 γ_i, β_i do not depend on k_1 and the VIX smiles generated by the model do not depend on k_1 either.

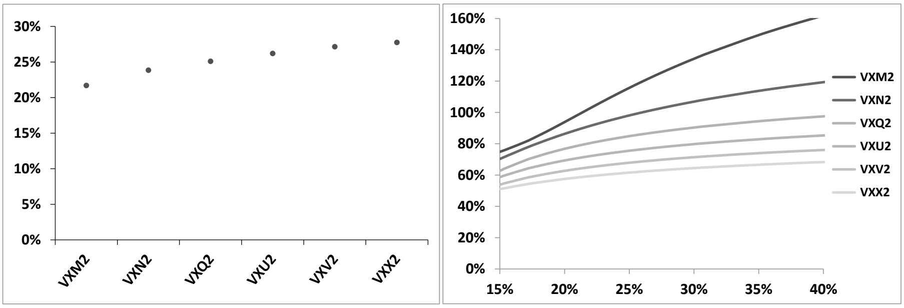


Figure 7.12: Left: VIX futures as of June 8, 2012 for expiries ranging from June 20 (VXM2) to November 21 (VXX2). Right: Smiles of VIX futures.

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.[41] It also shows the instantaneous volatility of VIX future VXV2, which expires on October 17, 2012. This instantaneous volatility, σ_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 $x_t^{T_i}$.

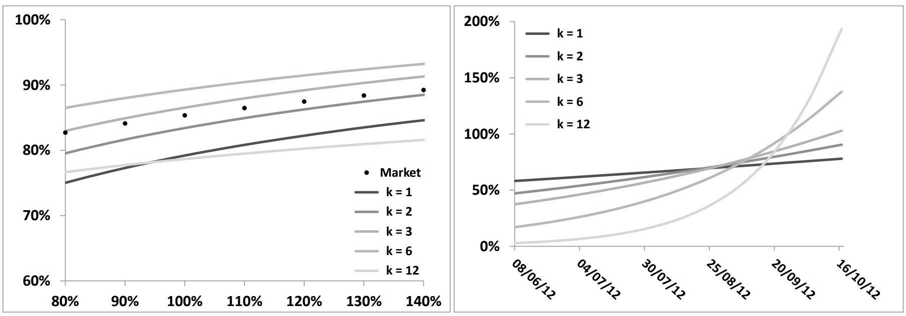


Figure 7.13: Left: VXX smiles as of June 8, 2012 for the December 21st maturity, for different values of k_1, compared to the market smile. Right: instantaneous volatility of the VIX future expiring on October 17, 2012 (VXV2) for different values of k_1.

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-

[41] 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.[42]

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 σ, and that their correlation is constant, equal to ρ. The instantaneous volatility of the VXX, $\sigma_X(t)$, is given for $t \in [T_{i-1}, T_i]$ by:

$$\sigma_X^2(t) \;=\; \sigma^2 \left(w^2(T_i - t) + w^2(T_{i+1} - t) + 2\rho w(T_i - t) w(T_{i+1} - t) \right)$$

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 $[T_{i-1}, T_i]$:

$$\widehat{\sigma}_X^2 \;=\; \frac{1}{\Delta} \int_{T_{i-1}}^{T_i} \sigma_X^2(t) dt \;=\; \frac{2+\rho}{3} \sigma^2$$

Consider the two cases $\rho_{\min} = 80\%$, $\rho_{\max} = 100\%$. The ratio of implied volatilities $\frac{\widehat{\sigma}_X^{\max}}{\widehat{\sigma}_X^{\min}}$ 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

[42] For $k_1 = 3$ the volatility-of-volatility adjustment factors ζ_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.[43]

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:

$$\widehat{\sigma}^2_{T_i,T_{i+\Delta}}(t) \;=\; E_t\left[\,\left(F^i_{T_i}\right)^2\,\right] \tag{7.96}$$

We know from Section 3.1.3 that any European payoff on $F^i_{T_i}$ can be replicated by a static position consisting of cash, forwards (or futures) and vanilla options on $F^i_{T_i}$. The decomposition in (3.6) applied to the function $f(S) = S^2$ reads:

$$S^2 \;=\; S_*^2 \;+\; 2S_*\left(S - S_*\right) + 2\int_0^{S_*} (K-S)^+\, dK + 2\int_{S_*}^{\infty} (S-K)^+\, dK \tag{7.97}$$

where S_* is arbitrary. We now apply this identity to $S = F^i_{T_i}$ with $S_* = F^i_t$ 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:

$$\widehat{\sigma}^2_{T_i,T_{i+\Delta}}(t) \;=\; \left(F^i_t\right)^2 + 2\int_0^{F^i_t} \mathcal{P}^{Ki}_t dK + 2\int_{F^i_t}^{\infty} \mathcal{C}^{Ki}_t dK \tag{7.98}$$

where $\mathcal{P}^{Ki}_t$ (resp. $\mathcal{C}^{Ki}_t$) 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 $\left(F^i_T - F^i_t\right)$ – 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 $\widehat{\sigma}^2_{T_i,T_{i+\Delta}}(t)$ 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, $\widehat{\sigma}_{T_i,T_{i+\Delta}}$ 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.

[43] 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 $\widehat{\sigma}_{\mathrm{VS},T}$ is small, so that we can use VSs *in lieu* of log contracts.[44]

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.

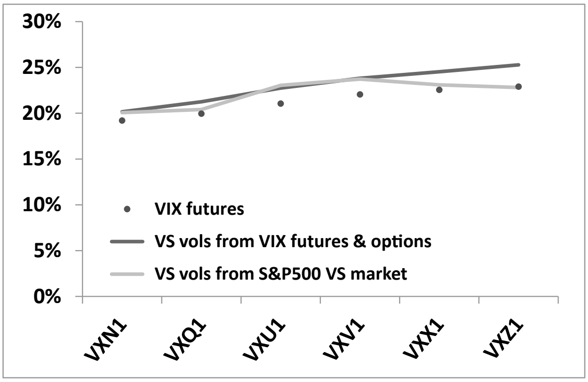


Figure 7.14: VIX futures, VS volatilities as generated by (7.98) and VS volatilities as derived from the S&P 500 VS market, all in 365 convention.

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 $\widehat{\sigma}^2_{T_i,T_{i+\Delta}}(t)_{\text{VIX mkt}} - \widehat{\sigma}^2_{T_i,T_{i+\Delta}}(t)_{\text{VS mkt}}$. 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 –

[44] As discussed in Section 5.5, there are valid reasons why $\widehat{\sigma}_T$ and $\widehat{\sigma}_{\mathrm{VS},T}$ could be different. Typically, because sellers of VSs would lose on large drawdowns of the underlying index, we expect that $\widehat{\sigma}_{\mathrm{VS},T} > \widehat{\sigma}_T$. It so happens that usually – see below – equivalent log-contract volatilities derived from the VIX market lie higher than $\widehat{\sigma}_{\mathrm{VS},T}$. In case $\widehat{\sigma}_{\mathrm{VS},T} > \widehat{\sigma}_T$, 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[45] – 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 $\left(F^i_{T_i}\right)^2$ 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 $\left(F^i_{T_i}\right)^2$ in the opposite case.[46]
- At $t = T_i$ we need to unwind the forward VS position at an implied VS volatility that is exactly equal to $\widehat{\sigma}_{\mathrm{VIX},T_i}$. 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 $F^i_{T_i}$. 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 arbitraged 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 arbitraged 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,

[45] There are three business days from Friday to Wednesday.

[46] 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 $F^i_{T_i}$.

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 $F_t^i, i = 1\ldots7$. 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.

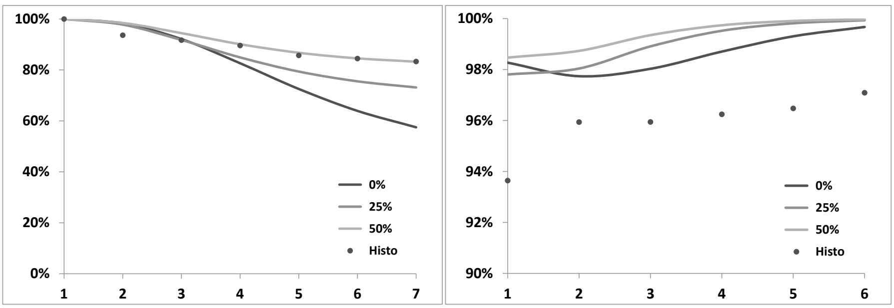


Figure 7.15: Left: correlations $\rho(F_t^1, F_t^i)$ of the first 7 VIX futures with the first future in the two-factor model for different values of the factor/factor correlation, and as observed in reality (dots), as a function of i. Right: correlations $\rho(F_t^i, F_t^{i+1})$ of contiguous futures.

We have used three parameter sets for the two-factor model, all calibrated to the benchmark form (7.40) for $\nu_T^{\mathrm{B}}(t)$, page 228, with $\sigma_0 = 100\%$, $\tau_0 = 3$ months, $\alpha = 0.4$, characterized by different levels of correlation ρ_{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.

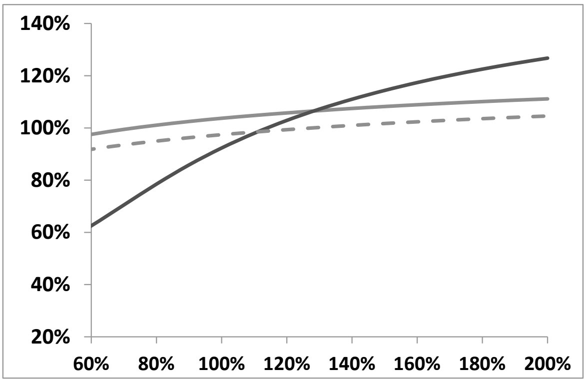


Figure 7.16: Implied volatilities of options on realized variance. The inception date is June 14, 2011 and the maturity is January 20, 2012. Implied volatilities of the realized variance are shown as a function of *volatility* moneyness using Set II parameters and three configurations for $\gamma^i, \beta^i, \zeta^i$: 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 ζ^i as derived from calibration 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 ζ 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 ξ_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 ν, 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].[47]

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 ξ_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:

$$\xi_t^i \;\equiv\; \widehat{\sigma}^2_{T_i,T_{i+1}}(t) \;=\; \frac{(T_{i+1}-t)\widehat{\sigma}^2_{T_{i+1}}(t) \;-\; (T_i-t)\widehat{\sigma}^2_{T_i}(t)}{T_{i+1}-T_i}$$

[47] 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 ρ_1 and ρ_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:

$$\rho_2(S_2) \;=\; \int E\big[\rho_{12}(S_2|S_1,\bullet)\,|S_1\big]\rho_1(S_1)\,dS_1, \quad \forall S_2$$

Still, the presence of the $\bullet$ state variables (X_t^1, X_t^2, in the two-factor model) affords considerable freedom in selecting ρ_{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 ξ_T^i are driftless.[48] We can thus recycle the two-factor model and, mimicking (7.28), write the SDE of ξ_t^i as:

$$\begin{aligned} d\xi_t^i &= (2\nu_i)\xi_t^i\,\alpha_{\theta_i}\left((1-\theta_i)\,e^{-k_1(T_i-t)}dW_t^1 \;+\; \theta_i e^{-k_2(T_i-t)}dW_t^2\right) \\ \alpha_{\theta_i} &= 1/\sqrt{(1-\theta_i)^2+\theta_i^2+2\rho_{12}\theta_i\,(1-\theta_i)} \end{aligned} \tag{7.99}$$

where index i for parameters θ and ν keeps track of the forward variance ξ^i they apply to.

While θ and ν depend on i, we use the same values for k_1, k_2, ρ_{12} for all intervals – otherwise we lose the two-dimensional Markov representation of the ξ_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:

$$\begin{aligned} \xi_t^i &= \xi_0^i e^{\omega_i x_t^{T_i} - \frac{\omega_i^2}{2}\chi(t,T_i)} \\ x_t^{T_i} &= \alpha_{\theta_i}\left[(1-\theta_i)\,e^{-k_1(T_i-t)}X_t^1 \;+\; \theta_i e^{-k_2(T_i-t)}X_t^2\right] \end{aligned} \tag{7.100}$$

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 accross 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.

[48] The exposure to ξ^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:

$$\begin{cases} \widehat{\sigma}_{T_i,T_{i+1}}(t) &= \sqrt{\xi_t^i} \\ \widehat{\sigma}_{T_i,T_{i+1}}(t) &= \sqrt{\frac{1}{T_{i+1}-T_1}\int_{T_i}^{T_{i+1}} \xi_t^\tau \, d\tau} \end{cases}$$

The SDE of forward variance $\widehat{\sigma}^2_{T_i,T_{i+1}}(t)$ is given, in both models, by:

$$\frac{d\widehat{\sigma}^2_{T_i,T_{i+1}}}{\widehat{\sigma}^2_{T_i,T_{i+1}}} = 2\nu_i\alpha_{\theta_i}\left(\left(1-\theta_i\right)e^{-k_1(T_i-t)}dW_t^1 + \theta_i e^{-k_2(T_i-t)}dW_t^2\right) \tag{7.101a}$$

$$\frac{d\widehat{\sigma}^2_{T_i,T_{i+1}}}{\widehat{\sigma}^2_{T_i,T_{i+1}}} = 2\nu\alpha_\theta\Big((1-\theta)A_i^1(t)e^{-k_1(T_i-t)}dW_t^1 + \theta A_i^2(t)e^{-k_2(T_i-t)}dW_t^2\Big) \tag{7.101b}$$

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 ξ_t^τ, but let us demand that they coincide for all t, for forward variances ξ_t^τ equal to their initial values ξ_0^τ, 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, ρ are identical in both models, where ρ is the correlation between W_t^1 and W_t^2.

The conditions on $\nu_i,\ \theta_i$ read:

$$\begin{aligned} \nu_i\alpha_{\theta_i}\left(1-\theta_i\right) &= \nu\alpha_\theta(1-\theta)A_i^1(0) \\ \nu_i\alpha_{\theta_i}\theta_i &= \nu\alpha_\theta\theta A_i^2(0) \end{aligned}$$

This yields:

$$\begin{cases} \theta_i &= \frac{\theta A_i^2(0)}{\theta A_i^2(0)+(1-\theta)A_i^1(0)} \\ \nu_i &= \nu\,\frac{\alpha_\theta\theta}{\alpha_{\theta_i}\theta_i}A_i^2(0) \end{cases} \tag{7.102}$$

Provided θ_i and ν_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 ξ_t^τ equal to their initial values ξ_0^τ. Consequently:

- instantaneous volatilities of VS volatilities of maturities T_i

- instantaneous correlations of forward VS volatilities $\widehat{\sigma}_{T_i,T_{i+1}}$ and $\widehat{\sigma}_{T_j T_{j+1}}$

are identical as well, in both models, for forward variances ξ_t^τ equal to their initial values ξ_0^τ.

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 ξ^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

$$A_i^n(0) = \frac{1 - e^{-k_n(T_{i+1}-T_i)}}{k_n\,(T_{i+1} - T_i)}$$

θ_i and ν_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

$$\frac{1}{2\widehat{\sigma}_{T,T+\Delta}(0)}\left(\widehat{\sigma}^2_{T,T+\Delta}(T) \; - \widehat{\sigma}^2_{T,T+\Delta}(0)\right)^+$$

for various values of Δ.

The expiry T of the swaption is 1 year, and the discrete model is mapped for each value of Δ 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, $\sigma_0 = 125\%$.

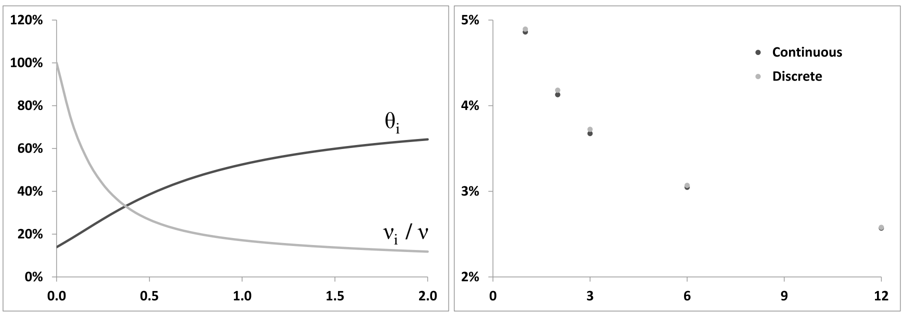


Figure 7.17: Left: θ_i and $\frac{\nu_i}{\nu}$ as a function of Δ (years). Right: prices of a 1-year ATM VS swaption in both continuous and discrete models, as a function of the maturity of the underlying VS volatility (months).

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.

ν	θ	k_1	k_2	ρ_{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.[49] 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:

$$\xi_t^i \;=\; \xi_0^i \Big((1-\gamma_i)\, e^{\omega_i x_t^{T_i} - \frac{\omega_i^2}{2}\chi(t,T_i)} + \gamma_i e^{\beta_i \omega_i x_t^{T_i} - \frac{(\beta_i\omega_i)^2}{2}\chi(t,T_i)} \Big)$$

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:

$$F_{T_i}^i \;=\; \widehat{\sigma}_{\mathrm{VIX},T_i}(T_i) \;=\; \widehat{\sigma}_{T_i,T_i+\Delta}(T_i)$$

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 $T_{i+1} - T_i = \Delta$.[50]

[49] In the continuous version of the model, instantaneous forward variances ξ_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.

[50] 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: $\widehat{\sigma}_{T_i+\Delta,T_{i+1}+\Delta}(T_{i+1}) \equiv \sqrt{\frac{\Delta\widehat{\sigma}^2_{T_{i+1},T_{i+1}+\Delta}(T_{i+1}) - (T_i+\Delta-T_{i+1})\widehat{\sigma}^2_{T_i,T_i+\Delta}(T_i)}{T_{i+1}-T_i}}$.

A parametric form

Since VIX futures are driftless – just like forward variances ξ^τ – we can use the discrete two-factor model directly on VIX futures and, mirroring (7.99) write:

$$\begin{aligned} dF_t^i &= \nu_i F_t^i \, \alpha_{\theta_i} \left((1-\theta_i)\, e^{-k_1(T_i-t)} dW_t^1 \;+\; \theta_i e^{-k_2(T_i-t)} dW_t^2 \right) \\ \alpha_{\theta_i} &= 1/\sqrt{(1-\theta_i)^2 + \theta_i^2 + 2\rho_{12}\theta_i (1-\theta_i)} \end{aligned} \tag{7.103}$$

where we use ν_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):

$$F_t^i \;=\; F_0^i \left((1-\gamma_i)\, e^{\omega_i x_t^{T_i} - \frac{\omega_i^2}{2}\chi(t,T_i)} + \gamma_i e^{\beta_i \omega_i x_t^{T_i} - \frac{(\beta_i\omega_i)^2}{2}\chi(t,T_i)} \right) \tag{7.104}$$

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:

$$F_t^i \;=\; F_0^i f^i(t, x_t^{T_i}) \tag{7.105}$$

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):

$$\frac{df^i}{dt} + \frac{\eta_i^2(T_i - t)}{2} \frac{d^2 f^i}{dx^2} \;=\; 0 \tag{7.106}$$

where $\eta_i^2\,(T_i - t)$ is the instantaneous variance of $x_t^{T_i}$. η_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 $x_{T_i}^{T_i} \to f^i(T_i, x_{T_i}^{T_i})$ 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:

$$\begin{aligned}
\mathcal{D}^K \;&=\; P\big(F_{T_i}^i < K\big) \;=\; P\big(f^i(T_i, x_{T_i}^{T_i}) < k\big) \\
&=\; \mathcal{N}_i\big(f^{i^{-1}}(T_i, k)\big)
\end{aligned}$$

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 $f^{i^{-1}}(T_i, k) = \mathcal{N}_i^{-1}\left(\mathcal{D}^K\right)$. 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.[51]

Characterizing the dynamics of VIX futures

What kind of dynamics does (7.105) generate for F_t^i? Making use of (7.106):

$$\frac{dF_t^i}{F_t^i} \;=\; \frac{d\ln f^i}{dx}\big(t, x_t^{T_i}\big)\,dx_t^{T_i}$$

The instantaneous volatility of F^i is thus given by:

$$\operatorname{vol}(F^i) \;=\; \eta\,(T_i - t)\left|\frac{d\ln f^i}{dx}\right|_{x = f^{i^{-1}}(t, F_t^i)} \tag{7.107}$$

[51] 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 mappping 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.[52] 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:

$$F^i_{T_i} \;=\; \widehat{\sigma}_{T_i,T_i+\Delta}(T_i)$$

What if we also need $\widehat{\sigma}_{T_i,T_i+\Delta}\,(t)$ for $t \leq T_i$?

VIX futures are given by:

$$F^i_{T_i} \;=\; F^i_0 f^i(t, x_t^{T_i})$$

Moreover, forward *variances* $\widehat{\sigma}^2_{T_i,T_i+\Delta}$ are driftless. We can thus represent $\widehat{\sigma}_{T_i,T_i+\Delta}$ as:

$$\widehat{\sigma}_{T_i,T_i+\Delta}\,(t) \;=\; F^i_0\sqrt{g^i(t, x_t^{T_i})}$$

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-Ühlenbeck processes: X_t^1 and X_t^2.[53]

[52]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: $\frac{df^i}{dx}(t,x) = E_t[\frac{df^i}{dx}(T_i, x_{T_i}^{T_i}) \mid x_t^{T_i} = x]$. $\frac{df^i}{dx}(T_i,x) \geq 0\ \forall x$ then implies $\frac{df^i}{dx}(t,x) \geq 0\ \forall x, \forall t$.

[53]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 $E\left[h(f^i(T_i, x_{T_i}^{T_i}))\right]$, but $E\left[h(f^i(T_i, x_{T_i}^{T_i}))f^{i+1}(T_i, x_{T_i}^{T_{i+1}})\right]$. 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 cliquets 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:

$$dS_t \;=\; (r-q)S_t dt + \sigma^i\left(\frac{S_t}{S_{T_i}}\right) S_t dW_t^S \tag{7.108}$$

where the correlations of W_t^S with W_t^1 and W_t^2 are denoted by ρ_{SX^1} and ρ_{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 σ^i:

$$\sigma^i(s) \;=\; \sigma_0^i \frac{n^i}{n^i - 1} \frac{\left(n^i - 1\right) s^{\beta^i - 1}}{\left(n^i - 1\right) + s^{\beta^i - 1}} \tag{7.109}$$

which is parametrized by three numbers: σ_0^i, β^i and n^i.

- σ_0^i is the local volatility for $S = S_{T_i}$.

- β^i controls the skew of maturity T_{i+1}. $\left.\frac{d\sigma^i}{d\ln S}\right|_{S=S_{T_i}} = \sigma_0^i \frac{n^i-1}{n^i}(\beta^i - 1)$. For $\left(\beta^i - 1\right)$ 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:

$$\left.\frac{d\widehat{\sigma}_{K,T_{i+1}}}{d\ln K}\right|_{S_{T_i}} \simeq \frac{1}{2}\left.\frac{d\sigma^i}{d\ln S}\right|_{S_{T_i}} = \frac{\sigma_0^i}{2}\frac{n^i-1}{n^i}(\beta^i - 1) \tag{7.110}$$

- n^i prevents the divergence of σ 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.

σ_0^i, β^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:

$$\mathcal{S}_{T_{i+1}}(T_i) = \left(\frac{\widehat{\sigma}_{\mathrm{ATMF},T_{i+1}}(T_i)}{\widehat{\sigma}^i_{\text{ATMF, ref}}}\right)^{\gamma^i} \mathcal{S}^i_{\text{ref}} \tag{7.111}$$

  where $\mathcal{S}^i_{\text{ref}}$ and $\widehat{\sigma}^i_{\text{ATMF, ref}}$ are reference levels for ATMF skew and volatility. $\gamma^i, \mathcal{S}^i_{\text{ref}}$ and $\widehat{\sigma}^i_{\text{ATMF, ref}}$ 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 $\widehat{\sigma}_{\mathrm{ATMF},T_{i+1}}(T_i)$ is chosen, we need to determine two functions $\sigma_0^i\left(\right), \beta^i\left(\right)$ for each interval $[T_i, T_{i+1}]$ such that setting $\sigma_0^i = \sigma_0^i\left(\widehat{\sigma}_{T_i,T_{i+1}}(T_i)\right)$ and $\beta^i = \beta^i\left(\widehat{\sigma}_{T_i,T_{i+1}}(T_i)\right)$ produces the desired behavior.

The ability to choose different fixed ATMF skew $\mathcal{S}_i$ or different values of $\gamma^i, \mathcal{S}^i_{\text{ref}}$ and $\widehat{\sigma}^i_{\text{ATMF, ref}}$ for different intervals $[T_i, T_{i+1}]$ is not superfluous.

It does not make sense to offset the sensistivity 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 cliquets 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 σ_0^i and β^i.

Two examples

Consider an interval $[T_i, T_{i+1}]$. Functions $\sigma_0(\widehat{\sigma}_{T_i,T_{i+1}}(T_i))$ and $\beta(\widehat{\sigma}_{T_i,T_{i+1}}(T_i))$ – we omit the i index to lighten notation – are obtained as follows:

- Select discrete values σ_0^k of σ_0 spanning a sufficiently wide range.
- For each value of σ_0^k find the value β^k of β such that the ATMF skew has the desired value, either constant or specified by (7.111). For each trial value of β, 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}: $\widehat{\sigma}_{\mathrm{VS}}^k$.
- Store the couples $(\widehat{\sigma}_{\mathrm{VS}}^k, \sigma_0^k)$ and $(\widehat{\sigma}_{\mathrm{VS}}^k, \beta^k)$ and proceed to the next value of σ_0.
- Finally, interpolate the discrete couples $(\widehat{\sigma}_{\mathrm{VS}}^k, \sigma_0^k)$ to generate the function $\sigma_0(\widehat{\sigma}_{\mathrm{VS}})$, and likewise for $\beta(\widehat{\sigma}_{\mathrm{VS}})$.

Figure 7.18 shows functions $\sigma_0\,()$ and $\beta\,()$ such that the ATMF skew is fixed, equal to 5%. We use the difference between implied volatilities of the 95% and 105% strikes rather than $\left.\frac{d\widehat{\sigma}_{K,T_{i+1}}}{d\ln K}\right|_{S_{T_i}}$ as a measure of ATMF 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.

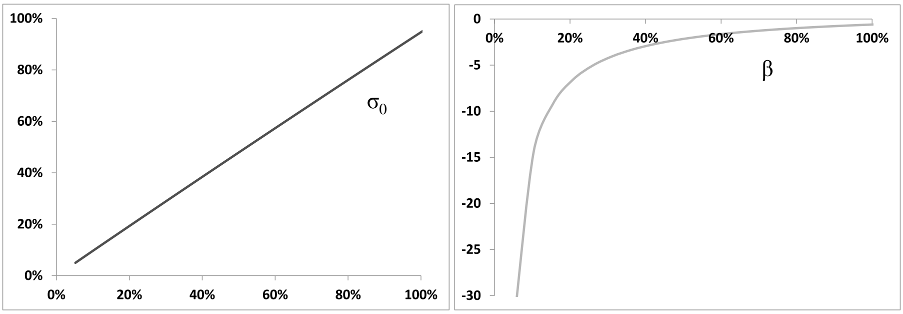
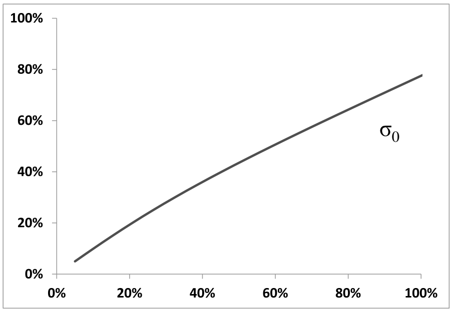


Figure 7.18: σ_0 (left) and β (right) as a function of $\widehat{\sigma}_{T_i,T_{i+1}}(T_i)$, such that the difference of implied volatilities of the 95% and 105% strikes is equal to 5%. We have taken $n_i = 3$ and $T_{i+1} - T_i = 1$ month.

Figure 7.19 shows functions $\sigma_0\,()$ and $\beta\,()$ such that the ATMF skew is of the form in (7.111) with $\gamma = 1$, $\mathcal{S}_{\mathrm{ref}} = 5\%$, $\widehat{\sigma}_{\text{ATMF, ref}} = 20\%$: the ATMF skew is proportional to the ATMF volatility.

The shapes of $\beta(\widehat{\sigma}_{\mathrm{VS}})$ are consistent with approximation (7.110) which implies that β is constant for a skew that is proportional to the ATMF volatility while $(\beta - 1)$



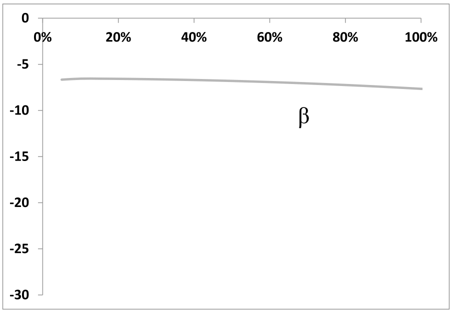


Figure 7.19: σ_0 (left) and β (right) as a function of $\widehat{\sigma}_{T_i,T_{i+1}}(T_i)$, such that the ATMF skew is of the form in (7.111) with $\gamma = 1$, $\mathcal{S}_{\text{ref}} = 5\%$, $\widehat{\sigma}_{\text{ATMF, ref}} = 20\%$. We have taken $n_i = 3$ and $T_{i+1} - T_i = 1$ month.

should be inversely proportional to the VS volatility in order to generate a skew that is independent on the level of volatility.[54]

- 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.

[54]For short maturities, approximation (3.28), page 121, shows that the VS implied volatility is equal to σ_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 σ_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 $\sigma_0(\xi^i_{T_i})$ and $\beta(\xi^i_{T_i})$ 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 $\xi^i_{T_i}$, (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 β a function of σ_0, with the dependence obtained in the calibration of the σ_0 and β functions with S_{T_i} set to the forward for maturity T_i, (b) generating the mapping $\sigma_0\left(\xi^i\right)$ for a range of values of S_{T_i}, thus obtaining in effect a mapping $\sigma_0\left(\xi^i, S_{T_i}\right)$. While not exact, step (a) ensures, in practice, that the target forward skew scenarios are obtained with good accuracy.[55]

- 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 αs^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 ξ^i_t
- 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 ξ^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 $\left(\beta^i - 1\right) \neq 0$ volatility becomes stochastic. We now derive an expression of the ATMF skew at order one in ν_i and in $(\beta^i - 1)$.

We assume a constant tenor Δ so that our schedule is given by $T_i = i\Delta$. We also use the same values for ν_i, θ_i, α. These numbers then only depend on tenor Δ, 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 σ_0, β.

We use a constant term structure of forward variances: $\xi^i_0 = \widehat{\sigma}^2$, and functions σ_0 and β 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 $\mathcal{S}_\Delta$.

[55] 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:

$$\mathcal{S}_T = \frac{1}{2\widehat{\sigma}^3 T}\int_0^T \frac{T-t}{T}\frac{\langle d\ln S_t \, d\widehat{\sigma}_T^2(t)\rangle_0}{dt}dt \tag{7.112}$$

where the 0 susbcript 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 ν and $(\beta-1)$.

We calculate $\mathcal{S}_T$ for maturities T that are multiples of tenor Δ: $T=N\Delta$. Using the definition of $\widehat{\sigma}_T(t)$, $(T-t)\,\widehat{\sigma}_T^2(t) = \int_t^T \xi_t^\tau d\tau$, where ξ_t^τ is the *instantaneous* forward variance for date τ. (7.112) can be rewritten as:

$$\begin{aligned}\mathcal{S}_{N\Delta} &= \frac{1}{2\widehat{\sigma}^3 T^2}\int_0^T \left\langle d\ln S_t \; d\Big(\int_t^T \xi_t^\tau d\tau\Big)\right\rangle_0 \\ &= \frac{1}{2\widehat{\sigma}^3 T^2}\sum_{i=0}^{N-1}\int_{T_i}^{T_{i+1}} \left\langle d\ln S_t \; d\Big(\int_t^{T_N} \xi_t^\tau d\tau\Big)\right\rangle_0\end{aligned}$$

For $\tau\in[T_k,T_{k+1}]$, $\xi_t^\tau = \xi_t^k$. Thus:

$$\int_t^{T_N}\xi_t^\tau d\tau = \int_t^{T_{i+1}}\xi_t^\tau d\tau + \Delta\sum_{j=i+1}^{N-1}\xi_t^j = (T_{i+1}-t)\widehat{\sigma}_{T_{i+1}}^2(t) + \Delta\sum_{j=i+1}^{N-1}\xi_t^j$$

$\mathcal{S}_T$ is given by:

$$\begin{aligned}\mathcal{S}_{N\Delta} = \frac{1}{2\widehat{\sigma}^3 T^2}\Bigg(&\Delta\sum_{i=0,\;j>i}^{N-1}\int_{T_i}^{T_{i+1}}\langle d\ln S_t \; d\xi_t^j\rangle_0 \\ &+\sum_{i=0}^{N-1}\int_{T_i}^{T_{i+1}}(T_{i+1}-t)\langle d\ln S_t \; d\widehat{\sigma}_{T_{i+1}}^2(t)\rangle_0\Bigg)\end{aligned} \tag{7.113}$$

The derivation of (7.112) utilizes the assumption that $E\left[\langle d\ln S_t \, d\widehat{\sigma}_T^2(t)\rangle \middle| \ln S\right]$ does not depend on S. We now verify that this holds at order one in ν and $(\beta-1)$.

- Consider the first line of (7.113). At order one in ν:

$$d\xi_t^j = (2\nu_\Delta)\xi_0^j\,\alpha_{\theta_\Delta}\Big((1-\theta_\Delta)\,e^{-k_1(T_j-t)}dW_t^1 + \theta_\Delta e^{-k_2(T_j-t)}dW_t^2\Big)$$

For the sake of obtaining the covariation at order one in ν and $(\beta-1)$, we take $dS_t = (r-q)S_t dt + \widehat{\sigma} S_t dW_t^S$ and get:

$$\left\langle d\ln S_t \; d\xi_t^j \right\rangle_0 = \\ (2\nu_\Delta)\, \widehat{\sigma}^3 \alpha_{\theta_\Delta} \left((1-\theta_\Delta)\, e^{-k_1(T_j-t)} \rho_{SX^1} \;+\; \theta_\Delta e^{-k_2(T_j-t)} \rho_{SX^2} \right) dt \tag{7.114}$$

where we have used that $\xi_0^j = \widehat{\sigma}^2$. $\left\langle d\ln S_t \; d\xi_t^j \right\rangle_0$ 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 $(\beta-1)$:

$$\sigma\left(\frac{S_t}{S_{T_i}}\right) = \sigma_0\left(\xi_{T_i}^i\right)\left(1 + \frac{n-1}{n}\left(\beta\left(\xi_{T_i}^i\right)-1\right)\ln\frac{S_t}{S_{T_i}}\right)$$

  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):

$$\sigma(t,S) = \overline{\sigma}(t) + \alpha(t)\ln\frac{S_t}{F_t} \tag{7.115}$$

  with:

$$\overline{\sigma}(t) = \sigma_0\left(\xi_{T_i}^i\right)\left(1 + \frac{n-1}{n}\left(\beta\left(\xi_{T_i}^i\right)-1\right)\ln\frac{F_t}{S_{T_i}}\right) \tag{7.116a}$$

$$\alpha(t) = \sigma_0\left(\xi_{T_i}^i\right)\frac{n-1}{n}\left(\beta\left(\xi_{T_i}^i\right)-1\right) \tag{7.116b}$$

  We know from Section 2.5.7 that, given a local volatility function of the form in (7.115), $\left\langle d\ln S_t \; d\widehat{\sigma}_{T_{j+1}}^2(t)\right\rangle_0$ does not depend on S_t, at order one in $\alpha(t)$.

  This covariation is already of order one in $\left(\beta\left(\xi_{T_i}^i\right)-1\right)$ thus need to be calculated at order *zero* in ν: taking the expectation of $\left\langle d\ln S_t \; d\widehat{\sigma}_{T_{j+1}}^2(t)\right\rangle_0$ with respect to $\xi_{T_i}^i$ simply amounts to setting $\xi_{T_i}^i = \widehat{\sigma}^2$.

  Thus the second contribution in (7.113) does not depend on S either, at order one in ν 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 $\widehat{\sigma}$, we have, using identity (7.112) for the case of a local volatility function of type (7.115):

$$\int_{T_i}^{T_{i+1}} (T_{i+1}-t)\left\langle d\ln S_t \; d\widehat{\sigma}_{T_{i+1}}^2(t)\right\rangle_0 = 2\widehat{\sigma}^3\Delta^2\mathcal{S}_\Delta$$

which allows us to rewrite (7.113) as:

$$\mathcal{S}_{N\Delta} \;=\; \frac{1}{N}\mathcal{S}_{\Delta} \;+\; \frac{1}{2\widehat{\sigma}^3 T^2}\Delta \sum_{i=0,\;j>i}^{N-1} \int_{T_i}^{T_{i+1}} \left\langle d\ln S_t \; d\xi_t^j \right\rangle_0 \tag{7.117}$$

Using (7.114) we have:

$$\begin{aligned}
&\frac{1}{2\widehat{\sigma}^3 T^2}\Delta \sum_{i=0,\;j>i}^{N-1} \int_{T_i}^{T_{i+1}} \left\langle d\ln S_t \; d\xi_t^j \right\rangle_0 \\
&= \frac{1}{2\widehat{\sigma}^3 T^2}\Delta 2\nu_\Delta \widehat{\sigma}^3 \alpha_{\theta_\Delta} \sum_{i=0,\;j>i}^{N-1} \int_{T_i}^{T_{i+1}} \\
&\quad \times \Big((1-\theta_\Delta)\, e^{-k_1(T_j-t)} \rho_{SX^1} + \theta_\Delta e^{-k_2(T_j-t)} \rho_{SX^2} \Big)\, dt \\
&= \nu_\Delta \alpha_{\theta_\Delta} \frac{1}{N^2} \sum_{i=0,\;j>i}^{N-1} \frac{1}{\Delta} \int_{T_i}^{T_{i+1}} \Big((1-\theta_\Delta)\, e^{-k_1(T_j-t)} \rho_{SX^1} + \theta_\Delta e^{-k_2(T_j-t)} \rho_{SX^2} \Big) dt \\
&= \nu_\Delta \alpha_{\theta_\Delta} \Big(\, (1-\theta_\Delta)\, \rho_{SX^1} \zeta\,(k_1\Delta, N) + \theta_\Delta\, \rho_{SX^2} \zeta\,(k_2\Delta, N) \,\Big)
\end{aligned}$$

where we have introduced function $\zeta\,(x, N)$ defined by:

$$\zeta\,(x, N) \;=\; \frac{1}{N^2} \sum_{i=0,\;j>i}^{N-1} \int_i^{i+1} \left(e^{-x}\right)^{j-u} du \;=\; \frac{e^x - 1}{x} \sum_{n=1}^{N-1} \frac{N-n}{N^2} \left(e^{-x}\right)^n \tag{7.118}$$

The final expression of the ATMF skew for maturity $T = N\Delta$, in the discrete two-factor model, at order one in ν and $(\beta - 1)$ is thus:

$$\mathcal{S}_{N\Delta} \;=\; \frac{1}{N}\mathcal{S}_{\Delta} \;+\; \frac{1}{2\widehat{\sigma}^3 T^2}\Delta \sum_{i=0,\;j>i}^{N-1} \int_{T_i}^{T_{i+1}} \left\langle d\ln S_t \; d\xi_t^j \right\rangle_0 \tag{7.119a}$$

$$\;=\; \frac{1}{N}\mathcal{S}_{\Delta} \;+\; \nu_\Delta \alpha_{\theta_\Delta} \Big(\, (1-\theta_\Delta)\, \rho_{SX^1} \zeta\,(k_1\Delta, N) + \theta_\Delta\, \rho_{SX^2} \zeta\,(k_2\Delta, N) \,\Big) \tag{7.119b}$$

- Expression (7.119) is an expansion of the ATMF skew at order one in $(\beta - 1)$ (first piece in (7.119b)) and ν (second piece). Note that $(\beta - 1)$ does not appear explicitly – only the ATMF future skew $\mathcal{S}_\Delta$ for tenor Δ 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 ν_Δ equal to 0.

When pricing a cliquet of period Δ, $\mathcal{S}_\Delta$ controls the forward-smile adjustment δP_2 while ν controls the volatility-of-volatility adjustment δP_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 foward skew for tenor $\Delta, \mathcal{S}_\Delta$, while the covariance of S_t with forward variances ξ_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 S_Δ.

- Imagine taking the limit $\Delta \to 0$. 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 $x \to 0, N \to \infty$ with Nx fixed, equal to, respectively, $k_1 T$ or $k_2 T$. Converting the sum in (7.118) in an integral, we get:

$$\begin{aligned}
\lim_{\substack{x\to 0,\, N\to\infty \\ Nx\,=\,kT}} \frac{e^x - 1}{x} \sum_{n=1}^{N-1} \frac{N-n}{N^2}\left(e^{-x}\right)^n &= \int_0^T \frac{T-t}{T^2} e^{-kt} dt \\
&= \frac{kT - \left(1 - e^{-kT}\right)}{(kT)^2}
\end{aligned}$$

  S_T is then given at order one in ν by:

$$\mathcal{S}_T = \nu\alpha_\theta \left((1-\theta)\,\rho_{SX^1} \frac{k_1 T - \left(1 - e^{-k_1 T}\right)}{(k_1 T)^2} + \theta\,\rho_{SX^2} \frac{k_2 T - \left(1 - e^{-k_2 T}\right)}{(k_2 T)^2} \right) \tag{7.120}$$

  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 ν. At order one in ν, 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,[56] which are mapped according to (7.102)
- $\sigma_0()$ and $\beta()$ chosen so that future smiles of maturity Δ exhibit a fixed ATMF skew, with $\mathcal{S}_\Delta = -0.5$ – see Figure 7.18. The latter value of $\mathcal{S}_\Delta$ corresponds to a difference of 5 points of implied volatility for 95% and 105% strikes for maturity Δ. We take $n_i = 3$, $\forall i$.
- a flat VS volatility $\widehat{\sigma}$ equal to 20%.

ν	θ	k_1	k_2	ρ_{12}	ρ_{SX^1}	ρ_{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 ($\mathcal{S}_\Delta = 0$) 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. $\mathcal{S}_T$ in Figure 7.20 is

[56] 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.

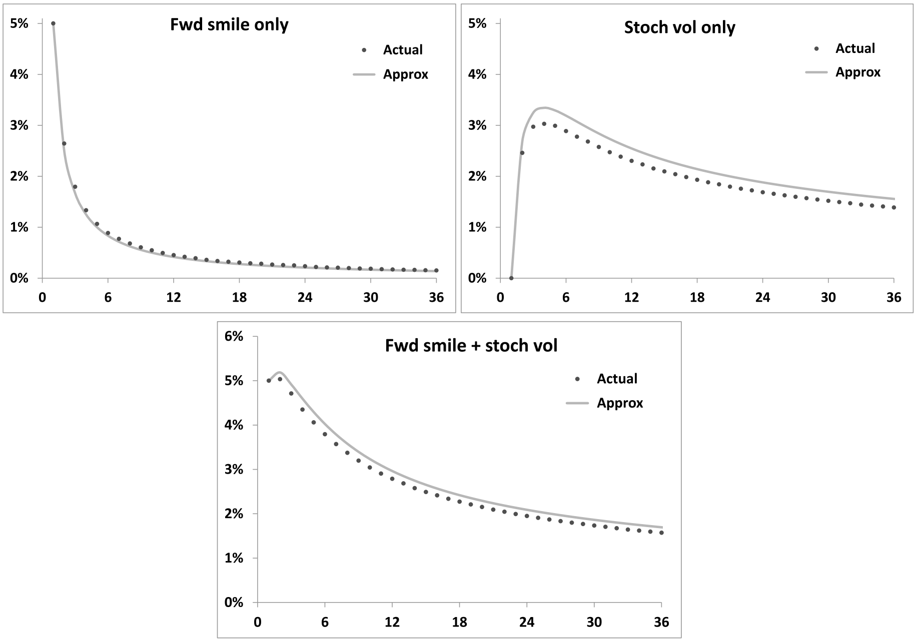


Figure 7.20: Top: $\mathcal{S}_T$, as the $95\%/105\%$ skew, obtained either with $\nu = 0$ (left) or with $\mathcal{S}_\Delta = 0$ (right) as a function of T (months), evaluated in a Monte-Carlo simulation of the discrete two-factor model with parameters in Table 7.8 (Actual) and as given by order-one formula (7.119b) (Approx). $\Delta = 1$ month.
Bottom: $\mathcal{S}_T$ when both forward smile ($\mathcal{S}_\Delta = -0.5$) and volatility of volatility ($\nu = 174\%$) are switched on.

non-monotonic as a function of T, however this depends on the relative magnitude of $\mathcal{S}_\Delta$ and ν_Δ – see Figure 7.21 where $\mathcal{S}_T$ is graphed for 3 different values of $\mathcal{S}_\Delta$.

This is illustrated also in Figure 7.22, which shows $\mathcal{S}_T$ for the same parameters as in Figure 7.8, except ν 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 Δ. In models thus specified, prices of cliquets 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.

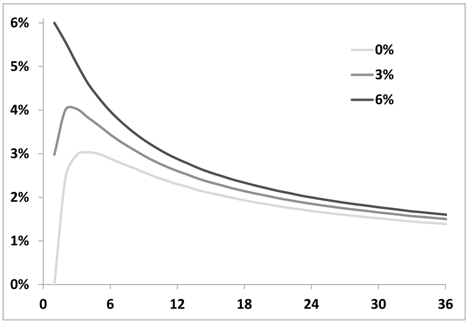


Figure 7.21: $\mathcal{S}_T$, as the $95\%/105\%$ skew, as a function of T (months), evaluated in a Monte-Carlo simulation of the discrete two-factor model with parameters in Table 7.8 for 3 different values of $\mathcal{S}_\Delta$, such that the 1-month $95\%/105\%$ ATMF skew is equal to 0, 3% and 6%.

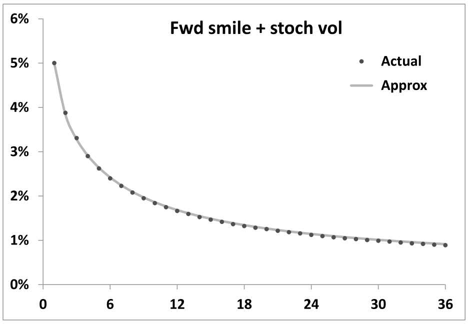


Figure 7.22: $\mathcal{S}_T$, as the $95\%/105\%$ skew, as a function of T (months), evaluated in a Monte-Carlo simulation of the discrete two-factor model with the same parameters as in Figure 7.20 except ν has been halved.

One first sets future smiles for maturity Δ, 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 cliquets, such as accumulators.

- For any choice of time scale Δ, 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 Δ 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:

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

This SDE admits a probabilistic interpretation. Its solution is given by:

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

under a dynamics for $S_t,\ \xi_t^u$ given by:

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

with λ_t^u and correlations between W_t^S and W_t^u such that:

$$\begin{aligned}\lim_{dt\to 0}\frac{1}{dt}E_t\left[d\ln S_t d\xi_t^u\right] &= \sqrt{\xi_t^t}\lambda_t^u\frac{1}{dt}E_t\left[dW_t^S dW_t^u\right] = \mu\left(t,u,\xi\right)\\ \lim_{dt\to 0}\frac{1}{dt}E_t\left[d\xi_t^u d\xi_t^{u'}\right] &= \lambda_t^u\lambda_t^{u'}\frac{1}{dt}E_t\left[dW_t^u dW_t^{u'}\right] = \nu\left(t,u,u',\xi\right)\end{aligned}$$

❧❧❧❧❧

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 ξ_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 ν is the volatility of a very short volatility. We introduce driftless processes x_t^T:

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

ξ_t^T is given by:

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

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:

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

This benchmark form, for given values of σ_0, τ_0, α, 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:

$$U_t = \frac{Q_t + (T-t)\,\widehat{\sigma}_T^2(t)}{T}$$

Q_t is the quadratic variation at t, and $\widehat{\sigma}_T(t)$ 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 $R_t = \frac{\widehat{\sigma}_T^2(t)}{U_t}$ 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 τ 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 $R_t = \frac{\widehat{\sigma}_{\tau T}^2(0)}{\widehat{\sigma}_T^2(0)}$, which produces the following amended expression for σ_{eff}, at inception:

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

▶ 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 $\sigma_{\rm eff}$ where κ is the (conditional) kurtosis of daily returns.

$$\sigma_{\rm eff}^2 \;=\; \frac{1}{T}\int_0^T 4\left(\frac{T-\tau}{T}\right)^2\left(\frac{\widehat{\sigma}_{\tau T}^2\,(0)}{\widehat{\sigma}_T^2\,(0)}\right)^2 \nu_T^2\,(\tau)\,d\tau \;+\; \frac{2+\kappa}{NT}$$

▶ 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 ξ_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 ξ_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 ξ_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 $\xi_t^i = \widehat{\sigma}^2_{T_i,T_{i+1}}(t)$, then construct a dynamics for S_t that complies with that of the ξ_t^i.

▶ Since the ξ^i are driftless, as their continuous counterparts ξ^T, the two-factor model can be employed for the ξ^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 ξ^i, (b) the density of $\frac{S_{T_{i+1}}}{S_{T_i}}$ is independent on S_{T_i} so that payoffs $\Sigma_i \omega_i f\left(\frac{S_{T_{i+1}}}{S_{T_i}}\right)$ 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 σ^i, (b) the correlation of S_t and forward variances ξ^i. Working at order one in both the slope of local volatility functions σ^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 σ^i decays as $\frac{1}{T}$. Numerical tests confirm the accuracy of the order-one expansion.

This page intentionally left blank