6.18 Including the Caplet Smile in the LFM

6.18.1 A Mini-tour on the Smile Problem

We have seen earlier that Black's formula for caplets is the standard in the cap market. This formula is consistent with the LFM, in that it comes as the expected value of the discounted caplet payoff under the related forward measure when the forward-rate dynamics is given by the LFM.

To fix ideas, let us consider again the time-0 price of a T_2-maturity caplet resetting at time T_1 $(0 < T_1 < T_2)$ with strike K and a notional amount of 1. Caplets and caps have been described more generally in Section 6.4. Let τ denote the year fraction between T_1 and T_2. Such a contract pays out the amount

$$\tau(F(T_1; T_1, T_2) - K)^+,$$

at time T_2, so that its value at time 0 is

$$P(0, T_2)\tau E_0^2[(F(T_1; T_1, T_2) - K)^+].$$

The dynamics for F in the above expectation under the T_2-forward measure is the lognormal LFM dynamics

$$dF(t; T_1, T_2) = \sigma_2(t)F(t; T_1, T_2)\, dW_t\,. \tag{6.73}$$

Lognormality of the T_1-marginal distribution of this dynamics implies that the above expectation results in Black's formula

$$\mathbf{Cpl}^{\text{Black}}(0, T_1, T_2, K) = P(0, T_2)\tau \text{Bl}(K, F_2(0), v_2(T_1))\ ,$$
$$v_2(T_1)^2 = \int_0^{T_1} \sigma_2^2(t)dt\,.$$

It is clear that in this derivation, the average volatility of the forward rate in $[0, T_1]$, i.e. $v_2(T_1)/\sqrt{T_1}$, does not depend on the strike K of the option. Indeed, in this formulation, volatility is a characteristic of the forward rate underlying the contract, and has nothing to do with the nature of the contract itself. In particular, it has nothing to do with the strike K of the contract.

Now take two different strikes K_1 and K_2. Suppose that the market provides us with the prices of the two related caplets $\mathbf{Cpl}^{\text{MKT}}(0, T_1, T_2, K_1)$ and $\mathbf{Cpl}^{\text{MKT}}(0, T_1, T_2, K_2)$. Both caplets have the same underlying forward rates and the same maturity.

Life would be simple if the market followed Black's formula in a consistent way. But is this the case? Does there exist a *single* volatility parameter $v_2(T_1)$ such that both

$$\mathbf{Cpl}^{\text{MKT}}(0, T_1, T_2, K_1) = P(0, T_2)\tau \text{Bl}(K_1, F_2(0), v_2(T_1))$$

and

$$\mathbf{Cpl}^{\text{MKT}}(0, T_1, T_2, K_2) = P(0, T_2)\tau \text{Bl}(K_2, F_2(0), v_2(T_1))$$

hold? The answer is a resounding "no". In general, market caplet prices do not behave like this. What one sees when looking at the market is that two *different* volatilities $v_2(T_1, K_1)$ and $v_2(T_1, K_2)$ are required to match the observed market prices if one is to use Black's formula:

$$\mathbf{Cpl}^{\text{MKT}}(0, T_1, T_2, K_1) = P(0, T_2)\tau \text{Bl}(K_1, F_2(0), v_2^{\text{MKT}}(T_1, K_1)),$$
$$\mathbf{Cpl}^{\text{MKT}}(0, T_1, T_2, K_2) = P(0, T_2)\tau \text{Bl}(K_2, F_2(0), v_2^{\text{MKT}}(T_1, K_2)).$$

In other terms, each caplet market price requires its own Black volatility $v_2^{\text{MKT}}(T_1, K)$ depending on the caplet strike K.

The market therefore uses Black's formula simply as a metric to express caplet prices as volatilities. The curve $K \mapsto v_2^{\text{MKT}}(T_1, K)/\sqrt{T_1}$ is the so called volatility smile of the T_1-expiry caplet. If Black's formula were consistent along different strikes, this curve would be flat, since volatility should not depend on the strike K. Instead, this curve is commonly seen to exhibit "smiley" or "skewed" shapes. The term skew is generally used for those structures where, for a fixed maturity, low-strikes implied volatilities are higher than high-strikes implied volatilities. The term smile is used instead to denote those structures where, again for a fixed maturity, the volatility has a minimum value around the current value of underlying forward rate.

Clearly, only some strikes $K = K_i$ are quoted by the market, so that usually the remaining points have to be determined through interpolation or through an alternative model. Interpolation in K, for a fixed expiry T_1, can be easy but it does not give any insight as to the underlying forward-rate dynamics compatible with such prices.

Indeed, suppose that we have a few market caplet prices for expiry T_1 and for a set of strikes $K = K_i$. By interpolation, we can obtain the price for every other possible K, i.e. we can build a function $K \mapsto \mathbf{Cpl}^{\text{MKT}}(0, T_1, T_2, K)$. Now, if this strike-$K$ price corresponds really to an expectation, we have

$$\mathbf{Cpl}^{\text{MKT}}(0, T_1, T_2, K) = P(0, T_2)\tau E_0^2(F(T_1; T_1, T_2) - K)^+ \tag{6.74}$$
$$= P(0, T_2)\tau \int (x - K)^+ p_2(x)\, dx, \tag{6.75}$$

where p_2 is the probability density function of $F_2(T_1)$ under the T_2-forward measure. If Black's formula were consistent, this density would be the log-normal density, coming for example from a dynamics such as (6.73). We have seen that this is not the case in the market. However, by differentiating the above integral twice with respect to K we see that, see also Breeden and Litzenberger (1978),

$$\frac{\partial^2 \mathbf{Cpl}^{\text{MKT}}(0, T_1, T_2, K)}{\partial K^2} = P(0, T_2)\tau p_2(K),$$

so that by differentiating the interpolated-prices curve we can find the density p_2 of the forward rate at time T_1 that is compatible with the given interpolated prices. However, the method of interpolation may interfere with the recovery of the density, since a second derivative of the interpolated curve is involved. Moreover, what kind of dynamics, alternative to (6.73), does the density p_2 come from?

A partial answer to these issues can be given the other way around, by starting from an alternative dynamics. Indeed, assume that

$$dF(t;T_1,T_2) = \nu(t, F(t;T_1,T_2))\, dW_t \tag{6.76}$$

under the T_2-forward measure, where ν can be either a deterministic or a stochastic function of $F(t;T_1,T_2)$. In the latter case we would be using a so called "stochastic-volatility model", where for example $\nu(t,F) = \xi(t)F$, with ξ following a second stochastic differential equation.

In this section, instead, we will concentrate on a deterministic $\nu(t,\cdot)$, leading to a so called "local-volatility model", such as for example $\nu(t,F) = \sigma_2(t)F^\gamma$ (CEV model), where $0 \le \gamma \le 1$ and where σ_2 is itself deterministic. We will then propose a new $\nu(t,\cdot)$ of our own, flexible enough for practical purposes.

We have seen above how the "true" forward-adjusted density p_2 of the underlying forward rate is linked to market prices through second-order differentiation. The problem we will face is finding a dynamics alternative to (6.73) and as compatible as possible with the density p_2 ideally associated with market prices. This will be done by fitting directly the prices implied by our alternative model to the market prices $\mathbf{Cpl}^{\text{MKT}}(0,T_1,T_2,K_i)$ for the considered strikes K_i, or equivalently by fitting the model implied volatilities to the implied volatilities $v_2^{\text{MKT}}(T_1,K_i)$ for the observed strikes. To understand this point, it may be helpful to explicitly clarify how an alternative dynamics such as (6.76) leads to a volatility smile to be fitted to the market smile. The alternative dynamics generates a smile, which is obtained as follows.

1. Set K to a starting value;
2. Compute the model caplet price

$$\Pi(K) = P(0,T_2)\tau E_0^2(F(T_1;T_1,T_2) - K)^+$$

   with F obtained through the alternative dynamics (6.76).
3. Invert Black's formula for this strike, i.e. solve

$$\Pi(K) = P(0,T_2)\tau \text{Bl}(K_1, F_2(0), v(K)\sqrt{T_1})$$

   in $v(K)$, thus obtaining the (average) model implied volatility $v(K)$.
4. Change K and restart from point 2.

The fact that the alternative dynamics is not lognormal implies that we obtain a curve $K \mapsto v(K)$ that is not flat. Clearly, one needs to choose $\nu(t,\cdot)$ flexible enough for this curve to be able to resemble or even match the corresponding volatility curves coming from the market. Indeed, the model implied volatilities $v(K_i)$ corresponding to the observed strikes have to be made as close as possible to the corresponding market volatilities $v_2^{\mathrm{MKT}}(T_1, K_i)/\sqrt{T_1}$, by acting on the coefficient $\nu(\cdot, F)$ in the alternative dynamics. We will address this problem in the following sections.

We finally point out that one has to deal, in general, with an implied-volatility surface, since we have a caplet-volatility curve for each considered expiry. The calibration issues, however, are essentially unchanged, apart from the obviously larger computational effort required when trying to fit a bigger set of data.

6.18.2 Modeling the Smile

We start with a little history and a few references.

Similarly to what happens in the equity or foreign-exchange markets, also in the interest-rate market non-flat structures are normally observed when plotting implied volatilities against strikes and maturities. For example, you may plot the curve $K \mapsto v_2^{\mathrm{MKT}}(T_1, K)/\sqrt{T_1}$ given in the earlier section and you may find it to have a smile-like shape in several cases.

As we have just seen in the mini-tour, modeling the dynamics of forward LIBOR rates as in the LFM leads to implied caplet (and hence cap) volatilities that are constant for each fixed caplet (cap) maturity. The LFM, therefore, can be used to exactly retrieve ATM cap prices, but it fails to reproduce non-flat volatility surfaces when a whole range of strikes is considered.

The above considerations suggest the need for an alternative model that is capable of suitably fitting the larger set of prices that is usually available to a trader. Several approaches can then be followed. A first approach is based on assuming (alternative) explicit dynamics for the forward-rate processes that immediately lead to volatility smiles or skews. Examples are the general CEV process of Cox (1975) and Cox and Ross (1976) or the displaced diffusion of Rubinstein (1983). A second approach is based on the assumption of a continuum of traded strikes and goes back to Breeden and Litzenberger (1978). Successive developments are due to Dupire (1994, 1997) and Derman and Kani (1994, 1998). However, the former approach does not provide sufficient flexibility to properly calibrate the whole volatility surface, whereas the latter has the major drawback that one needs to smoothly interpolate option prices between consecutive strikes in order to be able to differentiate them twice with respect to the strike.

In general the problem of finding a distribution that consistently prices all quoted options is largely undetermined, since there are infinitely many curves connecting (smoothly) finitely many points. A possible solution is then given by assuming a particular parametric distribution depending on several, time-dependent parameters. But the question remains of finding forward-rate dynamics consistent with the chosen parametric density. A possible solution to this issue is given by Brigo and Mercurio (2000a, 2000b) who find dynamics leading to a parametric distribution that may be flexible enough for practical purposes. Their resulting forward-rate process combines the parametric risk-neutral distribution approach with the alternative-dynamics approach.

In the following sections, we first introduce the forward-LIBOR model that can be obtained by displacing a given lognormal diffusion. We then describe the CEV model used by Andersen and Andreasen (2000) to model the evolution of the forward-rate process. We finally illustrate the lognormal-mixture approach proposed by Brigo and Mercurio (2000a, 2000b, 2001b).

6.18.3 The Shifted-Lognormal Case

A very simple way of constructing forward-rate dynamics that implies non-flat volatility structures is by shifting the generic lognormal dynamics analogous to (6.73). Indeed, let us assume that the forward rate F_j evolves, under its associated T_j-forward measure, according to

$$\begin{aligned} F_j(t) &= X_j(t) + \alpha, \\ dX_j(t) &= \beta(t) X_j(t)\, dW_t, \end{aligned} \tag{6.77}$$

where α is a real constant, β is a deterministic function of time and W is a standard Brownian motion. We immediately have that

$$dF_j(t) = \beta(t)(F_j(t) - \alpha)\, dW_t, \tag{6.78}$$

so that, for $t < T \leq T_{j-1}$, the forward rate F_j can be explicitly written as

$$F_j(T) = \alpha + (F_j(t) - \alpha) e^{-\frac{1}{2}\int_t^T \beta^2(u)\, du + \int_t^T \beta(u)\, dW_u}. \tag{6.79}$$

The distribution of $F_j(T)$, conditional on $F_j(t)$, $t < T \leq T_{j-1}$, is then a shifted lognormal distribution with density

$$p_{F_j(T)|F_j(t)}(x) = \frac{1}{(x-\alpha)U(t,T)\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{\ln \frac{x-\alpha}{F_j(t)-\alpha} + \frac{1}{2}U^2(t,T)}{U(t,T)} \right)^2 \right\}, \tag{6.80}$$

for $x > \alpha$, where

$$U(t,T) := \sqrt{\int_t^T \beta^2(u) du}. \tag{6.81}$$

The resulting model for F_j preserves the analytically tractability of the geometric Brownian motion X. Notice indeed that

$$P(t,T_j) E^j \{ [F_j(T_{j-1}) - K]^+ | \mathcal{F}_t \} = P(t,T_j) E^j \{ [X_j(T_{j-1}) - (K-\alpha)]^+ | \mathcal{F}_t \},$$

so that, for $\alpha < K$, the caplet price $\mathbf{Cpl}(t, T_{j-1}, T_j, \tau, N, K)$ associated with (6.78) is simply given by

$$\mathbf{Cpl}(t, T_{j-1}, T_j, \tau, N, K) = \tau N P(t, T_j)\text{Bl}(K - \alpha, F_j(t) - \alpha, U(t, T_{j-1})). \tag{6.82}$$

The implied Black volatility $\hat{\sigma} = \hat{\sigma}(K, \alpha)$ corresponding to a given strike K and to a chosen α is obtained by backing out the volatility parameter $\hat{\sigma}$ in Black's formula that matches the model price:

$$\begin{aligned} &\tau N P(t, T_j)\text{Bl}(K, F_j(t), \hat{\sigma}(K, \alpha)\sqrt{T_{j-1} - t}\,) \\ &= \tau N P(t, T_j)\text{Bl}(K - \alpha, F_j(t) - \alpha, U(t, T_{j-1})). \end{aligned}$$

We can now understand why the simple affine transformation (6.77) can be useful in practice. The resulting forward-rate process, in fact, besides having explicit dynamics and known marginal density, immediately leads to closed-form formulas for caplet prices that allow for skews in the caplet implied volatility. An example of the skewed volatility structure $K \mapsto \hat{\sigma}(K, \alpha)$ that is implied by such a model is shown in Figure 6.5.[1]

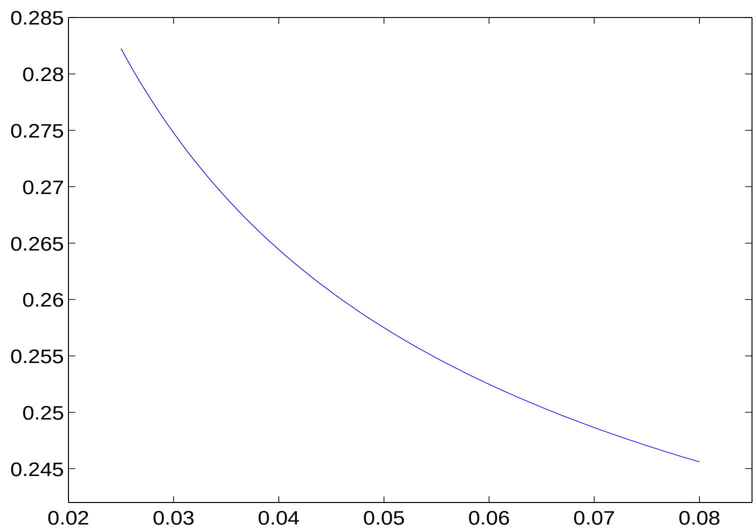

Fig. 6.5. Caplet volatility structure $\hat{\sigma}(K, \alpha)$ plotted against K implied, at time $t = 0$, by the forward-rate dynamics (6.78), where we set $T_{j-1} = 1$, $T_j = 1.5$, $\alpha = -0.015$, $\beta(t) = 0.2$ for all t and $F_j(0) = 0.055$.

Introducing a non-zero parameter α has two effects on the implied caplet volatility structure, which for $\alpha = 0$ is flat at the constant level $U(0, T_{j-1})$. First, it leads to a strictly decreasing ($\alpha < 0$) or increasing ($\alpha > 0$) curve.

[1] Such a figure shows a decreasing caplet-volatility curve. In real markets, however, different structures can be encountered too (smile-shaped, skewed to the right,...).

Second, it moves the curve upwards ($\alpha < 0$) or downwards ($\alpha > 0$). More generally, ceteris paribus, increasing α shifts the volatility curve $K \mapsto \hat{\sigma}(K, \alpha)$ down, whereas decreasing α shifts the curve up. The formal proof of these properties is straightforward. Notice, for example, that at time $t = 0$ the implied at-the-money ($K = F_j(0)$) caplet volatility $\hat{\sigma}$ satisfies

$$\mathrm{Bl}(F_j(0), F_j(0), \sqrt{T_{j-1}}\hat{\sigma}(F_j(0), \alpha)) = \mathrm{Bl}(F_j(0) - \alpha, F_j(0) - \alpha, U(0, T_{j-1})),$$

which reads

$$(F_j(0) - \alpha)\left[2\Phi\left(\frac{1}{2}U(0, T_{j-1})\right) - 1\right] = F_j(0)\left[2\Phi\left(\frac{1}{2}\sqrt{T_{j-1}}\hat{\sigma}(F_j(0), \alpha)\right) - 1\right].$$

When increasing α the left hand side of this equation decreases, thus decreasing the $\hat{\sigma}$ in the right-hand side that is needed to match the decreased left-hand side. Moreover, when differentiating (6.82) with respect to α we obtain a quantity that is always negative.

Shifting a lognormal diffusion can then help in recovering skewed volatility structures. However, such structures are often too rigid, and highly negative slopes are impossible to recover. Moreover, the best fitting of market data is often achieved for decreasing implied volatility curves, which correspond to negative values of the α parameter, and hence to a support of the forward-rate density containing negative values. Even though the probability of negative rates may be negligible in practice, many people regard this drawback as an undesirable feature.

The next models we illustrate may offer the properties and flexibility required for a satisfactory fitting of market data.

6.18.4 The Constant Elasticity of Variance (CEV) Model

Another classical model leading to skews in the implied caplet-volatility structure is the CEV model of Cox (1975) and Cox and Ross (1976). Recently, Andersen and Andreasen (2000) applied the CEV dynamics as a model of the evolution of forward LIBOR rates.

Andersen and Andreasen start with a general forward-LIBOR dynamics of the following type:

$$dF_j(t) = \phi(F_j(t))\sigma_j(t)\, dZ_j^j(t),$$

where ϕ is a general function. Andersen and Andreasen suggest as a particularly tractable case in this family the CEV model, where

$$\phi(F_j(t)) = [F_j(t)]^\gamma,$$

with $0 < \gamma < 1$. Notice that the “border” cases $\gamma = 0$ and $\gamma = 1$ would lead respectively to a normal and a lognormal dynamics.

The model then reads

$$dF_j(t) = \sigma_j(t)[F_j(t)]^\gamma\, dW_t, \quad F_j = 0 \text{ absorbing boundary when } 0 < \gamma < 1/2, \tag{6.83}$$

where we set $W = Z_j^j$, a one-dimensional Brownian motion under the T_j forward measure.

For $0 < \gamma < 1/2$ equation (6.83) does not have a unique solution unless we specify a boundary condition at $F_j = 0$. This is why we take $F_j = 0$ as an absorbing boundary for the above SDE when $0 < \gamma < 1/2$.[2]

Time dependence of σ_j can be dealt with through a deterministic time change. Indeed, by first setting

$$v(\tau, T) = \int_\tau^T \sigma_j(s)^2 ds$$

and then

$$\widetilde{W}(v(0,t)) := \int_0^t \sigma_j(s) dW(s),$$

we obtain a Brownian motion $\widetilde{W}$ with time parameter v. We substitute this time change in equation (6.83) by setting $f_j(v(t)) := F_j(t)$ and obtain

$$df_j(v) = f_j(v)^\gamma d\widetilde{W}(v), \quad f_j = 0 \text{ absorbing boundary when } 0 < \gamma < 1/2. \tag{6.84}$$

This is a process that can be easily transformed into a Bessel process via a change of variable. Straightforward manipulations lead then to the transition density function of f. By also remembering our time change, we can finally go back to the transition density for the continuous part of our original forward-rate dynamics. The continuous part of the density function of $F_j(T)$ conditional on $F_j(t)$, $t < T \leq T_{j-1}$, is then given by

$$\begin{aligned} p_{F_j(T)|F_j(t)}(x) &= 2(1-\gamma)k^{1/(2-2\gamma)}(uw^{1-4\gamma})^{1/(4-4\gamma)}e^{-u-w}I_{1/(2-2\gamma)}(2\sqrt{uw}), \\ k &= \frac{1}{2v(t,T)(1-\gamma)^2}, \\ u &= k[F_j(t)]^{2(1-\gamma)}, \\ w &= kx^{2(1-\gamma)}, \end{aligned} \tag{6.85}$$

with I_q denoting the modified Bessel function of the first kind of order q. Moreover, denoting by $g(y,z) = \frac{e^{-z}z^{y-1}}{\Gamma(y)}$ the gamma density function and by $G(y,x) = \int_x^{+\infty} g(y,z)dz$ the complementary gamma distribution, the probability that $F_j(T) = 0$ conditional on $F_j(t)$ is $G\left(\frac{1}{2(1-\gamma)}, u\right)$.

[2] Andersen and Andreasen (2000) also extend their treatment to the case $\gamma > 1$, while noticing that this can lead to explosion when leaving the T_j-forward measure (under which the process has null drift).

A major advantage of the model (6.83) is its analytical tractability, allowing for the above transition density function. This transition density can be useful, for example, in Monte Carlo simulations. From knowledge of the density follows also the possibility to price simple claims. In particular, the following explicit formula for a caplet price can be derived:

$$\begin{aligned}\mathbf{Cpl}(t,T_{j-1},T_j,\tau,N,K) = \tau N P(t,T_j)\Bigg[F_j(t)\sum_{n=0}^{+\infty} g(n+1,u)\,G\Big(c_n, kK^{2(1-\gamma)}\Big)\\ -K\sum_{n=0}^{+\infty} g\left(c_n,u\right)G\Big(n+1, kK^{2(1-\gamma)}\Big)\Bigg],\end{aligned} \tag{6.86}$$

where k and u are defined as in (6.85) and

$$c_n := n+1+\frac{1}{2(1-\gamma)}.$$

This price can be expressed also in terms of the non-central chi-squared distribution function we have encountered in the CIR model. Recall that we denote by $\chi^2(x;r,\rho)$ the cumulative distribution function for a non-central chi-squared distribution with r degrees of freedom and non-centrality parameter ρ, computed at point x. Then the above price can be rewritten as

$$\begin{aligned}\mathbf{Cpl}(t,T_{j-1},T_j,\tau,N,K) = \tau N P(t,T_j)\Bigg[F_j(t)\bigg(1-\chi^2\Big(2K^{1-\gamma};\frac{1}{1-\gamma}+2,2u\Big)\bigg)\\ -K\chi^2\Big(2u;\frac{1}{1-\gamma},2kK^{1-\gamma}\Big)\Bigg].\end{aligned} \tag{6.87}$$

As hinted at above, the caplet price (6.86) leads to skews in the implied volatility structure. An example of the structure that can be implied is shown in Figure 6.6. As previously done in the case of a geometric Brownian motion, an extension of the above model can be proposed based on displacing the CEV process (6.83) and defining accordingly the forward-rate dynamics. The introduction of the extra parameter determining the density shifting may improve the calibration to market data.

Finally, there is the possibly annoying feature of absorption in $F = 0$. While this does not necessarily constitute a problem for caplet pricing, it can be an undesirable feature from an empirical point of view. Also, it is not clear whether there could be some problems when pricing more exotic structures. As a remedy to this absorption problem, Andersen and Andreasen (2000) propose a "Limited" CEV (LCEV) process, where instead of $\phi(F) = F^\gamma$ they set

$$\phi(F) = F\ \min(\epsilon^{\gamma-1}, F^{\gamma-1})\ ,$$

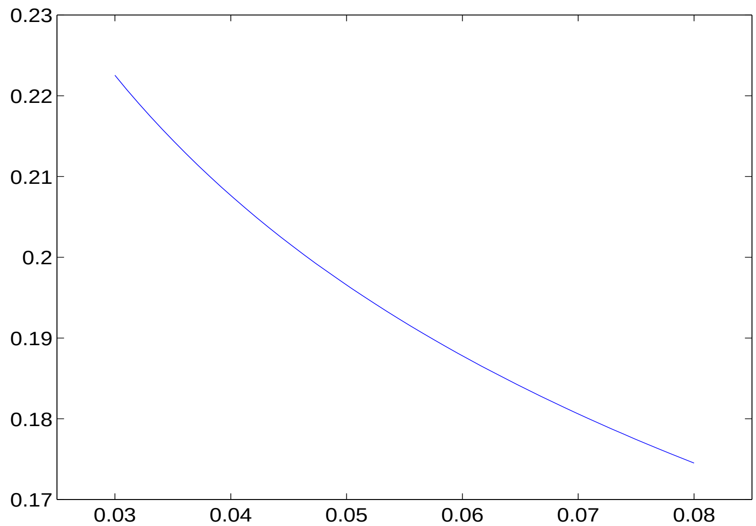

Fig. 6.6. Caplet volatility structure implied by (6.86) at time $t = 0$, where we set $T_{j-1} = 1$, $T_j = 1.5$, $\sigma_j(t) = 1.5$ for all t, $\gamma = 0.5$ and $F_j(0) = 0.055$.

where ϵ is a small positive real number. This function collapses the CEV diffusion coefficient F^γ to a (lognormal) level-proportional diffusion coefficient $F\epsilon^{\gamma-1}$ when F is small enough to make little difference (smaller than ϵ itself). Andersen and Andreasen (2000) compare the LCEV and CEV models as far as cap prices are concerned and conclude that the differences are small and tend to vanish when $\epsilon \to 0$. They also investigate, to some extent, the speed of convergence. A Crank-Nicholson scheme is used to compute cap prices within the LCEV model. As for the CEV model itself, Andersen and Andreasen allow for $\gamma > 1$ also in the LCEV case, with the difference that then ϵ has to be taken very large.

As far as the calibration of the CEV model to swaptions is concerned, approximated swaption prices based on "freezing the drift" and "collapsing all measures" are also derived (analogous to Rebonato's formula in the LFM). See Andersen and Andreasen (2000) for the details.