Part III

Term Structure Models

10

One-Factor Short Rate Models I

So far, our focus has been on vanilla models suitable for simple securities for which a change of measure allows the price to be expressed as an expectation of (a function of) a single random variable, typically a forward swap or Libor rate. However, many practically important securities, such as those that are callable or path-dependent, depend on interest rates in a substantially more complex manner, necessitating the construction of models for the dynamics of the *entire* discount curve — and not just a select few points on it. We have already, in Chapter 4, outlined the HJM theory that governs all dynamic discount curve models driven by vector-valued Brownian motions. The general HJM class with its infinite-dimensional Markovian dynamics is, however, too unwieldy to work with in practice, so it is of considerable interest to identify HJM model sub-classes that involve a finite number of Markov state variables only. We shall devote several chapters to this task, covering first the "classical" approach of writing down an explicit SDE for the short rate $r(t)$.

In our treatment of short rate models, we start out in this chapter with an in-depth analysis of the one-factor mean-reverting Gaussian model, providing a classical perspective on a model that we encountered in a modern HJM setting in Chapter 4. The chapter also covers the affine one-factor model, of which the Gaussian model is a special case. In Chapter 11, we generalize our discussion to arbitrary one-factor SDEs for the short rate, and finally, in Chapter 12, we introduce the class of multi-factor short rate models.

For derivatives pricing purposes, the short rate modeling approach has largely been superseded by newer approaches. Still, short rate models remain quite popular in empirical work, and a good understanding of these models provides a strong foundation for work with more sophisticated models.

10.1 The One-Factor Gaussian Short Rate Model

We recall that discount bond prices are given by the risk-neutral expectation

$$P(t,T) = \mathrm{E}_t^{\mathrm{Q}}\left(e^{-\int_t^T r(u)\,du}\right), \tag{10.1}$$

so knowledge of the risk-neutral dynamics for $r(t)$ is in principle sufficient to compute time t discount bond prices to all maturities $T > t$. In practice, the expectation in (10.1) may, of course, not be computable in closed form, so to make short rate models operational in practice we must look for the sub-class of models where (10.1) is either analytically tractable or, at the very least, amenable to fast numerical methods.

One approach for which (10.1) becomes particularly tractable is to model the short rate as a Gaussian random variable. The resulting *Gaussian short rate* (GSR) model has a long and distinguished history in the financial literature. While our applications focus leaves us little room for historical ruminations, we shall make a slight concession here, by developing the GSR model progressively from the historically important — yet ultimately impractical — special case in Ho and Lee [1986]. Our development of the model will also initially progress by classical "bottoms-up" means, developing the dynamics of the forward curve from an SDE for the short rate, rather than the other way around. Besides providing some historical perspective, our style of presentation involves several generally applicable techniques and should give the reader additional intuition about the mechanics of the models involved.

10.1.1 The Ho-Lee Model

10.1.1.1 Notations and First Steps

Starting from the fundamental assumption that the short rate $r(t)$ is adapted to a single Brownian motion $W(t)$, the simplest possible dynamics we can imagine is the martingale process $r(t) = r(0) + \sigma_r W(t)$, or

$$dr(t) = \sigma_r\, dW(t), \tag{10.2}$$

where $\sigma_r > 0$ is a constant and $W(t)$ is a Brownian motion in the risk-neutral measure Q. From the basic risk-neutral pricing relationship (10.1), the time t discount bond maturing at time T then must have the price

$$P(t,T) = \mathrm{E}_t\left(e^{-\int_t^T r(u)\,du}\right) = \mathrm{E}_t\left(e^{-r(0)(T-t)-\sigma_r\int_t^T W(u)\,du}\right), \tag{10.3}$$

where $\mathrm{E}_t = \mathrm{E}_t^{\mathrm{Q}}$ is the time t risk-neutral expectation operator.

Lemma 10.1.1. *If $r(t)$ follows (10.2) in the risk-neutral measure, then*

$$\mathrm{E}_t\left(e^{-\int_t^T r(u)\,du}\right) = \exp\left(-r(t)\,(T-t) + \frac{1}{6}\sigma_r^2(T-t)^3\right).$$

Proof. We notice that

$$r(u) = r(t) + \int_t^u \sigma_r \, dW(s), \quad u > t,$$

so that

$$-\int_t^T r(u)du = -r(t)\,(T-t) - \sigma_r \int_t^T \int_t^u dW(s)\,du.$$

The order of integration can be changed by Fubini's theorem (see Duffie [2001]), such that

$$\int_t^T \int_t^u dW(s)\,du = \int_t^T \int_s^T du\,dW(s) = \int_t^T (T-s)\,dW(s).$$

By the Ito isometry, it then follows that $-\int_t^T r(u)\,du$ is Gaussian with mean $-r(t)(T-t)$ and variance

$$\mathrm{Var}_t\left(\int_t^T r(u)du\right) = \sigma_r^2 \int_t^T (T-s)^2\,ds = \frac{1}{3}\sigma_r^2 (T-t)^3.$$

The result of the lemma then follows from basic moment properties of log-normal variables, see e.g. (1.22). □

Let us define a yield $y(t,T) = -\ln P(t,T)/(T-t)$, such that

$$y(t,T) = r(t) - \frac{1}{6}\sigma_r^2 (T-t)^2.$$

The yield curve shapes that can be produced by the simple model in (10.2) are rather primitive, as is evident from this expression. In particular, the yield curve is always downward-sloping in $T-t$ and $y_\infty = \lim_{T\to\infty} y(t,T) = -\infty$.

10.1.1.2 Fitting the Term Structure of Discount Bonds

The model presented above effectively has only two parameters — $r(0)$ and σ_r — with which one can attempt to fit the initial yield curve. It should be clear that this is insufficient to properly match observable discount bond prices, which effectively disqualifies the model from practical pricing applications. Fortunately, as realized in the paper Ho and Lee [1986], a remedy is quite straightforward[1]: simply introduce a deterministic function $a(t)$ and alter the model to be

$$r(t) = r(0) + a(t) + \sigma_r W(t), \quad a(0) = 0, \tag{10.4}$$

[1]The original paper by Ho and Lee was set exclusively in discrete time. The continuous-time version of the model developed here is, we feel, significantly more transparent.

such that

$$dr(t) = a'(t)\,dt + \sigma_r\,dW(t), \tag{10.5}$$

where $a'(t)$ is the first-order derivative of $a(t)$. To match the discount bond curve at time 0, $a(t)$ cannot be freely stipulated, but must be set as specified in Lemma 10.1.2 below.

Lemma 10.1.2. *Let $r(t)$ be given as in (10.4), and assume that discount bond prices at time 0, $P(0,T)$, are known for all $T > 0$. Set*

$$a(t) = f(0,t) - r(0) + \frac{1}{2}\sigma_r^2 t^2, \quad f(0,t) = -\frac{\partial \ln P(0,t)}{\partial t}.$$

Then, for any $T > 0$,

$$\mathrm{E}\left(e^{-\int_0^T r(u)\,du}\right) = P(0,T). \tag{10.6}$$

Proof. Applying Lemma 10.1.1, we get

$$\mathrm{E}\left(e^{-\int_0^t r(u)\,du}\right) = \exp\left(-r(0)t + \frac{1}{6}\sigma_r^2 t^3\right) \times \exp\left(-\int_0^t a(u)\,du\right),$$

from which it follows that (10.6) is satisfied if

$$-\int_0^t a(u)\,du = \ln P(0,t) + r(0)t - \frac{1}{6}\sigma_r^2 t^3.$$

Taking derivatives with respect to t yields

$$a(t) = -\frac{\partial \ln P(0,t)}{\partial t} - r(0) + \frac{1}{2}\sigma_r^2 t^2 = f(0,t) - r(0) + \frac{1}{2}\sigma_r^2 t^2.$$

□

The model (10.4) with $a(t)$ set as in Lemma 10.1.2 is known as the *Ho-Lee* model. We characterize the model further in the following proposition.

Proposition 10.1.3. *In the Ho-Lee model, the risk-neutral process for $r(t)$ is*

$$dr(t) = \left(\frac{\partial f(0,t)}{\partial t} + \sigma_r^2 t\right) dt + \sigma_r\,dW(t), \tag{10.7}$$

and bond prices at time t can be reconstituted from $r(t)$ through the expression

$$P(t,T) = \frac{P(0,T)}{P(0,t)} \exp\left(-\left(r(t) - f(0,t)\right)(T-t) - \frac{1}{2}\sigma_r^2 t\,(T-t)^2\right).$$

Proof. Equation (10.7) follows directly from (10.5) when $a(t)$ satisfies Lemma 10.1.2. To show the second part of the proposition, applying Lemma 10.1.1 to $r(t) - a(t)$ yields

$$
\begin{aligned}
&P(t,T) \\
&\quad = \exp\left(-(r(t)-a(t))(T-t)+\frac{1}{6}\sigma_r^2(T-t)^3\right) \\
&\qquad \times \exp\left(-\int_t^T a(u)\,du\right) \\
&\quad = \exp\left(-(r(t)-a(t))(T-t)+\frac{1}{6}\sigma_r^2(T-t)^3\right) \\
&\qquad \times \exp\left(\ln P(0,T)+r(0)T-\frac{1}{6}\sigma_r^2T^3-\ln P(0,t)-r(0)t+\frac{1}{6}\sigma_r^2t^3\right) \\
&\quad = \frac{P(0,T)}{P(0,t)}\exp\left(-(r(t)-a(t)-r(0))(T-t)+\frac{1}{2}\sigma_r^2(Tt^2-T^2t)\right).
\end{aligned}
$$

In this expression $-a(t)-r(0)=-f(0,t)-\sigma_r^2t^2/2$ from the definition of $a(t)$. The result follows. □

10.1.1.3 Analysis and Comparison with HJM Approach

To gain a better understanding of the Ho-Lee model, let us examine the dynamics for bonds and forward rates implied by the model. From Proposition 10.1.3, we get

$$
f(t,T)=-\frac{\partial \ln P(t,T)}{\partial T}=f(0,T)+r(t)-f(0,t)+\sigma_r^2t\,(T-t) \tag{10.8}
$$

and

$$
df(t,T)=dr(t)+\left(\sigma_r^2(T-2t)-\frac{\partial f(0,t)}{\partial t}\right)dt=\sigma_r^2(T-t)\,dt+\sigma_r\,dW(t). \tag{10.9}
$$

In similar fashion, we get

$$
dP(t,T)/P(t,T)=r(t)\,dt-\sigma_r(T-t)\,dW(t).
$$

In the notations of Section 4.4, we have thus established that forward rate volatilities in the Ho-Lee model are $\sigma_f(t,T)=\sigma_r$ and discount bond volatilities are $\sigma_P(t,T)=\sigma_r(T-t)$. Due to the constancy of $\sigma_f(t,T)$, random perturbations of the forward curve from movement in the dW term will thus be parallel[2], in the sense that all points on the forward curve will move by identical amounts. Discount bond volatilities, on the other hand, approach

[2]Due to the presence of the T-dependent "convexity adjustment" term $\sigma_r^2(T-t)$ in the drift of the forward rate process, net forward curve movements are not perfectly parallel. Were this the case, it is well-known that the model would be arbitrageable.

zero in linear fashion as $t \to T$, reflecting the pull to par phenomenon discussed earlier in Chapter 4.

Setting aside for a moment the question about whether the Ho-Lee model is a reasonable representation of the real world, let us make a brief interlude to point out that we could, in fact, have specified the model directly as an HJM model with $\sigma_f(t,T) = \sigma_r$ and a single Brownian motion. The HJM result, Lemma 4.4.1, then immediately establishes the drift in the SDE for $f(t,T)$ to be

$$\mu_f(t,T) = \sigma_r \int_t^T \sigma_r \, du = \sigma_r^2 \, (T-t) \, ,$$

consistent with (10.9) above. Integrating this equation establishes (10.8), from which the discount bond reconstitution formula in Proposition (10.1.3) follows. To establish (10.7), we simply write $r(t) = f(t,t)$ and differentiate:

$$dr(t) = df(t,T)|_{T=t} + \left. \frac{\partial f(t,T)}{\partial T} \right|_{T=t} dt = \sigma_r \, dW(t) + \left(\frac{\partial f(0,t)}{\partial t} + \sigma_r^2 t \right) dt,$$

where the second equality uses (10.8)–(10.9). Notice that arriving at Proposition 10.1.3 in this manner did not involve evaluation of any expectations.

The Ho-Lee model has several drawbacks that disqualifies it for most, if not all, pricing applications. We list some of them below.

- The constancy of forward rate volatilities as a function of forward rate maturity $(T-t)$ is unrealistic: long-dated forward rates are less volatile than short-dated ones.
- The constancy of forward rate volatilities as a function of calendar time t gives the model time-stationary dynamics, but also results in the model having far too few degrees of freedom to allow for calibration to quoted option prices.
- Spot and forward interest rates are Gaussian and can therefore become negative, which is unrealistic.
- The model has only one driving Brownian motion and instantaneous moves of all forward rates are therefore perfectly correlated, contrary to empirical evidence.

The last objection is common for all one-factor short rate models and will disqualify these models for the pricing of options that have strong payoff dependency on non-parallel moves of the yield curve, e.g. spread options (see Chapter 17). The possibility of generating negative rates also cannot be helped unless we abandon the Gaussian setting (which we shall do later in this chapter), but we *can* address the problems associated with using constant forward rate volatility. We turn to this problem next.

10.1.2 The Mean-Reverting GSR Model

10.1.2.1 The Vasicek Model

Many empirical studies find that interest rates exhibit *mean reversion*, in the sense that if an interest rate is high by historical standards, it will most likely fall in the future (and vice versa if the interest rate is low). To model this phenomenon, Vasicek [1977] assumed that the short rate follows a one-factor *Ornstein-Uhlenbeck* process in the risk-neutral measure:

$$dr(t) = \varkappa\,(\vartheta - r(t))\,dt + \sigma_r\,dW(t), \tag{10.10}$$

where $\varkappa, \vartheta, \sigma_r$ are positive constants. From results for the linear SDE in Section 1.6, it follows that the short rate can be written

$$r(t) = \vartheta + (r(0) - \vartheta)e^{-\varkappa t} + \sigma_r \int_0^t e^{-\varkappa(t-s)}\,dW(s). \tag{10.11}$$

It follows that $r(t)$ is a Gaussian random variable with moments

$$\mathrm{E}\,(r(t)) = \vartheta + (r(0) - \vartheta)\,e^{-\varkappa t}, \tag{10.12}$$

$$\mathrm{Var}\,(r(t)) = \frac{\sigma_r^2}{2\varkappa}\left(1 - e^{-2\varkappa t}\right). \tag{10.13}$$

As $t \to \infty$, the mean of the short rate approaches ϑ and the variance goes to $\sigma_r^2/(2\varkappa)$. Accordingly, ϑ is often known as the *long-term level* (or sometimes the *mean reversion level*) of the short rate. The speed at which the short rate can be expected to revert to its long-term level is determined by $\varkappa$, known as the *mean reversion speed*.

To establish a discount bond pricing formula in the Vasicek model, we use (10.11) to write

$$-\int_0^t r(u)\,du = -\vartheta t - (r(0) - \vartheta)\left(1 - e^{-\varkappa t}\right)/\varkappa - \sigma_r \int_0^t \int_0^u e^{-\varkappa(u-s)}\,dW(s)\,du.$$

Clearly $-\int_0^t r(u)\,du$ is Gaussian, with mean

$$-\vartheta t - (r(0) - \vartheta)\left(1 - e^{-\varkappa t}\right)/\varkappa.$$

To establish the variance, we follow the approach in Lemma 10.1.1 and reverse the order of integration in the stochastic integral, followed by an application of the Ito isometry. The result is

$$\begin{aligned}\mathrm{Var}\left(\sigma_r \int_0^t \int_0^u e^{-\varkappa(u-s)}\,dW(s)\,du\right) &= \sigma_r^2 \int_0^t e^{2\varkappa s}\left(\int_s^t e^{-\varkappa u}du\right)^2 ds \\ &= \frac{\sigma_r^2}{2\varkappa^3}\left(-e^{-2\varkappa t} + 4e^{-\varkappa t} + 2t\varkappa - 3\right).\end{aligned}$$

From the usual result for log-normal variables, it follows that discount bond prices in the Vasicek model can be computed as

$$\begin{aligned} P(0,t) &= \exp\left(\mathrm{E}\left(-\int_0^t r(u)\,du\right) + \frac{1}{2}\mathrm{Var}\left(-\int_0^t r(u)\,du\right)\right) \\ &= \exp\left(-\frac{1-e^{-\varkappa t}}{\varkappa} r(0) - \vartheta t + \frac{\vartheta}{\varkappa}\left(1-e^{-\varkappa t}\right)\right) \\ &\quad \times \exp\left(\frac{\sigma_r^2}{4\varkappa^3}\left(-e^{-2\varkappa t} + 4e^{-\varkappa t} + 2t\varkappa - 3\right)\right). \end{aligned}$$

More generally, we have the following proposition, the proof of which is straightforward.

Proposition 10.1.4. *Define*

$$\begin{aligned} B(t,T) &= \frac{1-e^{-\varkappa(T-t)}}{\varkappa}, \\ A(t,T) &= \left(\vartheta - \frac{\sigma_r^2}{2\varkappa^2}\right)(B(t,T) - (T-t)) - \frac{\sigma_r^2 B(t,T)^2}{4\varkappa}. \end{aligned}$$

Then, in the Vasicek model (10.10),

$$P(t,T) = \exp\left(A(t,T) - B(t,T)r(t)\right).$$

As we did for the model in Section 10.1.1.1, define $y(t,T) = -\ln P(t,T)/(T-t)$, and notice that now a finite limit exists,

$$y_\infty = \lim_{T\to\infty} y(t,T) = \vartheta - \sigma_r^2/\left(2\varkappa^2\right).$$

In the Vasicek model, three different yield curve shapes are possible.

Lemma 10.1.5. *Let $y(t,T) = -\ln P(t,T)/(T-t)$. Then*

- *If $r(t) > \vartheta$, then $y(t,T)$ decreases in $T-t$.*
- *If $r(t) < y_\infty - \sigma_r^2/(4\varkappa^2)$, then $y(t,T)$ increases in $T-t$.*
- *Otherwise, $y(t,T)$ first increases in $T-t$ and then decreases (i.e. $y(t,T)$ is humped).*

Proof. By straightforward, but tedious, calculus. □

While this is certainly an improvement over the martingale model we encountered in Section 10.1.1.1, the Vasicek model is still not capable of fitting the observable yield curve accurately enough for pricing applications. It should be obvious that the way to solve this problem is to mimic the step that lead to the Ho-Lee model in Section 10.1.1.2: introduce a deterministic function of time into the definition (10.11). That is, we write

$$r(t) = a(t) + \vartheta + (r(0) - \vartheta)e^{-\varkappa t} + \sigma_r \int_0^t e^{-\varkappa(t-s)}\,dW(s) = a(t) + r_{OU}(t),$$

where $a(t)$ is a deterministic function and $r_{OU}(t)$ is the short rate in the Vasicek model. The function $a(t)$ is determined from the condition that

$$\mathrm{E}\left(e^{-\int_0^t r_{OU}(u)\,du}\right) e^{-\int_0^t a(u)du} = P(0,t),$$

where the right-hand side is assumed given. Further development of this model proceeds as in Section 10.1.1.2, and results are easily imagined; we skip the analysis as the resulting model is a special case of the more general setup in Section 10.1.2.2 below. We do note, however, that the Vasicek model — both with and without adjustment to fit the initial yield curve — is easily shown to have forward rate and discount bond volatilities of

$$\sigma_f(t,T) = \sigma_r e^{-\varkappa(T-t)}, \quad \sigma_P(t,T) = \sigma_r \left(\frac{1-e^{-\varkappa(T-t)}}{\varkappa}\right).$$

Introduction of mean reversion into the model will thus introduce exponential decay in the term structure of forward rate volatilities. From an empirical standpoint this is considerably more appealing than the maturity-independent forward rate volatilities in the Ho-Lee model, and also in qualitative agreement with the fact that short- and medium-maturity interest rate options trade at higher implied volatilities[3] than do long-dated options. While this is a step up from the Ho-Lee model, the model still has too few degrees of freedom for many derivatives pricing applications, as the model will rarely calibrate well to observed prices of vanilla options (e.g. European swaptions and caps). We improve on this in the next section.

10.1.2.2 The General One-Factor GSR Model

The most general form of the one-factor GSR model is given by the SDE

$$dr(t) = \varkappa(t)\,(\vartheta(t) - r(t))\,dt + \sigma_r(t)\,dW(t), \tag{10.14}$$

i.e. we have now allowed all parameters in the Vasicek model to depend on time. While this model can be developed by classical means (see e.g. Hull and White [1994a] for, often laborious, details), it is significantly easier to work within an HJM setting. In fact, we already showed in Section 4.5.2 that short rate dynamics of the form in (10.14) must originate from a "separable" HJM model of the form

$$df(t,T) = \sigma_f(t,T)\left(\int_t^T \sigma_f(t,u)\,du\right) dt + \sigma_f(t,T)\,dW(t), \tag{10.15}$$

$$\sigma_f(t,T) = \sigma_r(t)\exp\left(-\int_t^T \varkappa(u)\,du\right).$$

Chapter 4 also proved the following result for the function $\vartheta(t)$.

[3] An exception to this observation is the humped volatility term structure that can often be observed in caplet markets. We return to this issue in Section 10.1.2.3.

Proposition 10.1.6. *For the general one-factor GSR model (10.14) to match the initial yield curve, we must have*

$$\vartheta(t) = \frac{1}{\varkappa(t)}\frac{\partial f(0,t)}{\partial t} + f(0,t) + \frac{1}{\varkappa(t)}\int_0^t e^{-2\int_u^t \varkappa(s)\,ds}\sigma_r(u)^2\,du.$$

Proof. Follows from Proposition 4.5.4, when $d = 1$. □

We notice the presence of $\partial f(0,t)/\partial t$ in the expression for $\vartheta(t)$ (a similar term was, of course, present in the Ho-Lee model) which can be a nuisance in applications where the initial forward curve is not smooth, as when we have used simple bootstrapping to construct the curve. To get rid of the term, we now switch variables, from $r(t)$ itself to $x(t) = r(t) - f(0,t)$. Dynamics for $x(t)$, as well as the bond reconstitution formula for (10.14) in terms of $x(t)$ are listed next.

Proposition 10.1.7. *Define*

$$x(t) \triangleq r(t) - f(0,t).$$

Then, for the model (10.14)–(10.15),

$$dx(t) = (y(t) - \varkappa(t)x(t))\,dt + \sigma_r(t)\,dW(t), \quad x(0) = 0, \tag{10.16}$$

where

$$y(t) = \int_0^t e^{-2\int_u^t \varkappa(s)ds}\sigma_r(u)^2\,du. \tag{10.17}$$

The bond reconstitution formula is

$$P(t,T) = \frac{P(0,T)}{P(0,t)}\exp\left(-x(t)G(t,T) - \frac{1}{2}y(t)G(t,T)^2\right), \tag{10.18}$$

$$G(t,T) = \int_t^T e^{-\int_t^u \varkappa(s)ds}\,du.$$

Proof. To simplify notation, define $K(t) = \int_0^t \varkappa(u)\,du$, and set $g(t) = \sigma_r(t)e^{K(t)}$, $h(t) = e^{-K(t)}$. Then $\sigma_f(t,T) = g(t)h(T)$ and, by integration of (10.15),

$$f(t,T) = f(0,T) + h(T)\int_0^t g(u)^2\int_u^T h(s)\,ds\,du + h(T)\int_0^t g(u)\,dW(u). \tag{10.19}$$

Set

$$x(t) = h(t)\int_0^t g(u)^2\int_u^t h(s)\,ds\,du + h(t)\int_0^t g(u)\,dW(u),$$

and note that, by the Leibniz rule for differentiation of an integral,

$$
\begin{aligned}
dx(t) =& h'(t)\left(\int_0^t g(u)^2 \int_u^t h(s)\,ds\,du\right)dt + h(t)^2\left(\int_0^t g(u)^2\,du\right)dt \\
&+ h(t)g(t)\,dW(t) + h'(t)\int_0^t g(u)\,dW(u)\,dt \\
=& \left(\frac{h'(t)}{h(t)}x(t) + y(t)\right)dt + h(t)g(t)\,dW(t) \\
=& \left(y(t) - \varkappa(t)x(t)\right)dt + \sigma_r(t)\,dW(t),
\end{aligned}
$$

where

$$
y(t) = h(t)^2 \int_0^t g(u)^2\,du
$$

was defined in (10.17). From (10.19) we have

$$
\begin{aligned}
f(t,T) &= f(0,T) + \frac{h(T)}{h(t)}x(t) + h(T)\int_0^t g(u)^2 \int_u^T h(s)\,ds\,du \\
&\quad - h(T)\int_0^t g(u)^2 \int_u^t h(s)\,ds\,du \\
&= f(0,T) + \frac{h(T)}{h(t)}x(t) + h(T)\int_t^T h(s)\,ds \int_0^t g(u)^2\,du \\
&= f(0,T) + \frac{h(T)}{h(t)}\left(x(t) + \frac{y(t)}{h(t)}\int_t^T h(s)\,ds\right),
\end{aligned}
$$

such that in particular $r(t) = f(t,t) = f(0,t) + x(t)$, as claimed earlier. Inserting the expression for $f(t,T)$ into the basic relation

$$
P(t,T) = \exp\left(-\int_t^T f(t,u)\,du\right)
$$

produces (10.18) after a few rearrangements. □

Remark 10.1.8. The discount bond dynamics for $P(t,T)$ are

$$
dP(t,T)/P(t,T) = r(t)\,dt - \sigma_P(t,T)\,dW(t), \quad \sigma_P(t,T) = \sigma_r(t)G(t,T).
$$

Remark 10.1.9. In the reconstitution formula (10.18), notice that

$$
G(t,T) = \left(G(0,T) - G(0,t)\right)e^{\int_0^t \varkappa(s)ds},
$$

a result that is often useful in grid-based numerical work (see Section 10.1.5).

Proposition 10.1.7 is an important result and shall serve as the foundation for most of the remaining discussion of Gaussian short rate models.

10.1.2.3 Time-Stationarity and Caplet Hump

A Gaussian HJM model is said to be *time-stationary* if the instantaneous volatility $\sigma_f(t,T)$ is only a function of $T-t$, i.e. the time *to* maturity rather than the time *of* maturity T. Time stationarity is an appealing feature, as it implies that the volatility term structure of forward rates will look the same in the future as it does today; in the absence of other information, this prediction is often very reasonable and in good agreement with empirical observation. In the setting of the one-factor GSR model, imposing time-stationarity will require us to set both $\sigma_r(t)$ and $\varkappa(t)$ to constants, such that

$$\sigma_f(t,T) = \sigma_r e^{-\varkappa(T-t)}. \tag{10.20}$$

In other words, the only time-stationary forward rate volatility term structure that can be constructed in the GSR model is an exponentially decaying one. In practice, however, it is quite common to observe (from the caplet market, say) forward rate volatility structures that have a marked "hump", with short-dated options trading at very low volatilities. This effect can largely be attributed to central bank activity, as the extreme short end of the forward curve tends to move primarily in response to central bank changes to funding rates. As such changes are relatively infrequent and normally quite predictable[4], short-dated forward rates are typically associated with relatively little uncertainty and, consequently, have low volatilities.

If we attempt to match a GSR model to a humped forward volatility structure, it follows from (10.20) that this cannot be done in a stationary manner and we are forced to let $\varkappa$ become a function of time. To see this, suppose that we at time 0 observe forward volatilities $\sigma_f(0,T) = b(T)$, where $b(T)$ is a humped function of T, i.e. $b(T)$ initially increases in T but ultimately decreases in T. Ideally, we would like to set $\sigma_f(t,T) = b(T-t)$, but this is not possible in the GSR setting, as explained above. To make the GSR model match $b(T)$ at time 0, we instead are forced to make $\varkappa$ a function of time, determined from the relation

$$\sigma_f(0,T) = \sigma_r e^{-\int_0^T \varkappa(u)du} = b(T), \quad \sigma_r = b(0).$$

Taking logarithms and differentiating gives

$$\varkappa(t) = -\frac{d\left(\ln b(t)\right)}{dt} = -\frac{b'(t)}{b(t)}.$$

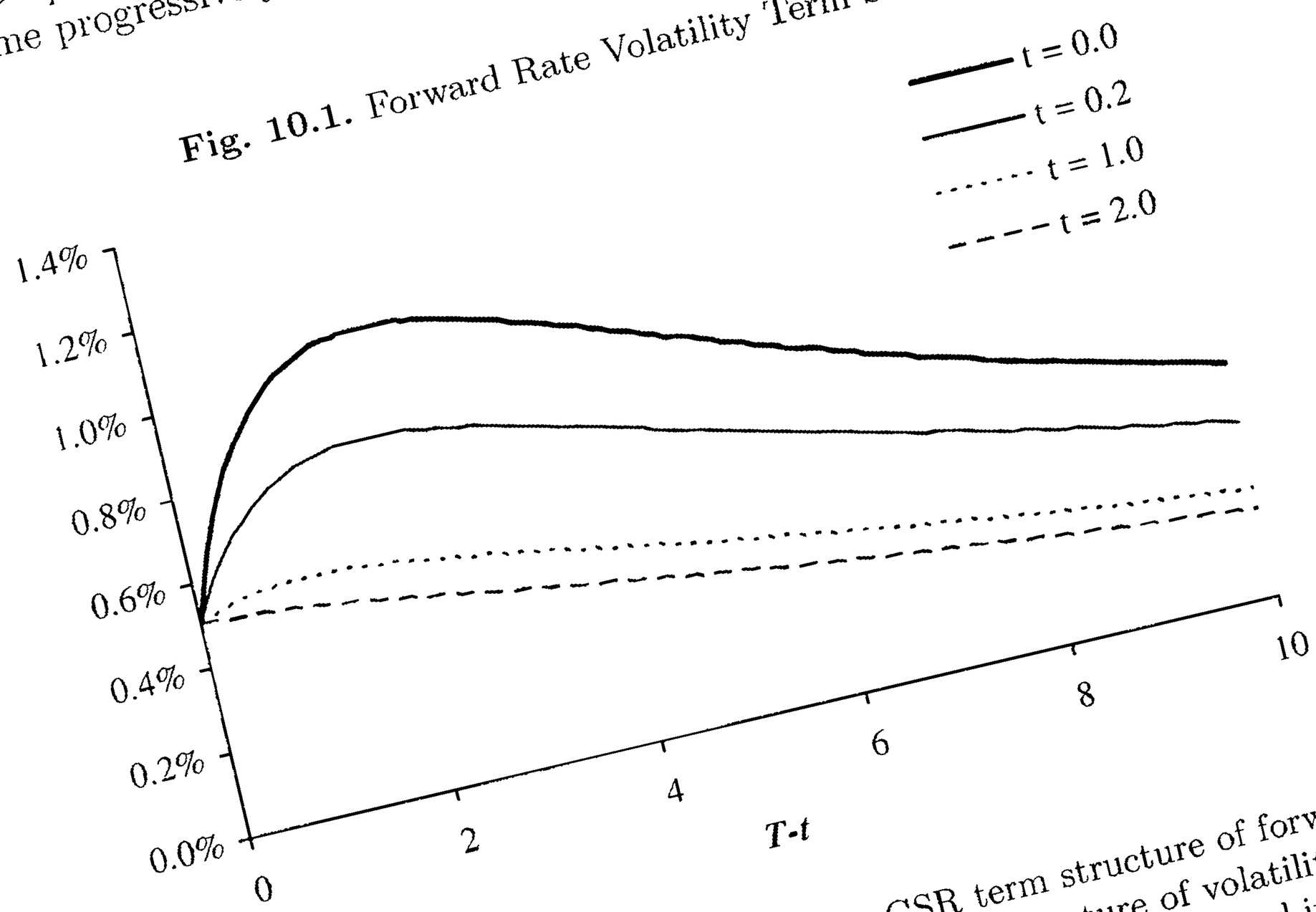
If t_p is the time t at which $b(t)$ reaches its peak (i.e. $b'(t_p) = 0$), it follows that $\varkappa(t)$ will be negative for all $t < t_p$ and positive for all $t > t_p$. At time $t > 0$, our so-calibrated GSR model will produce instantaneous forward volatilities of

[4] On occasion there is significant uncertainty in the market about the intentions of monetary authorities, in which case the caplet hump may disappear temporarily.

$$\sigma_f(t,T) = \sigma_r e^{-\int_t^T \varkappa(u)du} = \sigma_r b(T)/b(t).$$

Clearly $\sigma_f(t,T)$ is not stationary. In fact, once $t > t_p$ the model no longer produces a hump at all, as $b(T)/b(t)$ is monotonically decaying in $T-t$ for $t > t_p$. Figure 10.1 demonstrates this effect; notice also how volatilities become progressively lower as time t moves forward.

Fig. 10.1. Forward Rate Volatility Term Structure



Notes: The figure shows the evolution of the GSR term structure of forward rate volatility with t. The model fit to the original term structure of volatility ($t = 0$) was done solely through the mean reversion function $\varkappa(t)$, as described in the text.

The lesson to be learned from the examination above is essentially this: in the GSR model and, in fact, in all short rate models, one should resist making $\varkappa$ a function of time lest one is willing to accept strongly non-stationary evolution of forward rate volatilities. Yet, working with perfectly time-stationary GSR models is often too constraining in practical applications, as the resulting model has too few degrees of freedom in its volatility characteristics to calibrate against observed vanilla option prices in a meaningful way. Our recommendation for most applications is to freeze $\varkappa$ at a constant value, but to allow σ_r to be a function of time (see Section 13.1.8 for much more on mean reversion calibration). That is, we would set

$$\sigma_f(t,T) = \sigma_r(t)e^{-\varkappa(T-t)}.$$

While the resulting model is not time-stationary, it retains through time a persistent exponential shape of its instantaneous forward volatility structure.

The reader may at this point reasonably ask whether models exist that can produce a time-stationary hump in instantaneous forward volatilities. The answer is yes, but such models would generally need more than a single Markov variable to characterize moves in the yield curve. We return to this issue in Chapter 12 and, indeed, in many later chapters on multi-factor models.

10.1.3 European Option Pricing

In the general one-factor GSR model (10.14), suppose that we fix the mean reversion function $\varkappa(t)$ exogenously, e.g. based on empirical observations or from observation of typical decay speed of implied volatilities with option maturity[5]. The function $\vartheta(t)$ in (10.14) is then uniquely fixed by the initial forward curve, so to complete the specification of the model it remains to determine the function $\sigma_r(t)$. In pricing applications, this function is normally found by calibration of the GSR model to observed prices of liquid European options, such as caps and swaptions. While we shall postpone most of the intricacies of volatility calibration to later chapters, it should be clear that for a calibration to caps and swaptions to be efficient, we need computationally efficient methods for the valuation of these instruments.

In Section 4.5.1, we showed that for any Gaussian HJM model — whether the short rate is Markov or not — caplets can be priced by simple Black-Scholes formulas; see Proposition 4.5.2 for the details. Consequently, we here focus our attention on the pricing of swaptions. For concreteness, consider a payer swaption expiring at time T_0, with the underlying swap paying an annualized coupon c at times $T_1 < T_2 < \ldots < T_N$, with $T_1 > T_0$. We recall from Chapter 5 that the swaption payout at time T_0 is

$$V_{\text{swaption}}(T_0) = \left(1 - P(T_0, T_N) - c\sum_{i=0}^{N-1} \tau_i P(T_0, T_{i+1})\right)^+, \quad \tau_i = T_{i+1} - T_i. \tag{10.21}$$

10.1.3.1 The Jamshidian Decomposition

Our first approach is exact, and is based on a method developed by Jamshidian [1989]. The basic idea is to rewrite the swaption payout from an option on a sum of discount bonds to a sum of options on discount bonds. To develop the idea in detail, let us write $P(T_0, T_N) = P(T_0, T_N, x(T_0))$ to recognize the dependence of $P(T_0, T_N)$ on $x(T_0) = r(T_0) - f(0, T_0)$ through the reconstitution formula (10.18). We also define a "critical" value x^* for

[5] As argued above, normally we would pick $\varkappa(t)$ to be a constant. A more detailed examination of the estimation of mean reversions — and the role it plays in Bermudan swaption pricing — can be found in Chapter 13.

which the swap at time T_0 is exactly zero; x^* can be found by numerical root search on the equation

$$P(T_0, T_N, x^*) + c \sum_{i=0}^{N-1} \tau_i P(T_0, T_{i+1}, x^*) = 1. \tag{10.22}$$

Finally, define "strikes"

$$K_i = P(T_0, T_i, x^*), \quad i = 1, \ldots, N;$$

it follows that

$$K_N + c \sum_{i=0}^{N-1} \tau_i K_{i+1} = 1. \tag{10.23}$$

We are now ready to apply the Jamshidian "trick". Inspection of (10.18) shows that all zero-coupon bonds $P(T_0, T_i, x(T_0))$ are monotonically decreasing in $x(T_0)$, whereby the swaption only pays out a positive amount if $x(T_0) > x^*$. That is,

$$\begin{aligned}
&V_{\text{swaption}}(T_0) \\
&\quad = \left(1 - P(T_0, T_N, x(T_0)) - c \sum_{i=0}^{N-1} \tau_i P(T_0, T_{i+1}, x(T_0))\right) 1_{\{x(T_0) > x^*\}} \\
&\quad = \left(K_N + c \sum_{i=0}^{N-1} \tau_i K_{i+1} - P(T_0, T_N, x(T_0))\right. \\
&\qquad\qquad \left. - c \sum_{i=0}^{N-1} \tau_i P(T_0, T_{i+1}, x(T_0))\right) 1_{\{x(T_0) > x^*\}},
\end{aligned}$$

where the second equality follows from (10.23). Thus,

$$\begin{aligned}
V_{\text{swaption}}(T_0) &= (K_N - P(T_0, T_N, x(T_0))) 1_{\{x(T_0) > x^*\}} \\
&\quad + c \sum_{i=0}^{N-1} \tau_i (K_{i+1} - P(T_0, T_{i+1}, x(T_0))) 1_{\{x(T_0) > x^*\}} \\
&= (K_N - P(T_0, T_N, x(T_0)))^+ \\
&\quad + c \sum_{i=0}^{N-1} \tau_i (K_{i+1} - P(T_0, T_{i+1}, x(T_0)))^+ ,
\end{aligned} \tag{10.24}$$

where we used monotonicity of $P(T_0, T_i, x(T_0))$ on the last step. With this result, we have decomposed the swaption payout into $N + 1$ put options on zero-coupon bonds. Such options are easily valued using the formula from Proposition 4.5.1, allowing us to price the swaption in closed form.

10.1.3.2 Gaussian Swap Rate Approximation

While the Jamshidian approach above is perfectly adequate for many applications, its use of numerical root search and the need to price a potentially large amount of zero-coupon options can be cumbersome. One may wonder, then, whether perhaps a simpler approach is possible, given the simplicity of the dynamics of rates in the GSR framework. One obvious idea is to examine the SDE of the forward swap rate in an appropriate annuity measure, introducing approximations as needed to make the dynamics tractable. This idea shall be used many times in this book, often in combination with sophisticated techniques for simplification of the swap rate SDEs. Here, we have more modest aspirations and will be content with a simpler — yet still functional — approach. The reader shall consider this section a warm-up exercise for more accurate approximations to come, in particular in Sections 13.1.4 and 13.1.5 that also cover the GSR case.

We start by rewriting the swaption payout as

$$V_{\text{swaption}}(T_0) = A(T_0)\left(S(T_0) - c\right)^+,$$

where $A(t)$ and $S(t)$ are the swap annuity and forward swap rate, respectively, see (5.13)–(5.14):

$$A(t) \triangleq A_{0,N}(t) = \sum_{i=0}^{N-1} \tau_i P(t, T_{i+1}), \; S(t) \triangleq S_{0,N}(t) = \frac{P(t,T_0) - P(t,T_N)}{A(t)}.$$

Let Q^A be the measure induced by using $A(t)$ as the numeraire, such that

$$V_{\text{swaption}}(0) = A(0)\mathrm{E}^A\left(\left(S(T_0) - c\right)^+\right), \tag{10.25}$$

where E^A denotes expectation in measure Q^A. To evaluate (10.25), we need to determine the dynamics of $S(t)$ in Q^A. Lemma 4.2.4 establishes that $S(t)$ is a martingale under Q^A. From the reconstitution formula (10.18) we also know that $S(t)$ and $A(t)$ must be deterministic functions of $x(t)$:

$$S(t) = S\left(t, x(t)\right), \quad A(t) = A\left(t, x(t)\right),$$

so from Ito's lemma

$$dS(t) = q\left(t, x(t)\right)\sigma_r(t)\, dW^A(t), \quad q(t,x) = \frac{\partial}{\partial x}\frac{P(t,T_0,x) - P(t,T_N,x)}{\sum_{i=0}^{N-1}\tau_i P(t,T_{i+1},x)},$$

where W^A is a Q^A-Brownian motion and where we use (10.18) to express the $P(t,T_i)$'s as functions of x. Evaluating the partial derivatives yields

$$\begin{aligned} q(t,x) = &-\frac{P(t,T_0,x)G(t,T_0) - P(t,T_N,x)G(t,T_N)}{A(t,x)} \\ &+ \frac{S(t,x)}{A(t,x)}\sum_{i=0}^{N-1}\tau_i P(t,T_{i+1},x)G(t,T_{i+1}), \end{aligned} \tag{10.26}$$

where we recall that

$$G(t,T) = \int_t^T e^{-\int_t^u \varkappa(s)ds}\, du.$$

The function $q(t,x)$ can be experimentally verified to be close to a constant in x-direction so, as a good approximation, we can write

$$q(t,x(t)) \approx q(t,\overline{x}(t)), \tag{10.27}$$

where $\overline{x}(t)$ is some *deterministic* proxy for $x(t)$. With this, the option formula in the Normal model, see Remark 7.2.9, immediately leads to the following lemma:

Lemma 10.1.10. *Let $\overline{x}(t)$ be a deterministic function of time, and assume that (10.27) holds. Then*

$$V_{\text{swaption}}(0) \approx A(0)\left[(S(0)-c)\,\Phi(d) + \sqrt{v}\varphi(d)\right],$$

where

$$d = \frac{S(0)-c}{\sqrt{v}}, \quad v = \int_0^{T_0} q\left(t,\overline{x}(t)\right)^2 \sigma_r(t)^2\, dt. \tag{10.28}$$

It remains to choose $\overline{x}(t)$. An easy choice is to set $\overline{x}(t) = 0$, which will yield good precision if $\sigma_r(t)$ is not too high. What also works reasonably well is to simply evaluate $q(t,\overline{x})$ at the forward discount bond curve, i.e. replace $P(t,T_i,\overline{x})$ with $P(0,T_i)/P(0,t)$ in (10.26). More accurate choices for $\overline{x}(t)$, as well as refinements to the approximation (10.27), are developed in Sections 13.1.4 and 13.1.5.

10.1.4 Swaption Calibration

In a typical application of the model, the European option pricing formulas from Section 10.1.3 are used to calibrate the model, i.e. to find the volatility curve $\sigma_r(t)$ so as to match the market prices of one or more calibration targets, most often European swaptions.

Let us assume that we are given a collection of $N-1$ swaptions defined on a maturity grid $0 = T_0 < T_1 < \ldots < T_N$ such that the i-th swaption expires at times T_i, $i = 1,\ldots,N-1$. Such a collection[6] is often called a *swaption strip*. A common choice of swaption strip (used, for instance, for Bermudan swaptions) would have the underlying swaps for all swaptions

[6]Note that it is common to set $T_0 = 0$ when defining swaption strips, a convention that slightly clashes with the notation used above when deriving swaption formulas (where the swaption maturity $T_0 > 0$). We shall later, in Chapter 14, develop more formal notation for indexation, but for now trust the reader's ability to adapt generic swaption formulas to the swaption strip convention.

mature on the same date T_N. If this is the case, the strip is called the *coterminal* swaption strip.

With the mean reversion $\varkappa(t)$ fixed, we can make the important observation that the value of the swaption expiring at time T_i depends on the volatility curve $\sigma_r(s)$ for $s \in [0, T_i]$ *only*. This can be seen most clearly from the formula for v in Lemma 10.1.10, but is also evident from the pricing formula (10.24) and the fact that the discount bond reconstitution formula (10.18) for $P(t,T)$ does not depend on $\sigma_r(s)$ for $s \geq t$.

The special structure of volatility dependence allows us to perform calibration for one swaption at a time, replacing a potentially multi-dimensional optimization problem with a series of one-dimensional root searches. Assume that $\sigma_r(t)$ is piecewise flat on the maturity grid, with σ_i denoting the flat value on $[T_i, T_{i+1}]$. A possible algorithm based on the formula (10.24) would then work as follows.

1. Assume $\sigma_0, \ldots, \sigma_{i-1}$ have been found.
2. Set the value σ_i such that the model price of the $(i+1)$-th swaption, i.e. a swaption that expiries at T_{i+1}, is equal its market price, by numerically inverting (10.24) for σ_i while $\sigma_0, \ldots, \sigma_{i-1}$ are kept constant.
3. Repeat Step 2 for $i = 0, \ldots, N-2$.

At first glance, it may appear that the pricing formula from Lemma 10.1.10 will give rise to a linear system on $\sigma_0^2, \ldots, \sigma_{N-2}^2$, allowing us to execute Step 2 above by simple matrix inversion. The reality, however, is slightly more complex as the weight functions $q(\cdot,\cdot)$ also depend on the volatility curve $\sigma_r(\cdot)$ through $P(t,T)$'s dependence on $y(t)$ in (10.18). Nevertheless, even with the proper update of $y(t)$ through (10.17) in each step, the inversion in Step 2 above is simple fare for any one-dimensional root solver. Further details can be gleaned from Section 13.1.7 that discusses volatility calibration for the closely related quasi-Gaussian models.

We should note that the volatility calibration scheme above is not guaranteed to always work: a condition sometimes called a "volatility squeeze" may cause the inversion in Step 2 to fail if the market value of the T_{i+1}-expiry swaption is significantly below that of the swaption expiring at T_i. In practice, market data is rarely extreme enough for this to happen, and sometimes the problem can be cured by increasing the mean reversion speed $\varkappa(t)$. Some care must be exercised here, though, as the usage of unrealistically high mean reversions will impact the inter-temporal correlations of the model (see Chapter 13), which may lead to unrealistic prices for exotics options, as discussed in Chapter 18.

10.1.5 Finite Difference Methods

We round off our discussion of GSR models with some brief comments on numerical implementation. We start with finite difference methods here, and

turn to Monte Carlo applications in Section 10.1.6. Our discussion of both techniques is rather brief; for further analysis and alternatives we simply refer to Chapters 2 and 3.

10.1.5.1 PDE and Spatial Boundary Conditions

Our treatment of finite difference methods for the GSR model — and for short rate models in general, see Section 11.3.1 — essentially involves little outside of straightforward applications of schemes from Chapter 2. Still, let us start by noting that the algorithms we describe here nevertheless deviate quite markedly from the somewhat old-fashioned (and often suboptimal) tree-based schemes that abound in the short rate literature, even in recent work.

Consider a claim V with the terminal payout $V(T)$ that depends on the discount curve at time T. As the discount curve at time T can be reconstituted solely from knowledge of $x(T)$, we write $V(T) = V(T, x(T))$. By standard results (see Section 1.8), we write $V(t) = V(t, x(t))$, where $V(t, x)$ satisfies the PDE

$$\frac{\partial V}{\partial t} + (y(t) - \varkappa(t)x)\frac{\partial V}{\partial x} + \frac{1}{2}\sigma_r(t)^2\frac{\partial^2 V}{\partial x^2} = (x + f(0,t))\,V, \tag{10.29}$$

subject to a known terminal (payout) condition for $V(T, x)$. This PDE can be solved numerically using finite difference methods, e.g. the Crank-Nicolson method in Section 2.2.

In setting up the finite difference scheme, we require knowledge of spatial boundary conditions in the x-domain. In the absence of contractually agreed-upon boundary conditions (as would be the case for e.g. barrier options) one possibility is to set

$$\left.\frac{\partial^2 V}{\partial x^2}\right|_{x=x_{\min}} = \left.\frac{\partial^2 V}{\partial x^2}\right|_{x=x_{\max}} = 0, \tag{10.30}$$

as recommended in Section 2.2.2, where $x_{\max}$ and $x_{\min}$ are the grid boundaries. The boundaries are typically determined by probabilistic means, e.g.

$$x_{\max} = \mathrm{E}\left(x(T)\right) + \alpha\sqrt{\mathrm{Var}\left(x(T)\right)}, \quad x_{\min} = \mathrm{E}\left(x(T)\right) - \alpha\sqrt{\mathrm{Var}\left(x(T)\right)}, \tag{10.31}$$

for some confidence multiplier α. The moments required in this computation can be found from equations (10.12)–(10.13); see also (10.40)–(10.41).

While workable in practice, the specification (10.30) is not particularly accurate for many actual payout types. As a consequence, one often finds that α needs to be set quite large[7] (e.g. at values of 5–6, or larger) to

[7] An alternative approach that is advocated by some is to set the mean reversion $\varkappa$ to zero when determining $x_{\max}$ and $x_{\min}$, in which case α can be reduced.

prevent mis-specification errors at the boundary from affecting the solution at $(t, x) = (0, 0)$. This, in turn, implies that significant computational effort is spent in areas of the x-domain that are probabilistically insignificant. One way to improve on this situation is to rely on the PDE itself to generate boundary conditions, as described earlier in Section 2.2.2 (see also Section 9.4.4). We present the details of this idea in the next section.

10.1.5.2 Determining Spatial Boundary Conditions from PDE

We assume that the PDE (10.29) has been discretized on a spatial grid $\{x_j\}_{j=0}^{m+1}$, so that $V_j(t) = V(t, x_j)$, etc. Let us focus on establishing the boundary condition at $x_0 = x_{\min}$, say. Using a θ-method discretization scheme, as in Section 2.2, with an *upward* discretization of the x-derivatives we get, at some time step $[t, t+\delta]$,

$$\frac{V_0(t+\delta) - V_0(t)}{\delta} + \theta\mu(t, x_0)\frac{V_1(t) - V_0(t)}{x_1 - x_0} \tag{10.32}$$

$$+ (1-\theta)\,\mu(t+\delta, x_0)\frac{V_1(t+\delta) - V_0(t+\delta)}{x_1 - x_0}$$

$$+ \frac{\theta}{2}\sigma_r(t)^2\left\{\frac{V_2(t) - V_1(t)}{x_2 - x_1} - \frac{V_1(t) - V_0(t)}{x_1 - x_0}\right\}\frac{1}{\frac{1}{2}(x_2 - x_0)}$$

$$+ \frac{1-\theta}{2}\sigma_r(t+\delta)^2 \tag{10.33}$$

$$\times\left\{\frac{V_2(t+\delta) - V_1(t+\delta)}{x_2 - x_1} - \frac{V_1(t+\delta) - V_0(t+\delta)}{x_1 - x_0}\right\}\frac{1}{\frac{1}{2}(x_2 - x_0)}$$

$$= \theta\,(x_0 + f(0,t))\,V_0(t) + (1-\theta)\,(x_0 + f(0, t+\delta))\,V_0(t+\delta), \tag{10.34}$$

where $\mu(t, x) \triangleq y(t) - \varkappa(t)x$. This equation can be rearranged to write $V_0(t)$ as

$$V_0(t) = k_1(t)V_1(t) + k_2(t)V_2(t) + g_0\,(t+\delta)\,, \tag{10.35}$$

where $k_1(t)$ and $k_2(t)$ are easily computed functions of the process parameters, and where $g_0(t+\delta)$ is a function of $V_0(t+\delta)$, $V_1(t+\delta)$, and $V_2(t+\delta)$. We leave it to the reader to write out k_1, k_2, and g in detail. Applying similar principles, we get

$$V_{m+1}(t) = k_{m-1}(t)V_{m-1}(t) + k_m(t)V_m(t) + g_{m+1}\,(t+\delta)\,. \tag{10.36}$$

Comparing (10.35)–(10.36) with the equations (2.12)–(2.13), we see that the boundary conditions (10.35)–(10.36) can be incorporated into our usual tri-diagonal roll-back scheme by simply interpreting $\underline{f}(t, x_0) = g_0(t+\delta)$ and $\overline{f}(t, x_{m+1}) = g_{m+1}(t+\delta)$ in the scheme of Section 2.2. As we are rolling back in time (from $t+\delta$ to t) when using the finite difference equations, both $g_0(t+\delta)$ and $g_{m+1}(t+\delta)$ are known at time t, so this interpretation involves no difficulties.

10.1.5.3 Upwinding

For the PDE (10.29), notice that the condition (2.34) states that convection domination can cause spurious oscillations to creep into the finite difference scheme unless

$$|y(t) - \varkappa(t)x|\Delta_x \leq \sigma_r(t)^2, \tag{10.37}$$

for all x spanned by the finite difference grid. Since $\sigma_r(t)^2$ is typically a small number (around 0.001), it is not uncommon for this inequality to be violated at the edges of the finite difference grid (i.e. in the neighborhoods around x_0 and x_{m+1}) where the mean reversion pushes or pulls strongly at $x(t)$. To avoid numerical difficulties with the finite difference scheme, it is therefore recommended to apply the upwinding scheme in Section 2.6.1. While in principle this may reduce the spatial convergence order of the scheme, in practice the effect of upwinding on convergence is often minimal provided that the finite difference grid is dimensioned in such a way that (10.37) is only violated in a fairly small portion of the grid.

10.1.6 Monte Carlo Simulation

10.1.6.1 Exact Discretization

Consider the problem of pricing a derivative security that pays an amount $V(T)$ at time T, where $V(T)$ may be a function of the entire path of the discount curve over time interval $[0, T]$. Working in the risk-neutral measure, we are thus interested in computing

$$\begin{aligned} V(0) &= \mathrm{E}\left(V(T)e^{-\int_0^T r(u)du}\right) \\ &= P(0,T)\mathrm{E}\left(V(T;\{x(t) : 0 \leq t \leq T\})e^{-\int_0^T x(u)du}\right), \end{aligned} \tag{10.38}$$

where the second equality shifts variables to $x(t) = r(t) - f(0,t)$ and emphasizes the dependence of $V(T)$ on the entire path of $x(t)$. Recall from the discussion in connection with Proposition 10.1.7 that there are distinct advantages to working with the variable $x(t) = r(t) - f(0,t)$ rather than $r(t)$. In the GSR model, the dynamics for $x(t)$ are given by (10.16), i.e.

$$dx(t) = (y(t) - \varkappa(t)x(t))\,dt + \sigma_r(t)\,dW(t),\ y(t) = \int_0^t e^{-2\int_u^t \varkappa(u)du}\sigma_r(u)^2 du.$$

For the purpose of Monte Carlo pricing of (10.38), we discretize the time-interval into a schedule $t_0 < t_1 < \ldots < t_N$, with $t_0 = 0$ and $t_N = T$. The exact choice of the schedule depends on the particulars of the payout $V(T)$; if, say, $V(T)$ only depends on the yield curve at time T, it suffices to set $N = 1$. Now, we can solve the Gaussian SDE for $x(t)$ (see Section 1.6) to write

$$x(t_{i+1}) = e^{-\int_{t_i}^{t_{i+1}} \varkappa(u)du} x(t_i) + \int_{t_i}^{t_{i+1}} e^{-\int_s^{t_{i+1}} \varkappa(u)du} y(s)\, ds + \int_{t_i}^{t_{i+1}} e^{-\int_s^{t_{i+1}} \varkappa(u)du} \sigma_r(s)\, dW(s), \quad (10.39)$$

which we recognize as being a Gaussian random variable with moments

$$\mathrm{E}\left(x(t_{i+1})|x(t_i)\right) = e^{-\int_{t_i}^{t_{i+1}} \varkappa(u)du} x(t_i) + \int_{t_i}^{t_{i+1}} e^{-\int_s^{t_{i+1}} \varkappa(u)du} y(s)\, ds, \quad (10.40)$$

$$\mathrm{Var}\left(x(t_{i+1})|x(t_i)\right) = \int_{t_i}^{t_{i+1}} \left(e^{-\int_s^{t_{i+1}} \varkappa(u)du} \sigma_r(s)\right)^2 ds. \quad (10.41)$$

Advancement of $x(t)$ on the schedule can thus be done in bias-free manner, by writing

$$x(t_{i+1}) = \mathrm{E}\left(x(t_{i+1})|x(t_i)\right) + \sqrt{\mathrm{Var}\left(x(t_{i+1})|x(t_i)\right)} Z_i, \quad i = 0, \ldots, N-1,$$

where $Z_0, \ldots, Z_{N-1}$ is a sequence of independent standard Gaussian random variables.

For every date on the simulated path $x(t_0)$, ..., $x(t_N)$, we can use the reconstitution formula in Proposition 10.1.7 to reconstitute the entire discount curve, in turn allowing us to compute $V(T)$ on the path. To evaluate (10.38), it remains to simulate the quantity

$$I(T) = -\int_0^T x(u)\, du$$

on the path. Given $x(t_0)$, ..., $x(t_N)$, an obvious choice would be to compute $I(T)$ by quadrature (e.g. trapezoidal integration, or similar). As this inevitably introduces a discretization bias (see Andersen and Boyle [2000] for more analysis), it is preferable to use the following result.

Lemma 10.1.11. *Let $G(t,T)$ be as in Proposition 10.1.7. Given $I(t_i)$ and $x(t_i)$, $I(t_{i+1})$ is Gaussian with moments*

$$\mathrm{E}\left(I(t_{i+1})|I(t_i), x(t_i)\right) = I(t_i) - x(t_i)G(t_i, t_{i+1}) - \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} e^{-\int_s^u \varkappa(v)dv} y(s)\, ds\, du, \quad (10.42)$$

and

$$\mathrm{Var}\left(I(t_{i+1})|I(t_i), x(t_i)\right) = 2\int_{t_i}^{t_{i+1}} \int_{t_i}^{u} e^{-\int_s^u \varkappa(v)dv} y(s)\, ds\, du - y(t_i)G(t_i, t_{i+1})^2. \quad (10.43)$$

Also, we have

$$\begin{aligned}&\mathrm{Cov}\left(x(t_{i+1}), I(t_{i+1})|I(t_i), x(t_i)\right)\\&\qquad = \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} \sigma_r(s)^2 e^{-\int_s^u \varkappa(v)dv} e^{-\int_s^{t_{i+1}} \varkappa(v)dv} ds\, du.\end{aligned}$$

Proof. Straightforward but tedious calculations for Gaussian random variables. □

Over the time step $[t_i, t_{i+1}]$ we advance $I(t)$ according to the formula

$$I(t_{i+1}) = \mathrm{E}\left(I(t_{i+1})|I(t_i), x(t_i)\right) + \sqrt{\mathrm{Var}\left(I(t_{i+1})|I(t_i), x(t_i)\right)}\widetilde{Z}_i,$$

where $\widetilde{Z}_0, \ldots, \widetilde{Z}_{N-1}$ is a sequence of independent standard Gaussian random variables, and where the required moments of $I(t_{i+1})$ can be found in Lemma 10.1.11. To honor the covariance between the $x(t)$ and $I(t)$ processes, we require that the variables Z_i and $\widetilde{Z}_i$ are correlated:

$$\mathrm{Corr}\left(Z_i, \widetilde{Z}_i\right) = \frac{\mathrm{Cov}\left(I(t_{i+1}), x(t_{i+1})|I(t_i), x(t_i)\right)}{\sqrt{\mathrm{Var}\left(I(t_{i+1})|I(t_i), x(t_i)\right)}\sqrt{\mathrm{Var}\left(x(t_{i+1})|I(t_i), x(t_i)\right)}}.$$

As explained in Chapter 3, correlated Gaussian samples can be generated from uncorrelated samples through the Cholesky decomposition.

The scheme outlined above allows us to simulate bias-free paths of the variables $x(t)$ and $I(t)$, which in turn allows us to compute independent, unbiased samples of $V(T)e^{I(T)}$. Monte Carlo estimation of the expectation for $V(0)$ can then be performed in standard Monte Carlo fashion, by forming sample averages. The discretization scheme involves several time-integrals over dates in the observation schedule, many of them nested; it goes without saying that these integrals should be pre-computed before actual path simulations commence.

10.1.6.2 Approximate Discretization

For a quick-and-dirty implementation of the Gaussian model, we may elect to skip the algorithm in the previous section and instead apply one of the approximate discretization schemes in Section 3.2. As a starting point, we have the vector SDE

$$d\begin{pmatrix} x(t) \\ I(t) \end{pmatrix} = \begin{pmatrix} y(t) - \varkappa(t)x(t) \\ -x(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_r(t) \\ 0 \end{pmatrix} dW(t).$$

A plain-vanilla Euler scheme from Section 3.2.3 would write

$$\begin{pmatrix} \widehat{x}(t_{i+1}) \\ \widehat{I}(t_{i+1}) \end{pmatrix} = \begin{pmatrix} \widehat{x}(t_i) \\ \widehat{I}(t_i) \end{pmatrix} + \begin{pmatrix} y(t_i) - \varkappa(t_i)\widehat{x}(t_i) \\ -\widehat{x}(t_i) \end{pmatrix} \Delta_i + \begin{pmatrix} \sigma_r(t_i) \\ 0 \end{pmatrix} Z_i\sqrt{\Delta_i},$$

where $\Delta_i = t_{i+1} - t_i$ and Z_i is a standard Gaussian random variable. Unless $\varkappa$ is small, this scheme *cannot* be recommended due to the stability issues discussed in Section 3.2.3. As explained in Section 3.2.3.1, it is preferable to incorporate the fact that

$$\begin{aligned}\mathrm{E}\left(x(t_{i+1})|x(t_i)\right) &= e^{-\int_{t_i}^{t_{i+1}} \varkappa(u)du} x(t_i) + \int_{t_i}^{t_{i+1}} e^{-\int_s^{t_{i+1}} \varkappa(u)du} y(s)\,ds \\ &\approx e^{-\varkappa(t_i)(t_{i+1}-t_i)} x(t_i) + \frac{1 - e^{\varkappa(t_i)(t_{i+1}-t_i)}}{\varkappa(t_i)} y(t_i).\end{aligned}$$

That is, we write

$$\begin{pmatrix} \widehat{x}(t_{i+1}) \\ \widehat{I}(t_{i+1}) \end{pmatrix} = \begin{pmatrix} e^{-\varkappa(t_i)\Delta_i} x(t_i) + \frac{1-e^{\varkappa(t_i)\Delta_i}}{\varkappa(t_i)} y(t_i) \\ \widehat{I}(t_i) - \widehat{x}(t_i)\Delta_i \end{pmatrix} + \begin{pmatrix} \sigma_r(t_i) \\ 0 \end{pmatrix} Z_i \sqrt{\Delta_i}.$$

This scheme has first-order (weak) convergence. Higher-order schemes can be found in Section 3.2, but are essentially obsolete here: if truly low bias is critically important, we should use the unbiased scheme from Section 10.1.6.1.

10.1.6.3 Using other Measures for Simulation

The need to simulate $I(t)$ can be avoided entirely by a suitable change of the probability measure. Switching to the terminal measure Q^T (see Section 4.2.4), we rewrite (10.38) as

$$V(0) = P(0,T)\,\mathrm{E}^T\left(V(T)\right),$$

and observe that we now need to simulate $x(t)$ only in order to calculate the payoff. The dynamics of $x(t)$ under the terminal measure Q^T follow from (4.34),

$$\begin{aligned} dx(t) &= \left(y(t) - \varkappa(t)x(t)\right)dt + \sigma_r(t)\left(dW^T(t) - \sigma_P(t,T)\,dt\right) \\ &= \left(y(t) - \sigma_r(t)^2 G(t,T) - \varkappa(t)x(t)\right)dt + \sigma_r(t)\,dW^T(t), \end{aligned}$$

with $W^T(t)$ being a Q^T-Brownian motion. The dynamics remain Gaussian and Markov under the terminal measure, and hence $x(t)$ can be simulated bias-free on the time grid $\{t_i\}_{i=1}^N$.

An alternative to the terminal measure that is "closer" in some ways to the risk-neutral measure, yet still allows one to avoid the simulation of $I(t)$, is the spot measure from Section 4.2.3. We recall that this measure is associated with the discretely compounded money market account $B(t)$,

$$B(t) = P(t,t_{i+1}) \prod_{n=0}^{i} \frac{1}{P\left(t_n, t_{n+1}\right)}, \quad t \in \left(t_i, t_{i+1}\right].$$

Under the spot measure Q^B, we obtain

$$V(0) = \mathrm{E}^B \left(\prod_{n=0}^{N-1} P\left(t_n, t_{n+1}, x(t_n)\right) V(T) \right),$$

where we explicitly indicated the dependence of the discount bond on the state process $x(t)$. Notice that the random variable under the expectation operator is a function of $x(t)$ on the grid $\{t_i\}_{i=1}^N$ only. Moreover, over the interval $[t_n, t_{n+1}]$, the measure Q^B coincides with the T_{n+1}-forward measure, which gives us the dynamics of $x(t)$,

$$\begin{aligned} dx(t) = \left(y(t) - \sigma_r(t)^2 G\left(t, t_{n+1}\right) - \varkappa(t)x(t)\right) dt \\ + \sigma_r(t)\, dW^B(t), \quad t \in \left(t_n, t_{n+1}\right], \end{aligned}$$

with $W^B(t)$ a Q^B-Brownian motion. Again, we can generate a sample of $x(t_{n+1})$ from $x(t_n)$ in a bias-free manner. We refer the reader to Chapter 14 for more details on numeraire simulation strategies.

10.2 The Affine One-Factor Model

Earlier (in Section 10.1.1.3) we identified the non-zero probability of negative interest rates as one of the drawbacks of the one-factor GSR model. Another problem is the lack of interest rate dependence in the GSR short rate volatility, leaving the user with no means of controlling the volatility skew implied by the model. While there are different ways of addressing these issues, one type of model that can, in part at least, address both of these shortcomings of the GSR model is the *affine short rate model.* This model — or, rather, model *class* — constitutes a significant extension of the GSR model (which in fact is a member of the affine class), yet retains a high degree of analytical tractability. Originally introduced by Duffie and Kan [1996], the affine class of short rate models enjoys high popularity among practitioners and academics alike, particularly for econometric work. The affine models are also quite useful for derivatives pricing, although ultimately the constraints one need to impose on diffusion dynamics can be too strong for some applications.

10.2.1 Basic Definitions

10.2.1.1 SDE

Consider a time-homogeneous one-factor short rate process of the form

$$dr(t) = \varkappa\left(\vartheta - r(t)\right) dt + \sigma v\left(r(t)\right) dW(t), \tag{10.44}$$

where $W(t)$ is a Brownian motion in the risk-neutral measure, $\varkappa > 0$, $\sigma > 0$, and ϑ are constants, and $v(\cdot) : \mathbb{R} \to \mathbb{R}$ is a deterministic function of the level of the short rate. We notice that the drift of (10.44) is affine, i.e. linear, in $r(t)$. If the square of the diffusion term in (10.44) is also affine, we say that (10.44) is a time-homogeneous *affine one-factor short rate model*. Evidently, the function $v(r)$ is thus limited to the form

$$v(r) = \sqrt{\alpha + \beta r}, \tag{10.45}$$

for constants α and β. We notice that the special case of $\beta = 0$ produces the GSR model of Section 10.1.2.2, whereas the case $\alpha = 0$ produces a square-root type model similar to those encountered, for stochastic volatility modeling, in Chapter 8. The case $\alpha = 0$, $\beta = 1$ was first studied by Cox et al. [1985] and is known as the *Cox-Ingersol-Ross* (CIR) model.

10.2.1.2 Regularity Issues

Not all combinations of parameters in (10.44) and (10.45) produce a well-defined SDE. If $\beta = 0$ for all t, we must require that $\alpha \geq 0$ for all t to ensure that $v(r(t))$ is defined. If $\beta \neq 0$, for the square root in (10.45) to exist, we must ensure that the drift term in (10.44) has the same sign as β whenever $\alpha + \beta r(t) = 0$. That is,

$$\varkappa\beta\left(\vartheta + \alpha/\beta\right) \geq 0, \quad \beta \neq 0, \tag{10.46}$$

for all $t \geq 0$. Notice that if we wish for the volatility term in (10.45) to be strictly positive ($\alpha + \beta r(t) > 0$), we need to replace this condition with the stronger Feller condition (recall Proposition 8.3.1)

$$\varkappa\beta\left(\vartheta + \alpha/\beta\right) \geq \frac{1}{2}\beta^2\sigma^2.$$

For the CIR model the requirement that $r(t)$ stays strictly positive can be seen to translate into the classical condition $2\varkappa\vartheta \geq \sigma^2$.

For the purposes of modeling interest rates, it is most reasonable to assume that $\varkappa > 0$ to ensure that rates are mean-reverting rather than mean-fleeing, and that $\beta \geq 0$. In this case, the domain of the short rate becomes

$$r(t) \in [-\alpha/\beta, \infty), \quad \beta > 0, \tag{10.47}$$

and $r(t) \in (-\infty, \infty)$ for the case $\beta = 0$, $\alpha > 0$ (Gaussian model). Evidently, to keep $r(t)$ non-negative for all t, we need to set $\alpha \leq 0$, subject to the restriction that $-\alpha/\beta \leq r(0)$.

10.2.1.3 Volatility Skew

The parameters α and β in (10.44) effectively determine the volatility skew behavior of the affine model. If both parameters are non-negative, the affine

model can generate skews ranging from a Gaussian process ($\alpha \gg \beta$) to a square-root process ($\alpha \ll \beta$). In the usual language, for non-negative α, β, the skew "power" of the affine model thus ranges from 0 to 0.5. By allowing α to be negative, effective skew powers above 0.5 are possible, although the allowed range of the underlying process $r(t)$ will then be floored at some positive level, which may have undesirable side effects if α/β is not close to zero.

10.2.1.4 Time-Dependent Parameters

The SDE (10.44) does not depend on time and hence will generally not match the initial yield curve at time 0. As we did for the Gaussian model, we may extend the SDE to have time-dependent parameters, e.g.

$$dr(t) = \varkappa(t)\,(\vartheta(t) - r(t))\,dt + \sigma(t)\sqrt{\alpha + \beta r(t)}\,dW(t). \tag{10.48}$$

Notice that we have not introduced time-dependence in α and β, leaving the domain (10.47) unchanged[8]. Not all functions $\varkappa, \vartheta, \sigma$ produce a well-defined SDE; for instance, if $\varkappa(t)$ is positive (which is always the case in practice) and $\beta > 0$, then (10.46) shows that we need

$$\varkappa(t)\beta\vartheta(t) \geq -\alpha \tag{10.49}$$

in order for (10.48) to be well-defined.

Remark 10.2.1. For generality we allow $\varkappa(t)$ to be a function of time throughout. As argued in Section 10.1.2.3, however, it is often most reasonable to let $\varkappa(t)$ be a constant.

10.2.2 Discount Bond Pricing and Extended Transform

Starting from the time-dependent SDE (10.48), we now turn to the search for a discount bond reconstitution formula, i.e. a formula that allows us to compute the risk-neutral expectation

$$P(t,T) = \mathrm{E}_t\left(e^{-\int_t^T r(u)du}\right), \tag{10.50}$$

as an explicit function of $r(t)$. Rather than directly attacking (10.50), we turn to the more general problem of establishing the so-called *extended transform* $g(t,T;c_1,c_2)$ defined by

$$g(t,T;c_1,c_2) = \mathrm{E}_t\left(e^{-c_1 r(T) - c_2\int_t^T r(u)du}\right), \quad c_1, c_2 \in \mathbb{C}. \tag{10.51}$$

[8]The results of the next sections do, however, often generalize to time-dependence in α and β. See Remark 10.2.3, for example.

Notice how this generalizes the idea of the moment-generating function from Chapters 8 and 9. Also note that the knowledge of g allows us to find discount bond prices as a special case,

$$P(t,T) = g(t,T;0,1).$$

For the values of c_1 and c_2 for which g exists, we can use the following result, which is an extension of Proposition 9.1.2.

Proposition 10.2.2. *For the model (10.48), whenever the extended transform g in (10.51) is defined, it is given by*

$$g(t,T;c_1,c_2) = \exp\left(A(t,T;c_1,c_2) - B(t,T;c_1,c_2)r(t)\right), \tag{10.52}$$

where A and B satisfy a system of Riccati ODEs

$$\frac{dA}{dt} - \varkappa(t)\vartheta(t)B + \frac{1}{2}\sigma(t)^2\alpha B^2 = 0, \tag{10.53}$$

$$-\frac{dB}{dt} + \varkappa(t)B + \frac{1}{2}\sigma(t)^2\beta B^2 = c_2, \tag{10.54}$$

subject to the terminal conditions $A(T;T,c_1,c_2) = 0$, $B(T;T,c_1,c_2) = c_1$.

Proof. Follows that of Proposition 9.1.2 closely. □

Remark 10.2.3. If the parameters α and β are functions of time, Proposition 10.2.2 continues to hold if we simply replace α, β with $\alpha(t)$, $\beta(t)$ in the Riccati equations for A and B.

Proposition 10.2.2 establishes that the joint characteristic function of $r(T)$ and $\int_0^T r(u)\,du$ is known analytically for the affine model, a result that accounts for much of its popularity in the financial literature. Solution of the Riccati equations (10.53)–(10.54) can be done quickly and robustly by any number of standard ODE schemes, such as the Runge-Kutta method (see Press et al. [1992]). For the case where parameters are piecewise constant in time, establishing A and B in Proposition 10.2.2 can also be done analytically; see Section 10.2.2.1 below.

10.2.2.1 Constant Parameters

We now turn to establishing the extended transform g for the special case where all parameters in (10.48) are constants. As a warm-up case, we first list a result for the CIR case.

Proposition 10.2.4. *Consider the CIR model*

$$dr(t) = \varkappa(\vartheta - r(t))\,dt + \sigma\sqrt{r(t)}\,dW(t), \tag{10.55}$$

and let $g(t,T;c_1,c_2)$ *be defined as in (10.51). Set*

$$\gamma = \gamma(c_2,\sigma) = \sqrt{\varkappa^2 + 2\sigma^2 c_2}.$$

Then

$$g(t,T;c_1,c_2) = \exp\left(A_{\mathrm{CIR}}(t,T;\vartheta,\sigma,c_1,c_2) - B_{\mathrm{CIR}}(t,T;\vartheta,\sigma,c_1,c_2)r(t)\right),$$

where

$$\begin{aligned} A_{\mathrm{CIR}}(t,T;\vartheta,\sigma,c_1,c_2) &= \varkappa\vartheta\sigma^{-2}\left(\varkappa+\gamma(c_2,\sigma)\right)(T-t) \\ &\quad - 2\varkappa\vartheta\sigma^{-2}\ln\left(1+\frac{\left(\varkappa+\gamma(c_2,\sigma)-c_1\sigma^2\right)\left(e^{\gamma(c_2,\sigma)(T-t)}-1\right)}{2\gamma(c_2,\sigma)}\right), \end{aligned}$$

and

$$\begin{aligned} &B_{\mathrm{CIR}}(t,T;\vartheta,\sigma,c_1,c_2) = \\ &\quad \frac{(2c_2-\varkappa c_1)\left(1-e^{-\gamma(c_2,\sigma)(T-t)}\right)+\gamma(c_2,\sigma)c_1\left(1+e^{-\gamma(c_2,\sigma)(T-t)}\right)}{\left(\varkappa+\gamma(c_2,\sigma)+c_1\sigma^2\right)\left(1-e^{-\gamma(c_2,\sigma)(T-t)}\right)+2\gamma(c_2,\sigma)e^{-\gamma(c_2,\sigma)(T-t)}}. \end{aligned}$$

Proof. The result is a small extension of Proposition 8.3.8, and follows by direct solution of the ODEs (10.53)–(10.54). □

Armed with this result, it is straightforward to extend it to the general constant-parameter case

$$dr(t) = \varkappa(\vartheta - r(t))\,dt + \sigma\sqrt{\alpha+\beta r(t)}\,dW(t), \quad \beta > 0. \tag{10.56}$$

In particular, we notice that if $y(t) = \alpha + \beta r(t)$, then $y(t)$ follows the SDE

$$dy(t) = \beta\,dr(t) = \varkappa(\beta\vartheta + \alpha - y(t))\,dt + \beta\sigma\sqrt{y(t)}\,dW(t),$$

which is of the form (10.55). We also have

$$\begin{aligned} g(t,T;c_1,c_2) &= \mathrm{E}_t\left(\exp\left(-c_1 r(T) - c_2\int_t^T r(u)du\right)\right) \\ &= \mathrm{E}_t\left(\exp\left(-c_1\left(\frac{y(T)-\alpha}{\beta}\right) - c_2\int_t^T\left(\frac{y(u)-\alpha}{\beta}\right)du\right)\right) \\ &= e^{c_1\alpha/\beta}e^{c_2\alpha(T-t)/\beta}\mathrm{E}_t\left(\exp\left(-\frac{c_1}{\beta}y(T) - \frac{c_2}{\beta}\int_t^T y(u)du\right)\right). \end{aligned}$$

The expectation involved in the last equality can here be evaluated directly from Proposition 10.2.4, leading to the following lemma.

Lemma 10.2.5. *The extended transform for the constant parameter affine model (10.56) is*

$$g(t,T;c_1,c_2) = \exp\left(A(t,T;c_1,c_2) - B(t,T;c_1,c_2)r(t)\right),$$

where

$$\begin{aligned} A(t,T;c_1,c_2) &= c_1\alpha/\beta + c_2\alpha(T-t)/\beta + A_{\mathrm{CIR}}\left(t,T;\beta\vartheta+\alpha,\beta\sigma,\frac{c_1}{\beta},\frac{c_2}{\beta}\right) \\ &\quad - \alpha B_{\mathrm{CIR}}\left(t,T;\beta\vartheta+\alpha,\beta\sigma,\frac{c_1}{\beta},\frac{c_2}{\beta}\right), \\ B(t,T;c_1,c_2) &= \beta B_{\mathrm{CIR}}\left(t,T;\beta\vartheta+\alpha,\beta\sigma,\frac{c_1}{\beta},\frac{c_2}{\beta}\right), \end{aligned}$$

and the functions A_{CIR} and B_{CIR} are given in Proposition 10.2.4.

10.2.2.2 Piecewise Constant Parameters

We can use the results established in Section 10.2.2.1 to compute extended transforms for the case where we are given a time grid $0 = t_0 < t_1 < t_2 < \ldots$, on which all model parameters $\varkappa$ and σ can be assumed piecewise constant. The resulting recursive routine is a robust and efficient[9] alternative to Runge-Kutta solvers.

For simplicity of notation, let us define $g(t_i,t_j;c_1,c_2) = g_{i,j}(c_1,c_2)$, $A(t_i,t_j;c_1,c_2) = A_{i,j}(c_1,c_2)$, and so on. Then, from Proposition 10.2.2,

$$g_{i,j}(c_1,c_2) = e^{A_{i,j}(c_1,c_2) - r(t_i)B_{i,j}(c_1,c_2)}, \quad j > i, \tag{10.57}$$

and, using the law of iterated conditional expectations,

$$\begin{aligned} g_{i-1,j}(c_1,c_2) &= \mathrm{E}_{t_{i-1}}\left(e^{-c_1 r(t_i) - c_2\int_{t_{i-1}}^{t_i} r(u)du}\right) \\ &= \mathrm{E}_{t_{i-1}}\left(\mathrm{E}_{t_i}\left(e^{-c_1 r(t_i) - c_2\int_{t_{i-1}}^{t_i} r(u)du}\right)\right) \\ &= \mathrm{E}_{t_{i-1}}\left(e^{-c_2\int_{t_{i-1}}^{t_i} r(u)du}\mathrm{E}_{t_i}\left(e^{-c_1 r(t_j) - c_2\int_{t_i}^{t_j} r(u)du}\right)\right) \\ &= \mathrm{E}_{t_{i-1}}\left(e^{-c_2\int_{t_{i-1}}^{t_i} r(u)du} g_{i,j}(c_1,c_2)\right). \end{aligned}$$

Inserting (10.57) into the last equation then yields

[9] As pointed out in Section 9.1, depending on the level of accuracy required, the Runge-Kutta numerical solution of the ODEs can sometimes have higher computational efficiency.

$$g_{i-1,j}(c_1,c_2) = \mathrm{E}_{t_{i-1}}\left(e^{-c_2\int_{t_{i-1}}^{t_i} r(u)du} e^{A_{i,j}(c_1,c_2)-r(t_i)B_{i,j}(c_1,c_2)}\right)$$
$$= e^{A_{i,j}(c_1,c_2)} g_{i-1,i}\left(B_{i,j}\left(c_1,c_2\right),c_2\right).$$

Applying (10.57) to the right-hand side of this equation leads to

$$e^{A_{i-1,j}(c_1,c_2)-r(t_{i-1})B_{i-1,j}(c_1,c_2)}$$
$$= e^{A_{i,j}(c_1,c_2)} e^{A_{i-1,i}(B_{i,j}(c_1,c_2),c_2)-r(t_{i-1})B_{i-1,i}(B_{i,j}(c_1,c_2),c_2)},$$

or, finally,

$$A_{i-1,j}\left(c_1,c_2\right) = A_{i,j}\left(c_1,c_2\right) + A_{i-1,i}\left(B_{i,j}\left(c_1,c_2\right),c_2\right), \tag{10.58}$$
$$B_{i-1,j}(c_1,c_2) = B_{i-1,i}\left(B_{i,j}\left(c_1,c_2\right),c_2\right). \tag{10.59}$$

As parameters are constant on the time grid, the functions $A_{i-1,i}$ and $B_{i-1,i}$ can be computed in closed form from the results of Lemma 10.2.5. For a fixed j, (10.58)–(10.59) can be used in backward fashion to establish $A_{i,j}$ and $B_{i,j}$ for $i = j-1, j-2, \ldots, 0$; the recursion starts with an application of Lemma 10.2.5 to compute $A_{j-1,j}$ and $B_{j-1,j}$.

10.2.3 Discount Bond Calibration

10.2.3.1 Change of Variables

In the affine SDE (10.48), the role of the mean reversion level $\vartheta(t)$ is to calibrate the model to the initial term structure of discount bonds. As we discussed in the context of the GSR model, $\vartheta(t)$ will depend on the derivative $\partial f(0,t)/\partial t$ which may, for many curve construction algorithms, be irregular. For practical applications of affine models, it is therefore strongly recommended to follow the advice of Section 10.1.2.2 and rewrite the model in terms of a variable that measures the difference between $r(t)$ and $f(0,t)$. Let this variable be $x(t)$, defined as

$$x(t) = r(t) - f(0,t).$$

The SDE for $x(t)$ becomes

$$dx(t) = dr(t) - \frac{\partial f(0,t)}{\partial t}dt$$
$$= \left(\omega(t) - \varkappa(t)x(t)\right)dt + \sigma(t)\sqrt{\xi(t)+\beta x(t)}\,dW(t), \tag{10.60}$$

where $x(0) = 0$, $\xi(t) = \alpha + \beta f(0,t)$, and

$$\omega(t) = \varkappa(t)\vartheta(t) - \partial f(0,t)/\partial t - \varkappa(t)f(0,t).$$

The deterministic function $\omega(t)$ is likely to be smooth even if the forward curve is not.

Written in terms of $x(t)$, the extended transform in Proposition 10.2.2 becomes

$$\begin{aligned} g(t,T;c_1,c_2) =& e^{-c_1 f(0,T)-c_2\int_t^T f(0,u)du}\mathrm{E}_t\left(e^{-c_1 x(T)-c_2\int_t^T x(u)du}\right) \\ =& e^{-c_1 f(0,T)}\frac{P(0,T)^{c_2}}{P(0,t)^{c_2}}\exp(C(t,T;c_1,c_2)-x(t)B(t,T;c_1,c_2)), \end{aligned} \tag{10.61}$$

where B solves (10.54) and C can, after suitable translation of the results in Proposition 10.2.2 to the process (10.60), be written as the solution to the Riccati ODE:

$$\frac{dC}{dt}-\omega(t)B+\frac{1}{2}\sigma(t)^2\xi(t)B^2=0. \tag{10.62}$$

10.2.3.2 Algorithm for $\omega(t)$

We now assume (but see Section 10.2.5) that α and β have been fixed, and that $\varkappa(t)$ and $\sigma(t)$ are known for all values of $t\geq 0$. In the SDE (10.60) for $x(t)$, it only remains to establish the function $\omega(t)$, which shall be done to match observed discount bond prices at time 0.

To make matters more concise, let us set $b(t,T)=B(t,T;0,1)$ and $c(t,T)=C(t,T;0,1)$ such that, from the definition of $C(t,T)$,

$$P(t,T)=g\,(t,T;0,1)=\frac{P(0,T)}{P(0,t)}e^{c(t,T)-x(t)b(t,T)}. \tag{10.63}$$

The functions b and c obviously satisfy a Riccati system,

$$\frac{dc}{dt}-\omega(t)b+\frac{1}{2}\sigma(t)^2\xi(t)b^2=0, \tag{10.64}$$

$$-\frac{db}{dt}+\varkappa(t)b+\frac{1}{2}\sigma(t)^2\beta b^2=1, \tag{10.65}$$

where $c(T,T)=b(T,T)=0$.

Setting $t=0$ in equation (10.63) establishes the fundamental calibration requirement that $c(0,T)=c(0,T;\omega(\cdot))=0$ for all T which, combined with (10.64), defines a so-called *Volterra integral equation* for $\omega(\cdot)$. We can solve it on a time grid $t_0<t_1<t_2<\ldots<t_N$ by iterative bootstrapping of the equation $c(0,t_i;\omega(\cdot))=0$. Assuming that $\omega(\cdot)$ is piecewise constant at a level ω_i over the time bucket $(t_i,t_{i+1}]$, we can use the following algorithm.

1. As a pre-processing step, find $b(t_i,t_j)$ for all i,j, $j>i$, by solving (10.65). This does not depend on $\omega(\cdot)$.
2. For a given i, assume that ω_j is known for $j<i$.
3. Compute $\Theta(t_i)=\frac{1}{2}\int_0^{t_{i+1}}\sigma(s)^2\xi(s)b(s,t_{i+1})^2\,ds-\int_0^{t_i}\omega(s)b(s,t_{i+1})\,ds$.

4. Compute ω_i as the solution to $\Theta(t_i) - \omega_i \int_{t_i}^{t_{i+1}} b(s, t_{i+1})\, ds = 0$.
5. Repeat steps 2–4 for all $i = 0, 1, \ldots, N-1$.

Notice that no numerical root search is needed and that the computational complexity of the scheme is $O(N^2)$. By modifying Steps 3 and 4, other interpolation techniques can be supported, although stability issues might come into play. See also Press et al. [1992] for more general schemes to solve Volterra equations.

We should note that there may be cases where the algorithm above will fail, in the sense that the basic regularity condition (10.49) will prevent a valid solution for $\omega(\cdot)$ from existing. This is a fundamental issue with non-Gaussian affine short rate models, but is rarely observed as very strongly downward-sloping yield curves are required to trigger the problem (see the discussion in Hull and White [1994a]).

10.2.4 European Option Pricing

The short rate volatility function $\sigma(t)$ in the affine model (10.60) will normally be determined through calibration against swaptions and caps/floors. For such calibration to be computationally feasible it is, of course, important to establish fast methods for pricing European interest rate options.

For simple options such as caplets or, equivalently, options on zero-coupon bonds, the availability of the moment-generating function for the logarithm of the bond (see Proposition 10.2.2) allows for application of the Fourier methods[10] of Section 8.4. Extensions to swaption pricing through the Jamshidian approach of Section 10.1.3.1 is possible in principle, but the need to perform Fourier integration of a large number of Riccati ODE solutions makes this approach impractical. Several approximation techniques have been proposed in the literature; see, for instance, Collin-Dufresne and Goldstein [2002a] for a survey and details on a method based on Gram-Charlier expansions. Our preferred approach to swaption pricing in the affine model borrows the techniques of Section 10.1.3.2 to work out an approximation for the swap rate martingale dynamics. We shall outline one straightforward and quite accurate approach here; as was the case for the GSR model, we again will stop short of the full-blown projection techniques that will be introduced later in this book for more realistic candidates for actual trading applications.

Let us, as in Section 10.1.3.2, start out by rewriting the swaption payout as

$$V_{\text{swaption}}(T_0) = A(T_0)\,(S(T_0) - c)^+\,, \tag{10.66}$$

where

[10] For time-homogeneous models, closed-form pricing formulas for options on discount bonds exist for some models, including the CIR model (see Cox et al. [1985]).

$$A(t) = \sum_{i=0}^{N-1} \tau_i P(t, T_{i+1}), \quad S(t) = \frac{P(t, T_0) - P(t, T_N)}{A(t)}.$$

Let Q^A be the measure induced by using $A(t)$ as the numeraire; in this measure $S(t)$ is a martingale. By the reconstitution result (10.63) we have

$$dS(t) = \frac{\partial S(t)}{\partial x} \sigma(t) \sqrt{\xi(t) + \beta x(t)}\, dW^A(t), \tag{10.67}$$

where $W^A(t)$ is a Q^A-Brownian motion and

$$\frac{\partial S(t)}{\partial x} = -\frac{b(t, T_0)P(t, T_0) - b(t, T_N)P(t, T_N)}{A(t)} + \frac{S(t)}{A(t)} \sum_{i=0}^{N-1} \tau_i b(t, T_{i+1}) P(t, T_{i+1}).$$

The dynamics (10.67) are generally intractable, but $S(t)$ can — as was the case for the GSR model — be verified to often be well approximated by a linear function of $x(t)$, with slope and intercept being functions of time. Using a Taylor expansion around some point $\overline{x}$ (e.g. $\overline{x} = 0$, but see the discussion in Section 10.1.3.2), we can find $\zeta(t)$, $\chi(t)$ such that

$$S(t) \approx \zeta(t) + \chi(t) x(t),$$

and then (10.67) approximately reduces to an affine SDE for $S(t)$:

$$\begin{aligned} dS(t) &\approx \chi(t)\sigma(t)\sqrt{\xi(t) + \beta x(t)}\, dW^A(t) \\ &= \chi(t)\sigma(t)\sqrt{\xi(t) + \beta\left(\frac{S(t) - \zeta(t)}{\chi(t)}\right)}\, dW^A(t) \\ &= \sigma(t)\sqrt{\xi_s(t) + \beta_s(t) S(t)}\, dW^A(t). \end{aligned} \tag{10.68}$$

While valuation of the payout (10.66) cannot be accomplished in closed form when $S(t)$ follows the time-dependent affine SDE (10.68), we can always rely on transform-based methods. Indeed, it is evident that the characteristic function of $S(T_0)$ can be constructed by applying Proposition 10.2.2 and Remark 10.2.3 to (10.68), whereafter Theorem 8.4.3 gives us a way to calculate the required expected value in

$$V(0) = A(0)\mathrm{E}^{\mathrm{Q}^A}\left((S(T_0) - c)^+\right). \tag{10.69}$$

We trust that the reader can see how this would work, so we omit the details. Instead, we proceed to further simplify matters, through time averaging of parameters.

First, we wish to reduce (10.68) to the simplified form

$$dS(t) = \sigma(t)\sqrt{\beta_s(t)}\sqrt{\psi + S(t)}\, dW^A(t), \tag{10.70}$$

where ψ is a some constant. One approach for setting ψ is to simply match quadratic variance of $S(t)$ over $[0, T_0]$, i.e.

$$\int_0^{T_0} \sigma(t)^2 \beta_s(t)\psi\, dt = \int_0^{T_0} \sigma(t)^2 \beta_s(t)\xi_s(t)\, dt$$

or

$$\psi = \frac{\int_0^{T_0} \sigma(t)^2 \beta_s(t)\xi_s(t)\, dt}{\int_0^{T_0} \sigma(t)^2 \beta_s(t)\, dt}.$$

A more sophisticated alternative would be to rely on a small-noise expansion, as in Chapter 7. In any case, for the SDE (10.70), the expectation in (10.69) can be evaluated in closed form. To see this, simply define $y(t) \equiv \psi + S(t)$ and note that

$$dy(t) = \sigma(t)\sqrt{\beta_s(t)}\sqrt{y(t)}\, dW^A(t), \quad y(0) = \psi + S(0), \tag{10.71}$$

and

$$V(0) = A(0)\mathrm{E}^{\mathrm{Q}^A}\left((S(T_0) - c)^+\right) = A(0)\mathrm{E}^{\mathrm{Q}^A}\left((y(T_0) - c_y)^+\right), \tag{10.72}$$

with $c_y \triangleq \psi + c$. Since $y(t)$ in (10.71) is simply a (time-dependent) CEV process with CEV power 1/2, computation of the call option expectation in (10.72) can be carried out by the formulas in Section 7.2. Swaption prices produced this way are, in our experience, accurate and robust, and much more convenient to compute than by competing methods.

10.2.5 Swaption Calibration

As we showed in Section 10.1.4, calibration of the GSR model volatility to swaption prices is a matter of straightforward bootstrapping. Unfortunately, matters are more complicated for general affine models.

10.2.5.1 Basic Problem

To gain insight, let us first consider the simple problem of calibrating the model volatility function $\sigma(t)$ in (10.60) to match the time 0 price of a Δ-tenor zero-coupon bond option maturing at T. Assuming that the initial yield curve is known at time 0, how much volatility information is needed to price this option? The answer to this question depends on the specification of ξ and β.

If $\beta = 0$, we know that the function $b(T, T+\Delta)$ in the bond reconstitution formula (10.63) is independent of σ; see Proposition 10.1.7 (and adjust notation accordingly). It can also be verified that while $c(T, T+\Delta)$ in

(10.63) depends on the initial discount curve all the way to time $T + \Delta$, it only requires the specification of $\sigma(t)$ to time T. Further, the state of $x(T)$ only depends on $\sigma(t)$, $t < T$. In total, when $\beta = 0$, the discount bond option payout is only affected by $\{\sigma(t)\}_{0 \le t \le T}$, irrespective of the magnitude of the bond tenor Δ. This is also obvious from the reconstitution formula (10.18).

If $\beta \neq 0$, however, we see from (10.65) that $b(T, T + \Delta)$ depends[11] on the volatility $\{\sigma(t)\}_{0 \le t \le T+\Delta}$. This again makes $c(t, T+\Delta)$ depend on volatilities in $[t, T + \Delta]$, requiring the full knowledge of $\{\sigma(t)\}_{0 \le t \le T+\Delta}$ to price the option at time 0. This fact has implications for calibration to, say, swaption prices as regular bootstrapping techniques cannot be employed.

10.2.5.2 Calibration Algorithm

Consider now the situation where we wish to calibrate our volatility function $\sigma(t)$ to a swaption strip defined on a maturity grid $0 = T_0 < T_1 < \ldots < T_N$. Recall that a swaption strip consists of $N - 1$ swaptions expiring at times T_i, $i = 1, \ldots, N - 1$; we here assume that all swaptions are written on swaps that mature at time T_N (coterminal strip). According to the discussion above, pricing any one of these swaptions — even the short-dated ones — in an affine model will require knowledge of $\{\sigma(t)\}_{0 \le t \le T_N}$. As it would be too slow to calibrate volatilities by simultaneous, multi-dimensional root search on all levels $\sigma(T_i)$, $i = 0, 1, \ldots, N$, we instead notice that while, say, the swaption maturing on date T_i depends on volatilities everywhere on $[0, T_N]$, its dependence on the volatilities in $[0, T_i]$ is much stronger than on the volatilities in the interval $(T_i, T_N]$. Assuming that $\sigma(t)$ is piecewise constant on the maturity grid — with σ_i denoting the flat value on $(T_i, T_{i+1}]$ — we can use this observation to propose the following iterative calibration approach.

1. Start out by setting all σ_i, $i = 0, \ldots, N-1$, equal to a reasonably chosen constant, or equal to values approximated from a calibrated GSR[12] model.
2. Compute $\omega(\cdot)$ to match time 0 prices of the N discount bonds maturing on $T_1, T_2, \ldots, T_N$. One can use the algorithm in Section 10.2.3.2 for this.
3. Set the value σ_0 — but leave all other volatilities σ_i, $i = 1, \ldots, N - 1$, unchanged — such that the swaption maturing at time T_1 is priced correctly. We can use the pricing techniques in Section 10.2.4 for this.
4. Repeat Step 3 for $\sigma_1, \sigma_2, \ldots, \sigma_{N-2}$, always leaving future (but not past) points on the volatility curve unchanged.
5. Repeat Step 2 and recompute all swaption prices.
6. Repeat Steps 3–5 until all swaptions are priced within given tolerances.

---

[11] Recall that we solve $b(t, T + \Delta)$ backward in time from the known boundary condition at $t = T + \Delta$.

[12] For instance, if $\sigma_g(t)$ is the volatility function in the Gaussian model, then we can extract an estimate for $\sigma(t)$ from the relation $\sigma(t)\sqrt{\alpha + f(0,t)\beta} \approx \sigma_g(t)$.

Notice that in Step 4, altering σ_i will slightly distort the prices of swaptions maturing at dates earlier than T_i; this necessitates the iteration in Step 6.

We (re-)emphasize that the algorithm above, when applied to the Gaussian model, will converge within one iteration in Step 6. Finally, we note that the calibrated model needs to be checked against the regularity conditions discussed in Section 10.2.1.2; if conditions are violated, the problem may potentially be remedied by increasing α.

10.2.6 Quadratic One-Factor Model

In conclusion, let us consider an interesting special case of an affine class. A *quadratic Gaussian one-factor* model is obtained by specifying the short rate to be a quadratic function of a linear Gaussian process,

$$r(t) = \alpha(t) + \beta(t)y(t) + \gamma(t)y(t)^2, \tag{10.73}$$

where

$$dy(t) = -\varkappa(t)y(t)dt + \sigma(t)\,dW(t), \quad y(0) = 0. \tag{10.74}$$

While this is not immediately obvious, the model is indeed of affine type, albeit in *two* factors. If we denote $u(t) = y(t)^2$, we see that $r(t)$ is a linear function of the state vector $(y(t), u(t))$, which follows the SDEs

$$d\begin{pmatrix} y(t) \\ u(t) \end{pmatrix} = \begin{pmatrix} -\varkappa(t)y(t) \\ \sigma(t)^2 - 2\varkappa(t)u(t) \end{pmatrix} dt + \sigma(t)\begin{pmatrix} 1 \\ 2y(t) \end{pmatrix} dW(t), \tag{10.75}$$

which is affine.

We consider multi-dimensional quadratic Gaussian models in a fair amount of detail in Chapter 12, so we shall be suitably brief here. The affine connection makes it unsurprising that bond reconstruction formulas exist for the quadratic model. In fact, we have that zero-coupon discount bonds are exponentials of a quadratic function of $y(t)$,

$$P(t,T) = P(t,T;y(t)) = e^{a(t,T)-b(t,T)y(t)-c(t,T)y(t)^2}$$

with the coefficients a, b, c satisfying Riccati ODEs.

In some ways the parameterization (10.73)–(10.74) is more convenient than the general affine specification. For example, with the discount bonds known functions of a *Gaussian* factor $y(t)$, a swap rate — or a swap value — is a known function of a Gaussian random variable, which allows us to price a swaption by a simple one-dimensional Gaussian integration. We return to this topic in Chapter 12.

10.2.7 Numerical Methods for the Affine Short Rate Model

Much of the material on numerical methods for the GSR model applies to the affine short rate processes, so we shall be brief. Turning first to finite difference methods, let us again emphasize that the spatial variable should be set to be $x(t) = r(t) - f(0,t)$ rather than $r(t)$ itself. The dynamics for $x(t)$ can be found in (10.60) and lead to the general derivatives pricing PDE

$$\frac{\partial V}{\partial t} + (\omega(t) - \varkappa(t)x)\frac{\partial V}{\partial x} + \frac{1}{2}\sigma(t)^2(\xi(t) + \beta x)\frac{\partial^2 V}{\partial x^2} = (x + f(0,t))V,$$

which can be solved by standard methods, given appropriate terminal and boundary conditions. We refer to Section 10.1.5 for general guidelines. Dimensioning of the spatial dimensions of the finite difference grid by probabilistic means will require estimates for the mean and variance of $x(T)$, with T being the terminal horizon. We can compute these from the moment-generating functions established earlier, or, perhaps more easily, by approximating the SDE for $x(t)$ as being approximately Gaussian. If $r(t)$ is close to a CIR process, we can also use the analytical moment results established in Corollary 8.3.3. When establishing the terminal boundary function (i.e. the option payout), we can rely on the reconstitution formulas in (10.63) to turn values of x in the finite difference lattice into the discount bond prices that are required to evaluate the payout.

As for Monte Carlo methods, many of the principles of Section 10.1.6 continue to apply, and we can draw on material in Chapter 8 to design schemes to advance $x(t)$ through time. To elaborate a bit on this, suppose that we are interested in advancing $x(t)$ from time t_i to time t_{i+1}. Assume that all parameters in (10.60) are piecewise constant, such that

$$dx(t) \approx \varkappa_i(q_i - x(t))\,dt + \sigma_i\sqrt{\xi_i + \beta x(t)}\,dW(t), \quad t \in [t_i, t_{i+1}],$$

where[13] $\varkappa_i = \varkappa(t_i)$, $q_i = \omega(t_i)/\varkappa(t_i)$, $\sigma_i = \sigma(t_i)$, and $\xi_i = \xi(t_i)$. Defining $y(t) = \xi_i + \beta x(t)$, it follows that we can approximate $x(t_{i+1}) \approx (y(t_{i+1}) - \xi_i)/\beta$, where

$$dy(t) = \varkappa_i(\beta q_i + \xi_i - y(t))\,dt + \beta\sigma_i\sqrt{y(t)}\,dW(t), \quad y(t_i) = \xi_i + \beta x(t_i). \tag{10.76}$$

Simulation of this SDE, however, was discussed in detail in Section 9.5 where a number of practical algorithms were introduced. We should notice that typical parameterizations of (10.76) will rarely violate the Feller condition, making this SDE considerably easier to deal with numerically than the stochastic volatility applications in Section 9.5. Additional material on Monte Carlo simulation of generic short rate processes — most of which also applies to affine processes — can be found in Section 11.3.3.

[13] Alternatively, we can also set $\varkappa_i = (\varkappa(t_i) + \varkappa(t_{i+1}))/2$, and so forth.

11

One-Factor Short Rate Models II

While the affine specification (including the Gaussian case) of Chapter 10 is, without doubt, the most popular one-factor short rate model in practice, quite a few other models have been proposed in the literature. In this chapter we cover the most important of these models, paying special attention to the case where the short rate is log-normal. We also briefly discuss some issues in the econometric estimation of short rate models, and introduce the important concept of *unspanned stochastic volatility*.

As most of the models introduced in this chapter have no analytical bond reconstitution formulas, their calibration to the initial term structure requires numerical work. Accordingly, the second half of this chapter is dedicated to numerical methods for pricing and, especially, calibration of models based on generic short rate SDE. Particularly useful in this regard is the discussion in Section 11.3.2 on efficient finite difference schemes based on the important concept of *forward induction*.

11.1 Log-Normal Short Rate Models

Given the pervasiveness of the log-normal Black-Scholes model in derivatives pricing theory, it should come as no surprise that many authors have attempted to specify one-factor short rate models where the dynamics of $r(t)$ are of the form $dr(t) = O(dt) + \sigma_r(t)r(t)\,dW(t)$ with deterministic $\sigma_r(t)$. This section reviews this class of models which, somewhat surprisingly, turns out to have a number of rather severe drawbacks.

11.1.1 The Black-Derman-Toy Model

The *Black-Derman-Toy* (BDT) model was originally specified in a discrete-time binomial setting in Black et al. [1990], but subsequent research has shown the continuous-time limit of the model to be

$$r(t) = U(t)e^{\sigma_r(t)W(t)}, \tag{11.1}$$

where $U(t)$ and $\sigma_r(t)$ are deterministic functions, and $W(t)$ is a scalar Brownian motion in the risk-neutral measure. Notice that $r(t)$ is here an outright function of $W(t)$, a property sometimes known as *path independence* in the short rate dynamics. The following lemma rewrites the BDT model in more familiar terms.

Lemma 11.1.1. *Let $r(t)$ be given by (11.1), and let the prime denote a time derivative. Then*

$$d\ln r(t) = \left(\vartheta(t) + \frac{\sigma_r'(t)}{\sigma_r(t)}\ln r(t)\right)dt + \sigma_r(t)\,dW(t), \tag{11.2}$$

$$dr(t)/r(t) = \left(\vartheta(t) + \frac{1}{2}\sigma_r(t)^2 + \frac{\sigma_r'(t)}{\sigma_r(t)}\ln r(t)\right)dt + \sigma_r(t)\,dW(t),$$

where

$$\vartheta(t) = U'(t)/U(t) - \ln(U(t))\sigma_r'(t)/\sigma_r(t).$$

Proof. Set $y(t) = \ln r(t) = \ln U(t) + \sigma(t)W(t)$ and apply Ito's lemma to get

$$\begin{aligned} dy(t) &= \left(\frac{d\ln U(t)}{dt} + \sigma_r'(t)\,W(t)\right)dt + \sigma_r(t)\,dW(t) \\ &= (U'(t)/U(t) + (y(t) - \ln U(t))\,\sigma_r'(t)/\sigma_r(t))\,dt + \sigma_r(t)\,dW(t), \end{aligned}$$

so that (11.2) follows. A second application of Ito's lemma to $r(t) = e^{y(t)}$ then gives the result for $dr(t)$. □

Of the various specifications of the BDT model, the most convenient is probably (11.2) which describes the logarithm of $r(t)$ as a mean-reverting Gaussian process with a mean reversion speed

$$\varkappa(t) = -\frac{\sigma_r'(t)}{\sigma_r(t)}.$$

As $\ln r(t)$ is Gaussian, $r(t)$ is log-normal. In the formulation (11.2), $\vartheta(t)$ can be considered a free parameter, the value of which can — for a fixed volatility function $\sigma_r(t)$ — be determined by calibrating the model to the initial yield curve. As the BDT model has no known bond reconstitution formula, this fit must be done numerically. The original presentation in Black et al. [1990] outlines such a routine, based on brute-force backward induction in a binomial tree; this algorithm is, however, computationally inefficient and should *not* be used. For a much faster approach, see Sections 11.3.2.1 and 11.3.2.2.

Besides the lack of the bond reconstitution formula, the BDT model is plagued by a number of issues that makes it unsuitable for practical applications. One of the issues is explained in Section 11.1.3. In addition, it

is problematic that the mean reversion speed of the model is beyond user control and is fully determined from the short rate volatility and its time derivative. In particular, for those values of t where $\sigma_r(t)$ grows in t, the mean reversion will be negative, i.e. the model will imply "mean-fleeing" behavior. It should be obvious that this feature of the model is undesirable.

11.1.2 Black-Karasinski Model

In order to rectify the problems surrounding the mean reversion in the BDT model, Black and Karasinski [1991] (BK) took the obvious step of introducing a model where mean reversion for $\ln r(t)$ is exogenously specified. In other words, we write

$$d\ln r(t) = \varkappa(t)\,(\vartheta(t) - \ln r(t))\,dt + \sigma_r(t)\,dW(t). \tag{11.3}$$

Equivalently, we may write

$$r(t) = e^{x(t)},$$

where $x(t)$ is a standard mean-reverting Gaussian process; accordingly, the BK short rate dynamics are straightforward to simulate in the Monte Carlo method. The BK dynamics generalize and improve those of the BDT model but still do not allow for a closed-form discount bond reconstitution formula.

11.1.3 Issues in Log-Normal Models

The BK model — and its special case, the BDT model — have short rates that are log-normally distributed. A similar distribution of rates would arise for risk-neutral dynamics of geometric Brownian motion type

$$dr(t) = \mu_r(t)r(t)\,dt + \sigma_r(t)r(t)\,dW(t). \tag{11.4}$$

A time-homogeneous version of this model was considered in Rendleman and Bartter [1980]; for the time-homogeneous case (very complicated) formulas[1] for discount bond prices exist, see Dothan [1978] and Hogan and Weintraub [1993]. The model (11.4) has no mean reversion, and cannot be recommended for practical applications, however.

A common problem shared by all the log-normal short rate models (11.2), (11.3), (11.4) is the fact that the expected value of the inverse of future discount bond price is infinite, i.e.

$$\mathrm{E}_t\left(\frac{1}{P(t',T)}\right) = \infty, \quad t < t' < T. \tag{11.5}$$

[1] A computationally efficient recursive procedure for bond pricing can be found in Hansen and Jørgensen [1998].

This result is formally shown in Hogan and Weintraub [1993], but is hardly surprising since the expectation of e^{cX}, $c > 0$, is well-known to be infinite when X is a log-normal random variable[2]. An important corollary of this result is listed below, originally due to Sandmann and Sondermann [1997].

Corollary 11.1.2. *Define a forward Libor rate*

$$L(t,T) = \frac{1}{\tau}\left(\frac{P(t,T)}{P(t,T+\tau)} - 1\right),$$

where $\tau > 0$ is some accrual factor. Assuming that (11.5) holds, then also

$$\mathrm{E}_t\left(L(T,T)\right) = \infty, \quad T > t, \tag{11.6}$$

and

$$\mathrm{E}_t\left(e^{\int_t^T r(u)du}\right) = \infty. \tag{11.7}$$

Proof. As $1 + \tau L(T,T) = 1/P(T,T+\tau)$, equation (11.6) follows directly from (11.5). To show (11.7), we use Jensen's inequality[3] to show

$$\frac{1}{P(t',T)} = \frac{1}{\mathrm{E}_{t'}\left(e^{-\int_{t'}^T r(u)du}\right)} \leq \mathrm{E}_{t'}\left(e^{\int_{t'}^T r(u)du}\right). \tag{11.8}$$

Taking expectations conditional on the time t filtration yields (11.7). □

Both (11.6) and (11.7) have unfortunate economic consequences. Formula (11.7) predicts that the expected return of investing in the continuously compounded money market account for a finite period of time is infinite; and (11.6) predicts that all Libor futures rates should be infinite[4]. A related problem was discussed in the context of log-normal HJM models in Section 4.5.3.

11.1.4 Sandmann-Sondermann Transformation

The problems outlined in Corollary 11.1.2 are a significant drawback of log-normal short rate models, and one that should disqualify their use in many applications. On the other hand, market data may dictate that interest rates should, in fact, be log-normal. This is not as big a dilemma as it may appear, as there are ways to build models with a strong log-normal flair, yet avoiding (11.5). One way is to use HJM models of the type

$$df(t,T) = O(dt) + \sigma(t)r(t)\,dW(t),$$

[2]Even though the log-normal distribution has finite moments of all orders, the moment-generating function is infinite at any positive argument.

[3]For details on Jensen's inequality, see the proof of Proposition 11.1.3.

[4]Recall from Section 4.1.2 that the risk-neutral expectation of a random variable must, in the absence of arbitrage, equal its traded futures price.

where f is, as always, the instantaneous forward rate. We return to this type of models in Chapter 13. An interesting alternative was proposed by Sandmann and Sondermann [1997] and in essence involves writing the log-normal dynamics for a *discretely compounded rate* $r_d(t)$, rather than the infinitesimal (continuously compounded) rate $r(t)$. Specifically, we relate r_d and r through the expression

$$e^{r(t)\delta} = 1 + r_d(t)\delta \quad \Longrightarrow \quad r(t) = \delta^{-1} \ln\left(1 + r_d(t)\delta\right), \tag{11.9}$$

where $\delta > 0$ is some finite compounding interval. Sandmann and Sondermann [1997] set $\delta = 1$, but any finite positive value will, in fact, do.

The effect of working with discretely compounded rates is summarized in the following result.

Proposition 11.1.3. *Let $r_d(t)$ be log-normal for all t. Then, for $t < t' < T$,*

$$\mathrm{E}_t\left(L(T,T)\right) < \infty, \tag{11.10}$$

$$\mathrm{E}_t\left(e^{\int_{t'}^{T} r(u)du}\right) < \infty. \tag{11.11}$$

Proof. As the proof of Proposition 11.1.3 is quite instructive, we give full details. From (11.8), to prove both (11.10) and (11.11) it suffices to show that the expectation of $e^{\int_{t'}^{T} r(u)du}$ is finite. For this, let us recall that Jensen's inequality for integrals states that for a real-valued function g and a concave function φ

$$\varphi\left(\int_{t'}^{T} g(u)f(u)\,du\right) \geq \int_{t'}^{T} \varphi\left(g(u)\right) f(u)\,du, \tag{11.12}$$

provided that $f(u)$ is non-negative and $\int_{t'}^{T} f(u)\,du = 1$. If φ is convex, the inequality is reversed. Now write

$$\delta^{-1}\int_{t'}^{T} \ln\left(1 + r_d(u)\delta\right) du = \delta^{-1}\int_{t'}^{T} \frac{1}{T-t'} \ln\left(\left(1 + r_d(u)\delta\right)^{T-t'}\right) du$$

and apply (11.12) to the right-hand side with $f(u) = 1/(T-t')$, $\varphi(u) = \ln(u)$, and $g(u) = (1 + r_d(u)\delta)^{T-t'}$ to show that

$$\delta^{-1}\ln\left(\int_{t}^{T} \frac{1}{T-t'}\left(1 + r_d(u)\delta\right)^{T-t'} du\right) \geq \delta^{-1}\int_{t'}^{T} \ln\left(1 + r_d(u)\delta\right) du. \tag{11.13}$$

From (11.13) it then follows immediately that

$$
\begin{aligned}
&\mathrm{E}_t\left(\exp\left(\int_{t'}^{T} r(u)\,du\right)\right)\\
&\quad=\mathrm{E}_t\left(\exp\left(\int_{t'}^{T} \delta^{-1}\ln\left(1+r_d(u)\delta\right)\,du\right)\right)\\
&\quad\le\mathrm{E}_t\left(\exp\left(\delta^{-1}\ln\left(\int_{t'}^{T}\frac{1}{T-t'}\left(1+r_d(u)\delta\right)^{T-t'}du\right)\right)\right)\\
&\quad=\mathrm{E}_t\left(\left(\frac{1}{T-t'}\int_{t'}^{T}\left(1+r_d(u)\delta\right)^{T-t'}du\right)^{1/\delta}\right).
\end{aligned}
$$

Assume that $0<\delta<1$, and set $f(u)=1/(T-t')$, $\varphi(u)=u^{1/\delta}$, and $g(u)=(1+r_d(u)\delta)^{T-t'}$. By Jensen's inequality (here (11.12) is reversed, since φ is now convex) we must have

$$
\left(\frac{1}{T-t'}\int_{t'}^{T}\left(1+r_d(u)\delta\right)^{T-t'}du\right)^{1/\delta}\le\frac{1}{T-t'}\int_{t'}^{T}\left(1+r_d(u)\delta\right)^{(T-t')/\delta}du. \tag{11.14}
$$

Therefore

$$
\mathrm{E}_t\left(e^{\int_{t'}^{T} r(u)du}\right)\le\frac{1}{T-t'}\mathrm{E}_t\left(\int_{t'}^{T}\left(1+r_d(u)\delta\right)^{(T-t')/\delta}du\right),\quad 0<\delta<1. \tag{11.15}
$$

Since finite powers of log-normal random variables have finite expected values, (11.11) has been shown for $0<\delta<1$. For $\delta\ge 1$ we have $1/\delta\le 1$, and

$$
\begin{aligned}
&\mathrm{E}_t\left(\left(\frac{1}{T-t'}\int_{t'}^{T}\left(1+r_d(u)\delta\right)^{T-t'}du\right)^{1/\delta}\right)\\
&\qquad\qquad\le\left(\mathrm{E}_t\left(\frac{1}{T-t'}\int_{t'}^{T}\left(1+r_d(u)\delta\right)^{T-t'}du\right)\right)^{1/\delta},
\end{aligned}
$$

and (11.11) follows from the same arguments. □

Comparison of Corollary 11.1.2 and Proposition 11.1.3 shows that models of the BK type

$$
d\ln r_d(t)=\varkappa(t)\left(\vartheta(t)-\ln r_d(t)\right)dt+\sigma(t)\,dW(t) \tag{11.16}
$$

and geometric Brownian motion models

$$
dr_d(t)/r_d(t)=\mu(t)\,dt+\sigma(t)\,dW(t), \tag{11.17}
$$

become significantly more reasonable when the dynamics are written in $r_d(t)$, rather than $r(t)$. We invite the reader to apply Ito's lemma to the

transformation (11.9) to uncover the r-dynamics for the models (11.16)–(11.17).

Remark 11.1.4. When applied to the HJM model class, the "trick" of shifting from continuously compounded to discretely compounded rates lays the foundation for the class of so-called *Libor market* models. We return to these models in Chapters 14 and 15.

11.2 Other Short Rate Models

11.2.1 Power-Type Models and Empirical Model Estimation

A natural extension of the Gaussian and affine short rate SDEs involves retaining the linear mean-reverting drift term of these models, but using a general power function in the diffusion term. That is, we write

$$dr(t) = \varkappa(t)\,(\vartheta(t) - r(t))\,dt + \sigma(t)r(t)^p dW(t), \quad p > 0. \tag{11.18}$$

The time-inhomogeneous Gaussian and CIR models correspond to the choices $p = 0$ and $p = 1/2$, respectively. The special case of (11.18) where $p = 1$ was suggested in Brennan and Schwartz [1980] and Courtadon [1982], and is quite similar to the BK model — to the extent that the case $p = 1$ shares[5] the unfortunate properties of the BK model listed in Corollary 11.1.2.

The general model (11.18) is similar to the CEV model described in Chapter 7, and manipulation of the parameter p may allow for a better fit of the model to observed volatility smiles in interest rate options. Due to its intractability — no bond reconstitution formula exists for $p \notin \{0, 1/2\}$ — the model is, however, rarely used in derivatives pricing applications. (For related, but significantly more tractable, models with power-type diffusion terms, see Chapter 13.) Starting with the influential article by Chan et al. [1992] (CKLS), the specification (11.18) has, however, been quite popular in econometric work. As the CKLS paper is one of the most frequently cited references on estimation of one-factor short rate models, let us make a (very) brief foray into econometrics, to review the CKLS conclusions and some of the criticism their work has subsequently drawn.

The CKLS estimation procedure is based on US Treasury bond data from the period 1964–1989. Assuming that $r(t)$ can be approximated by the one-month yield on US Treasury bonds, they estimate eight models with parameter restrictions, and one model with no restrictions. Generally speaking, the empirical results indicate that the value of the parameter p is the most important in determining whether a model is accepted or rejected.

[5] The proof of this statement is straightforward, and follows from a standard comparison theorem for SDEs (see p. 293 of Karatzas and Shreve [1997]).

The unrestricted estimate of p is close to 1.5, and values less than around $p = 1$ are rejected in their tests. The Vasicek and CIR models are, for instance, rejected, whereas the Brennan-Schwarz/Courtadon model (with $p = 1$) is accepted.

The fact that the CKLS estimates suggest that $p \geq 1$ is surprising in light of the generally downward-sloping volatility skews in fixed income derivatives, and also raises considerable questions about model regularity, as one would expect from Corollary 11.1.2. Indeed, as shown in Honore [1998b], $r(t)$ will a.s. explode (i.e. reach ∞) in finite time if $p > 1$. Beyond this, we notice that the CKLS analysis has received criticism on a number of procedural and data-related fronts. For instance, Honore [1998a] (and quite a few others) point out that the 1 month Treasury yield may be an unreliable proxy for the short rate. Repeating the analysis with a carefully computed value of $r(t)$ (obtained by exploiting the fact that a one-factor model implies a one-to-one correspondence between any discount yield and the short rate), Honore revises the CKLS estimate for p significantly downward, to around $p = 0.8$. Bliss and Smith [1997] also point out that the data used by CKLS covers the period October 1979 — September 1982, when the US Federal Reserve followed unusual monetary policies ("The Fed Experiment"). Properly accounting for this, Bliss and Smith revise the CKLS estimate for p down to around 1.0. Applying different estimation methods and different data sets, Andersen and Lund [1997] and Christensen et al. [2001] estimate p to around 0.7 and 0.8, respectively.

Moving away from observations of only a short rate proxy, Gibbons and Ramaswamy [1993] test the ability of the CIR model to simultaneously describe the evolution in four zero-coupon rates; with data covering the same period as the CKLS study, they accept the hypothesis $p = 1/2$. Finally, to muddy waters even further, Ait-Shalia [1996] points out (in an analysis that has subsequently been criticized as lacking robustness in Chapman and Pearson [2000]) that the simple linear drift term in (11.18) is fundamentally misspecified and should be adjusted to include non-linear terms such as $1/r(t)$ and $r(t)^2$.

By now, it should be clear to the reader that the problem of estimating short rate models is not close to being conclusively solved, despite an impressively long list of papers associated to it. In much contemporary empirical research, the importance of the choice of p is generally downplayed, with the affine class ($p = 1/2$) enjoying considerable current popularity due to its analytical tractability.

11.2.2 The Black Shadow Rate Model

As described in Chapter 10, one drawback of the Gaussian short rate model class is the implication that interest rates can become negative with positive probability. As long as investors can stuff their mattresses with currency, (nominal) interest rates must, however, always remain non-negative

to preclude arbitrage. In practice, the probability of negative rates may be small enough to ignore, but as argued in Rogers [1996] prices of certain contingent claims may be highly sensitive to even a remote probability of negative rates, in which case the Gaussian model should obviously be avoided. Possible model alternatives with non-negative rates include the log-normal models in Section 11.1, as well as the affine models in Section 10.2.

Rather than altering the underlying model, an alternative "fix" of the Gaussian model involves simply taking the positive part of the Gaussian short rate process, i.e. writing

$$r(t) = (r^*(t))^+ , \tag{11.19}$$

where $r^*(t)$ is a Gaussian process

$$dr^*(t) = \varkappa(t)\,(\vartheta(t) - r^*(t))\;dt + \sigma_r(t)\,dW(t). \tag{11.20}$$

This approach was first proposed in Black [1995] and Rogers [1995]. The form of (11.19) suggests an analogy where the interest rate $r(t)$ is an *option*, granting a choice between an underlying *shadow short rate* $r^*(t)$ and zero. In other words, whenever an interest rate product has a negative rate, we invest our money in currency instead.

The truncation in (11.19) may at first glance appear rather crude. For instance, the process for $r(t)$ retains full volatility $\sigma_r(t)$ as long as $r(t) > 0$, and can then suddenly get extinguished completely for potentially long stretches of time. In contrast, alternative models with non-negative rates such as BDT, BK, and CIR generally all imply that the interest rate volatility will gradually vanish (linearly for BDT/BK, as a square-root for CIR) as the short rate tends to zero. Interestingly, there is some empirical evidence (from the US in the 1930's and Japan in the 1990's) that suggests that very low interest rates are often accompanied by higher absolute rate volatility than standard models would predict, see Goldstein and Keirstead [1997]. Such evidence may lend some credibility to models such as (11.19).

The process (11.19) is not analytically tractable, and numerical methods (such as those of Section 11.3) must be applied to price discount bonds and other fixed income securities. For the case where all parameters in (11.20) are constants — i.e. $r^*(t)$ follows a Vasicek model — Gorovoi and Linetsky [2004] list[6] a complicated eigenfunction expansion series for discount bond prices. Of course, as the constant-parameter model will not be able to match the current yield curve, the result in Gorovoi and Linetsky [2004] has limited uses in practical applications.

Finally, the reader may very well ask whether perhaps (11.19) could be replaced by a reflecting or absorbing barrier at zero, or by the application of

[6] The authors also develop eigenfunction expansions for cases where the shadow rate $r^*(t)$ follow time-homogeneous diffusions more complicated than the Vasicek model.

a suitable transformation such as $r(t) = r^*(t)^2$. The latter idea was discussed in Section 10.2.6 and the former is investigated in Goldstein and Keirstead [1997] where eigenfunction expansions are derived for the time-homogeneous case. In Black [1995], the author objects to reflecting barriers on economic grounds.

11.2.3 Spanned and Unspanned Stochastic Volatility: the Fong and Vasicek Model

In previous chapters, we demonstrated how to incorporate stochastic volatility as a mechanism to model the volatility smile in vanilla models. One wonders how to proceed with such a construction for term structure models. We postpone much of this discussion to later chapters, but shall here take the opportunity to discuss what constitutes a *true* stochastic volatility model for interest rate evolution, and why the short rate framework is *not* particularly amendable to stochastic volatility extensions. More specifically, we shall introduce the notion of *spanned* and *unspanned* stochastic volatility. For concreteness, our discussion focuses on a model proposed by Fong and Vasicek [1991] which, despite initial appearances, is in fact not a true stochastic volatility model.

The Fong-Vasicek (FV) model is characterized by risk-neutral SDEs

$$\begin{aligned} dr(t) &= \varkappa_r \left(\vartheta_r - r(t)\right) dt + \sqrt{z(t)}\, dW_1(t), \\ dz(t) &= \varkappa_z \left(\vartheta_z - z(t)\right) dt + \eta \sqrt{z(t)}\, dW_2(t), \end{aligned}$$

where W_1 and W_2 are correlated Brownian motions, $\langle dW_1(t), dW_2(t)\rangle = \rho\, dt$, and $\varkappa_r, \vartheta_r, \varkappa_z, \vartheta_z, \eta$ are positive constants. We recognize the FV model as being essentially a time-homogeneous GSR model augmented by a stochastic variance process $z(t)$; the process for $z(t)$ is identical to that of the Heston model (see Chapter 8). Bond prices in the FV model can be shown to satisfy

$$P(t, t+\delta) = e^{A(\delta) + r(t)B(\delta) + z(t)C(\delta)}, \tag{11.21}$$

for deterministic functions A, B, C satisfying a coupled system of ODEs. Details about these ODEs[7] and their rather complicated analytical solution can be found in Selby and Strickland [1995]. For our purposes here, what matters is not the precise form of A, B, and C, but rather the fact that $P(t,T)$ in the FV model is a deterministic function of the two state variables $r(t)$ and $z(t)$. As a consequence, one can theoretically hedge out exposure to both $r(t)$ and $z(t)$ by simply taking positions in two discount bonds with different maturities. Equivalently, given observations at time t of the prices of two discount bonds with different maturities, we can invert (11.21) to uncover the current values of the two variables $r(t)$ and $z(t)$.

[7] The ODEs are easily derived by substituting the right-hand side of (11.21) into the no-arbitrage PDE for $P(t,T)$ in the FV model.

When a "stochastic volatility" variable $z(t)$ can be hedged by positions in discount bonds, we say that $z(t)$ is *spanned* by the discount curve. If all stochastic volatility variables are spanned by the discount curve, moves in, say, implied volatilities of caps and swaptions would always be accompanied by moves in the yield curve, making vega hedging theoretically unnecessary. In reality, however, there is much evidence that interest rate option volatilities cannot be perfectly hedged by trading only discount bonds; see Casassus et al. [2005] for a review of the literature. Formally, this implies that the volatilities of discount bond prices depend on a vector of random state variables $(z_1(t), \ldots, z_n(t))$ that are not included in the state variables used in reconstitution formulas for the discount curve. That is,

$$\frac{\partial P(t,T)}{\partial z_i(t)} = 0, \quad i = 1, \ldots, n, \tag{11.22}$$

yet, for $Q(t,T) \equiv \mathrm{E}_t(d(P(t,T))^2)/dt$,

$$\frac{\partial Q(t,T)}{\partial z_i(t)} \neq 0, \quad i = 1, \ldots, n. \tag{11.23}$$

Random variables satisfying (11.22)–(11.23) are said to represent *unspanned* stochastic volatility (USV). Whenever we talk about true stochastic volatility in this book, we always refer to models with USV, i.e. to models that prescribe moves in rate volatility that cannot be inferred from moves in the level and the shape of the discount curve.

A detailed, and highly recommended, account of USV can be found in Collin-Dufresne and Goldstein [2002b]. Among many results, the authors prove that in a time-homogeneous setting, bivariate affine models for the term structure of interest rates cannot exhibit USV. This explains our results for the FV model, and also demonstrates that several other classical stochastic-volatility models (e.g. Longstaff and Schwartz [1992]) are not in the USV class. Further analysis of a number of stochastic volatility models proposed in the literature (a surprising number of which do not, in fact, allow for USV) can be found in Collin-Dufresne and Goldstein [2002b] and Casassus et al. [2005].

11.3 Numerical Methods for General One-Factor Short Rate Models

The models described in Sections 11.1 and 11.2 all (with the exception of the Fong-Vasicek model) involve risk-neutral SDEs for the short rate of the type

$$dr(t) = \mu_r(t, r(t))\, dt + \sigma_r(t, r(t))\, dW(t), \tag{11.24}$$

for certain user-specified deterministic functions $\mu_r(t,r)$ and $\sigma_r(t,r)$. As discussed, many of these models do not allow for closed-form expressions

relating discount bonds to the state of $r(t)$, necessitating the application of numerical methods for even simple tasks such as calibrating the model to the initial yield curve. While this issue ultimately should make one pause when it comes to deciding whether a model is suited for practical applications, we still want to cover some techniques that are useful in handling generic models such as (11.24). Some of the techniques we shall discuss, e.g. calibration through forward induction, have broad applicability.

11.3.1 Finite Difference Methods

As always, we set $x(t) = r(t) - f(0,t)$, such that the generic pricing PDE for a derivative $V = V(t,x)$ becomes

$$\frac{\partial V}{\partial t} + \mu_x(t,x)\frac{\partial V}{\partial x} + \frac{1}{2}\sigma_x(t,x)^2\frac{\partial^2 V}{\partial x^2} = (x + f(0,t))\,V, \tag{11.25}$$

where we have defined

$$\mu_x(t,x) = \mu_r\left(t, x + f(0,t)\right) - \frac{\partial f(0,t)}{\partial t}, \quad \sigma_x(t,x) = \sigma_r\left(t, x + f(0,t)\right). \tag{11.26}$$

Given terminal conditions for $V(T,x)$, as well as suitable boundary conditions in the x-domain, we can solve this equation by the generic finite difference methods of Chapter 2. For later use, let us quickly recall that a standard θ-method finite difference scheme on an equidistant x-grid $\{x_j\}_{j=0}^{m+1}$ would result in a matrix scheme of the type

$$\begin{aligned}&[\mathbf{I} - \theta\Delta_t\mathbf{A}\left((1-\theta)t_{i+1} + \theta t_i\right)]\,\widehat{\mathbf{V}}(t_i) = \\ &\qquad [\mathbf{I} + (1-\theta)\Delta_t\mathbf{A}\left((1-\theta)t_{i+1} + \theta t_i\right)]\,\widehat{\mathbf{V}}(t_{i+1}) + \mathbf{B}(t_i, t_{i+1}),\end{aligned} \tag{11.27}$$

where $\widehat{\mathbf{V}}(t) = (\widehat{V}(t,x_1), \ldots, \widehat{V}(t,x_m))^\top$ with $\widehat{V}(t,x)$ denoting the approximation to the true solution $V(t,x)$, $\mathbf{A}(t)$ is a tri-diagonal matrix, and $\mathbf{B}(t_i, t_{i+1})$ is a vector representing any boundary conditions that cannot be folded into the matrix $\mathbf{A}$. We solve the matrix system (11.27) backward in time, starting from a given value of $\widehat{\mathbf{V}}(T)$. Determination of the spatial boundaries in the x-domain — that is, the values of x_0 and x_{m+1} — can, as always, be set by probabilistic arguments, through estimation of the first and second moment of $x(T)$. While these estimates should be targeted to the specifics of the model at hand, if all else fails we can always rely on Gaussian estimate, e.g. something like

$$x_0 = \mu_x\left(0,0\right)T - \alpha\sigma_x\left(0,0\right)\sqrt{T}, \quad x_{m+1} = \mu_x\left(0,0\right)T + \alpha\sigma_x\left(0,0\right)\sqrt{T},$$

where α is some confidence interval multiplier (e.g. 4 or 5).

The discussion in Section 10.1.5.2 about x-domain boundary conditions apply to (11.27) as well, but determining the terminal condition $\widehat{\mathbf{V}}(T)$ can

be problematic[8], as option payouts will often involve several discount factors at time T (e.g. to price a swap or to compute a Libor rate). As there is generally no way of computing such discount bonds analytically for the model (11.24), we are forced to compute the discount bond prices themselves by finite difference methods. To compute the value of the discount bond $P(T, T^*, x)$, $T^* > T$, at time T we:

1. Extend the finite difference grid to time T^*.
2. Set the boundary condition $P(T^*, T^*, x_j) = 1$, $j = 0, \ldots, m+1$.
3. On some suitable time grid, use (11.27) on P to step backward to time T.
4. Use the finite difference estimates $\widehat{P}(T, T^*, x_j)$, $j = 0, \ldots, m+1$, to fill in $\widehat{\mathrm{V}}(T)$.

To the extent that $\mathrm{V}(T)$ involves multiple discount bonds, we perform the algorithm above on all of the required discount bonds; the grid must then be extended to the maturity of the longest-dated discount bond needed in the payout computation. In some cases this can dramatically increase computation times relative to models where closed-form discount bond reconstitution formulas exist. For instance, for a 3 month option on a 30 year swap, a model with an analytical formula for discount bonds (e.g. the affine short rate model) would require us only to build the finite difference grid out to 3 months; when such a formula does not exist, we are forced to use a 30 year finite difference grid.

11.3.2 Calibration to Initial Yield Curve

Assume that the volatility function $\sigma_r(t, r)$ has been fixed, but $\mu_r(t, r) = \mu_r(t, r; \vartheta(t))$ has a free time-dependent parameter $\vartheta(t)$, the value of which we wish to set in such a way that the initial discount curve $P(0, T)$, $T > 0$, is correctly recovered by the model. Suppose that we assume, as is common, that $\vartheta(t)$ is piecewise constant on some time grid $0 = t_0 < t_1 < \ldots < t_N$, with ϑ_{i-1} denoting the value of ϑ that applies on the interval $[t_{i-1}, t_i)$. A brute force approach to the calibration of $\vartheta(t)$ could proceed as follows.

1. Assuming that $\vartheta_0, \ldots, \vartheta_{i-2}$ are known, make a guess[9] for ϑ_{i-1}.
2. Setting the terminal boundary value to $V(t_i, x) = 1$ for all x, use the backward finite difference grid algorithm (11.27) to compute the value of $P(0, t_i)$.
3. If the computed value of $P(0, t_i)$ equals that quoted in the market stop; otherwise return to Step 1.

[8] The same holds for exercise values of callable securities.

[9] An initial guess for ϑ_{i-1} could be $\vartheta_{i-1} = \vartheta_{i-2}$. Subsequent guesses would be performed by a root-search algorithm.

This approach is similar to an algorithm proposed in Black et al. [1990] (albeit the authors worked only with binomial trees) and involves very high computational costs as the numerical search for each of the parameters $\vartheta_0, \vartheta_1, \ldots$ involves solving a full finite difference grid in each loop. Specifically, assume that on average each search for ϑ_i involves M iterations over steps 1–3 above. With N ϑ-values to find, the computational effort of a finite difference grid with m spatial steps (say) is $O(Nm)$; it follows that the total computational cost for the calibration of $\vartheta(t)$ is

$$O\left(MN^2m\right),$$

which is often prohibitively expensive in practice.

11.3.2.1 Forward Induction

In the setting of binomial and trinomial trees, the brute-force BDT algorithm was markedly improved upon in Jamshidian [1991a], using a technique known as *forward induction*. The basic idea is to work with a forward equation, rather than with the backward equation (11.27). Two varieties of this approach are feasible in a finite difference setting, depending on whether the forward equation is developed by direct discretization of the continuous-time Fokker-Planck equation, or by rearrangement of a discretized backward equation. We cover the former approach here, and the latter in Section 11.3.2.2 below.

Let $G(t, x; s, y)$, $s \geq t$, denote the time t value of a security that pays out a Dirac delta amount iff $x(s) = y$, given that $x(t) = x$. Clearly then

$$P(0,T) = \int_{-\infty}^{\infty} G(0,0;T,y)\,dy. \tag{11.28}$$

We already saw this financial contract, the so-called Arrow-Debreu security, at the end of Section 1.8. Its price G — being the value of a perfectly valid derivative contract — certainly satisfies a backward Kolmogorov equation

$$\frac{\partial G}{\partial t} + \mu_x(t,x;\vartheta(t))\frac{\partial G}{\partial x} + \frac{1}{2}\sigma_x(t,x)^2\frac{\partial^2 G}{\partial x^2} = (x + f(0,t))\,G, \quad (s,y) \text{ fixed}, \tag{11.29}$$

with the terminal value condition $G(s,y;s,y) = \delta(y)$. Clearly $G(t,z;s,y)$ is closely related to the transition density $p(t,z;s,y)$ defined in Section 1.8, and can be expected to satisfy a forward Kolmogorov equation, too. The correct equation is

$$\begin{aligned} -\frac{\partial G}{\partial s} - \frac{\partial}{\partial y}\left(\mu_x(s,y;\vartheta(s))G\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma_x(t,y)^2 G\right) \\ = (y + f(0,t))\,G, \quad (t,x) \text{ fixed}, \end{aligned} \tag{11.30}$$

with the terminal value condition $G(s,y;s,y) = \delta(y)$. Clearly $G(t,z;s,y)$ is closely related to the transition density $p(t,z;s,y)$ defined in Section 1.8, and can be expected to satisfy a forward Kolmogorov equation, too. The correct equation is

subject to the initial condition $G(t,x;t,x) = \delta(x)$. This PDE is identical to the Fokker-Planck equation listed in Section 1.8, except for the fact that the right-hand side is not zero, but contains a discounting term.

Fixing $(t,x) = (0,0)$, the PDE (11.30) can be discretized by finite difference methods in standard fashion, although we keep in mind that the PDE is to be solved *forward* in time from its initial (Dirac) condition, rather than backward. In most cases[10], the appropriate spatial boundary conditions for the finite difference solution $\widehat{G}$ are

$$\widehat{G}(0,0;s,y_0) = \widehat{G}(0,0;s,y_{m+1}) = 0,$$

which assumes that y_0 and y_{m+1} have been set such that the probability density at these levels is negligible. In a discrete setting, the initial condition $G(t,x;t,x) = \delta(x)$ is translated to $\widehat{G}(0,0;0,y) = (\Delta y)^{-1} 1_{\{y=0\}}$ for Δy suitably defined, in agreement with the averaging principles of Section 2.5.2. Due to the strongly discontinuous initial condition, Rannacher stepping (see Section 2.5) should always be used.

We are now ready to state our revised algorithm for calibration of $\vartheta(t)$, working again with the assumption that $\vartheta(t)$ is piecewise constant on a time grid $0 = t_0 < t_1 < \ldots < t_N$. We assume that $\vartheta_0, \ldots, \vartheta_{i-2}$ have been found, as has been $\widehat{G}(0,0;t_{i-1},y)$ for all $y \in \{y_j\}_{j=1}^m$ in the finite difference grid.

1. Make a guess for ϑ_{i-1}.
2. Solve (11.30) one time step forward, to time t_i, and save $\widehat{G}(0,0;t_i,y_j)$, $j = 1,\ldots,m$.
3. Compute the discount bond price $P(0,t_i) = \sum_j \widehat{G}(0,0;t_i,y_j)(y_j - y_{j-1})$.
4. If the computed value of $P(0,t_i)$ equals that quoted in the market stop; otherwise return to Step 1.

The cost of Steps 2 and 3 in this algorithm is $O(m)$, where m is the number of points in the y-direction of the finite difference grid. The total computational effort of this algorithm is therefore

$$O(MmN),$$

where M is the average number of root search iterations over Steps 1–4 above. We recall that the effort of the brute-force backward equation approach was $O(MmN^2)$, so the use of forward induction saves us a factor of N. As N is often of the order of $N = 100$, these savings can be considerable. In typical applications, M is often in the order $M = 2$ to 4, so calibration of the model (11.24) to the initial discount curve should only be a few times slower than pricing by finite difference methods a single option maturing at time t_N (the cost of which we recall to be $O(mN)$).

[10]For some models, the density can grow to infinity at the boundary, notably in the CIR model for $x \to 0+$ when the Feller condition is violated. Should that be the case, a more careful analysis of boundary conditions is required, see e.g. Lucic [2008].

11.3.2.2 Forward-from-Backward Induction

The backward and forward Kolmogorov equations (11.29)–(11.30) are consistent in the continuous-time limit, but not necessarily so when discretized, finite difference style. As a result, the function $\vartheta(t)$ uncovered from the algorithm in Section 11.3.2.1 will generally not allow a finite difference grid based on the backward equation (11.25) to recover the initial term structure of discount bonds without errors, even if discretized on t- and x-grids that are identical to those used for the calibration of $\vartheta(t)$. As long as the time line is sufficiently finely spaced, the errors tend to be very small, however, and rarely a cause for concern. Nevertheless, it should be noted that it is, in fact, possible to restate the forward induction algorithm in such a way that the algorithm becomes precisely compatible with the brute-force backward equation approach we discussed earlier.

To develop this approach, we start out with the discretized backward equation (11.27) and rearrange it to yield

$$\widehat{\mathbf{V}}(t_i) = \mathbf{T}_i^{i+1}\widehat{\mathbf{V}}(t_{i+1}) + \mathbf{G}_i^{i+1}, \tag{11.31}$$

where, with $\mathbf{A}_i^{i+1} \triangleq \mathbf{A}((1-\theta)t_{i+1} + \theta t_i)$,

$$\begin{aligned}\mathbf{T}_i^{i+1} &\triangleq \left[\mathbf{I} - \theta\Delta_t\mathbf{A}_i^{i+1}\right]^{-1}\left[\mathbf{I} + (1-\theta)\Delta_t\mathbf{A}_i^{i+1}\right]\\ &= \mathbf{I} + \left[\mathbf{I} - \theta\Delta_t\mathbf{A}_i^{i+1}\right]^{-1}\mathbf{A}_i^{i+1}\Delta_t,\end{aligned}$$

and

$$\mathbf{G}_i^{i+1} = \left[\mathbf{I} - \theta\Delta_t\mathbf{A}_i^{i+1}\right]^{-1}\mathbf{B}(t_i, t_{i+1}).$$

Repeated application of (11.31) yields, for some l,

$$\widehat{\mathbf{V}}(0) = \mathbf{T}_0^l\widehat{\mathbf{V}}(t_l) + \mathbf{G}_0^l, \tag{11.32}$$

where $\mathbf{T}_0^l$ and $\mathbf{G}_0^l$ can be found iteratively from the equations

$$\mathbf{T}_0^{i+1} = \mathbf{T}_0^i\mathbf{T}_i^{i+1}, \quad \mathbf{T}_0^0 = \mathbf{I}, \tag{11.33}$$

$$\mathbf{G}_0^{i+1} = \mathbf{T}_0^i\mathbf{G}_i^{i+1} + \mathbf{G}_0^i, \quad \mathbf{G}_0^0 = \mathbf{0}. \tag{11.34}$$

Before we proceed, let us introduce some notation. First, we let $\mathbf{1}_j$ denote an m-dimensional column vector with j-th element equal to 1 and all other elements equal to zero. Also we set $\mathbf{1}$ to mean a column vector with all elements equal to 1. Consider now the zero-coupon bond maturing at time t_i, and assume that for the grid $\{x_j\}_{j=0}^{m+1}$ the initial value of $x = 0$ sits in position β, i.e. $x_\beta = 0$. By the definition of a discount bond, (11.32) allows us to write

$$P(0, t_i) = \mathbf{1}_\beta^\top\left(\mathbf{T}_0^i\mathbf{1} + \mathbf{G}_0^i\right) = \mathbf{1}^\top\mathbf{D}_0^i + g_0^i, \tag{11.35}$$

where $\mathbf{D}_0^i$ is an m-dimensional vector and g_0^i is a scalar:

$$\mathbf{D}_0^i = \left(\mathbf{T}_0^i\right)^\top \mathbf{1}_\beta, \quad g_0^i = \left(\mathbf{G}_0^i\right)^\top \mathbf{1}_\beta.$$

Assuming that the influence from the finite difference grid boundary is small, we evidently have

$$P(0, t_i) \approx \sum_{j=1}^{m} \left(\mathbf{D}_0^i\right)_j,$$

where $(\mathbf{D}_0^i)_j$ denotes the j-th element of $\mathbf{D}_0^i$. Comparison with (11.28) shows that $\mathbf{D}_0^i$ can be interpreted as the discrete-time Arrow-Debreu security vector for maturity t_i (up to the scaling Δx). From (11.33)–(11.34) we have

$$\mathbf{D}_0^{i+1} = \left(\mathbf{T}_0^i \mathbf{T}_i^{i+1}\right)^\top \mathbf{1}_\beta = \left(\mathbf{T}_i^{i+1}\right)^\top \mathbf{D}_0^i, \quad \mathbf{D}_0^0 = \mathbf{1}_\beta, \tag{11.36}$$

$$g_0^{i+1} = \left(\mathbf{T}_0^i \mathbf{G}_i^{i+1} + \mathbf{G}_0^i\right)^\top \mathbf{1}_\beta = \left(\mathbf{G}_i^{i+1}\right)^\top \mathbf{D}_0^i + g_0^i, \quad g_0^0 = 0. \tag{11.37}$$

Recalling the definition of $\mathbf{T}_i^{i+1}$, it follows that

$$\begin{aligned}
\mathbf{D}_0^{i+1} &= \left(\mathbf{T}_i^{i+1}\right)^\top \mathbf{D}_0^i \\
&= \left[\mathbf{I} + (1-\theta)\Delta_t \mathbf{A}_i^{i+1}\right]^\top \left[\mathbf{I} - \theta\Delta_t \left(\mathbf{A}_i^{i+1}\right)^\top\right]^{-1} \mathbf{D}_0^i \\
&\triangleq \left[\mathbf{I} + (1-\theta)\Delta_t \left(\mathbf{A}_i^{i+1}\right)^\top\right] \mathbf{Y}_0^i,
\end{aligned} \tag{11.38}$$

where $\mathbf{Y}_0^i$ is a vector satisfying a tri-diagonal matrix equation

$$\left[\mathbf{I} - \theta\Delta_t \left(\mathbf{A}_i^{i+1}\right)^\top\right] \mathbf{Y}_0^i = \mathbf{D}_0^i. \tag{11.39}$$

With this, we are ready to state our revised fitting algorithm for $\vartheta(t)$. We assume that $\vartheta_0, \ldots, \vartheta_{i-2}$ have been found, as has $\mathbf{D}_0^{i-1}$ and g_0^{i-1}.

1. Make a guess for ϑ_{i-1}.
2. Solve (11.39) for $\mathbf{Y}_0^{i-1}$.
3. Compute $\mathbf{D}_0^i$ from (11.38); and g_0^i from (11.37).
4. Compute the discount bond price $P(0, t_i)$ from (11.35).
5. If the computed value of $P(0, t_i)$ equals that quoted in the market, then stop; otherwise return to Step 1.

The computational efforts for Steps 2, 3, and 4 are all $O(m)$, so the algorithm above is of the same computational complexity as the algorithm in Section 11.3.2.1.

11.3.2.3 Yield Curve and Volatility Calibration

Volatility calibration of a general short rate model of the type (11.24) is a rather involved affair. The typical scheme — moving model volatilities around until the prices of calibration targets match the market — is beset with complications such as

- Bond reconstitution formulas are unavailable so the model needs to be numerically recalibrated to the initial yield curve after each update of the model volatilities.
- All bonds used in the payoffs of calibration targets need to be computed numerically for each volatility update.
- Bond and calibration target values depend on the entire volatility function, making decoupling of individual target calibrations difficult.

These difficulties, however, have not prevented some major investment banks from risk-managing large derivatives portfolios with such a setup. As we cannot possibly do justice to all the tricks that would be required to make this operational, we content ourselves with presenting a mere outline of a possible algorithm.

We consider the model of the type (11.24) but, for notational convenience, write the volatility term in a separable form

$$dr(t) = \mu_r\left(t, r(t); \vartheta(t)\right) dt + \sigma_r(t)\psi\left(r(t)\right) dW(t). \tag{11.40}$$

Here, the purely-time-dependent function $\sigma_r(t)$ is used to calibrate to swaptions, and $\vartheta(t)$ is used to match the initial yield curve. Implicitly, $\vartheta(t)$ depends on $\sigma_r(t)$.

As in Section 10.1.4, we assume that we are given a collection of $N-1$ swaptions defined on a maturity grid $0 = T_0 < T_1 < \ldots < T_N$ such that the i-th swaption expires at times T_i, $i = 1, \ldots, N-1$. For concreteness assume that all underlying swaps mature on T_N. We further assume that $\sigma_r(t)$ is discretized in a piecewise constant manner on the maturity grid, with σ_i denoting the flat value on $[T_i, T_{i+1})$.

Before discussing calibration, let us outline an efficient algorithm for pricing all swaptions in the calibration set given a collection of volatilities $\sigma_0, \ldots, \sigma_{N-1}$.We implicitly assume that the model is rewritten using a state variable x as in (11.25), and the x-domain is discretized with $\{x_j\}_{j=0}^{m+1}$. As the algorithm involves both forward and backward induction, it is important to follow the approach of Section 11.3.2.2 and use a forward algorithm that is fully compatible with the backward one.

1. Update the volatility function of the model with the new values $\sigma_0, \ldots, \sigma_{N-1}$.
2. Using the forward induction algorithm from Section 11.3.2.2, calibrate $\vartheta(t)$ for all $t \in [0, T_N]$.
3. On Step 2, Arrow-Debreu prices $\widehat{G}(0, 0; T_i, \cdot)$ are calculated; *save them* for all $i = 1, \ldots, N-1$.
4. For each $n = N, \ldots, 2$:
   a) Create a new copy of the finite difference grid.
   b) Populate a payoff $P(T_n, T_n) = 1$ at time T_n.
   c) Calculate $P(T_i, T_n)$ from $P(T_{i+1}, T_n)$ by backward induction for $i = n-1, \ldots, 1$.

5. For each $i = 1, \ldots, N-1$:
   a) Create the T_i-expiry swaption payoff from $P(T_i, T_n)$, $n = i+1, \ldots, N$, calculated on Step 4.
   b) Integrate the payoff against $\widehat{G}(0, 0; T_i, \cdot)$ stored on Step 3.
   c) This gives the value of the T_i-expiry swaption.

This algorithm describes how to map a set of model volatilities $\sigma_0, \ldots, \sigma_{N-1}$ into model prices of calibration targets. In principle, one can now perform a multi-dimensional optimization to match swaption prices from the model to the market. Given that the number of swaptions could be large — 30 or 40 would not be uncommon — and each valuation is rather expensive, the resulting algorithm, while not necessarily completely impractical, would require significant computational resources.

A further improvement entails adopting the iterative bootstrap algorithm outlined in Section 10.2.5. We recall that the main premise of this algorithm approach was that the value of the T_i-expiry swaption depended on $\sigma_r(s)$ for $s \in [0, T_i]$ in a much stronger way than on $\sigma_r(s)$ for $s \in [T_i, T_N]$; this tends to also be true for the model of the type (11.40) for a wide range of volatility specifications. To incorporate this observation into the algorithm above, we would work our way forward from $i = 0$ and determine σ_i by one-dimensional root-search to match the market value of the T_{i+1}-expiry swaption; all values σ_j, $j \neq i$ would be kept constant in the root search. The bootstrap loop would work its way from $i = 0$ to $i = N-1$ and would be repeated a few times until convergence, in the manner described in Section 10.2.5. We trust that the reader gets the idea and can fill in remaining details.

11.3.2.4 The Dybvig Parameterization

In some models, it may be the case that the SDE (11.24) can be reformulated as

$$r(t) = s(t) + \vartheta(t), \tag{11.41}$$

where $\vartheta(t)$ is a free time-dependent function to be fitted against discount bond prices, and $s(t)$ satisfies an SDE

$$ds(t) = \mu_s\left(t, s(t)\right) dt + \sigma_s\left(t, s(t)\right) dW(t). \tag{11.42}$$

We notice that

$$P(0,T) = \mathrm{E}\left(e^{-\int_0^T r(u)du}\right) = e^{-\int_0^T \vartheta(u)du}\mathrm{E}\left(e^{-\int_0^T s(u)du}\right),$$

such that

$$\int_0^T \vartheta(u)du = \ln\left(\frac{\mathrm{E}\left(e^{-\int_0^T s(u)du}\right)}{P(0,T)}\right). \tag{11.43}$$

To the extent that the numerator in the right-hand side of (11.43) is easy to compute — e.g. if the SDE for $s(t)$ permits a closed-form solution — calibration of $\vartheta(t)$ can conveniently be found by direct differentiation, see (11.44).

The specification in (11.41) was proposed in Dybvig [1997] and is, as we have already seen in Section 10.1.1.2, quite natural in the context of Gaussian models. For non-Gaussian models, a "fudge" approach in (11.41) may be less desirable, as the domain of $r(t)$ is hard to control. For instance, suppose that $s(t)$ is a time-homogeneous CIR process

$$ds(t) = \varkappa\,(s_0 - s(t))\,dt + \sigma\sqrt{s(t)}\,dW(t),$$

which is guaranteed to produce only non-negative values of $s(t)$. The combined process[11] $r(t) = s(t) + \vartheta(t)$, however, will have domain $r(t) \in [\vartheta(t), \infty)$ which is rather awkward as $\vartheta(t)$ is largely out of the user's control. This is reflected in the SDE for $r(t)$, where now $\vartheta(t)$ enters into the volatility term:

$$dr(t) = \varkappa\,(s_0 + \vartheta(t) + \vartheta'(t)/\varkappa - r(t))\,dt + \sigma\sqrt{r(t) - \vartheta(t)}\,dW(t),$$

where $\vartheta'(t) = d\vartheta(t)/dt$. By affecting the short rate volatility, the interpretation of $\vartheta(t)$ as only serving to fit the yield curve can no longer be maintained. This conclusion holds not only for CIR models, but for all models where σ_s depends on $s(t)$, since

$$dr(t) = \mu_s\,(t, r(t) - \vartheta(t))\,dt + \vartheta'(t)\,dt + \sigma_s\,(t, r(t) - \vartheta(t))\,dW(t).$$

While sometimes very convenient, the Dybvig parameterization should consequently be approached with considerable care.

11.3.2.5 Link to HJM Models

By construction the Dybvig parameterization in Section 11.3.2.4 ensures that the model is calibrated to the initial forward curve $f(0,t)$, $t > 0$. As the resulting model is driven by Brownian motions, we know that it must be in the HJM class. An interesting question arises: what is the type of an HJM model that is defined through the Dybvig procedure. To answer this, let $s(t)$ satisfy (11.42) and define

$$Q(t,T,s) = \mathrm{E}\left(e^{-\int_t^T s(u)\,du}\,\middle|\, s(t) = s\right).$$

From (11.43), we then get

[11] This model has been advocated by Brigo and Mercurio [2001] as an easy-to-implement alternative to a true time-dependent CIR process. For reasons explained above, this model has certain drawbacks that require careful evaluation.

$$\vartheta(t) = \frac{\partial}{\partial t} \ln \left(\frac{Q(0,t,s(0)}{P(0,t)} \right) \tag{11.44}$$

and

$$P(t,T) = e^{-\int_t^T \vartheta(u)\,du} Q\left(t,T,s(t)\right) = \frac{Q\left(0,t,s(0)\right)}{Q\left(0,T,s(0)\right)} \frac{P(0,T)}{P(0,t)} Q\left(t,T,s(t)\right).$$

As

$$f(t,T) = -\frac{\partial}{\partial T} \ln P(t,T) = -\frac{\partial}{\partial T} Q\left(t,T,s(t)\right) + \vartheta(T),$$

it follows that

$$df(t,T) = O(dt) + \sigma_f(t,T)\,dW(t),$$

where

$$\begin{aligned}\sigma_f(t,T) &= -\frac{\partial^2}{\partial T \partial s} Q\left(t,T,s(t)\right) \sigma_s\left(t,s(t)\right) \\ &= -\frac{\partial^2}{\partial T \partial s} Q\left(t,T,f(t,t) - \vartheta(t)\right) \sigma_s\left(t,f(t,t) - \vartheta(t)\right).\end{aligned}$$

At time t, the forward rate volatility structure generated by the short rate model evidently depends on the forward curve at time t (through $f(t,t)$) as well as the function $\vartheta(t)$. As $\vartheta(t)$ depends (through (11.44)) on the forward curve at time 0, it follows that the HJM dynamics here generally have "memory" of the initial condition at time 0. If one were to alter these initial conditions, the form of HJM dynamics would fundamentally change.

Looking back to Section 10.1.2.2 where the Gaussian short rate model was developed from a separable HJM model, no dependence in the HJM dynamics on initial conditions arose. Hence, it is clear that not *all* short rate models generate "memory" in the HJM dynamics. This raises the obvious question: under which circumstances will a finite-dimensional Markov HJM model have no dependence in its dynamics of the initial forward curve at time 0? The answer to this is listed in Filipovic and Teichmann [2004] which shows that "essentially" all such models must be time-inhomogeneous *affine* models. There are considerable technical details involved in the exact statement of the result, all of which can be found in Filipovic and Teichmann [2004].

11.3.2.6 The Hagan and Woodward Parameterization

Many of the short rate models considered so far have a free time-dependent parameter in the *drift* of the state variable dynamics; the forward induction algorithm can then be used to set this parameter to match the initial yield curve. Hagan and Woodward [1999a] propose an interesting twist on the idea, where the free parameter is introduced into the *numeraire*.

Hagan and Woodward [1999a] start with an observation that only two ingredients are required to create an interest rate model:

- A set of stochastic processes that drive the evolution of interest rates.
- A positive-valued process that is used as a numeraire.

Once the numeraire is specified, the values of all instruments are recovered by the standard pricing formula, see Chapter 1. Critically, the numeraire does not need to be the money market account, or any other "identifiable" security such as a discount bond or an annuity — a positive process is all that is required (so in fact we need not a numeraire but a *deflator* as defined in Section 1.3, but in this section we use the two terms interchangeably).

We define the stochastic process that drives interest rates by a general one-dimensional[12] process

$$dx(t) = \mu_x(t, x(t))\,dt + \sigma_x(t, x(t))\,dW(t), \quad x(0) = 0. \tag{11.45}$$

Furthermore, we choose the deflator to be a function of the state variable $x(t)$. Without loss of generality, we specify

$$N(t) = \frac{1}{P(0,t)} e^{h(t,x(t))+a(t)}. \tag{11.46}$$

Here $h(t,x)$ is user-specified, and $a(t)$ is used to fit the model to the initial yield curve. It is often natural to normalize the parameters such that

$$h(t,0) = 0, \quad a(0) = 0.$$

Once the deflator is specified, we assume that (11.45) is, in fact, given under the measure Q^N associated with this deflator.

Let E^N denote expectations in measure Q^N. The time t price of a T-maturity discount bond in this model, as a function of the state variable x, is given by

$$\begin{aligned} P(t,T,x) &= N(t)\mathrm{E}^N\left(N(T)^{-1}\middle|\, x(t) = x\right) \\ &= \frac{P(0,T)}{P(0,t)} e^{h(t,x)+a(t)-a(T)} \mathrm{E}_t^N\left(e^{-h(T,x(T))}\right). \end{aligned} \tag{11.47}$$

Consistency with the initial yield curve requires

$$P(0,T,x(0)) = P(0,T), \quad T \geq 0,$$

and we obtain the following condition on $a(T)$,

$$a(T) = \ln \mathrm{E}^N\left(e^{-h(T,x(T))}\right), \quad T \geq 0. \tag{11.48}$$

Hagan and Woodward [1999a] show that if the model (11.45)–(11.46) is consistent with the initial yield curve, i.e. the condition (11.48) is satisfied, the model is in fact arbitrage free.

[12] Multi-dimensional extensions are possible.

To obtain $a(t)$ from (11.48), one should not solve for $\mathrm{E}^N(e^{-h(T,x(T))})$ with a backward PDE. Instead, similarly to Section 11.3.2, one should use forward PDE to obtain $p(t,x)$, the density $\mathrm{Q}^N(x(t) \in dx)/dx$, for $(t,x) \in \mathbb{R}^+ \times \mathbb{R}$. The forward Kolmogorov equation states that

$$-\frac{\partial p}{\partial t}(t,x) - \frac{\partial}{\partial x}(\mu_x(t,x)p(t,x)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma_x(t,x)^2 p(t,x)) = 0,\ p(0,x) = \delta(x).$$

Once $p(t,x)$ is determined, we obtain

$$a(T) = \ln \int e^{-h(T,x)} p\,(T,x)\ dx, \quad T \geq 0,$$

where the integral is taken over the range of the random variable $x(T)$. Interestingly, the calibration of $a(t)$ is independent of the initial yield curve $P(0,T)$, $T \geq 0$. The calibration is somewhat faster than the forward induction algorithm of Section 11.3.2.1 for general short rate models as it requires only a *single* forward pass of the finite difference scheme.

Zero-coupon discount bonds are obtained via (11.47), i.e., generally, numerically. For special choices of $\mu_x(t,x)$, $\sigma_x(t,x)$ and $h(t,x)$, closed-form formulas could be available.

Let us define $\gamma(t,x;T)$ by

$$\gamma\,(t,x(t);T) = \mathrm{E}_t^N\left(e^{-h(T,x(T))}\right),$$

so that

$$P\,(t,T,x) = \frac{P\,(0,T)}{P\,(0,t)} e^{h(t,x)+a(t)-a(T)} \gamma\,(t,x;T)\,.$$

Then instantaneous forward rates are given by

$$f\,(t,T,x) = f\,(0,T) + a'(T) - \frac{\partial \ln \gamma\,(t,x;T)}{\partial T}.$$

Applying Ito's lemma to $e^{-h(T,x(T))}$ and setting $T = t$, we observe that the short rate $r(t) = f(t,t,x(t)) = r(t,x(t))$ is given by

$$r\,(t,x) = f\,(0,t) + a'(t) - \left.\frac{\partial \ln \gamma\,(t,x;T)}{\partial T}\right|_{T=t} \tag{11.49}$$

$$= f\,(0,t) + a'(t) + \frac{\partial h\,(t,x)}{\partial t} + \mu_x(t,x)\frac{\partial h\,(t,x)}{\partial x} + \frac{1}{2}\sigma_x(t,x)^2\left(\frac{\partial^2 h\,(t,x)}{\partial x^2} - \left(\frac{\partial h\,(t,x)}{\partial x}\right)^2\right). \tag{11.50}$$

To make matters more concrete, let us now specialize to the case where $h(t,x) = h(t)x$ and $\mu_x(t,x) = 0$ in the general framework (11.45)–(11.46). One can show that this class of models includes the one-factor Gaussian

short rate model and the affine models. If one relaxes the requirement that $\mu_x(t,x) \equiv 0$, then the BK model is also in the class. With this restricted parameterization (11.50) yields, after ignoring small convexity terms,

$$r(t) \approx f(0,t) + a'(t) + h'(t)x(t), \tag{11.51}$$

and, approximately,

$$dr(t) \approx (\vartheta_r(t) - \varkappa_r(t)r(t))\, dt + \sigma_r(t, r(t))\, dW(t),$$

where

$$\varkappa_r(t) = -\frac{h''(t)}{h'(t)}, \tag{11.52}$$

with $\vartheta_r(t)$, $\sigma_r(t,r)$ appropriately defined. Hence, the numeraire scaling $h(t)$ can be conceptually linked to the mean reversion parameter for the short rate.

As a practical example of the general approach, Hagan and Woodward [1999a] propose the following class of "$\beta - \eta$" models:

$$dx(t) = \lambda(t)\,(1 + \beta x(t))^{\eta}\, dW(t),$$
$$N(t) = \frac{1}{P(0,t)} e^{h(t)x(t)+a(t)}.$$

For this specification, the transition density of $x(t)$ is known in closed form, allowing for (more or less) analytical calibration to the initial yield curve. The parameters β, η are used to match the skew of the volatility smile. Note the resemblance between the volatility term in this model and that of a vanilla displaced-CEV model in Section 7.2.4. As pointed out in that section, adding a displacement to the CEV function does not significantly alter the range of available volatility smiles. Hence, should this approach be pursued, we recommend the specialization of the $\beta - \eta$ model with $\eta = 1$, and perhaps an extension of the skew parameter β to be time dependent. Specifying a constant mean reversion $\varkappa_r(t) \equiv \varkappa_r$ and solving (11.52) for $h(t)$ (which we normalize, conveniently and without loss of generality, by $h(0) = 0$ and $h'(0) = 1$), we obtain

$$dx(t) = \lambda(t)\,(1 + \beta(t)x(t))\, dW(t),$$
$$N(t) = \frac{1}{P(0,t)} \exp\left(\frac{1 - e^{-\varkappa_r t}}{\varkappa_r} x(t) + a(t)\right).$$

In conclusion, we note a fairly strong resemblance between the Dybvig parameterization and the approach of Hagan and Woodward. Indeed, comparing (11.51) and (11.41) we see that $a'(t)$ in the former plays pretty much the same role as $\vartheta(t)$ in the latter. The initial yield curve fit conditions, (11.48) and (11.43), are also rather similar. Hence, our words of caution with regards to the Dybvig parameterization apply here as well.

11.3.3 Monte Carlo Simulation

11.3.3.1 SDE Discretization

For the purposes of securities pricing by Monte Carlo methods, we are generally interested in advancing not only $r(t)$ through time, but also the inverse of the money market numeraire

$$\exp\left(-\int_0^t r(u)du\right) = P(0,t)\exp\left(-\int_0^t x(u)du\right) \triangleq P(0,t)Y(t), \quad (11.53)$$

where we recall that $x(t) \equiv r(t) - f(0,t)$. Our starting point can be[13] the vector SDE

$$d\begin{pmatrix} x(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} \mu_x(t, x(t)) \\ -x(t)Y(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_x(t, x(t)) \\ 0 \end{pmatrix} dW(t), \quad (11.54)$$

where the functions μ_x and σ_x were defined in (11.26) above.

In general, simulation of (11.54) requires usage of discretization methods, several of which were introduced in Chapter 3. For instance, the *Euler scheme* for (11.54) would advance the SDE for time t_i to t_{i+1} according to

$$\begin{pmatrix} \widehat{x}_{i+1} \\ \widehat{Y}_{i+1} \end{pmatrix} = \begin{pmatrix} \widehat{x}_i \\ \widehat{Y}_i \end{pmatrix} + \begin{pmatrix} \mu_x(t_i, \widehat{x}_i) \\ -\widehat{Y}_i\widehat{x}_i \end{pmatrix} \Delta_i + \begin{pmatrix} \sigma_x(t_i, \widehat{x}_i) \\ 0 \end{pmatrix} Z_i\sqrt{\Delta_i}, \ \Delta_i = t_{i+1} - t_i,$$

where $\widehat{x}_i = \widehat{x}(t_i)$, $\widehat{Y}_i = \widehat{Y}(t_i)$, and $Z_i \sim \mathcal{N}(0,1)$ is a sample from a standard Gaussian distribution. The Euler scheme is of (weak) convergence order one in the time step. To improve this, a second-order Milstein scheme can be constructed by Ito-Taylor expanding (11.54) to second order, using the technique in Section 3.2.6.3. The construction is tedious but straightforward (see Chapter IV in Andersen [1996] for the details), so we skip it and only show the final result

$$\begin{aligned} \widehat{x}_{i+1} = \widehat{x}_i &+ \left(\mu_x(t_i, \widehat{x}_i) - \frac{1}{2}\mathcal{L}_1\sigma_x(t_i, \widehat{x}_i)\right)\Delta_i + \sigma_x(t_i, \widehat{x}_i)Z_i\sqrt{\Delta_i} \\ &+ \frac{1}{2}\left(\mathcal{L}_1\mu_x(t_i, \widehat{x}_i) + \mathcal{L}_0\sigma_x(t_i, \widehat{x}_i)\right)Z_i\Delta_i\sqrt{\Delta_i} \\ &+ \frac{1}{2}\mathcal{L}_0\mu_x(t_i, \widehat{x}_i)\Delta_i^2 + \frac{1}{2}\mathcal{L}_1\sigma_x(t_i, \widehat{x}_i)Z_i^2\Delta_i, \end{aligned} \quad (11.55)$$

$$\widehat{Y}_{i+1} = \widehat{Y}_i\left(1 - \widehat{x}_i\Delta_i + \frac{1}{2}\left(\widehat{x}_i^2 - \mu_x(t_i, \widehat{x}_i)\right)\Delta_i^2 - \frac{1}{2}\sigma_x(t_i, \widehat{x}_i)Z_i\Delta_i\sqrt{\Delta_i}\right), \quad (11.56)$$

where we have introduced differential operators

[13] Instead of discretizing $Y(t)$, we could also discretize $I(t) = \ln Y(t)$, as in Section 10.1.6.1. For variation we use $Y(t)$ in this section.

$$\mathcal{L}_0 = \frac{\partial}{\partial t} + \mu_x(t,x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma_x(t,x)^2\frac{\partial^2}{\partial x^2}, \quad \mathcal{L}_1 = \sigma_x(t,x)\frac{\partial}{\partial x}.$$

The Milstein scheme (11.55)–(11.56) is rather formidable-looking, and its practical efficiency tends to be quite model-dependent. Still, using an affine model as a test case Andersen [1996] shows that the Milstein scheme outperforms the Euler scheme handily, even after taking the additional computational burden of (11.55)–(11.56) into consideration. Schemes with order higher than two can be constructed along similar principles, but will, in our experience, rarely be worth the hassle. We also remind the reader that higher-order schemes can be constructed by Richardson extrapolation, as discussed in Section 3.2.7. Andersen [1996] reports modest gains for a third-order scheme constructed by Richardson extrapolation of the Milstein scheme above.

At this point, let us note that for European-style securities paying a single cash flow at time T, the ideas of Section 10.1.6.3 can be applied here, and the burden of simulating $Y(t)$ could be avoided by a change to the T-forward measure Q^T. For securities that pay intermediate cash flows, however, matters are more complicated as these flows must effectively be future-valued to time T. For instance, a random coupon c paid at time $T' < T$ will require us to compute the numeraire-deflated value $c/P(T',T)$. But here, unfortunately, the quantity $P(T',T)$ is generally not known analytically at time T' as a function of the model state variables. Of course, without affecting the economics of the trade one could invest the proceeds c into a money market account β at time T', yielding the payout $c\beta(T)/\beta(T')$ at time T. Evaluating this payout, however, would again require us to keep track of $Y(t)$, at least on the interval $[T',T]$. This problem, however, can be avoided by using the spot measure instead of the forward measure, as outlined in Section 10.1.6.3. Much more material about numeraire simulation strategies can be found in Chapter 14.

Finally, a note on variance reduction for short rate model simulation. A systematic discussion of variance reduction techniques for short rate models can be found in Chapter IV of Andersen [1996] and Andersen and Boyle [2000]. Most of the methods discussed in these sources can be found in the survey of Section 3.4 and shall not be repeated. We do, however, highlight here the particularly useful idea of applying importance sampling based on information extracted from a tractable (e.g. Gaussian or affine) approximation to the short rate SDE. We postpone the discussion of this technique, which relies on the material in Section 3.4.4.3, to Chapter 25.

11.3.3.2 Practical Issues with Monte Carlo Methods

As was the case for finite difference methods (see discussion in Section 11.3.1), whenever an explicit bond reconstitution formula is lacking, the effort and complexity required to price derivatives by Monte Carlo methods

increase significantly. For instance, consider applying Monte Carlo methods to the pricing of short-dated expiry T call option on a long-dated maturity T^* discount bond. This price (V) is computed as a risk-neutral expectation

$$\begin{aligned} V(0) &= P(0,T)\mathrm{E}\left(Y(T)^{-1}\left(P(T,T^*)-K\right)^+\right) \\ &= P(0,T)\mathrm{E}\left(Y(T)^{-1}\left(\mathrm{E}_T\left(e^{-\int_T^{T^*} x(u)du-\int_T^{T^*} f(0,u)du}\right)-K\right)^+\right) \\ &= P(0,T^*)\mathrm{E}\left(Y(T)^{-1}\left(\mathrm{E}_T\left(e^{-\int_T^{T^*} x(u)du}\right)-K\frac{P(0,T)}{P(0,T^*)}\right)^+\right). \end{aligned}$$

An immediate problem is here the fact that the inner time T expectation

$$\mathrm{E}_T\left(e^{-\int_T^{T^*} x(u)du}\right)$$

is not explicitly known as a function of $x(T)$, but must itself be computed by numerical methods. A brute-force approach involves estimating the expectation by Monte Carlo methods, launching a "simulation-within-a-simulation" at time T. The computational expense involved in such a scheme would most likely be prohibitive. Alternatives involve using a regression on a space of basis functions to estimate the function

$$Q(T,T^*,x) = \mathrm{E}\left(e^{-\int_T^{T^*} x(u)du}\middle| x(T)=x\right);$$

we discuss this approach in some detail in Chapter 18.

Alternatively, we can always estimate $Q(T,T^*,x)$ by finite difference methods, as in Section 11.3.1. Combining finite difference methods and Monte Carlo methods for the purposes of pricing a European option on a discount bond makes little practical sense, of course, as we would always prefer finite difference methods for this payout. For path-dependent options, however, this idea may in fact be the best way of computing option prices. Loosely, such a scheme would use a finite difference grid to pre-compute zero-coupon bond prices $P(t_i,\cdot,x)$ on a grid $x \in \{x_j\}_{j=0}^{m+1}$, at all dates t_i, $i = 1,2,\ldots,N$, required by the path-dependent payout function considered. When paths for $x(t)$ in a subsequent step are generated by Monte Carlo simulation, interpolation of the N discount bond price vectors of dimension $m+2$ available in the finite difference grid would allow us to compute rapidly discount bond prices $Q(t_i,\cdot,x(t_i))$, $i = 1,2,\ldots$, at all relevant dates. We can use the schemes in Section 11.3.3.1 for the purpose of drawing paths of $x(t)$. If, however, we wish to make the dynamics for $x(t)$ perfectly consistent with the finite difference grid, we can use the forward induction techniques of Sections 11.3.2.1 and 11.3.2.2 to work out the (discrete) transition probabilities for $x(t)$ implied by the finite difference grid. Paths for $x(t)$ can then be generated directly from these probabilities. With this approach, we only draw values of $x(t)$ on the spatial grid $\{x_j\}_{j=0}^{m+1}$ of the finite difference grid, and therefore never have to apply interpolation methods when looking up discount bond prices.

11.A Appendix: Markov-Functional Models

The purpose of this appendix is to give a brief account of the class of *Markov-functional* (MF) models. We only consider the one-factor case. Extensions to higher dimensions are possible, but practical implementation challenges tend to increase substantially for dimensions higher than one. MF models were introduced in Kennedy et al. [2000] and while their popularity is generally waning, they are still used in some banks.

11.A.1 State Process and Numeraire Mapping

We have already observed in Section 11.3.2.6 that to define an arbitrage-free interest rate model we really only need two ingredients: a stochastic process that drives the evolution of interest rates, and a functional form that maps that process into a numeraire. The development of Markov-functional models normally starts with specializing this setup to a numeraire taken to be the discount bond $P(\cdot, T^*)$ to the final maturity of interest T^*, and a Markov stochastic process that is Gaussian in the corresponding terminal measure Q^* (see Section 4.2.4). Assuming arbitrarily that $x(t)$ is a Q^*-martingale, we write

$$dx(t) = \sigma(t)e^{\varkappa t}\, dW^*(t), \quad x(0) = 0, \tag{11.57}$$

where W^* is a one-dimensional Brownian motion in the terminal measure, and where we for simplicity have assumed that the mean reversion $\varkappa$ is constant. The role of $\varkappa$ is to control inter-temporal correlations in the model; see Section 13.1.8.1 for the importance of this. The transition density of $x(t)$ in Q^* is, trivially,

$$\begin{aligned} p\left(y, t; z, s\right) &\triangleq \mathrm{Q}^*\left(x(s) \in [z, z+dz] | x(t) = y\right)/dz \\ &= \frac{1}{\sqrt{2\pi}v(t,s)} \exp\left(\frac{-\left(z-y\right)^2}{2v(t,s)^2}\right), \quad s > t, \end{aligned}$$

where

$$v(t,s)^2 = \int_t^s \sigma(u)^2 e^{2\varkappa u}\, du.$$

In the spirit of Section 11.3.2.6, we define $P(t, T^*)$ to be a deterministic function of the state variable process $x(t)$,

$$P(t, T^*) = P(t, T^*, x(t)), \quad P(t, T^*, x) = H\left(t, x\right), \tag{11.58}$$

for some exogenously given function $H : \mathbb{R}^2 \to [0, 1]$. As we recall, this is sufficient to define all discount bonds in the model since, for any $0 \le t < T \le T^*$,

$$P(t, T) = P(t, T^*)\mathrm{E}_t^*\left(\frac{1}{P(T, T^*)}\right). \tag{11.59}$$

where E^* denotes expectation in measure Q^*. This allows us to express all discount bonds as functions of $x(t)$:

$$\begin{aligned} P(t,T) &= P(t,T,x(t)), \\ P(t,T,x) &= P(t,T^*,x)\mathrm{E}^*\left(\frac{1}{P(T,T^*,x(T))}\middle| x(t)=x\right) \\ &= H(t,x)\int_{-\infty}^{\infty}\frac{p(x,t;z,T)}{H(T,z)}\,dz. \end{aligned} \tag{11.60}$$

The formula (11.59) can be specialized to $t=0$, yielding

$$P(0,T) = P(0,T^*)\mathrm{E}^*\left(\frac{1}{P(T,T^*)}\right) = P(0,T^*)\mathrm{E}^*\left(\frac{1}{H(T,x(T))}\right), \tag{11.61}$$

which constitutes a *no-arbitrage condition* on the mapping function $H(\cdot,\cdot)$. This condition is often used to choose a particular function $H(\cdot,\cdot)$ from a given parametric family; compare this to condition (11.48) in Section 11.3.2.6.

In practice, the numeraire mapping function $H(t,x)$ in (11.58) is often specified only indirectly, through definition of functional forms for market rates, such as Libor or swap rates. The following two sections explore variations on this idea.

11.A.2 Libor MF Parameterization

Let us assume that a tenor structure

$$0 = T_0 < T_1 < \ldots < T_N = T^*, \quad \tau_n = T_{n+1} - T_n,$$

is given, and define spanning forward Libor rates by

$$L_n(t) \triangleq L(t,T_n,T_{n+1}) = \left(\frac{P(t,T_n)}{P(t,T_{n+1})} - 1\right)\tau_n^{-1}, \quad n = 0,\ldots,N-1, \tag{11.62}$$

(see (4.2)). It turns out that, if we can specify the mapping of the state process $x(\cdot)$ into Libor rates on their fixing dates, $L_n(T_n)$, for all $n = 1,\ldots,N-1$:

$$L_n(T_n) = l_n(x(T_n)), \quad n = 1,\ldots,N-1,$$

then this is sufficient to recover the numeraire-mapping function $H(T_n,\cdot)$, $n = 1,\ldots,N$, on tenor dates[14] and consequently define the MF model by

[14] With this approach, the numeraire-mapping function is undefined for times that are not in the tenor structure. The "discrete" nature of the resulting model is one of the common criticisms of the MF approach. Pragmatically, it means that *all* dates of interest for a particular derivative security should be added to the tenor structure, or interpolation schemes not unlike those considered in Section 15.1 need to be designed.

(11.59). We show this by induction on the fixing time T_n, for $n = N-1, \ldots, 1$. The starting point of the induction follows directly from (11.62) as we have

$$H(T_{N-1}, x) = P(T_{N-1}, T^*, x) = P(T_{N-1}, T_N, x) = (1 + \tau_{N-1} l_{N-1}(x))^{-1}. \tag{11.63}$$

For the induction step, let us assume that $H(T_i, x)$ are known for $i = n+1, \ldots, N-1$. By (11.59) we have

$$\frac{P(T_n, T_{n+1})}{P(T_n, T^*)} = \mathrm{E}^*_{T_n}\left(\frac{1}{P(T_{n+1}, T^*)}\right)$$

which implies that

$$\frac{1}{H(T_n, x)} = (1 + \tau_n l_n(x))\, \mathrm{E}^*_{T_n}\left(\frac{1}{H(T_{n+1}, x(T_{n+1}))}\middle| x(T_n) = x\right), \tag{11.64}$$

and the statement follows.

The consistency condition (11.61) is often used to select a particular function $l_n(\cdot)$ from a parametric family, for each n. To explain this, let us first consider what functional forms for $l_n(\cdot)$ are typically used. Suppose we desire to build a model where Libor rates on the tenor structure are close to log-normal[15]. Then, with $v_n = v(0, T_n)$, we fundamentally would want something like

$$L_n(T_n) \approx L_n(0) \exp\left(k_n x(T_n) - \frac{1}{2} k_n^2 v_n^2\right), \tag{11.65}$$

where

$$k_n = \frac{e^{-\varkappa T_n} - e^{-\varkappa T_{n+1}}}{\varkappa T_n},$$

to hold for all n. The particular form of k_n's, as well as the volatility parameterization (11.57), are strongly inspired by the Gaussian short rate model, see Proposition 10.1.7. (To see the connection more clearly the reader should note that the state variable $x(t)$ here is related to $x(t)$ in Proposition 10.1.7 by a multiplicative scaling of $e^{\varkappa t}$; compare (11.57) to (10.16) and disregard $y(t)$ in the latter.) Note that the quantity

$$T_n^{-1/2} k_n v_n$$

after calibration should be close to the implied Black volatility of a caplet maturing at time T_n. To preclude arbitrage, (11.65) cannot be used as is for all n (since only $L_{N-1}(t)$ is a martingale in Q^*), so we could, for instance, add a "convexity multiplier" c_n and write

$$l_n(x; c_n) = c_n L_n(0) \exp\left(k_n x - \frac{1}{2} k_n^2 v_n^2\right).$$

[15] It is straightforward to extend the arguments to the case of displaced log-normal Libor rates, say. We leave this to the reader.

While v_n's (or, equivalently, the mean reversion $\varkappa$ and the model volatility $\sigma(t)$) can be treated as free constants to be calibrated to option prices, we would use (11.61) to set c_n's such that the initial yield curve is replicated by the model. It is trivial to see that $c_{N-1} = 1$ and

$$H_{N-1}(T_{N-1}, x) = (1 + \tau_{N-1} l_{N-1}(x; 1))^{-1},$$

wherefore other c_n's may be obtained as solutions to

$$\mathrm{E}^*_{T_n}\left((1 + \tau_n l_n(x(T_n); c_n))\, \mathrm{E}^*_{T_n}\left(\frac{1}{H(T_{n+1}, x(T_{n+1}))}\middle| x(T_n)\right)\right) = P(0, T_n)$$

for $n = N-2, \ldots, 1$.

11.A.3 Swap MF Parameterization

Defining an MF model in terms of Libor rates is especially convenient if the model is meant to price a security that depends primarily on Libor rates (on their fixing dates), e.g. a TARN (see Section 5.15 and Chapter 20). In particular, the Libor MF parameterization allows one direct control over volatilities and other distributional characteristics of Libor rates, which makes it fairly straightforward to set up a calibration scheme that is suitable for the security (see e.g. Section 20.1.3). Libor rates, however, are not always the primary driving factors; for example, prices of Bermudan swaptions are arguably more directly linked to distributions of swap rates (see Section 19.2). Fortunately, MF models can be formulated in terms of swap rates as well, as we shall now demonstrate. It is worth noting that the relationship between Libor and swap MF models is similar to that between Libor and swap *market* models which we explore in Section 15.4.

For concreteness let us consider a set of so-called "core" swap rates,

$$S_n(t) \triangleq S_{n,N-n}(t) = \frac{P(t, T_n) - P(t, T^*)}{A_n(t)}, \tag{11.66}$$

$$A_n(t) \triangleq A_{n,N-n}(t) = \sum_{i=n}^{N-1} \tau_i P(t, T_{i+1}),$$

$n = 1, \ldots, N-1$, where we used the notations (4.8), (4.10) for $A_{k,m}$ and $S_{k,m}$. We assume that the core swap rates on their fixing dates are specified as deterministic functions $s_n(x)$ of the state process,

$$S_n(T_n) = s_n(x(T_n)), \quad n = 1, \ldots, N-1.$$

As for the Libor-based specification above, we claim that the knowledge of the functions $\{s_n(\cdot)\}_{n=1}^{N-1}$ (together with the dynamics of the state process $x(t)$) is sufficient to define the numeraire mapping $H(\cdot, \cdot)$ and, therefore, an arbitrage-free model of interest rates.

The proof also proceeds by induction. As we have that $S_{N-1}(T_{N-1}) = L_{N-1}(T_{N-1})$, the starting point of induction is given by (11.63), i.e.

$$H\left(T_{N-1}, x\right) = \left(1 + \tau_{N-1} s_{N-1}(x)\right)^{-1}.$$

For the induction step $n+1 \to n$, we note from (11.66) that

$$\begin{aligned}\frac{1}{P(T_n, T^*)} &= 1 + S_n(T_n)\frac{A_n(T_n)}{P(T_n, T^*)} \\ &= 1 + S_n(T_n)\sum_{i=n}^{N-1}\tau_i \frac{P(T_n, T_{i+1})}{P(T_n, T^*)}\end{aligned}$$

and so we have (compare to (11.64))

$$\frac{1}{H(T_n, x)} = 1 + s_n(x)\sum_{i=n}^{N-1}\tau_i \mathrm{E}^*_{T_n}\left(\frac{1}{H(T_{i+1}, x(T_{i+1}))}\middle| x(T_n) = x\right).$$

As in the Libor specification, we can choose functions $\{s_n(\cdot)\}$ to approximate log-normal (or displaced log-normal) distribution of swap rates. Also in direct analogy to the Libor case, we typically have some no-arbitrage conditions to satisfy, usually by means of setting some parameters in the specific functional form of $\{s_n(\cdot)\}$. We leave these details for the reader to explore.

11.A.4 Non-Parametric Calibration

So far we defined Markov-functional models by specific parametric mappings of the state process into Libor and swap rates. Originally, however, the class of models was introduced in a non-parametric way (see Hunt and Kennedy [2000] for a typical treatment), where mapping functions are deduced from market prices of caplets or swaptions across all strikes. While we typically prefer the parametric approach (for reasons we touch upon below), let us nevertheless quickly review the non-parametric method for completeness.

Through equation (11.60), we can turn the payout of any T-maturity security that depends on the state of the yield curve into a function of $x(T)$, $g(x(T); K, T)$ say, where K is some payout parameter (virtually always a strike). The time 0 price of this security is

$$\begin{aligned}V(0; K, T) &= P(0, T^*)\mathrm{E}^*\left(\frac{g\left(x(T)\right)}{P(T, T^*)}\right) \\ &= P(0, T^*)\int_{-\infty}^{\infty} p\left(0, 0; z, T\right)\frac{g(z; K)}{H\left(T, z\right)}\,dz. \qquad (11.67)\end{aligned}$$

Assuming that $H(T, x)$ is invertible in x, we may write (11.67) as

$$V(0;K,T) = P(0,T^*)\int_{-\infty}^{\infty} p\,(0,0;z,T)\,q\,(H(T,z);K,T)\;dz, \tag{11.68}$$

where

$$q\,(H(T,z);K,T) \triangleq \frac{g(z;K,T)}{H\,(T,z)}.$$

If $V(0;K,T_n)$ is known[16] for a continuum of parameters (strikes) K, (11.68) defines an integral equation that may allow one to uncover the function $H(T_n,\cdot)$ (or, often more conveniently, $l_n(\cdot)$ or $s_n(\cdot)$).

Solution of (11.68) is typically done for a fixed number of M strikes, with $H(T,x)$ solved for on a grid $\{x_j\}_{j=1}^M$. In practice, this procedure is difficult to make fully robust, and the numerical solution is often prone to instabilities at long maturities, even if sophisticated special-purpose numerical techniques are employed (see Hunt and Kennedy [2000] for such techniques, many of which rely on the fact that polynomials can be integrated exactly against the Gaussian density). Even when numerically stable, a non-parametric solution for $H(\cdot,\cdot)$ may imply unrealistic evolution of the volatility smile through time, a general feature of local volatility models as explained in Section 7.1.3. To avoid these issues we may either pre-smooth the option prices used for calibration purposes (e.g. by best-fitting a CEV or a displaced log-normal model to the market smile), or, preferably in our opinion, we may use a low-dimensional parametric form for $H(t,x)$ as in Sections 11.A.2 and 11.A.3.

11.A.5 Numerical Implementation

Numerical securities valuation in an MF model is typically quite simple as the state process is Gaussian. Let us assume that the function $H(t,x)$ has been established, and consider, say, implementation of the model in a finite difference grid. Let the derivative value function be $V(t,x(t))$, and set $V^*(t,x) = V(t,x)/H(t,x)$ (such that $V(0,0) = P(0,T^*)V^*(0,0)$). As $V^*(t,x)$ must be a Q^*-martingale, we can write

$$\frac{\partial V^*}{\partial t} + \frac{1}{2}\sigma(t)^2 e^{2\varkappa t}\frac{\partial^2 V^*}{\partial x^2} = 0, \tag{11.69}$$

subject to appropriate terminal and intermediate jump payout conditions. In evaluating terminal and intermediate payout conditions, we would typically need to apply the numerical expression (11.60) to establish the state of the yield curve. We should note that the MF literature generally prefers to use Gaussian integration methods (rather than standard PDE solvers) to evaluate the PDE (11.69), see Hunt and Kennedy [2000] for details.

[16] We could use market prices for this, or we could use option prices computed from a vanilla model that we wish for the MF model to emulate.

11.A.6 Comments and Comparisons

The one-factor MF model competes with a number of models in this book, especially the quasi-Gaussian class in Chapter 13. The quasi-Gaussian model allows for arbitrary local volatility (as does the MF model), but has closed-form formulas that allow for reconstituting the term structure of discount bonds *analytically* from the underlying state variable, rather than through numerical integration (see (11.60)). In addition, the quasi-Gaussian model is substantially more "direct" in its modeling of the forward curve and has an easy-to-state term structure of instantaneous forward rate volatilities. This, in turn, makes the model more transparent in its causality structure — especially when it comes to the evolution of the volatility term structure and smile — and often makes it easy to devise good closed-form approximations for swaption and cap prices. As a consequence, calibration of quasi-Gaussian models to option and bond prices is virtually always much faster than for MF models. In addition, quasi-Gaussian models are quite straightforward to extend to high dimensions and to stochastic volatility dynamics; these extensions are far more difficult[17] for MF models. On the flip side, a quasi-Gaussian model involving a single Brownian motion involves a *two-dimensional* state vector process, which makes derivatives pricing by finite difference methods slower than for MF models[18]. For most applications, total computation time of calibration and valuation is, however, less for the quasi-Gaussian model.

Due to its flexibility, extensibility, transparency, and ease of numerical implementation (no integration tricks are required), we generally prefer the quasi-Gaussian model over the MF model, and consequently dedicate an entire chapter to the former — and only this appendix to the latter. For those interested in learning more about MF models than what we offered here, Kennedy et al. [2000] and Hunt and Kennedy [2000] are good starting points.

[17] Indeed, we are unaware of the existence of any published MF models with stochastic volatility.

[18] As we shall see in Chapter 13, one component in the state vector is locally deterministic (i.e. it involves no Brownian motion term), so in a sense the quasi-Gaussian model has a state process dimension of "one-and-a-half", allowing for significant speed-ups in the numerical implementation.

12

Multi-Factor Short Rate Models

Short rate models with only a single driving Brownian motion imply that the instantaneous correlation between forward rates at different maturities is one, a prediction that is demonstrably contrary to reality, as we show in Chapter 14. While many standard securities are, as it turns out, only weakly affected by correlations across the term structure of forward rates, this may not be the case for exotic securities, especially the ones that depend in a non-linear way on the spread between rates of different maturities. Indeed, as a general rule all derivatives that have payouts[1] exhibiting significant convexity to non-parallel moves of the forward curve must *not* be priced in a one-factor model.

In this chapter, we proceed to extend the material from Chapters 10 and 11 to cover the case of multiple driving Brownian motions. This will allow us to properly deal with securities that depend on non-parallel forward curve moves, and will also entail more subtle benefits, including the ability to model non-monotonic volatility term structures in fully time-stationary fashion. The role of the traditional multi-factor short rate models in modern derivatives pricing is, even more so than for the one-factor models, increasingly limited, as more sophisticated multi-factor frameworks have emerged over the last decade. We shall have ample opportunity to address these developments in future chapters, but a brief treatment here of the multi-factor short rate model class is still worthwhile.

As multi-factor short rate models are typically substantially more demanding to handle numerically than are one-factor models, analytical tractability is key to making multi-factor models operational. For instance, a completely generic SDE specification of a multi-factor model (along the

[1] The judgment of whether a security is convex in forward rate twists and tilts can often be quite difficult. Some securities that one might guess should be sensitive to forward rate correlation in fact only display material sensitivity to forward rate *auto*-correlation. Bermudan swaptions are a good example; see Chapter 19 for more details.

lines of Section 11.3) will require significant computational effort to calibrate to market yields and volatilities, rarely leading to a usable result. As a consequence, we here elect to stay entirely in the realm of models that will allow discount bonds to be priced in closed form from the state variables of the model.

This chapter is broken into three parts. The first part develops the multi-factor Gaussian model in considerable generality, in the process demonstrating a number of features and techniques that apply to all short rate models. The second, much shorter, part provides a brief description of the multi-factor affine class, and the third part considers a particular class of quadratic-affine models that are well-suited for practical applications. For a fuller treatment of the multi-factor affine and affine-quadratic models, we refer the reader to Duffie et al. [2000], Duffie and Kan [1996], Duffie et al. [2003], Leippold and Wu [2002] and Ahn et al. [2002].

12.1 The Gaussian Model

As was the case for the one-factor Gaussian model, the multi-factor Gaussian model can be developed in two different ways: the "classical" way (from the bottom up) and the "modern" way (from a separability condition). As either technique leads to useful insights, we here show both.

12.1.1 Development from Separability Condition

A general d-factor Gaussian model can be written as

$$dP(t,T)/P(t,T) = r(t)\,dt - \sigma_P(t,T)^\top\,dW(t),$$

where $\sigma_P(t,T)$ is a bounded d-dimensional function of time, and $W(t)$ a d-dimensional Brownian motion in the risk-neutral measure Q. Written in terms of instantaneous forward rates, we get, from the HJM results in Chapter 4,

$$\begin{aligned} df(t,T) &= \sigma_f(t,T)^\top \sigma_P(t,T)\,dt + \sigma_f(t,T)^\top\,dW(t) \\ &= \sigma_f(t,T)^\top \int_t^T \sigma_f(t,u)\,du\,dt + \sigma_f(t,T)^\top\,dW(t). \end{aligned} \tag{12.1}$$

This model is generally not Markovian, unless we impose additional restrictions. A relevant result is the following.

Proposition 12.1.1. *Assume that $\sigma_f(t,T)$ is "separable", in the sense that it can be written as*

$$\sigma_f(t,T) = g(t)h(T), \tag{12.2}$$

where g is a $d \times d$ deterministic matrix-valued function, and h is a d-dimensional deterministic vector. Then

$$f(t,T) = f(0,T) + \Omega(t,T) + h(T)^\top z(t),$$

where $\Omega(t,T)$ is a deterministic scalar given in (12.6) and $z(t)$ is a d-dimensional random vector satisfying

$$dz(t) = g(t)^\top dW(t), \quad z(0) = 0. \tag{12.3}$$

In particular, we have

$$r(t) = f(0,t) + \Omega(t,t) + h(t)^\top z(t). \tag{12.4}$$

Proof. Inserting (12.2) into (12.1) and integrating over time, we get

$$f(t,T) = f(0,T) + \Omega(t,T) + h(T)^\top z(t), \tag{12.5}$$

where $z(t) = \int_0^t g(u)^\top dW(u)$ and

$$\Omega(t,T) = h(T)^\top \int_0^t g(s)^\top g(s) \int_s^T h(u)\, du\, ds. \tag{12.6}$$

□

Notice that the discount bond price volatility for the model in Proposition 12.1.1 becomes simply

$$\sigma_P(t,T) = g(t) \int_t^T h(u)\, du.$$

12.1.1.1 Mean-Reverting State Variables

Proposition 12.1.1 demonstrates that if (12.2) is satisfied, then the forward curve can be reconstructed from d Gaussian martingale variables $z_i(t)$, $i = 1, \ldots, d$, with joint SDE (12.3). The choice of d state variables is, however, not unique, and may in fact have disadvantages in a numerical implementation since often the components of $g(t)$ grow exponentially with time. As a result, it is common to shift variables to explicitly have a mean-reverting drift. To demonstrate one particular construction, set

$$H(t) = \operatorname{diag}(h(t)) = \begin{pmatrix} h_1(t) & 0 & \ddots & 0 \\ 0 & h_2(t) & \ddots & \ddots \\ \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & 0 & h_d(t) \end{pmatrix}. \tag{12.7}$$

Assuming that for all t we have $h_i(t) \neq 0$, $i = 1, \ldots, d$, then $H(t)$ is invertible, and we can define a diagonal $d \times d$ matrix $\varkappa(t)$ by

$$\varkappa(t) = -\frac{dH(t)}{dt} H(t)^{-1}. \tag{12.8}$$

Let us also set

$$x(t) = H(t) \int_0^t g(s)^\top g(s) \int_s^t h(u)\, du\, ds + H(t) z(t), \tag{12.9}$$

$$y(t) = H(t) \left(\int_0^t g(s)^\top g(s)\, ds \right) H(t). \tag{12.10}$$

Notice that $x(t)$ is a d-dimensional random vector, and $y(t)$ is a deterministic $d \times d$ symmetric matrix. It is easily verified that $y(t)$ solves the ODE

$$dy(t)/dt = H(t) g(t)^\top g(t) H(t) - \varkappa(t) y(t) - y(t) \varkappa(t).$$

Proposition 12.1.2. *Let the forward rate volatility be separable, as in Proposition 12.1.1. Let $\varkappa(t)$, $x(t)$ and $y(t)$ be defined as in (12.8)–(12.10), and assume that $H(t) = \operatorname{diag}(h(t))$ is invertible. Also define $\mathbf{1} = (1, 1, \ldots, 1)^\top \in \mathbb{R}^d$. Then*

$$dx(t) = (y(t)\mathbf{1} - \varkappa(t) x(t))\, dt + \sigma_x(t)^\top dW(t), \quad \sigma_x(t) = g(t) H(t),$$

and, with $M(t,T) \triangleq H(T) H(t)^{-1} \mathbf{1}$,

$$f(t,T) = f(0,T) + M(t,T)^\top \left(x(t) + y(t) \int_t^T M(t,u)\, du \right). \tag{12.11}$$

In particular, we have

$$r(t) = f(t,t) = f(0,t) + \mathbf{1}^\top x(t) = f(0,t) + \sum_{i=1}^d x_i(t).$$

Proof. Applying the Leibniz integration rule to the definition of $x(t)$ yields

$$\begin{aligned}
dx(t) &= \left[\frac{dH(t)}{dt} \int_0^t g(s)^\top g(s) \int_s^t h(u)\, du\, ds \right] dt \\
&\quad + \left[H(t) \int_0^t g(s)^\top g(s) h(t)\, ds \right] dt + \frac{dH(t)}{dt} z(t)\, dt + H(t)\, dz(t) \\
&= \frac{dH(t)}{dt} H(t)^{-1} x(t)\, dt + \left[H(t) \left(\int_0^t g(s)^\top g(s)\, ds \right) H(t) \mathbf{1} \right] dt \\
&\quad + H(t)\, g(t)^\top\, dW(t) \\
&= (y(t)\mathbf{1} - \varkappa(t) x(t))\, dt + \sigma_x(t)^\top dW(t).
\end{aligned}$$

Using the forward curve reconstitution formula in Proposition 12.1.1, we get

$$
\begin{aligned}
f(t,T) &= f(0,T) + h(T)^\top \int_0^t g(s)^\top g(s) \int_s^T h(u)\,du\,ds + h(T)^\top z(t) \\
&= f(0,T) + \mathbf{1}^\top H(T) \int_0^t g(s)^\top g(s) \int_s^T h(u)\,du\,ds + \mathbf{1}^\top H(T) z(t) \\
&= f(0,T) + \mathbf{1}^\top \left(H(T) \int_0^t g(s)^\top g(s) \int_s^T h(u)\,du\,ds \right) \\
&\qquad + \mathbf{1}^\top \left(H(T)H(t)^{-1}x(t) - H(T) \int_0^t g(s)^\top g(s) \int_s^t h(u)\,du\,ds \right) \\
&= f(0,T) + \mathbf{1}^\top H(T)H(t)^{-1}x(t) \\
&\qquad + \mathbf{1}^\top H(T)H(t)^{-1}y(t)H(t)^{-1} \int_t^T h(u)\,du.
\end{aligned}
$$

The result (12.11) follows from the definition of $M(t,T)$, the symmetry of $H(t)$, and the fact that $H(t)^{-1}h(u) = M(t,u)$. □

If $H(t)$ fails to be invertible, it must be because one or more of the elements in $h(t)$ are (locally) equal to zero. From Proposition 12.1.1, if this is the case it follows that some of the z_i's must be locally redundant, in turn demonstrating that the model is not truly d-dimensional for all t. As this strongly hints at a mis-specification, the invertibility condition in the proposition above is not a strong one.

In Propositions 12.1.1 and 12.1.2, reconstitution of the discount curve from the Markov state variables is done through the instantaneous forward curve. Obviously, we can also proceed to write explicit expressions for discount bond prices. For instance, using Proposition 12.1.2 we get:

Corollary 12.1.3. *In the setting of Proposition 12.1.2, define*

$$
G(t,T) = \int_t^T M(t,u)\,du.
$$

Then

$$
P(t,T) = \frac{P(0,T)}{P(0,t)} \exp\left(-G(t,T)^\top x(t) - \frac{1}{2} G(t,T)^\top y(t) G(t,T) \right).
$$

Proof. From the expression (12.11) for $f(t,T)$, we get

$$
\begin{aligned}
P(t,T) &= \exp\left(-\int_t^T f(t,u)\,du \right) \\
&= \exp\left(-\int_t^T f(0,u)\,du - \left(\int_t^T M(t,u)^\top du \right) x(t) \right) \\
&\qquad \times \exp\left(-\int_t^T M(t,u)^\top y(t) \int_t^u M(t,s)\,ds\,du \right),
\end{aligned}
$$

so that

$$P(t,T) = \frac{P(0,T)}{P(0,t)} \times \exp\left(-G(t,T)^\top x(t) - \int_t^T M(t,u)^\top y(t) \int_t^u M(t,s)\,ds\,du\right).$$

But here

$$\int_t^T M(t,u)^\top y(t) \int_t^u M(t,s)\,ds\,du = \int_t^T \frac{\partial G(t,u)}{\partial u}^\top y(t) G(t,u)\,du.$$

As $y(t)$ is symmetric, standard matrix calculus shows that

$$\begin{aligned}\frac{\partial}{\partial u}\left(G(t,u)^\top y(t) G(t,u)\right) &= \frac{\partial G(t,u)}{\partial u}^\top y(t) G(t,u) + G(t,u)^\top y(t) \frac{\partial G(t,u)}{\partial u} \\ &= 2\frac{\partial G(t,u)}{\partial u}^\top y(t) G(t,u),\end{aligned}$$

such that, finally,

$$\begin{aligned}\int_t^T \frac{\partial G(t,u)}{\partial u}^\top y(t) G(t,u)\,du &= \frac{1}{2}\int_t^T \frac{\partial}{\partial u}\left(G(t,u)^\top y(t) G(t,u)\right) du \\ &= \frac{1}{2} G(t,T)^\top y(t) G(t,T).\end{aligned}$$

□

Let us examine some of the matrices involved in the multi-dimensional Gaussian model. As $\varkappa(t)$ is diagonal, we must have

$$\varkappa(t) = \operatorname{diag}\left((\varkappa_1(t), \varkappa_2(t), \ldots, \varkappa_d(t))^\top\right),$$

in which case (12.8) implies that

$$h(t) = \left(e^{-\int_0^t \varkappa_1(s)ds}, e^{-\int_0^t \varkappa_2(s)ds}, \ldots, e^{-\int_0^t \varkappa_d(s)ds}\right)^\top. \tag{12.12}$$

Each element in the forward volatility vector $\sigma_f(t,T)$ is a time-weighted average of these d exponentiated integrals. Also, we note that

$$\begin{aligned}M(t,T) &= H(T)H(t)^{-1}\mathbf{1} \\ &= \left(e^{-\int_t^T \varkappa_1(s)ds}, e^{-\int_t^T \varkappa_2(s)ds}, \ldots, e^{-\int_t^T \varkappa_d(s)ds}\right)^\top,\end{aligned}$$

and

$$
\begin{aligned}
y(t) &= \int_0^t H(t)H(s)^{-1}\sigma_x(s)^\top \sigma_x(s)H(s)^{-1}H(t)\,ds \\
&= \int_0^t \operatorname{diag}\left(M(s,t)\right)\sigma_x(s)^\top \sigma_x(s)\operatorname{diag}\left(M(s,t)\right)ds.
\end{aligned}
$$

As all quantities in the dynamics for $x(t)$ and in the reconstitution formula for $f(t,T)$ and $P(t,T)$ evidently can be computed from knowledge of the d deterministic mean reversions $\varkappa_1(t), \varkappa_2(t), \ldots, \varkappa_d(t)$ and the $d \times d$ volatility matrix $\sigma_x(t)$, it follows that specification of $\varkappa(t)$ and $\sigma_x(t)$ fully determines our d-dimensional Gaussian model.

A brief comment about computing the bond reconstitution formula in Corollary 12.1.3 is in order. The vector $G(t,T)$ takes the form

$$
G(t,T) = \left(\int_t^T e^{-\int_t^u \varkappa_1(s)\,ds}du, \ldots, \int_t^T e^{-\int_t^u \varkappa_d(s)\,ds}du\right)^\top,
$$

where the individual components can, importantly, be rewritten as

$$
\begin{aligned}
\int_t^T e^{-\int_t^u \varkappa_i(s)\,ds}du &= \left(\int_0^T e^{-\int_0^u \varkappa_i(s)\,ds}du - \int_0^t e^{-\int_0^u \varkappa_i(s)\,ds}du\right) e^{\int_0^t \varkappa_i(s)\,ds} \\
&\triangleq \left(\Lambda_i(T) - \Lambda_i(t)\right) e^{\int_0^t \varkappa_i(s)\,ds}. \qquad (12.13)
\end{aligned}
$$

In implementations, we would typically pre-cache the $2d$ scalar functions

$$
\Lambda_1(t), \Lambda_2(t), \ldots, \Lambda_d(t), \exp\left(\int_0^t \varkappa_1(s)\,ds\right), \ldots, \exp\left(\int_0^t \varkappa_d(s)\,ds\right)
$$

on a suitable time grid, allowing subsequent discount bond pricing to be done quickly and conveniently for arbitrary t and T.

Remark 12.1.4. The risk-neutral process for the discount bond $P(t,T)$ is log-normal

$$
dP(t,T)/P(t,T) = r(t)\,dt - G(t,T)^\top \sigma_x(t)^\top\, dW(t).
$$

12.1.1.2 Further Changes of Variables

Going back to Proposition 12.1.2, we note that its form is rather convenient as it writes the short rate $r(t)$ as its forward value $f(0,t)$ plus a straight sum of d Gaussian mean-reverting variables, with each variable having a drift depending only on itself (since $\varkappa(t)$ is diagonal). This representation is, however, just one of many. If we allow the expression for $r(t)$ to be somewhat more complicated, then we are, for instance, free to use any mean reversion matrix — diagonal or not — that we would like. Before stating

this result, we need a little extra notation. Specifically, let us consider the generic homogeneous ODE system

$$\frac{dp(t)}{dt} = -Q(t)p(t),$$

where $Q(t)$ is a deterministic $d \times d$ matrix and p a d-dimensional (column) vector. It is well-known that the solution to this equation can always be represented as

$$p(T) = J_Q(T)p(0), \tag{12.14}$$

where $J_Q(T)$ is a $d \times d$ deterministic matrix satisfying

$$\frac{dJ_Q(t)}{dt} = -Q(t)J_Q(t). \tag{12.15}$$

The matrix $J_Q(t)$ is computable by classical ODE methods[2] and satisfies the boundary condition $J_Q(0) = I$, where I is the identity matrix. For the special case where Q is independent of time, we have

$$J_Q(t) = \exp(-Qt),$$

as one would expect[3]. For later use, let us notice that, in general,

$$\frac{d\left(J_Q(t)^{-1}\right)}{dt} = \left(J_Q(t)^{-1}\right)Q(t). \tag{12.16}$$

Lemma 12.1.5. *In the setup of Proposition 12.1.1, let us introduce some $d \times d$ mean reversion matrix $k(t)$ and assume that $J_k(t)$ (see (12.15)) exists and is invertible for all t. Then*

$$r(t) = f(0,t) + \Omega(t,t) + h(t)^\top J_k(t)^{-1}x(t),$$

where

$$dx(t) = -k(t)x(t)\,dt + \sigma_x(t)^\top dW(t), \quad \sigma_x(t) = g(t)J_k(t)^\top.$$

Proof. Set

$$x(t) = J_k(t)z(t),$$

such that $z(t) = J_k(t)^{-1}x(t)$ and, from (12.15),

$$dx(t) = -k(t)x(t)\,dt + J_k(t)g(t)^\top dW(t).$$

[2]Some readers may recognize $J_Q(T)$ as the *product integral* of $-Q(t)$ on $[0,T]$; see Dollard and Friedman [1979]. In the probability literature, the product integral is often referred to as the *fundamental matrix*, see Arnold [1974] or Karatzas and Shreve [1997].

[3]Recall that the exponential of a square matrix A is defined as $e^A = \sum_{k=0}^{\infty} A^k/k!$.

The result for $r(t)$ follows directly from Proposition 12.1.1. □

The lemma shows that we can incorporate essentially *any* mean reversion matrix k into the basic martingale setup in Proposition 12.1.1 by proper scaling of i) the weighting of the state variables in the expression of $r(t)$; and ii) the volatility matrix $g^\top$. For numerical applications, the best choice of mean reversion is typically one that leaves both $k(t)$ and $\sigma_x(t)$ close to constant.

12.1.2 Classical Development

The traditional approach to specification of a multi-dimensional short rate model does not go through a separability condition, but instead involves postulating that $r(t)$ is an affine function of a set of state variables satisfying a linear system of SDEs. That is, one would write

$$r(t) = b_q(t) + c_q(t)^\top q(t), \tag{12.17}$$

where $b_q(t) \in \mathbb{R}$ and $c_q(t) \in \mathbb{R}^d$ are deterministic, and the d-dimensional vector-valued process $q(t)$ satisfies the risk-neutral SDE

$$dq(t) = k(t)\,(m(t) - q(t))\,dt + \sigma(t)\,dW(t), \tag{12.18}$$

with $m(t) \in \mathbb{R}^d$ and $k(t), \sigma(t) \in \mathbb{R}^{d\times d}$ all being deterministic. Using the definition of $J_k(t)$ given above, we can solve (12.18) explicitly.

Lemma 12.1.6. *Let $q(t)$ be as given in (12.18). Then*

$$q(t) = J_k(t)\left(q(0) + \int_0^t J_k(s)^{-1}k(s)m(s)\,ds + \int_0^t J_k(s)^{-1}\sigma(s)\,dW(s)\right), \tag{12.19}$$

i.e. $q(t)$ has a d-dimensional Gaussian distribution, with mean

$$\mu_q(t) = J_k(t)\left(q(0) + \int_0^t J_k(s)^{-1}k(s)m(s)\,ds\right),$$

and covariance matrix

$$\Sigma_q(t) = J_k(t)\left(\int_0^t J_k(s)^{-1}\sigma(s)\sigma(s)^\top \left(J_k(s)^{-1}\right)^\top ds\right) J_k(t)^\top.$$

Proof. Set

$$u(t) = J_k(t)^{-1}q(t),$$

and observe, from (12.16), that

$$du(t) = J_k(t)^{-1}k(t)m(t)\,dt + J_k(t)^{-1}\sigma(t)\,dW(t).$$

Setting $q(t) = J_k(t)u(t)$ and observing that $u(0) = q(0)$ leads to the result in the lemma. □

Given Lemma 12.1.5 above, we would expect the class of models spanned by specification (12.17)–(12.18) to be identical to that of the separability condition in Proposition 12.1.1. For completeness, let us make this connection explicit.

Lemma 12.1.7. *Let $r(t)$ and $q(t)$ be as in (12.17) and (12.18), and define the martingale process*

$$dz(t) = \sigma_z(t)^\top dW(t), \quad z(0) = q(0), \; \sigma_z(t) = \sigma(t)^\top \left(J_k(t)^{-1}\right)^\top.$$

Then

$$r(t) = b_z(t) + c_z(t)^\top z(t),$$

where

$$b_z(t) = b_q(t) + c_q(t)^\top J_k(t) \int_0^t J_k(s)^{-1} k(s) m(s)\, ds,$$
$$c_z(t) = J_k(t)^\top c_q(t).$$

Proof. If we set

$$z(t) = J_k(t)^{-1} q(t) - w(t), \quad w(t) \triangleq \int_0^t J_k(s)^{-1} k(s) m(s)\, ds,$$

then Ito's lemma shows that

$$dz(t) = J_k(t)^{-1} \sigma(t)\, dW(t).$$

Notice that

$$q(t) = J_k(t)\left(z(t) + w(t)\right);$$

insertion of this expression into (12.17) proves the lemma. □

We emphasize that the form of the expression for $r(t)$ in Lemma 12.1.7 is identical to Proposition 12.1.1 once we align notation:

$$b_z(t) = f(0,t) + \Omega(t,t), \quad c_z(t) = h(t), \quad g(t) = \sigma_z(t) = \sigma(t)^\top \left(J_k(t)^{-1}\right)^\top.$$

Besides confirming that the classical approach is, indeed, equivalent to the approach in Section 12.1.1, we also note from Proposition 12.1.1 and Lemma 12.1.5 that we are free to change state variables to something other than $q(t)$ or $z(t)$.

12.1.2.1 Diagonalization of Mean Reversion Matrix

While we are looking at the traditional approach to multi-factor Gaussian models, let us for later use consider a standard question about this model class: if the model for $r(t)$ is time-homogeneous and the mean reversion matrix $k(t) = k$ is non-diagonal, can we transform the state variables in

such a way that the model remains time-homogeneous but has a *diagonal* mean reversion matrix? We know from Lemma 12.1.5 that such a change of variables is always possible if we can accept that the resulting model is not time-homogeneous. To retain time-homogeneity, however, we need to impose some regularity on k, as we show below.

Proposition 12.1.8. *Consider the model (12.17) and (12.18) with parameters c_q, k, σ being independent of time. Assume that k is diagonalizable,*

$$k = LKL^{-1},$$

where K is a $d \times d$ diagonal matrix. Set $c_Q = L^\top c_q$ and $\sigma_Q = L^{-1}\sigma$. Then

$$r(t) = b_q(t) + c_Q^\top Q(t),$$

where

$$dQ(t) = K\left(L^{-1}m(t) - Q(t)\right)\,dt + \sigma_Q\,dW(t).$$

Proof. Follows immediately from the variable transformation $Q(t) = L^{-1}q(t)$. □

In Proposition 12.1.8, we emphasize that the new mean reversion matrix K, as well as the volatility σ_Q and the scaling vector c_Q all are independent of time. We also remind the reader that a sufficient condition for k to be diagonalizable is that k has d distinct real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$; in this case $K = \mathrm{diag}((\lambda_1, \lambda_2, \ldots, \lambda_d)^\top)$. See also Section 3.1.3.

A closely related question is as follows: if the model (12.17) for $r(t)$ has a constant, non-diagonal mean reversion matrix k (but is otherwise not necessarily time-homogeneous), under which circumstances can we write $r(t) = f(0,t) + \mathbf{1}^\top x(t)$ where $x(t)$ has a constant *diagonal* mean reversion matrix $\varkappa$? From Proposition 12.1.2, we know that this re-write is generally possible if we allow $\varkappa$ to depend on t. For $\varkappa$ to additionally be constant, the following result suffices.

Proposition 12.1.9. *Consider the model (12.17) and (12.18) with k independent of time and diagonalizable, i.e.*

$$k = LKL^{-1},$$

where K is a constant $d \times d$ diagonal matrix. Assume that $B(t) = \mathrm{diag}(e^{-Kt}L^\top c_q)$ is invertible. Then

$$r(t) = f(0,t) + \mathbf{1}^\top x(t),$$

where

$$dx(t) = (y(t)\mathbf{1} - Kx(t))\,dt + \sigma_x(t)^\top dW(t),$$

with

$$\sigma_x(t) = \sigma(t)^\top \left(L^{-1}\right)^\top \mathrm{diag}\left(L^\top c_q\right),$$
$$y(t) = \int_0^t e^{-K(t-s)}\sigma_x(s)^\top \sigma_x(s) e^{-K(t-s)}ds.$$

Proof. Using the same steps as in Proposition 12.1.8 above, we know that

$$\begin{aligned} r(t) &= b_q(t) + c_Q^\top Q(t), \\ dQ(t) &= K\left(L^{-1}m(t) - Q(t)\right)\,dt + L^{-1}\sigma(t)\,dW(t). \end{aligned}$$

An application of Ito's lemma to $e^{Kt}Q(t)$ reveals that, in the notation of Section 12.1.1, this model is characterized by

$$h(t) = e^{-K^\top t}c_Q = e^{-Kt}L^\top c_q, \quad g(t) = \sigma(t)^\top \left(L^{-1}\right)^\top e^{Kt}.$$

As K is diagonal, $\varkappa(t)$ in (12.8) becomes

$$\varkappa(t) = \operatorname{diag}\left(Ke^{-Kt}L^\top c_q\right)\operatorname{diag}\left(e^{-Kt}L^\top c_q\right)^{-1} = K,$$

and

$$g(t)H(t) = \sigma(t)^\top (L^{-1})^\top e^{Kt}\operatorname{diag}(e^{-Kt}L^\top c_q) = \sigma(t)^\top (L^{-1})^\top \operatorname{diag}(L^\top c_q).$$

The result follows from Proposition 12.1.2. □

12.1.3 Correlation Structure

As discussed earlier, one important motivation for the introduction of a multi-factor interest rate model is the ability to control correlations among various points on the forward curve. Let $\rho(t, T_1, T_2)$ denote the time t instantaneous correlation between the forward rates $f(t, T_1)$ and $f(t, T_2)$. From the representation in Proposition 12.1.1, we get the following result, the proof of which is straightforward.

Lemma 12.1.10. *Let the model for $r(t)$ be as in Proposition 12.1.1. Then*

$$\rho(t, T_1, T_2) = \frac{h(T_1)^\top g(t)^\top g(t) h(T_2)}{\sqrt{h(T_1)^\top g(t)^\top g(t) h(T_1)}\sqrt{h(T_2)^\top g(t)^\top g(t) h(T_2)}}.$$

In a practical model, we generally would strongly prefer for this correlation structure to be *time-stationary*, in the sense that ρ does not depend outright on t, but only on time to maturity $T_1 - t$ and $T_2 - t$, i.e.

$$\rho(t, T_1, T_2) = \rho\left(T_1 - t, T_2 - t\right).$$

This restriction, if enforced, imposes a number of constraints on the model parameters; we return to this topic in Section 12.1.4.2 below.

While on the topic of correlation, let us remark that multi-factor Gaussian models are sometimes specified with the use of correlated Brownian motions. This setup, of course, can be translated to our setting quite easily. Specifically suppose that we start with a setup similar to that of (12.17) and (12.18), but now write

$$dq(t) = k(t)\,(m(t) - q(t))\;dt + \sigma(t)\,dW^*(t),$$

where W^* is a d-dimensional vector of correlated Brownian motions. Let $R(t)$ be the relevant correlation matrix of increments to $W^*(t)$ and let

$$R(t) = C(t)C(t)^\top.$$

for a square root matrix $C(t)$. Then, from results in Section 3.1.2.1, we may write $dW^*(t) = C(t)\,dW(t)$ for a standard (uncorrelated) vector-valued Brownian motion $W(t)$, and thereby

$$dq(t) = k(t)\,(m(t) - q(t))\;dt + \sigma(t)\,C(t)\,dW(t).$$

It follows, not surprisingly, that we can incorporate correlation in Brownian increments by a simple rotation of the volatility matrix (by $C(t)$).

12.1.4 The Two-Factor Gaussian Model

Having now outlined the general theory, let us make matters more concrete (and more practical) by focusing on the important case of $d = 2$.

12.1.4.1 Some Basics

In practical applications, a reasonable way to specify a two-dimensional model would be to set, in the notation of Proposition 12.1.1,

$$h(t) = \begin{pmatrix} e^{-\int_0^t \varkappa_1(u)du} \\ e^{-\int_0^t \varkappa_2(u)du} \end{pmatrix}, \quad g(t) = \begin{pmatrix} \sigma_{11}(t)e^{\int_0^t \varkappa_1(u)du} & \sigma_{12}(t)e^{\int_0^t \varkappa_2(u)du} \\ \sigma_{21}(t)e^{\int_0^t \varkappa_1(u)du} & \sigma_{22}(t)e^{\int_0^t \varkappa_2(u)du} \end{pmatrix}.$$

We may, without loss of generality, assume that $g(t)$ is lower diagonal[4], i.e. we can set $\sigma_{12}(t) = 0$ always. In this case, from Proposition 12.1.2 we have

$$r(t) = f(0,t) + x_1(t) + x_2(t),$$

where $x(t) = (x_1(t), x_2(t))^\top$ satisfies (with $x(0) = 0$)

$$dx(t) = (y(t)\mathbf{1} - \varkappa(t)x(t))\,dt + \sigma_x(t)^\top dW(t), \quad \sigma_x(t) = \begin{pmatrix} \sigma_{11}(t) & 0 \\ \sigma_{21}(t) & \sigma_{22}(t) \end{pmatrix}, \tag{12.20}$$

and where $\varkappa(t) = \mathrm{diag}((\varkappa_1(t), \varkappa_2(t))^\top)$ and $y(t)$ is a deterministic 2×2 matrix.

We notice that the instantaneous correlation between $x_1(t)$ and $x_2(t)$ is

$$\rho_x(t) = \frac{\sigma_{22}(t)\sigma_{21}(t)}{\sqrt{\sigma_{11}(t)^2 + \sigma_{21}(t)^2}\sqrt{\sigma_{22}(t)^2}},$$

[4]If $g(t)$ is not lower diagonal, we can always change variables (via a Cholesky decomposition, say) to make it so.

so for convenience we may, as in Section 12.1.3, rewrite our model as

$$dx(t) = (y(t)1 - \varkappa(t)x(t))\,dt + \sigma_x^*(t)\,dW^*(t), \tag{12.21}$$

where $\langle dW_1^*(t), dW_2^*(t)\rangle = \rho_x(t)\,dt$ and $\sigma_x^*(t)$ is diagonal with non-negative elements,

$$\sigma_x^*(t) = \begin{pmatrix} \sqrt{\sigma_{11}(t)^2 + \sigma_{21}(t)^2} & 0 \\ 0 & |\sigma_{22}(t)| \end{pmatrix} \triangleq \operatorname{diag}\left((\sigma_1(t), \sigma_2(t))^\top\right).$$

The term $y(t)$ in (12.21) can be computed by numerical integration from the results in Proposition 12.1.2.

From Corollary 12.1.3, the reconstitution formula for the yield curve is

$$f(t,T) = f(0,T) + M(t,T)^\top\left(x(t) + y(t)G(t,T)\right), \tag{12.22}$$

$$P(t,T) = \frac{P(0,T)}{P(0,t)} \exp\left(-G(t,T)^\top x(t) - \frac{1}{2}G(t,T)^\top y(t)G(t,T)\right),$$

where

$$G(t,T) = \int_t^T M(t,u)\,du, \quad M(t,T) = \left(e^{-\int_t^T \varkappa_1(u)\,du}, e^{-\int_t^T \varkappa_2(u)\,du}\right)^\top.$$

The specification (12.21)–(12.22) is, we feel, the most intuitive representation of the two-factor Gaussian short rate model. For a complete model specification, we evidently must specify 5 functions of time: $\rho_x(t)$, $\varkappa_1(t)$, $\varkappa_2(t)$, $\sigma_1(t)$, and $\sigma_2(t)$. As discussed earlier, however, we may want to ensure that the model has a time-stationary correlation structure, in the sense defined in Section 12.1.3. We turn to this in Section 12.1.4.2 below.

12.1.4.2 Variance and Correlation Structure

According to Proposition 12.1.1, it follows that the model (12.21) has the forward rate process

$$df(t,T) = O(dt) + \begin{pmatrix} \sigma_1(t)e^{-\int_t^T \varkappa_1(u)\,du} \\ \sigma_2(t)e^{-\int_t^T \varkappa_2(u)\,du} \end{pmatrix}^\top dW^*(t), \tag{12.23}$$

where we recall that $\langle dW_1^*(t), dW_2^*(t)\rangle = \rho_x(t)\,dt$. From this representation, or from the results in Section 12.1.3, we get the following lemma.

Lemma 12.1.11. *For the model (12.21), let*

$$b(t,T_1,T_2) = 1 + \rho_x(t)\frac{\sigma_2(t)}{\sigma_1(t)}\left(e^{-\int_t^{T_1}(\varkappa_2(u)-\varkappa_1(u))du} + e^{-\int_t^{T_2}(\varkappa_2(u)-\varkappa_1(u))du}\right)$$
$$+ \left(\frac{\sigma_2(t)}{\sigma_1(t)}\right)^2 e^{-\int_t^{T_1}(\varkappa_2(u)-\varkappa_1(u))du - \int_t^{T_2}(\varkappa_2(u)-\varkappa_1(u))du}.$$

Then

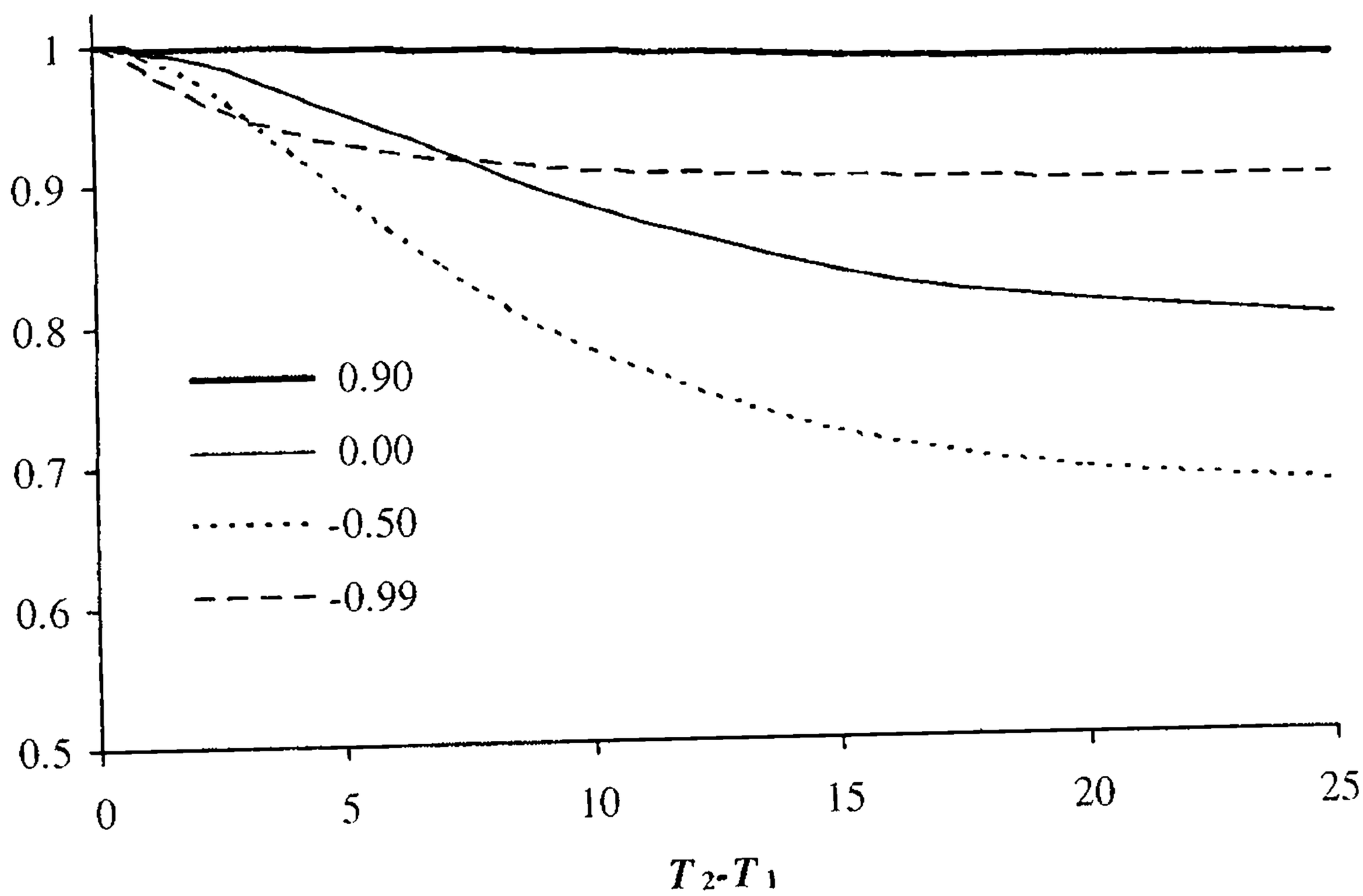
$$\mathrm{Var}_t\left(df(t,T)\right) = \sigma_1(t)^2 e^{-2\int_t^T \varkappa_1(u)du} b(t,T,T),$$
$$\rho(t,T_1,T_2) = \mathrm{Corr}_t(df(t,T_1), df(t,T_2)) = \frac{b(t,T_1,T_2)}{\sqrt{b(t,T_1,T_1)b(t,T_2,T_2)}}. \tag{12.24}$$

In particular, $\rho(t,T_1,T_2)$ *is* time-stationary *if* $\rho_x(t)$, $\varkappa_2(t) - \varkappa_1(t)$, *and* $\sigma_2(t)/\sigma_1(t)$ *are all constant.*

We emphasize that if we chose to make our correlation structure time-stationary, then Lemma 12.1.11 shows that only two functions of time ($\sigma_1(t)$ and $\varkappa_1(t)$, say) and three constants (ρ_x, $\varkappa_2(t) - \varkappa_1(t)$, and $\sigma_2(t)/\sigma_1(t)$) may be specified freely. Notice that if either i) $\rho_x = 1$; ii) $\varkappa_2(t) - \varkappa_1(t) = 0$; or iii) $\sigma_2(t)/\sigma_1(t) = 0$; then $\rho(t,T_1,T_2) = 1$, i.e. the model is reduced to having only one factor.

Figure 12.1 below shows some examples of the types of correlation term structures that can be generated in our two-factor Gaussian model.

Fig. 12.1. Forward Rate Correlation Term Structure



Notes: For the model (12.21), the figure graphs $\rho(0,T_1,T_2)$ from Lemma 12.1.11 against $T_2 - T_1$, using four different values of the parameter ρ_x. Other parameters were: $T_1 = 0.1$, $\varkappa_1 = 0.1$, $\varkappa_2 = 0.25$, $\sigma_1 = 0.025$, and $\sigma_2 = 0.02$.

Note the fact that the forward rate correlation is *not* necessarily a monotonic function of ρ_x. For parameterization purposes, it is often helpful

to consider $\rho(t,t,\infty)$, the correlation between the short rate and a (very) long-dated forward rate. From (12.24) we get, assuming time-homogeneity in the correlation structure,

$$\rho(t,t,\infty) = \frac{1+\rho_x c_\sigma}{\sqrt{1+2\rho_x c_\sigma + c_\sigma^2}}, \quad c_\sigma = \sigma_2(t)/\sigma_1(t),$$

an expression that does not depend on mean reversion speeds. Given either ρ_x or c_σ, an exogenous specification of $\rho(t,t,\infty)$ allows us to back out c_σ or ρ_x, respectively.

12.1.4.3 Volatility Hump

Besides allowing us to properly model the correlation between various points of the forward curve, the use of a two-factor Gaussian model has another benefit relative to a one-factor model: the ability to produce a time-stationary, non-monotonic term structure of forward rate volatilities. We recall from Section 10.1.2.3 that this was not possible in a one-factor model, where non-constant mean reversion was required to produce a caplet volatility "hump". To provide some details, assume that $\varkappa_1$ and $\varkappa_2$ are fixed at non-negative constant values, with at least one being positive. Also assume that σ_1 and σ_2 are fixed at constant positive values. From Lemma 12.1.11 we have

$$\begin{aligned}\mathrm{Var}_t\left(df(t,T)\right) &= \sigma_1^2 e^{-2\varkappa_1(T-t)} + \sigma_2^2 e^{-2\varkappa_2(T-t)} + 2\rho_x\sigma_1\sigma_2 e^{-(\varkappa_1+\varkappa_2)(T-t)} \\ &\triangleq g(T-t),\end{aligned}$$

where

$$\frac{1}{2}\frac{dg(\tau)}{d\tau} = -\varkappa_1\sigma_1^2 e^{-2\varkappa_1\tau} - \varkappa_2\sigma_2^2 e^{-2\varkappa_2\tau} - \rho_x\sigma_1\sigma_2(\varkappa_1+\varkappa_2)e^{-(\varkappa_1+\varkappa_2)\tau}.$$

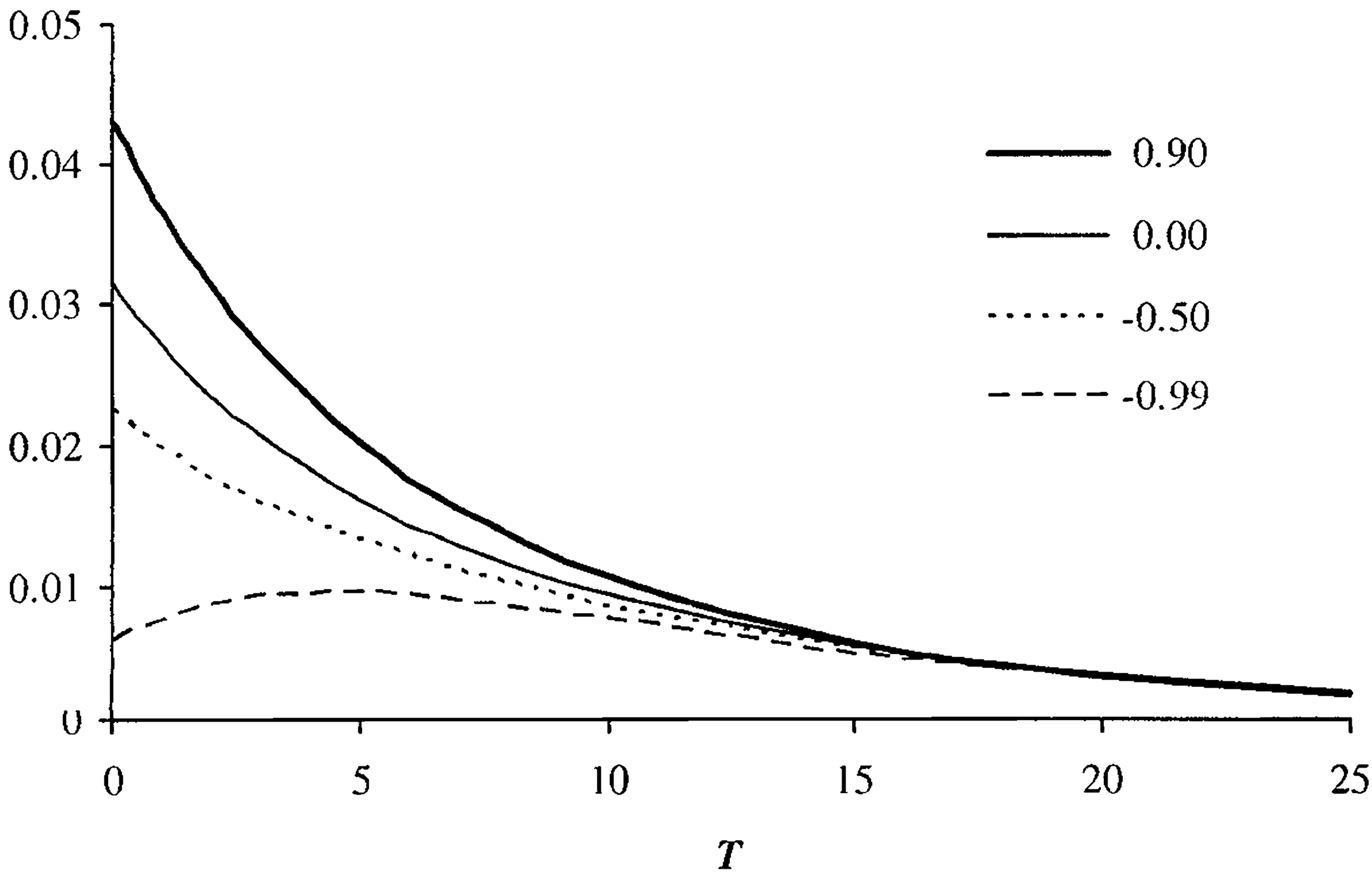
For positive values of ρ_x, the forward rate variance term structure will thus always be downward-sloping ($dg(\tau)/\partial\tau \leq 0$). However, if we set ρ_x sufficiently negative, there may be intermediate values for $\tau = T-t$ for which the variance will increase in τ; Figure 12.2 shows an example.

12.1.4.4 Another Formulation of the Two-Factor Model

To round off our treatment of the two-factor Gaussian model, we note that the model traditionally has been developed and presented in a manner quite different from ours. Indeed, a common starting point (e.g. Hull and White [1994b]) for the model is the doubly mean-reverting form:

$$\begin{aligned} dr(t) &= \varkappa_r\left(\vartheta(t)+\varepsilon(t)-r(t)\right)dt + \sigma_r\, dW_r(t), \\ d\varepsilon(t) &= -\varkappa_\varepsilon\varepsilon(t)\,dt + \sigma_\varepsilon\, dW_\varepsilon(t), \end{aligned} \tag{12.25}$$

Fig. 12.2. Forward Rate Volatility Term Structure



Notes: For the model (12.21), the figure graphs $\sqrt{\mathrm{Var}(df(0,T))}$ against T, using four different values of the parameter ρ_x. Other parameters were: $\varkappa_1 = 0.1$, $\varkappa_2 = 0.25$, $\sigma_1 = 0.025$, and $\sigma_2 = 0.02$.

where $\vartheta(t)$ is deterministic and $\langle dW_r(t), dW_\varepsilon(t)\rangle = \rho\, dt$, for some constant ρ. The model can be extended to time-dependent σ_r and σ_ε, but we omit this for the sake of simplicity.

To write (12.25) in terms that are more compatible with our notation, let $q(t) = (q_1(t), q_2(t))^\top$ be defined by

$$\begin{aligned} dq(t) &= \begin{pmatrix} \varkappa_r & -\varkappa_r \\ 0 & \varkappa_\varepsilon \end{pmatrix} \left(\begin{pmatrix} \vartheta(t) \\ 0 \end{pmatrix} - q(t) \right) dt + \begin{pmatrix} \sigma_r & 0 \\ \sigma_\varepsilon \rho & \sigma_\varepsilon \sqrt{1-\rho^2} \end{pmatrix} dW(t) \\ &\triangleq k\,(m(t) - q(t))\, dt + \sigma(t)\, dW(t), \end{aligned} \tag{12.26}$$

where the elements of $W(t) = (W_1(t), W_2(t))^\top$ are independent. Clearly, if we set

$$r(t) = q_1(t), \quad \varepsilon(t) = q_2(t), \tag{12.27}$$

we replicate the model (12.25) above. But (12.26)–(12.27) is of the form in Section 12.1.2, with $b_q(t) = 0$ and $c_q(t) = (1,0)^\top$, so we can use Propositions 12.1.8 and 12.1.9 to re-write the model in alternative formats. Some relevant characterizations are listed below.

Lemma 12.1.12. *Assume that $\varkappa_r \neq \varkappa_\varepsilon$. The model (12.25) can be written as*

$$r(t) = Q_1(t) + Q_2(t) \frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon},$$

where

$$dQ(t) = \begin{pmatrix} \varkappa_r & 0 \\ 0 & \varkappa_\varepsilon \end{pmatrix} \left(\begin{pmatrix} \vartheta(t) \\ 0 \end{pmatrix} - Q(t) \right) dt$$
$$+ \begin{pmatrix} \sigma_r - \rho\varkappa_r \frac{\sigma_\varepsilon}{\varkappa_r - \varkappa_\varepsilon} & -\varkappa_r \sigma_\varepsilon \frac{\sqrt{1-\rho^2}}{\varkappa_r - \varkappa_\varepsilon} \\ \rho\sigma_\varepsilon & \sigma_\varepsilon\sqrt{1-\rho^2} \end{pmatrix} dW(t).$$

Proof. If $\varkappa_r \neq \varkappa_\varepsilon$, the matrix k can be diagonalized as $k = LKL^{-1}$, where

$$L = \begin{pmatrix} 1 & \frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon} \\ 0 & 1 \end{pmatrix}, \quad K = \begin{pmatrix} \varkappa_r & 0 \\ 0 & \varkappa_\varepsilon \end{pmatrix}, \quad L^{-1} = \begin{pmatrix} 1 & -\frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon} \\ 0 & 1 \end{pmatrix}.$$

The lemma then follows from Proposition 12.1.8. □

Lemma 12.1.13. *Assume that $\varkappa_r \neq \varkappa_\varepsilon$ and let K be given in Lemma 12.1.12. The model (12.25) can be written as*

$$r(t) = f(0,t) + x_1(t) + x_2(t),$$

where

$$dx(t) = (y(t)\mathbf{1} - Kx(t))\,dt + \sigma_x^\top dW(t), \tag{12.28}$$
$$\sigma_x = \begin{pmatrix} \sigma_r - \rho\sigma_\varepsilon \frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon} & \rho\sigma_\varepsilon \frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon} \\ -\sigma_\varepsilon\sqrt{1-\rho^2}\frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon} & \sigma_\varepsilon\sqrt{1-\rho^2}\frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon} \end{pmatrix},$$

and

$$y(t) = \int_0^t e^{-K(t-s)}\sigma_x^\top \sigma_x e^{-K(t-s)} ds.$$

The forward rate volatility is

$$\sigma_f(t,T) = \begin{pmatrix} e^{-\varkappa_r(T-t)}\sigma_r + \left(e^{-\varkappa_\varepsilon(T-t)} - e^{-\varkappa_r(T-t)}\right)\rho\sigma_\varepsilon \frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon} \\ \left(e^{-\varkappa_\varepsilon(T-t)} - e^{-\varkappa_r(T-t)}\right)\sqrt{1-\rho^2}\sigma_\varepsilon \frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon} \end{pmatrix}.$$

Proof. The representation for $dx(t)$ follows directly from Proposition 12.1.9, after applying Lemma 12.1.12. The forward rate volatility can be recovered from Proposition 12.1.1. □

In Lemma 12.1.13, we note that $y(t)$ can be written in closed form; we leave this as an exercise to the reader. We also point out that in the dynamics for $x(t)$, we may translate back to the original correlated Brownian motions W_r and W_ε by an inverse Cholesky decomposition, writing

$$d\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix} d\begin{pmatrix} W_r(t) \\ W_\varepsilon(t) \end{pmatrix}.$$

Inserting this expression into (12.28) gives the simpler expression

$$dx(t) = (y(t)\mathbf{1} - Kx(t))\,dt + \begin{pmatrix} \sigma_r & -\frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon}\sigma_\varepsilon \\ 0 & \frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon}\sigma_\varepsilon \end{pmatrix} d\begin{pmatrix} W_r(t) \\ W_\varepsilon(t) \end{pmatrix}. \tag{12.29}$$

By the same token, we can write

$$df(t,T) = O(dt) + \begin{pmatrix} \sigma_r e^{-\varkappa_r(T-t)} \\ \frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon}\sigma_\varepsilon(e^{-\varkappa_\varepsilon(T-t)} - e^{-\varkappa_r(T-t)}) \end{pmatrix}^\top d\begin{pmatrix} W_r(t) \\ W_\varepsilon(t) \end{pmatrix}. \tag{12.30}$$

The discount bond reconstitution formula for the model (12.25) can be recovered from Corollary 12.1.3, using the representation in Lemma 12.1.13. The reconstitution formula can, of course, be re-stated in terms of the original q_i variables, should we desire to do so.

Finally, let us note that the special case of $\sigma_r = 0$ may be useful as a way to model the fact that central bank activities (which govern the dynamics of the short end of the forward curve) are often largely predictable. For the case $\sigma_r = 0$, notice that we get, from (12.29),

$$\begin{aligned} dx(t) &= O(dt) + \frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon}\sigma_\varepsilon \begin{pmatrix} -1 \\ 1 \end{pmatrix} dW_\varepsilon(t), \\ df(t,T) &= O(dt) + \frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon}\sigma_\varepsilon \left(e^{-\varkappa_\varepsilon(T-t)} - e^{-\varkappa_r(T-t)}\right) dW_\varepsilon(t). \end{aligned} \tag{12.31}$$

In other words, the two state variables x_1 and x_2 here become perfectly anti-correlated, and the forward curve dynamics are reduced to depending on only one Brownian motion. Despite this, notice that the model still requires two state variables (x_1 and x_2); this is a consequence of having two mean reversions in the forward rate volatility.

12.1.5 Multi-Factor Statistical Gaussian Model

The single-Brownian-motion, two-state model discussed at the end of Section 12.1.4 highlights an interesting and important interpretation of the model parameters. Let us re-write the model slightly to make our point. We have, from (12.31), that

$$df(t, t+\tau) = O(dt) + l(\tau)\,dz(t),$$

where we have denoted

$$\begin{aligned} l(\tau) &= \frac{\varkappa_r}{\varkappa_r - \varkappa_\varepsilon}\left(e^{-\varkappa_\varepsilon \tau} - e^{-\varkappa_r \tau}\right), \\ dz(t) &= \sigma_\varepsilon\, dW_\varepsilon(t). \end{aligned} \tag{12.32}$$

We can interpret $z(t)$ as the (single) factor that affects the movements of the forward rate curve $\{f(t, t+\tau)\}_{\tau \geq 0}$, and the function $l(\tau)$ as the response function, or a *loading*, whose value at time τ determines the impact of the

factor shock on a rate of tenor τ. Note that in this parameterization, the loading is a function of the tenor τ only, i.e. is time-homogeneous. This opens up the possibility of linking it to the statistically-estimated properties of the movements of the yield curve, the connection that we shall explore momentarily[5]. First, however, let us develop some technical tools.

The exponential functions $\{e^{-\varkappa\tau}\}_{\varkappa\in\mathbb{R}}$ are dense in the space of all continuous functions. Hence, any continuous function $l(\tau)$ can be approximated by a linear combination of exponential functions to an arbitrary degree of precision. Assume such a function (recycling the notations) $l(\tau)$ is given. Then, we can find a set of coefficients $\{v_i\}_{i=1}^n$ and exponents $\{\varkappa_i\}_{i=1}^n$ such that, approximately,

$$l(\tau) \approx \sum_{i=1}^{n} v_i e^{-\varkappa_i \tau}. \tag{12.33}$$

A moment of reflection on (12.32) and (12.33) shows that a model with the loading (12.33) could be represented as an n-state, single-Brownian-motion Gaussian model; we formalize this result as a proposition.

Proposition 12.1.14. *Let*

$$df\,(t, t+\tau) = O\,(dt) + l(\tau)\,dz(t), \tag{12.34}$$

$$l(\tau) = \sum_{i=1}^{n} v_i e^{-\varkappa_i \tau}, \quad dz(t) = \sigma_1(t)\,dW_1(t),$$

where $W_1(t)$ is a one-dimensional Brownian motion and $\sigma_1(t)$ is a one-dimensional function of time. Then this model admits a Markovian representation in n state variables

$$r(t) = f\,(0,t) + \sum_{i=1}^{n} x_i(t),$$

where, with $x(t) = (x_1(t), \ldots, x_n(t))^\top$ and $\varkappa = \mathrm{diag}((\varkappa_1, \ldots, \varkappa_n)^\top)$, we have

$$dx(t) = (y(t)\mathbf{1} - \varkappa x(t))\,dt + \sigma_x(t)^\top\,dW(t),$$

$$\sigma_x(t) = \sigma_1(t)\begin{pmatrix} v_1 & v_2 & \cdots & v_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad W(t) = (W_1(t), 0, \ldots, 0)^\top,$$

and

$$y(t) = H(t)\left(\int_0^t \sigma_1(s)^2 H(s)^{-1} U H(s)^{-1} ds\right) H(t),$$

$$H(t) = \mathrm{diag}\left((e^{-\varkappa_1 t}, \ldots, e^{-\varkappa_n t})^\top\right),$$

$$U = \{v_k v_j\}_{k,j=1}^n,$$

[5]The material in this section is largely inspired by Balasanov [1996].

Proof. From (12.34),

$$df\,(t,T) = O\,(dt) + \sum_{i=1}^{n} e^{-\varkappa_i T}\left(e^{\varkappa_i t}\sigma_1(t) v_i\, dW_1(t)\right).$$

Hence the model can be written in a separable form with

$$h(t) = \left((e^{-\varkappa_1 t}, \ldots, e^{-\varkappa_n t})^\top\right), \quad g(t) = \sigma_x(t) H(t)^{-1},$$

where $H(t)$ and $\sigma_x(t)$ are given in the statement of the proposition. The result follows from Proposition 12.1.2, definition (12.10) for $y(t)$, and the fact that

$$\sigma_x(t)^\top \sigma_x(t) = \sigma_1(t)^2 U,$$

where U is the $n \times n$ matrix defined above. □

The model (12.34) allows for an essentially arbitrary loading $l(\tau)$, but employs only one factor to describe the dynamics of the yield curve, a restriction that we can easily relax. Suppose we believe that m factors are needed to describe the dynamics of an interest rate curve. Also assume that we are given m loadings, each describing the (linear) response of the forward rate curve to a given factor. Approximating each loading by a linear combination of exponentials, we arrive at a model of the form

$$df\,(t, t+\tau) = O\,(dt) + \sum_{j=1}^{m} l_j\,(\tau)\; dz_j(t), \tag{12.35}$$

$$l_j(\tau) = \sum_{i=1}^{n_j} v_{j,i} e^{-\varkappa_{j,i}\tau}, \quad dz_j(t) = \sigma_j(t)\, dW_j(t),$$

where W_j's are independent Brownian motions. By a simple (but laborious, and left as an exercise to the reader) extension of Proposition 12.1.14, the model (12.35) can be shown to be Markovian in a total of

$$n = \sum_{j=1}^{m} n_j$$

state variables.

Proposition 12.1.15. *The model (12.35) admits a Markovian representation*

$$r(t) = f\,(0,t) + \mathbf{1}^\top x(t),$$

where $x(t) = (x_1(t), \ldots, x_n(t))^\top$ *satisfies*

$$dx(t) = (y(t)\mathbf{1} - \varkappa x(t))\; dt + \sigma_x(t)^\top\, dW(t),$$

with

$$\varkappa = \operatorname{diag}\left((\varkappa_{1,1}, \ldots, \varkappa_{1,n_1}, \varkappa_{2,1}, \ldots, \varkappa_{2,n_2}, \ldots, \varkappa_{m,n_m})^\top\right),$$
$$\sigma(t) = \operatorname{diag}\left((\sigma_1(t), \ldots, \sigma_1(t), \sigma_2(t), \ldots, \sigma_2(t), \ldots, \sigma_m(t))^\top\right),$$
$$h(t) = \left(e^{-\varkappa_{1,1}t}, \ldots, e^{-\varkappa_{1,n_1}t}, e^{-\varkappa_{2,1}t}, \ldots, e^{-\varkappa_{2,n_2}t}, \ldots, e^{-\varkappa_{m,n_m}t}\right)^\top,$$

and

$$\sigma_x(t) = v\sigma(t),$$

where

$$v = \begin{pmatrix} v_{1,1} & \cdots & v_{1,n_1} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & v_{2,1} & \cdots & v_{2,n_2} & \cdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & v_{m,1} & \cdots & v_{m,n_m} \end{pmatrix}.$$

Here $W(t)$ is an m-dimensional vector of independent Brownian motions

$$W(t) = (W_1(t), W_2(t), \ldots, W_m(t))^\top,$$

and

$$y(t) = H(t)\left(\int_0^t H(s)^{-1}\sigma(s)v^\top v\sigma(s)H(s)^{-1}ds\right)H(t), \;\; H(t) = \operatorname{diag}(h(t)).$$

The representation (12.35) allows us to link the interest rate model parameterization to statistical properties of the movements of the yield curve. To demonstrate, let us fix N_τ, the number of tenors of interest, and specify a set of tenors $\{\tau_1, \ldots, \tau_{N_\tau}\}$. Then we can observe from history how the vector of rates $f(t) = (f(t, t+\tau_1), \ldots, f(t, t+\tau_{N_\tau}))^\top$ changed over time. With the application of principal components (PC) analysis[6] we can identify a set of m, $m \leq N_\tau$, factors $\zeta(t) = (\zeta_1(t), \ldots, \zeta_m(t))^\top$, and m loadings $\lambda_j = (\lambda_j(\tau_1), \ldots, \lambda_j(\tau_{N_\tau}))^\top$, $j = 1, \ldots, m$, that we can use to represent

$$\Delta f(t) \approx \sum_{j=1}^{m} \lambda_j \, \Delta\zeta_j(t); \tag{12.36}$$

here Δ is a time-differencing operator, i.e. the day-to-day change of a given quantity. The PC analysis guarantees that the m factors will be optimal in the sense that the m factors will explain the largest possible variability of the vector of rates. As we shall see in Chapter 14, we can typically use a value m that is much smaller than N_τ, allowing for a significant reduction in dimension of the model. The loading vectors λ_j here define the shapes of the forward curve shocks associated with each factor.

[6] PC analysis was introduced in Section 3.1.3; its application to statistical analysis of interest rate curve movements will be described in details later, in Chapter 14.

Once we have identified loading vectors λ_j, $j = 1, \ldots, m$, the transition from (12.36) to (12.35) is merely an interpolation exercise where the functions $l_j(\tau)$ are extracted from the vectors λ_j by tenor-interpolation and a best-fit approximation with a linear combination of exponentials. After this step, the model can be represented, and efficiently implemented, in Markovian state variables as outlined in Proposition 12.1.15. The remaining factor volatility parameters $\sigma_j(t)$, $j = 1, \ldots, m$, may, for instance, be found by calibrating the model to market prices of caps/swaptions; see Section 12.1.6 for swaption pricing formulas that would be needed for such a calibration.

As we demonstrate later in Chapter 14, in the model (12.35) it is often sufficient to choose m to be 3 or 4, i.e. the yield curve movements through time are usually well explained by 3 to 4 factors. For each loading, the number of exponential terms required to match its shape is, typically, between 2 and 4 — the higher the loading number, the more complicated its shape typically is, and the more exponential terms are required. So, the overall number of state variables in the Markovian representation is typically around 10 or so.

The combination of statistical and risk neutral calibration, where some parameters (loadings) are obtained from historical data and others (factor volatilities) are market-implied, is an appealing characteristic of the model (12.35) and the parameterization strategy outlined above. Ultimately, however, (12.35), being nearly time-homogeneous, does not constitute a setup flexible enough for a precise calibration to all, or the majority of, market-quoted swaptions. In particular, historical loading shapes are often at odds with those consistent with the implied swaption volatilities. While, perhaps, not fully suitable as a model for interest rate exotics, (12.35) is still useful in settings where incorporation of historical information into pricing is of primary importance, such as risk management applications (such as VaR calculations, see Section 22.3), proprietary trading, or mortgage bonds valuation.

Remark 12.1.16. As an implementation note, we observe that working with instantaneous forward rates in the historical setting is inconvenient. Fortunately, (12.35) can be integrated in τ to obtain a similar linear representation for continuously compounded forward yields (see (4.1)), for which historical analysis can be performed more easily.

12.1.6 Swaption Pricing

While the parameterization (12.35) is just one of many for multi-factor interest rate models, it demonstrates a common strategy of specifying forward rate correlations exogenously and then calibrating the overall levels of model volatilities to European swaptions. We elaborate more on such calibration ideas in Section 12.1.7 and in Chapter 14. To make these ideas operational, we need to establish efficient pricing formulas for European swaptions. For

concreteness, we consider a payer swaption maturing at time $T_0 > 0$, with the underlying swap paying an annualized coupon c at times $T_1 < T_2 < \ldots < T_N$. The swaption payout at time T_0 is thereby

$$V_{\text{swaption}}(T_0) = \left(1 - P(T_0, T_N) - c\sum_{i=0}^{N-1} \tau_i P(T_0, T_{i+1})\right)^+, \quad \tau_i = T_{i+1} - T_i. \tag{12.37}$$

12.1.6.1 Two-Dimensional Jamshidian Decomposition

Let us consider to what extent we can use the Jamshidian decomposition in Section 10.1.3.1 in the multi-dimension case. For simplicity, we focus on the two-factor case, and later indicate how to extend our arguments to higher dimensions. Throughout we work with the parameterization[7] in Section 12.1.1.1, i.e. $r(t) = f(0,t) + x_1(t) + x_2(t)$, where x_1 and x_2 are state variables satisfying[8]

$$dx_1(t) = (\vartheta_1(t) - \varkappa_1(t)x_1(t))\,dt + \sigma_{11}(t)\,dW_1(t) + \sigma_{12}(t)\,dW_2(t), \tag{12.38}$$
$$dx_2(t) = (\vartheta_2(t) - \varkappa_2(t)x_2(t))\,dt + \sigma_{21}(t)\,dW_1(t) + \sigma_{22}(t)\,dW_2(t). \tag{12.39}$$

Expressions for $\vartheta_1(t) = y_{11}(t) + y_{12}(t)$ and $\vartheta_2(t) = y_{21}(t) + y_{22}(t)$ can be found in Section 12.1.1.1. We recall, in particular, the reconstitution formula

$$P(T, T+\Delta) = \frac{P(0, T+\Delta)}{P(0,T)} e^{A(T,T+\Delta) - x_1(T)G_1(T,T+\Delta) - x_2(T)G_2(T,T+\Delta)}, \tag{12.40}$$

where A, G_1, G_2 are known deterministic functions given in Corollary 12.1.3.

We shall first need to establish the following result.

Lemma 12.1.17. *Consider a put option on a discount bond, i.e. a derivative with T_0 payout*

$$p_i(T_0; K) = (K - P(T_0, T_i))^+, \quad T_i > T_0.$$

Let E^{T_0} denote expectation in the T_0-forward measure Q^{T_0}, and define the x_2-conditional option price

$$p_i(0; K, x_2) = P(0, T_0)\mathrm{E}^{T_0}\left(p_i(T_0; K) | x_2(T_0) = x_2\right).$$

Then

---

[7] This choice is made largely for reasons of familiarity. We indicate later (at the very end of this section) how the choice of different state variables may streamline the method.

[8] We here use $\sigma(t) = \sigma_x(t)^\top$.

$$p_i(0;K,x_2) = P(0,T_i)e^{A(T_0,T_i)-x_2G_2(T_0,T_i)} \times \left(K^*\Phi(-d_+) - e^{\Omega(T_0,T_i,x_2)}\Phi(-d_-)\right),$$

where

$$d_\pm = \frac{\Omega(T_0,T_i,x_2) - \ln K^* \pm \frac{1}{2}G_1^2(T_0,T_i)s_1(T_0,x_2)^2}{G_1(T_0,T_i)s_1(T_0,x_2)},$$

$$K^* = \frac{P(0,T_0)}{P(0,T_i)}e^{-A(T_0,T_i)+x_2G_2(T_0,T_i)}K,$$

$$\Omega(T_0,T_i,x_2) = -\mu_1(T_0,x_2)G_1(T_0,T_i) + \frac{1}{2}G_1^2(T_0,T_i)s_1(T_0,x_2)^2,$$

and $\mu_1(T_0,x_2)$ *and* $s_1(T_0,x_2)$ *are given in (12.41)–(12.42) below.*

Proof. A discount bond $P(t,T)$ has risk-neutral dynamics of the form (see Remark 12.1.4)

$$dP(t,T)/P(t,T) = r(t)\,dt - \sigma_{P,1}(t,T)\,dW_1(t) - \sigma_{P,2}(t,T)\,dW_2(t),$$

where

$$\sigma_{P,1}(t,T) = G_1(t,T)\sigma_{11}(t) + G_2(t,T)\sigma_{21}(t),$$
$$\sigma_{P,2}(t,T) = G_1(t,T)\sigma_{12}(t) + G_2(t,T)\sigma_{22}(t).$$

From standard results (see Chapter 4), we know that $dW^{T_0}(t) = dW(t) + \sigma_P(t,T_0)\,dt$, where $\sigma_P(t,T_0) = (\sigma_{P,1}(t,T_0),\sigma_{P,2}(t,T_0))^\top$, is a Brownian motion in Q^{T_0}. The Q^{T_0} dynamics for $x_1(t)$ and $x_2(t)$ therefore are

$$dx_1(t) = \left(\vartheta_1^{T_0}(t) - \varkappa_1(t)x_1(t)\right)\,dt + \sigma_{11}(t)\,dW_1^{T_0}(t) + \sigma_{12}(t)\,dW_2^{T_0}(t),$$
$$dx_2(t) = \left(\vartheta_2^{T_0}(t) - \varkappa_2(t)x_2(t)\right)\,dt + \sigma_{21}(t)\,dW_1^{T_0}(t) + \sigma_{22}(t)\,dW_2^{T_0}(t),$$

where

$$\vartheta_1^{T_0}(t) = \vartheta_1(t) - \sigma_{11}(t)\sigma_{P,1}(t,T_0) - \sigma_{12}(t)\sigma_{P,2}(t,T_0),$$
$$\vartheta_2^{T_0}(t) = \vartheta_2(t) - \sigma_{21}(t)\sigma_{P,1}(t,T_0) - \sigma_{22}(t)\sigma_{P,2}(t,T_0).$$

In measure Q^{T_0}, $x_1(T_0)$ and $x_2(T_0)$ are jointly Gaussian, with moments

$$\mathrm{E}^{T_0}\left(x_i(T_0)\right) = \int_0^{T_0} e^{-\int_s^{T_0}\varkappa_i(u)\,du}\vartheta_i^{T_0}(s)\,ds, \quad i=1,2,$$

$$\mathrm{Var}^{T_0}\left(x_1(T_0)\right) = \int_0^{T_0} e^{-\int_s^{T_0}2\varkappa_1(u)\,du}\left(\sigma_{11}(s)^2 + \sigma_{12}(s)^2\right)\,ds,$$

$$\mathrm{Var}^{T_0}\left(x_2(T_0)\right) = \int_0^{T_0} e^{-\int_s^{T_0}2\varkappa_2(u)\,du}\left(\sigma_{21}(s)^2 + \sigma_{22}(s)^2\right)\,ds,$$

$$\mathrm{Cov}^{T_0}\left(x_1(T_0),x_2(T_0)\right) = \int_0^{T_0} e^{-\int_s^{T_0}(\varkappa_1(u)+\varkappa_2(u))\,du} \times \left(\sigma_{11}(s)\sigma_{21}(s) + \sigma_{12}(s)\sigma_{22}(s)\right)\,ds.$$

Conditional upon $x_2(T_0)$, $x_1(T_0)$ must therefore be Gaussian with moments

$$\begin{aligned}\mathrm{E}^{T_0}\left(x_1(T_0)|x_2(T_0)=x_2\right) &= \mathrm{E}^{T_0}\left(x_1(T_0)\right) + \frac{\mathrm{Cov}^{T_0}\left(x_1(T_0),x_2(T_0)\right)}{\mathrm{Var}^{T_0}\left(x_2(T_0)\right)}x_2 \\ &\triangleq \mu_1(T_0,x_2),\end{aligned} \tag{12.41}$$

and

$$\begin{aligned}\mathrm{Var}^{T_0}\left(x_1(T_0)|x_2(T_0)=x_2\right) &= \mathrm{Var}^{T_0}\left(x_1(T_0)\right) - \frac{\mathrm{Cov}^{T_0}\left(x_1(T_0),x_2(T_0)\right)^2}{\mathrm{Var}^{T_0}\left(x_2(T_0)\right)} \\ &\triangleq s_1(T_0,x_2)^2.\end{aligned} \tag{12.42}$$

Using (12.40) we get, after a little rearrangement,

$$\begin{aligned}p_i(0;K,x_2) &= P(0,T_i)e^{A(T_0,T_i)-x_2G_2(T_0,T_i)} \\ &\quad\times \mathrm{E}^{T_0}\left(\left(K^* - e^{-x_1(T_0)G_1(T_0,T_i)}\right)^+\middle| x_2(T_0)=x_2\right),\end{aligned}$$

where K^* was defined above. The result of the lemma now follows from the standard results for log-normal random variables. □

Conditional on a value for $x_2(T_0)$, the payer swaption payout is monotonically increasing in $x_1(T_0)$, allowing for application of the Jamshidian decomposition to break the (conditional) swaption price into a sum of (conditional) discount bond options. A subsequent numerical integration against the density of $x_2(T_0)$ will then uncover the unconditional swaption price.

To formally state our result for the swaption price, define a function $x_1^*(x_2)$ as the solution to the equation $V_{\text{swaption}}(T_0,x_1^*(x_2),x_2)=0$, or

$$1 - P(T_0,T_N;x_1^*(x_2),x_2) - c\sum_{i=0}^{N-1}\tau_i P(T_0,T_{i+1};x_1^*(x_2),x_2) = 0,$$

where the functions $P(T_0,T_i;x_1^*(x_2),x_2)$ are given in Corollary 12.1.3. Given $x_1^*(x_2)$, we also define x_2-dependent strikes

$$K_i(x_2) = P(T_0,T_i;x_1^*(x_2),x_2), \quad i=1,\ldots,N. \tag{12.43}$$

Proposition 12.1.18. *In the two-factor Gaussian model (12.38)–(12.39), let K_i be given by (12.43), and let x_2-conditional discount bond put prices be given as in Lemma 12.1.17. Then, the swaption in (12.37) has price*

$$\begin{aligned}V_{\text{swaption}}(0) = \int_{-\infty}^{\infty} &\frac{p_N\left(0;K_N(x_2),x_2\right) + c\sum_{i=0}^{N-1}\tau_i p_{i+1}\left(0;K_{i+1}(x_2),x_2\right)}{\sqrt{\mathrm{Var}^{T_0}\left(x_2(T_0)\right)}} \\ &\times \phi\left(\frac{x_2 - \mathrm{E}^{T_0}\left(x_2(T_0)\right)}{\sqrt{\mathrm{Var}^{T_0}\left(x_2(T_0)\right)}}\right)dx_2,\end{aligned}$$

where $\phi(x)$ is the standard Gaussian density. The moments $\mathrm{E}^{T_0}(x_2(T_0))$ and $\mathrm{Var}^{T_0}(x_2(T_0))$ are given in Lemma 12.1.17.

Proof. Let $V(T_0, x_2)$ denote the swaption price at time T_0, conditional on $x_2(T_0) = x_2$. If $x_2(T_0) = x_2$, we note that the swaption only pays out a positive amount if $x_1(T_0) > x_1^*(x_2)$. Following the argument in Section 10.1.3.1, we can then easily decompose the swaption payout as follows,

$$\begin{aligned} V(T_0, x_2) &= \left(1 - P(T_0, T_N; x_1(T_0), x_2) - c\sum_{i=0}^{N-1} \tau_i P(T_0, T_{i+1}; x_1(T_0), x_2)\right) \\ &\quad \times 1_{\{x_1(T_0) > x_1^*(x_2)\}} \\ &= (K_N(x_2) - P(T_0, T_N; x_1(T_0), x_2))^+ \\ &\quad + c\sum_{i=0}^{N-1} \tau_i \left(K_{i+1}(x_2) - P(T_0, T_{i+1}; x_1(T_0), x_2)\right)^+ . \end{aligned}$$

Clearly, then

$$\begin{aligned} V_{\text{swaption}}(0) &= P(0, T_0)\mathrm{E}^{T_0}\left(V_{\text{swaption}}(T_0)\right) \\ &= P(0, T_0)\int_{\mathbb{R}} \mathrm{E}^{T_0}\left(V(T_0, x_2(T_0))\right) \mathrm{Q}^{T_0}\left(x_2(T_0) \in dx_2\right), \end{aligned}$$

and the result follows from the observation that $x_2(T_0)$ is Gaussian in measure Q^{T_0}, with moments given in Lemma 12.1.17. □

The technique behind Proposition 12.1.18 extends in straightforward fashion to dimension $d > 2$, with the "unconditioning" step involving numerical integration against a $(d-1)$-dimensional Gaussian density. This is rarely practical — especially since the integrand involves root-search to establish trigger levels for exercise — so in a real application we would typically never use Jamshidian decomposition, but instead introduce fast approximations. We list one such approximation in the next section.

As indicated earlier (in footnote 7), it is possible to make the derivation of the two-dimensional Jamshidian decomposition a little smoother by choosing another set of Markov state variables. To sketch how one might proceed, notice that, in the T_0-forward measure,

$$\begin{aligned} dP(t, T_0, T)/P(t, T_0, T) &= -\left(G(t, T) - G(t, T_0)\right)^\top \sigma_x(t)^\top dW^{T_0}(t) \\ &= -\left(\Lambda(T) - \Lambda(T_0))\right)^\top H(t)\sigma_x(t)^\top dW^{T_0}(t), \end{aligned} \tag{12.44}$$

where $\Lambda(t) = (\Lambda_1(t), \Lambda_2(t))^\top$ is the two-dimensional vector given in (12.13), and $H(t)$ is the 2×2 diagonal matrix

$$H(t) = \mathrm{diag}\left(\exp\left(\int_0^t \varkappa_1(s)\,ds\right), \exp\left(\int_0^t \varkappa_2(s)\,ds\right)\right).$$

Defining the two-dimensional Gaussian process

$$dz(t) = H(t)\sigma_x(t)^\top dW^{T_0}(t),$$

it follows from (12.44) that we can express forward bonds as closed form expressions of the Q^{T_0}-martingale process $z(t)$. The Q^{T_0}-dynamics for $z(t)$ are simpler than those of $x(t)$ (listed in the proof of Lemma 12.1.17), making subsequent manipulations easier.

12.1.6.2 Gaussian Swap Rate Approximation

We now return to the d-dimensional setting of Section 12.1.1. As in Section 5.10, we start by rewriting the swaption payout to

$$V_{\text{swaption}}(T_0) = A(T_0)\left(S(T_0) - c\right)^+,$$

where $A(t) = A_{0,N}(t)$ and $S(t) = S_{0,N}(t)$ are the swap annuity and par rate, respectively:

$$A(t) = \sum_{i=0}^{N-1} \tau_i P(t, T_{i+1}), \quad S(t) = \frac{P(t,T_0) - P(t,T_N)}{A(t)}.$$

Let Q^A be the measure induced by using $A(t)$ as the numeraire, such that

$$V_{\text{swaption}}(0) = A(0)\mathrm{E}^A\left(\left(S(T_0) - c\right)^+\right), \tag{12.45}$$

where E^A denotes expectation in measure Q^A. We know that $S(t)$ is a martingale in Q^A and, due to our Markov setting, a deterministic function of $x(t)$, i.e. $S(t) = S(t, x(t))$. It follows from Ito's lemma that

$$dS(t) = q\left(t, x(t)\right)^\top \sigma_x(t)^\top dW(t), \tag{12.46}$$

where $q(t,x)$ is a d-dimensional column vector with elements

$$q_j\left(t,x\right) = \frac{\partial S(t)}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{P(t,T_0,x) - P(t,T_N,x)}{\sum_{i=0}^{N-1}\tau_i P(t,T_{i+1},x)}, \quad j = 1,\ldots,d.$$

From the reconstitution formula in Corollary 12.1.3 we can evaluate the partial derivatives explicitly, yielding

$$\begin{aligned} q_j(t,x) = &-\frac{P(t,T_0,x)G_j(t,T_0) - P(t,T_N,x)G_j(t,T_N)}{A(t,x)} \\ &- \frac{S(t,x)\sum_{i=0}^{N-1}\tau_i P(t,T_{i+1},x)G_j(t,T_{i+1})}{A(t,x)}, \end{aligned}$$

where we recall that

$$G_j(t,T) = \int_t^T e^{-\int_t^u \varkappa_j(s)ds}\, du.$$

The functions q_j can be experimentally verified to be close to a constant[9] so, as a good approximation, we can write

$$q_j(t, x(t)) \approx q_j(t, \overline{x}(t)), \quad j = 1, \ldots, N, \tag{12.47}$$

where $\overline{x}(t)$ is some *deterministic* proxy for the random vector $x(t)$. A reasonable approach is to set $\overline{x}(t) = 0$, but see Chapter 13 for refinements. In any case, with the approximation (12.47), the following swaption pricing formula is easily proven.

Lemma 12.1.19. *Let $\overline{x}(t)$ be a deterministic function of time, and assume that (12.47) holds. Then*

$$V_{\text{swaption}}(0) = A(0)\left[(S(0) - c)\,\Phi(d) + \sqrt{v}\phi(d)\right],$$

where

$$d = \frac{S(0) - c}{\sqrt{v}}, \quad v = \int_0^{T_0} \left\| q\left(t, \overline{x}(t)\right)^\top \sigma_x(t)^\top \right\|^2 dt.$$

Proof. Follows directly from the Bachelier pricing formula (7.16), expression for the swap rate volatility (12.46), and approximation (12.47). □

12.1.7 Calibration and Parameterization via Benchmark Rates

With the swaption formulas developed in the previous sections, we have reached a point where we can entertain ideas of how to calibrate a multi-factor Gaussian model. As one would expect, the basic principles of such a calibration are, for the most part, identical for all multi-factor models, and a detailed discussion is best postponed to later chapters when our model catalog is more complete. Nevertheless, it is useful to present here a few ideas that will align the Gaussian multi-factor model with later material, particular that in Chapter 14.

Using the fundamental model setup from Section 12.1.1, we first observe that the parameters $g(t)$ and $h(t)$ of the Gaussian model (in the parameterization from Proposition 12.1.1) affect both volatilities and correlations of market rates. Thus, should we desire to recover $g(t)$ and $h(t)$ via calibration to market instruments, we would need both caplet/swaptions (for overall level of volatility) and spread options[10] (for correlation) as calibration targets. The former are much more liquid than the latter, so it would be

[9] Reflecting a rather intuitive fact that in a Gaussian model a swap rate is approximately Gaussian.

[10] Or other derivatives with first-order correlation dependence. See Chapter 17 for details on yield curve spread options.

beneficial to be able to separate volatility and correlation calibration; such separation of different model parameters is a good idea anyway, for any model. The parameterization of Proposition 12.1.1 is somewhat awkward for this purpose, so let us attempt to reparameterize the model in quantities that are more closely related to market observations.

It is intuitively obvious that a model with d stochastic factors is fully specified by volatilities and correlations of d (different) forward rates. Let us select d *benchmark tenors* $\delta_1 < \ldots < \delta_d$, and define d *benchmark rates* $f_i(t) = f(t, t+\delta_i)$, $i = 1, \ldots, d$. Notice that these forward rates are defined with "sliding" tenors, to encourage time-homogeneity. Let $\lambda_i(t)$ be the instantaneous volatility of the rate $f_i(t)$, $i = 1, \ldots, d$, and let $\chi_{i,j}(t)$ be the instantaneous correlation between rates $f_i(t)$ and $f_j(t)$. To recover model parameters from this data, we first observe that the instantaneous covariance matrix of the vector

$$f(t) = (f_1(t), \ldots, f_d(t))^\top$$

is given by $R^f(t) = \{\lambda_i(t)\lambda_j(t)\chi_{i,j}(t)\}_{i,j=1}^d$. On the other hand, in the model parameterized as in Proposition 12.1.1, the instantaneous covariance matrix is given by (see (12.5))

$$H^f(t)g(t)^\top g(t)H^f(t)^\top,$$

where the $d \times d$ matrix $H^f(t)$ is obtained by "stacking" vectors $h(t+\delta_i)$ together,

$$H^f(t) = \begin{pmatrix} h(t+\delta_1)^\top \\ \vdots \\ h(t+\delta_d)^\top \end{pmatrix}.$$

Let us assume that the vector $h(t)$ is directly parameterized, for instance by the specification of d different (and typically constant) mean reversion parameters $\varkappa_i$, $i = 1, \ldots, d$, to be applied in (12.12). It follows that the matrix $g(t)$ can then be recovered by solving

$$H^f(t)g(t)^\top = C^f(t)^\top, \tag{12.48}$$

where $C^f(t)$ is such that

$$R^f(t) = C^f(t)^\top C^f(t).$$

While this completely determines the model parameterization, (12.48) is not the only way to specify $g(t)$. For computational reasons we could, for example, decide to determine $g(t)$ by fitting *correlation* rather than *covariance*, as suggested in Andreasen [2005]. Defining $X^f(t) = \{\chi_{i,j}(t)\}$ and $D^f(t)$ by $X^f(t) = D^f(t)^\top D^f(t)$, we then obtain the matrix $g(t)$ by solving

$$H^f(t)g(t)^\top = \operatorname{diag}\left((\lambda_1(t), \ldots, \lambda_d(t))^\top\right) D^f(t)^\top. \tag{12.49}$$

The computational advantage of (12.49) over (12.48) is that the matrix $D^f(t)$ does not depend on $\lambda_i(t)$'s and, hence, does not need to be recomputed after each update of the volatilities $\lambda_i(t)$. (Such frequent updates happen, for example, during volatility calibration to quoted option prices.) The disadvantage of (12.49) is that the *true* instantaneous correlation matrix of the vector $f(t)$ is not going to be exactly equal to $X^f(t)$. (Similar issues are discussed at length in Section 14.3.4.) This is less of a concern if $X^f(t)$ itself is fit to the observed prices of correlation-dependent instruments such as spread options.

The ruminations above can be used to design a relatively straightforward calibration routine, a sketch of which is listed below. For full details, we again refer to Chapter 14.

1. Specify $h(t)$ via the mean reversion parameterization of Section 12.1.1.1, using d different constant mean reversions. The choice of mean reversions defines the interpolation of volatilities and correlations, i.e. how the volatilities/correlations of non-benchmark rates are obtained from those of the benchmark ones. As such, mean reversions have rather limited impact on prices of exotic derivatives, as volatilities and correlations of benchmark rates — presumably chosen not at random but to represent the primary risk factors for a given pricing problem — are controlled directly in our setup. With that in mind, we advocate choosing mean reversions in such a way as to improve the numerical properties of the algorithm. In particular, as an inversion of the matrix $H^f(t)$ is implicit in (12.48) (or (12.49)), we suggest using mean reversions that are sufficiently different from each other so that the matrix $H^f(t)$ has a better-behaved inverse. Besides stable numerics, a good choice of mean reversions will generate volatility factors that are fundamentally consistent with observed swaption quotes, in the sense that calibrated volatilities $\lambda(t)$ will be close to time-stationary. Working with a four-factor model, Andreasen [2005] suggests the following values:

$$\varkappa(t) = \mathrm{diag}((0.015, 0.15, 0.30, 1.20)^\top),$$
$$\{\delta_1, \delta_2, \delta_3, \delta_4\} = \{6m, 2y, 10y, 30y\},$$

   which gives us a good example to follow.
2. Populate the correlation matrix $\{\chi_{i,j}(t)\}$ of benchmark rates. This may be done through a smooth functional form, as described in Chapter 14. Most often a time-homogeneous specification should be used. The parameters of the functional form may be found from historical analysis or through calibration to market prices of spread options.
3. Calibrate benchmark rate volatilities against swaptions, using the results of Section 12.1.6. For a discussion of optimization techniques and relevant calibration norms, see Chapter 14.
4. Recover $g(t)$, the diffusion matrix of factors, using (12.48) or (12.49).

With the function $h(t)$ and the correlation matrix $\{\chi_{i,j}(t)\}$ pre-specified, the model has enough parameters to calibrate d swaption strips (see Section 10.1.4). In most applications, we recommend choosing d strips with constant swap tenors matching benchmark tenors $\delta_1, \ldots, \delta_d$. An alternative would be to do a best-fit calibration to all available options (i.e., a global calibration).

12.1.8 Monte Carlo Simulation

Monte Carlo methods for the d-dimensional Gaussian model are straightforward, as all state variables are jointly Gaussian. To demonstrate, we adopt the parameterization in Section 12.1.1.1 and consider pricing a security that pays an amount $V(T)$ at time T, where $V(T)$ may be a function of the entire path of the discount curve on $[0, T]$. Working in the risk-neutral measure, we must compute

$$\begin{aligned} V(0) &= \mathrm{E}^{\mathrm{Q}}\left(V(T)e^{-\int_0^T r(u)du}\right) \\ &= P(0,T)\mathrm{E}^{\mathrm{Q}}\left(V(T;\{x(t):0\le t\le T\})e^{-\int_0^T \mathbf{1}^\top x(u)du}\right), \end{aligned} \tag{12.50}$$

where we have used the relation $r(t) = f(0,t) + \mathbf{1}^\top x(t)$, and also have emphasized the dependence of $V(T)$ on the entire path of $x(\cdot)$. We assume that the determination of $V(T)$ involves observations of the yield curve on a discrete schedule $\{t_i\}_{i=0}^N$ with $t_0 = 0$ and $t_N = T$.

We recall from Proposition 12.1.2 that the risk-neutral dynamics for $x(t) = (x_1(t), \ldots, x_d(t))^\top$ are

$$dx(t) = (y(t)\mathbf{1} - \varkappa(t)x(t))\,dt + \sigma_x(t)^\top dW(t), \quad x(0) = 0,$$

for deterministic vectors/matrices $y(t)$, $\varkappa(t)$, $\sigma_x(t)$. Observe that $x(t_{i+1})$, conditional on $x(t_i)$, is d-dimensional Gaussian with mean

$$\mathrm{E}^{\mathrm{Q}}\left(x(t_{i+1})|x(t_i)\right) = e^{-\int_{t_i}^{t_{i+1}} \varkappa(u)du}x(t_i) + \int_{t_i}^{t_{i+1}} e^{-\int_s^{t_{i+1}} \varkappa(u)du}y(s)\mathbf{1}\,ds,$$

and covariance matrix

$$\mathrm{Var}\left(x(t_{i+1})|x(t_i)\right) = \int_{t_i}^{t_{i+1}} e^{-\int_s^{t_{i+1}} \varkappa(u)du}\sigma_x(s)\sigma_x(s)^\top e^{-\int_s^{t_{i+1}} \varkappa(u)^\top du}ds. \tag{12.51}$$

Let the square root of the covariance matrix be denoted C.

Advancement of $x(\cdot)$ from t_i to t_{i+1} can now proceed in an obvious fashion:

1. Draw d independent Gaussian samples $Z = (Z_1, Z_2, \ldots, Z_d)^\top$.
2. Compute $Z^* = CZ$.

3. Set $x(t_{i+1}) = \mathrm{E}^{\mathrm{Q}}(x(t_{i+1})|x(t_i)) + Z^*$.

At each time on the simulated path $x(t_0), \ldots, x(t_N)$, we can use the reconstitution formula in Corollary 12.1.3 to reconstruct the entire discount curve, in turn allowing us to compute $V(T)$ on the path. To evaluate (12.50), it remains to simulate the quantity

$$I(T) = -\int_0^T \mathbf{1}^\top x(u)\, du$$

on the path. Clearly $I(t)$ is a Gaussian process, so we can work out the moments of $I(t_{i+1})$ conditional on $I(t_i)$ and $x(t_i)$, allowing for bias-free joint time-stepping of $x(t)$ and $I(t)$ on the schedule $\{t_i\}_{i=0}^N$. The analysis proceeds as in Section 10.1.6.1, and is omitted for brevity. In practice, we find that it is often more convenient to change the measure (as in Section 10.1.6.3) or to compute $I(T)$ by numerical integration, i.e.

$$I(T) \approx -\mathbf{1}^\top \sum_{i=1}^N x(t_i).$$

While the last method obviously introduces some amount of bias, this is generally of little concern and can be controlled as needed through the insertion of extra dates in the schedule.

Finally, we remind the reader once again that all time integrals involved in the Monte Carlo discretization scheme above should be pre-computed before actual path simulations commence.

12.1.9 Finite Difference Methods

We finish our treatment of the multi-factor Gaussian short rate model with a brief discussion of finite difference applications. For this, let us consider a claim V the terminal payout of which can be computed solely from knowledge of the yield curve at time T. We assume that the yield curve is driven by a multi-factor Gaussian model, in the form described in Section 12.1.1.1. In this model, the discount curve at time T can be reconstituted solely from knowledge of the Markovian state variable vector $x(T)$, so we may write $V(T) = V(T, x(T))$. By standard results from Chapter 1, for $t < T$ we then have $V(t) = V(t, x(t))$ where $V(t, x)$ satisfies a d-dimensional PDE:

$$\mathcal{L}V - \left(f(0,t) + \mathbf{1}^\top x\right) V = 0, \tag{12.52}$$

where $\mathcal{L}$ is a partial differential operator,

$$\mathcal{L}V = V_t + \varkappa(t)\left(y(t)\mathbf{1} - x(t)\right)^\top V_x + \frac{1}{2}\mathrm{tr}\left(V_{xx}\sigma_x(t)^\top \sigma_x(t)\right).$$

In the definition of $\mathcal{L}$, $\mathrm{tr}(\cdot)$ is the matrix trace operator, V_x is a d-dimensional vector of first-order spatial derivatives, and V_{xx} a $d \times d$ matrix of second-order spatial derivatives. The d-dimensional PDE (12.52) is subject to a given terminal condition $V(T, x)$, the computation of which typically would involve usage of the discount bond reconstitution formulas in Corollary 12.1.3.

Numerical solution of (12.52) is practically feasible provided that d does not exceed 3 or 4, say. We recommend the Craig-Sneyd scheme in Section 2.12. As this is a splitting scheme with only one dimension being computed non-explicitly in each split step, applications of the one-dimensional side-boundary conditions of Section 10.1.5 carry over in straightforward fashion. A similar comment holds for the application of upwinding in the edges of the finite difference grid.

12.2 The Affine Model

12.2.1 Introduction

In Section 10.2, we introduced the one-factor affine short rate model through the short rate SDE

$$\begin{aligned} dr(t) &= \mu\left(t, r(t)\right)\, dt + \sigma\left(t, r(t)\right)\, dW(t) \\ &= \varkappa(t)\left(\vartheta(t) - r(t)\right)\, dt + \sqrt{\alpha + \beta r(t)}\, dW(t), \end{aligned}$$

for deterministic parameters $\varkappa(t)$, $\vartheta(t)$, α, β, subject to certain regularity conditions. The fact that both $\mu(t, r)$ and $\sigma(t, r)^2$ were linear (that is, affine) in r allowed for a discount bond price formula that was exponentially affine in r. We are now interested in examining how we can extend this one-factor setup to a multi-dimensional one. Our discussion of this extension shall be quite brief though, as the general multi-factor affine class is of fairly limited practical relevance in securities pricing applications. A subset of affine models — those that can be rewritten as linear-quadratic interest rate models — have some uses in practice and shall receive a fuller treatment in Section 12.3.

12.2.2 Basic Model

Let us consider a time-homogeneous[11] Markov system of state variables

$$dx(t) = \mu\left(x(t)\right)\, dt + \sigma\left(x(t)\right)\, dW(t), \tag{12.53}$$

[11] As is standard practice in much of the literature on multi-factor affine models, for notational simplicity we here assume that μ and σ do not depend on time t. See the comments at the end of Section 12.2.4 for extensions to time-dependent parameters.

where $x(t) = (x_1(t), \ldots, x_d(t))^\top$ has state space $D \subseteq \mathbb{R}^d$; W is a d-dimensional Brownian motion in the risk-neutral probability measure Q; and where $\mu : D \to \mathbb{R}^d$ and $\sigma : D \to \mathbb{R}^d \times \mathbb{R}^d$ have sufficient regularity for (12.53) to have a unique solution. We further write

$$r(t) = F(x(t)) \tag{12.54}$$

for some deterministic function $F : D \to \mathbb{R}$.

For the one-dimensional affine model in Section 10.2, we concluded that discount bonds were *exponential affine* in the underlying state variables. It is of interest to establish the circumstances for which something similar holds for (12.53)–(12.54) That is, we wish to determine the form that μ, σ, and F must take such that $P(t,T)$ can be written[12]

$$P(t,T) = \mathrm{E}\left(\exp\left(\int_t^T F(x(u))\,du\right)\right) = \exp(A(T-t) - B(T-t)^\top x(t)),$$

for deterministic functions $A : \mathbb{R} \to \mathbb{R}$ and $B : \mathbb{R} \to \mathbb{R}^d$. The following result is shown in Duffie and Kan [1996].

Proposition 12.2.1. *Suppose that in (12.53)–(12.54) μ, $\sigma\sigma^\top$ and F are affine functions of x. Then discount bond prices are exponential affine.*

Remark 12.2.2. Under additional non-degeneracy assumptions, Duffie and Kan [1996] show that the converse of Proposition 12.2.1 holds. However, many interesting models, including those of Chapter 13, violate these conditions.

The proof of Proposition 12.2.1 follows the line of attack of Proposition 10.2.2 and is omitted for brevity.

Duffie and Kan [1996] demonstrate that if $\sigma\sigma^\top$ is affine one may, under some mild non-degeneracy conditions, rearrange the dynamics for $x(t)$ such that $\sigma(t)$ is *diagonal.* That is, we may write

$$dx(t) = (a - bx(t))\,dt + \Sigma \begin{pmatrix} \sqrt{v_1(x(t))} & 0 & \ddots & 0 \\ 0 & \sqrt{v_2(x(t))} & \ddots & \ddots \\ \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & 0 & \sqrt{v_d(x(t))} \end{pmatrix} dW(t), \tag{12.55}$$

where $a \in \mathbb{R}^d$ and $b, \Sigma \in \mathbb{R}^d \times \mathbb{R}^d$ and

$$v_i(x) = \alpha_i + \beta_i^\top x, \quad i = 1, \ldots, d, \tag{12.56}$$

[12] Throughout this section, E denotes expectation in the risk-neutral measure Q.

with α_i scalar and $\beta_i \in \mathbb{R}^d$. The representation (12.55) is convenient in practical work and going forward shall, together with an affine short rate specification

$$r(t) = \xi_0 + \xi^\top x(t) = \xi_0 + \sum_{i=1}^{d} \xi_i x_i(t), \tag{12.57}$$

constitute our working definition for a multi-factor affine model.

12.2.3 Regularity Issues

As was the case for the one-factor affine model, there are strong restrictions on which values of $\Sigma, \alpha_i, \beta_i, a, b$ allow for valid solutions to (12.55). To state these restrictions, we first establish the process for $v_i(x(t))$ to be

$$d\left(v_i(x(t))\right) = \beta_i^\top dx(t) = \beta_i^\top \left(a - bx(t)\right) dt + \beta_i^\top \Sigma v\left(x(t)\right) dW(t), \tag{12.58}$$

where $v(x(t))$ is the diagonal matrix in (12.55). It is clear that the volatility process $v_i(x(t))$ must be non-negative for all i and t, i.e. the valid domain for $x(t)$ is

$$D = \left\{x \in \mathbb{R}^d : v_i(x) \geq 0, \quad i = 1, \ldots, d\right\}. \tag{12.59}$$

Clearly, when $x(t)$ is on the boundary $\partial_i D = \{x \in D : v_i(x) = 0\}$, we must ensure that i) the drift of $v_i(x(t))$ is positive; and ii) the instantaneous variance is zero. The first condition implies that, for all i,

$$\beta_i^\top \left(a - bx\right) \geq 0, \quad \forall x \in \partial_i D.$$

The second condition requires that, for all i,

$$\beta_i^\top \Sigma v\left(x\right) = 0, \quad \forall x \in \partial_i D.$$

This evidently requires that for $j = 1, \ldots, d$ either $v_j(x) = k v_i(x)$ for some constant k; or the j-th element of the row vector $\beta_i^\top \Sigma$ is zero, i.e. $(\beta_i^\top \Sigma)_j = 0$. As the constant k can be absorbed into the definition of Σ, we may simplify the first condition to $v_j(x) = v_i(x)$. These results motivate the following theorem, the detailed proof of which can be found in Duffie and Kan [1996].

Theorem 12.2.3. *Consider the SDE (12.55) with domain D given in (12.59) and assume that for all $i = 1, \ldots, d$,*

1. $\beta_i^\top (a - bx) \geq 0$ *for all x such that* $v_i(x) = 0$.
2. *For all $j = 1, \ldots, d$, if* $(\beta_i^\top \Sigma)_j \neq 0$ *then* $v_i(x) = v_j(x)$.

Then there exists a unique strong solution to (12.55) with $x(t) \in D$. If additionally $x(0)$ is such that $v_i(x(0)) > 0$ for all i, then also $v_i(x(t)) > 0$ provided that we replace Condition 1 with the stronger criterion

1.* $\beta_i^\top (a - bx) \geq \beta_i^\top \Sigma \Sigma^\top \beta_i / 2$ *for all x such that* $v_i(x) = 0$.

Remark 12.2.4. We recognize the criterion 1* above as a multi-variate generalization of the *Feller condition*, first encountered in Section 8.3.

We should emphasize that the regularity conditions outlined in Theorem 12.2.3 for the affine state variable SDE to be well-defined are quite strong and rule out many seemingly reasonable model specifications. Section 12.2.5 list some concrete models that satisfy the conditions of the theorem.

12.2.4 Discount Bond Prices

As advertised earlier, the main advantage of affine multi-factor models is the existence of discount bond reconstitution formulas that involve only the solution of ordinary Riccati ODEs. To develop this result, let

$$P(t,T,x) = \mathrm{E}\left(e^{-\int_t^T r(u)du} \middle| x(t) = x \right),$$

with $r(t)$ computed from (12.57) and the dynamics of the state variable vector $x(t)$ given in (12.55). From the Feynman-Kac result, we must have

$$\mathcal{L}P - \left(\xi_0 + \xi^\top x\right) P = 0, \tag{12.60}$$

where, using the same notation as in Section 12.1.9, $\mathcal{L}$ is a partial differential operator

$$\mathcal{L}P = P_t + (a - bx)^\top P_x + \frac{1}{2}\mathrm{tr}\left(P_{xx}\Sigma\upsilon(x)\upsilon(x)^\top\Sigma^\top\right).$$

Earlier results from Section 10.2.2 strongly hint that we should look for a solution of the form

$$P(t,T,x) = e^{A(T-t)-B(T-t)^\top x}, \tag{12.61}$$

where $A : \mathbb{R} \to \mathbb{R}$ and $B : \mathbb{R} \to \mathbb{R}^d$ are unknown deterministic functions. Inserting this solution into (12.60) and using the "matching principle"[13] we get the following result.

Proposition 12.2.5. *The solution to (12.60) is*

$$P(t,T,x) = \exp\left(A(T-t) - B(T-t)^\top x\right),$$

where the real-valued function $A(\tau)$ and the vector-valued function $B(\tau)$ satisfy the system of Riccati ODE equations

$$\frac{d}{d\tau}B(\tau) = -b^\top B(\tau) - \frac{1}{2}\beta^\top \mathrm{diag}(\Sigma^\top B(\tau))\Sigma^\top B(\tau) + \xi,$$
$$\frac{d}{d\tau}A(\tau) = -a^\top B(\tau) + \frac{1}{2}\alpha^\top \mathrm{diag}(\Sigma^\top B(\tau))\Sigma^\top B(\tau) - \xi_0,$$

[13] If $a + b^\top x = c + d^\top x$ holds for a non-empty open set, then $a = c$ and $b = d$.

with initial conditions $A(0) = 0$, $B(0) = 0$, *where* $\alpha = (\alpha_1, \ldots, \alpha_d)^\top$ *and the i-th row of matrix* β *is given by* $\beta_i^\top$, $i = 1, \ldots, d$. *The ODE system can be written component-wise as*

$$\begin{aligned}\frac{d}{d\tau}B_i(\tau) &= -\sum_{j=1}^{d} b_{j,i}B_j(\tau) \\ &\quad - \frac{1}{2}\sum_{k=1}^{d}\beta_{k,i}\left(\sum_{j=1}^{d}\Sigma_{j,k}B_j(\tau)\right)^2 + \xi_i, \quad i = 1, \ldots, d, \\ \frac{d}{d\tau}A(\tau) &= -\sum_{j=1}^{d} a_j B_j(\tau) + \frac{1}{2}\sum_{k=1}^{d}\alpha_k\left(\sum_{j=1}^{d}\Sigma_{j,k}B_j(\tau)\right)^2 - \xi_0.\end{aligned}$$

While analytical solution of the ODEs in Proposition 12.2.5 is sometimes possible (see Sections 12.2.5.2 and 12.2.5.3), in general one has to rely on numerical solution. The Runge-Kutta algorithm is a good choice for this.

An application of Proposition 12.2.5 reveals that discount bond price dynamics are

$$dP(t,T)/P(t,T) = r(t)\,dt - B(T-t)^\top \Sigma \upsilon\left(x(t)\right)\,dW(t),$$

and that forward rate dynamics are

$$df(t,T) = O(dt) + \frac{\partial B(T-t)^\top}{\partial T}\Sigma \upsilon\left(x(t)\right)\,dW(t),$$

where the drift term can be computed from the HJM results in Section 4.4. It follows that

$$\begin{aligned}&\mathrm{Corr}\left(df(t,T_1), df(t,T_2)\right) \\ &= \frac{Y(T_1-t,x(t))^\top Y(T_2-t,x(t))}{\sqrt{Y(T_1-t,x(t))^\top Y(T_1-t,x(t))}\sqrt{Y(T_2-t,x(t))^\top Y(T_2-t,x(t))}},\end{aligned}$$

where

$$Y\left(T-t,x\right) = \upsilon\left(x\right)\Sigma^\top\frac{\partial B(T-t)}{\partial T}.$$

Unlike the forward rate correlations computed earlier for the multi-factor Gaussian model, we notice that in the affine model $\mathrm{Corr}(df(t,T_1), df(t,T_2))$ generally depends on the random variable $x(t)$ and hence is *stochastic*.

Before moving on to concrete model examples, let us note that it is possible to extend the affine model to have time-dependent coefficients. In this case, the bond price equation would be

$$P(t,T,x) = \exp\left(A(t,T) - \sum_{i=1}^{d} B_i(t,T)x_i\right),$$

where, for $i = 1, 2, \ldots, d$,

$$\frac{d}{dt}B_i(t,T) = \sum_{j=1}^{d} b_{j,i}(t)B_j(t,T) + \frac{1}{2}\sum_{k=1}^{d}\beta_{k,i}(t)\left(\sum_{j=1}^{d}\Sigma_{j,k}(t)B_j(t,T)\right)^2 - \xi_i(t),$$

$$\frac{d}{dt}A(t,T) = \sum_{j=1}^{d} a_j(t)B_j(t,T) - \frac{1}{2}\sum_{k=1}^{d}\alpha_k(t)\left(\sum_{j=1}^{d}\Sigma_{j,k}(t)B_j(t,T)\right)^2 + \xi_0(t),$$

subject to $A(T,T) = B_1(T,T) = \ldots = B_d(T,T) = 0$. A certain amount of time-dependence would always be required in order to calibrate the model to the initial yield curve and to observed option prices.

12.2.5 Some Concrete Models

12.2.5.1 Fong-Vasicek Model

In Section 11.2.3 we encountered the two-factor model

$$\begin{aligned} dr(t) &= \varkappa_r\left(\vartheta_r - r(t)\right)dt + \sqrt{z(t)}\,dW_1(t), \\ dz(t) &= \varkappa_z\left(\vartheta_z - z(t)\right)dt + \eta\sqrt{z(t)}\left(\rho\,dW_1(t) + \sqrt{1-\rho^2}\,dW_2(t)\right). \end{aligned}$$

This model can be folded into the framework in Section 12.2.2 by writing $r(t) = x_1(t)$ (i.e., $\xi_0 = \xi_2 = 0$, $\xi_1 = 1$ in (12.57)), $z(t) = x_2(t)$, and

$$\begin{aligned} d\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} &= \left(\begin{pmatrix} \varkappa_r\vartheta_r \\ \varkappa_z\vartheta_z \end{pmatrix} - \begin{pmatrix} \varkappa_r & 0 \\ 0 & \varkappa_z \end{pmatrix}\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}\right)dt \\ &\quad + \begin{pmatrix} 1 & 0 \\ \eta\rho & \eta\sqrt{1-\rho^2} \end{pmatrix}\begin{pmatrix} \sqrt{x_2(t)} & 0 \\ 0 & \sqrt{x_2(t)} \end{pmatrix} d\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix}. \end{aligned}$$

Clearly this is of the form (12.55). It is easy to verify that the restrictions in Theorem 12.2.3 are all satisfied here.

12.2.5.2 Longstaff-Schwartz Model

Longstaff and Schwartz [1992] have proposed a two-factor extension of the CIR model we encountered in Section 10.2. In the language of (12.55)–(12.57),

the Longstaff-Schwartz (LS) model can be written as $r(t) = x_1(t) + x_2(t)$ ($\xi_0 = 0$, $\xi_1 = \xi_2 = 1$) with risk-neutral dynamics of the form

$$d\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \left(\begin{pmatrix} \varkappa_1 \vartheta_1 \\ \varkappa_2 \vartheta_2 \end{pmatrix} - \begin{pmatrix} \varkappa_1 & 0 \\ 0 & \varkappa_2 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \right) dt + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} \sqrt{x_1(t)} & 0 \\ 0 & \sqrt{x_2(t)} \end{pmatrix} d\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix}. \tag{12.62}$$

Again, it is easy to verify that the regularity conditions of Theorem 12.2.3 hold for (12.62). We notice that here the two state variables $x_1(t)$ and $x_2(t)$ are independent and both are time-homogeneous CIR. The independence assumption ensures that the analytical results from Section 10.2.2.1 can be used to solve the Riccati ODEs in Proposition 10.2.4 analytically. For completeness, we list the result below.

Lemma 12.2.6. *For the LS model (12.62), discount bond prices can be computed by*

$$P(t,T) = \exp\left(A_1(T-t) + A_2(T-t) - B_1(T-t)x_1(t) - B_2(T-t)x_2(t)\right),$$

where, for $i = 1, 2$,

$$A_i(\tau) = \varkappa_i \vartheta_i \sigma_i^{-2} (\varkappa_i + \gamma_i)\tau - 2\varkappa_i \vartheta_i \sigma_i^{-2} \ln\left(1 + \frac{(\varkappa_i + \gamma_i)(e^{\gamma_i \tau} - 1)}{2\gamma_i}\right),$$

$$B_i(\tau) = \frac{2(1 - e^{-\gamma_i \tau})}{(\varkappa_i + \gamma_i)(1 - e^{-\gamma_i \tau}) + 2\gamma_i e^{-\gamma_i \tau}},$$

with $\gamma_i = \sqrt{\varkappa_i^2 + 2\sigma_i^2}$.

Proof. From independence

$$P(t,T) = \mathrm{E}\left(e^{-\int_t^T r(u)\,du}\right) = \mathrm{E}\left(e^{-\int_t^T x_1(u)\,du}\right) \mathrm{E}\left(e^{-\int_t^T x_2(u)\,du}\right).$$

Application of the result in Proposition 10.2.4 (with $c_1 = 0$ and $c_2 = 1$) then proves the lemma. □

We should note that it is common to reparameterize the LS model in terms of $r(t)$ and $v(t) \triangleq d\mathrm{Var}_t(dr(t))/dt$, particularly when performing time series estimation. To quickly demonstrate the basic idea, notice that

$$dr(t) = \varkappa_1 (\vartheta_1 - x_1(t))\,dt + \varkappa_2 (\vartheta_2 - x_2(t))\,dt + \sigma_1 \sqrt{x_1(t)}\,dW_1(t) + \sigma_2 \sqrt{x_2(t)}\,dW_2(t),$$

such that

$$v(t) = \sigma_1^2 x_1(t) + \sigma_2^2 x_2(t). \tag{12.63}$$

Combining (12.63) with the equation $r(t) = x_1(t) + x_2(t)$ allows us to write, provided $\sigma_1 \neq \sigma_2$,

$$x_1(t) = \frac{\sigma_2^2 r(t) - v(t)}{\sigma_2^2 - \sigma_1^2}, \quad x_2(t) = \frac{v(t) - \sigma_1^2 r(t)}{\sigma_2^2 - \sigma_1^2}.$$

From this, it is possible to eliminate $x_1(t)$ and $x_2(t)$ from the SDEs for $r(t)$ and $v(t)$, resulting in a Markov model with $r(t)$ and $v(t)$ being the only state variables. We leave the details to the reader (or see Longstaff and Schwartz [1992]).

As shown in Longstaff and Schwartz [1992], the two-factor time-homogeneous specification (12.62) allows one to produce a substantially richer set of yield curve shapes than an ordinary one-factor CIR model. Of course, without the introduction of time-dependence in one or more parameters[14] (or application of the Dybvig "trick" from Section 11.3.2.4), the model will still never be able to perfectly fit the market-observed yield curve. The question of how to make an LS model acceptable for derivatives pricing purposes (which would necessarily involve further time-dependence and a scheme to allow for calibration to option prices) is of limited interest to us here, so we skip it and just point to Sections 10.2 and 12.3 for some general ideas. See also Clewlow and Strickland [1994] where some practical issues in parameter estimation for the LS model are discussed.

As a final remark, let us mention that the time-homogeneous LS model allows for an analytical pricing formula for European options on discount bonds. As both $x_1(t)$ and $x_2(t)$ are non-central chi-square random variables (see Section 8.3), the pricing formulas involve a two-dimensional non-central chi-square distribution, the practical computation of which is discussed in Chen and Scott [1992]. As time-homogeneous specifications are of little interest to the applications in this book, we do not list the pricing formulas, but simply refer to Longstaff and Schwartz [1992] for the details.

12.2.5.3 Multi-Factor CIR Models

A d-factor extension of the two-factor model in Section 12.2.5.2 would involve a system of decoupled SDEs of the form

$$dx_i(t) = \varkappa_i \left(\vartheta_i - x_i(t)\right) dt + \sigma_i \sqrt{x_i(t)}\, dW_i(t), \quad i = 1, \ldots, d, \tag{12.64}$$

with $r(t) = \sum_i x_i(t)$ and all Brownian motions $W_1(t), \ldots, W_d(t)$ mutually independent. The model satisfies the regularity conditions in Theorem 12.2.3 and it is clear from results in Section 8.3 that the resulting model will imply a non-negative short rate process. In fact, the short rate process will be strictly positive provided that there exists at least one $i \in \{1, \ldots, d\}$ for which the Feller condition $2\varkappa_i\vartheta_i \geq \sigma_i^2$ is satisfied. Discount bond pricing in the model (12.64) can be done analytically by re-using the one-factor results in Section 10.2.2.1, in the same manner as in Lemma 12.2.6. We leave the details to the reader. The uncoupled nature of the SDEs for the

[14] Longstaff and Schwartz [1993] suggest making $\varkappa_2$ a function of time.

various $x_i(t)$ in (12.64) is rather convenient as it allows us to reuse analytical and numerical techniques from Section 10.2. For instance, Monte Carlo simulation of (12.64) can proceed by simply simulating each $x_i(t)$ according to the scheme discussed in Section 10.2.7.

12.2.6 Brief Notes on Option Pricing

Pricing of contingent claims with no path-dependence can be done via the PDE (12.60), the solution of which would often proceed by finite difference methods, at least if the dimension d is modest. See Section 12.1.9 for further details. When model dimension is high or the payout is path-dependent, Monte Carlo methods are required. In some cases (as in Section 12.2.5.3 above), Monte Carlo discretization of d-dimensional affine models is a straightforward application of one-dimensional schemes.

To calibrate affine models to market option data, it is, as always, important to have fast schemes for swaption pricing. Without going into details, we note that the ideas laid out earlier in Section 10.2 for the one-factor models may be applied to the multi-factor affine models as well. As our treatment of (and interest in) affine models is rather cursory, we just refer the reader to material in Section 12.3 and Chapter 13. We should also note the existence of several dedicated swaption approximations in the literature for affine models; see Collin-Dufresne and Goldstein [2002a] for an example and further references.

12.3 The Quadratic Gaussian Model

A sub-class of affine models called *quadratic Gaussian* (QG) models is particularly attractive for practical applications. While currently more familiar to academics than to practitioners (see Chen et al. [2004], Alm et al. [2002], Assefa [2007]) the quadratic Gaussian models have several appealing properties: they are Markovian in a finite number of state variables, the state variables are Gaussian facilitating fast simulation, and the models in this class are capable of generating volatility smiles that can be parameterized in an intuitive way.

A QG model is obtained by generalizing (12.4) to include a quadratic term:

$$r(t) = z(t)^\top \gamma(t) z(t) + h(t)^\top z(t) + a(t), \tag{12.65}$$

where $\gamma(t)$ is a $d \times d$ symmetric matrix and, as before, $h(t)$ is a d-dimensional vector. The scalar function $a(t)$ is used to fit the initial yield curve. The state variable vector $z(t)$ follows (12.3), i.e.

$$dz(t) = g(t)^\top dW(t), \quad z(0) = 0, \tag{12.66}$$

with $W(t)$ a Brownian motion under the risk-neutral measure. Just like a linear Gaussian model, a quadratic model can be expressed in terms of mean-reverting state variables:

$$r(t) = x(t)^\top \widetilde{\gamma}(t) x(t) + \mathbf{1}^\top x(t) + a(t), \tag{12.67}$$

where

$$\widetilde{\gamma}(t) = H(t)^{-1} \gamma(t) H(t)^{-1},$$

and the transformed state variables $x(t)$, defined by $x(t) = H(t)z(t)$, follow

$$dx(t) = -\varkappa(t)x(t)\,dt + (g(t)H(t))^\top\, dW(t),$$

with $H(t)$ defined by (12.7) and $\varkappa(t)$ defined by (12.8). Other linear transformations, along the lines of Section 12.1.1.2, are also possible. Such representations may provide certain advantages for model implementation and numerical methods, as we often prefer to keep the diffusion term as constant in time as possible. Nevertheless, to reduce notational clutter we here stick to the driftless form (12.65)–(12.66).

We briefly saw a one-dimensional quadratic Gaussian model in Section 10.2.6; here, following Piterbarg [2009a], we study the multi-dimensional version.

12.3.1 Quadratic Gaussian Models are Affine

To show that the model defined by (12.65)–(12.66) is indeed affine, we introduce a vector of extra state variables $u(t)$ of length d^2, whose elements are pairwise products of the coordinates of $z(t)$, i.e.

$$u(t) = (z_1(t)z_1(t), z_1(t)z_2(t), z_1(t)z_3(t), \ldots, z_d(t)z_d(t))^\top.$$

Then, clearly, $r(t)$ is a linear function of $(z(t), u(t))$, and the coefficients of the SDE for the matrix $z(t)z(t)^\top$ are linear in $z(t)$:

$$d\left(z(t)z(t)^\top\right) = g(t)^\top\, dW(t)\, z(t)^\top + z(t)\, dW(t)^\top g(t) + \left(g(t)^\top g(t)\right) dt.$$

As we can write $u(t)$ by "unwrapping" the rows of $z(t)z(t)^\top$ into a vector, it follows that the coefficients of the SDE for $du(t)$ are linear in $z(t)$.

The analysis above makes it clear that the combined state variable vector $(z(t), u(t))$ has multi-factor affine dynamics. Represented as a standard affine model, the quadratic model would evidently require a total of $d(d+1)$ state variables (rather than just d), so there is often good reason *not* to use such a representation explicitly. We note in passing that the quadratic parameterization satisfy the regularity constraints from Section 12.2.3 by construction.

12.3.2 The Basics

Since the QG model (12.65)–(12.66) is affine, it should come as no surprise that bond reconstruction formulas are available.

Proposition 12.3.1. *In the quadratic Gaussian model (12.65)–(12.66), zero-coupon discount bonds are exponentials of quadratic forms,*

$$\begin{aligned} &-\ln P(t,T) \\ &\quad = z(t)^\top \gamma(t,T)\, z(t) + h(t,T)^\top z(t) + a(t,T) - \ln\left(P(0,T)/P(0,t)\right), \end{aligned}$$

with $\gamma(t,T)$, $h(t,T)$ satisfying Riccati equations

$$\begin{aligned} -\frac{d}{dt}\gamma(t,T) + 2\gamma(t,T)\, g(t)g(t)^\top \gamma(t,T) &= \gamma(t), \\ -\frac{d}{dt}h(t,T) + 2\gamma(t,T)\, g(t)g(t)^\top h(t,T) &= h(t), \end{aligned} \tag{12.68}$$

with terminal conditions $\gamma(T,T) = 0$, $h(T,T) = 0$.

Proof. By the same arguments as Proposition 12.2.5. □

Remark 12.3.2. The function $a(t,T)$ also satisfies a Riccati equation; however, we find that it is better to determine it from the no-arbitrage condition $P(0,t)\mathrm{E}^t(P(t,T)) = P(0,T)$, where E^t is the expected value operator under the t-forward measure Q^t, so that

$$a(t,T) = \ln \mathrm{E}^t\left(\exp\left(-z(t)^\top \gamma(t,T)\, z(t) - h(t,T)^\top z(t)\right)\right). \tag{12.69}$$

To calculate $a(\cdot,T)$ in (12.69), we need to know the distribution of $z(t)$ under the t-forward measure Q^t; this distribution will also be of general use in option pricing. From Proposition 12.3.1,

$$\frac{dP(t,T)}{P(t,T)} = O(dt) - \left(2z(t)^\top \gamma(t,T) + h(t,T)^\top\right) g(t)^\top dW(t),$$

so by Girsanov's theorem (Theorem 1.5.1),

$$dW^T(t) = dW(t) + g(t)\left(2\gamma(t,T)\, z(t) + h(t,T)\right) dt \tag{12.70}$$

is a Brownian motion under the T-forward measure Q^T. We use this fact to obtain the following result.

Proposition 12.3.3. *In the quadratic Gaussian model (12.65)–(12.66), the dynamics of the state process $z(t)$ in the T-forward measure Q^T are given by*

$$dz(t) = \left(m^T(t) - k^T(t)z(t)\right) dt + g(t)^\top dW^T(t), \tag{12.71}$$

where

$$k^T(t) = 2g(t)^\top g(t)\gamma(t,T), \quad m^T(t) = -g(t)^\top g(t)h(t,T),$$

and $W^T(t)$ *is a* Q^T*-Brownian motion. In particular,* $z(t)$ *is a Gaussian process under any* T*-forward measure, and is given in the integrated form by*

$$z(s) = J_{k^T}(s)\left[J_{k^T}(t)^{-1}z(t) + \int_t^s J_{k^T}(u)^{-1}m^T(u)\,du \right.$$
$$\left. + \int_t^s J_{k^T}(u)^{-1}g(u)^\top\,dW^T(u)\right], \quad (12.72)$$

where the matrix-valued function $J_{k^T}(t)$ *is defined by (12.15), i.e. satisfies an ODE*

$$\frac{d}{dt}J_{k^T}(t) = -2g(t)^\top g(t)\gamma(t,T)\,J_{k^T}(t), \quad J_{k^T}(0) = I. \quad (12.73)$$

Proof. The equation (12.71) follows from (12.70). Integrating a linear Gaussian SDE (12.71) we obtain (12.72), see Lemma 12.1.6 or Karatzas and Shreve [1997]. □

As $z(t)$ is Gaussian under any forward measure, its distribution is fully specified by its first and second moments.

Proposition 12.3.4. *In the quadratic Gaussian model (12.65)–(12.66), the conditional moments of the Gaussian state process* $z(t)$ *under the* T*-forward measure* Q^T*,*

$$m^T(t,s,z) \triangleq \mathrm{E}^T(z(s)|\,z(t)=z), \quad (12.74)$$
$$\nu^T(t,s,z) \triangleq \mathrm{Var}^T(z(s)|\,z(t)=z) \quad (12.75)$$
$$= \mathrm{E}^T\Big(\big(z(s)-m^T(t,s,z)\big)\big(z(s)-m^T(t,s,z)\big)^\top\Big|\,z(t)=z\Big),$$

are given by

$$m^T(t,s,z) = J_{k^T}(s)J_{k^T}(t)^{-1}z - J_{k^T}(s)\int_t^s J_{k^T}(u)^{-1}g(u)^\top g(u)h(u,T)\,du, \quad (12.76)$$

$$\nu^T(t,s,z) = J_{k^T}(s)\left(\int_t^s J_{k^T}(u)^{-1}g(u)^\top g(u)\left(J_{k^T}(u)^{-1}\right)^\top du\right)J_{k^T}(s)^\top, \quad (12.77)$$

where $J_{k^T}(t)$ *is defined by (12.73).*

Proof. Follows immediately from (12.72). □

To compute the function $a(t,T)$ via (12.69) we need to know the moment-generating function of a quadratic form of a Gaussian vector.

Proposition 12.3.5. *Let Z be a K-dimensional Gaussian vector with mean m and variance ν. Let Q be a symmetric $K \times K$ matrix and u a K-dimensional vector. Define*

$$\Psi(u, Q; m, \nu) \triangleq \ln \mathrm{E}\left(\exp\left(Z^\top Q Z + u^\top Z\right)\right).$$

If $\det(I - 2Q\nu) > 0$, *then*

$$\Psi(u, Q; m, \nu) = \frac{1}{2}\left(2m^\top Q + u^\top\right)\nu\left(I - 2Q\nu\right)^{-1}\left(2Qm + u\right) + m^\top Q m + u^\top m - \frac{1}{2}\ln\left(\det\left(I - 2Q\nu\right)\right).$$

Proof. In Appendix 12.A. □

In addition to the proof of Proposition 12.3.5, Appendix 12.A contains a number of technical results useful for working with quadratic forms of Gaussian vectors.

The QG model is Markovian in d state variables (the vector z) and it should not be surprising that, with the help of the quadratic term, it can generate a genuine U-shaped volatility smile (see Figure 12.3). The state vector follows a Gaussian process, and it can be simulated at minimal computational cost (see Section 12.3.7). While these properties make the quadratic model both flexible and numerically efficient, its practical usefulness ultimately hinges on our ability to parameterize it in a sensible and intuitive way. Such a task is not trivial given that the generic time-dependent quadratic term $\gamma(t)$ is, essentially, unconstrained. While the richness of the model allows for a potentially large number of parameterization strategies, we here have little interest in exhaustive classification and content ourselves with presenting just one possible — and quite reasonable, we think — approach.

12.3.3 Parameterization

12.3.3.1 Smile Generation

To devise a parameterization strategy for the QG model, it is useful to understand the mechanism by which it generates a volatility smile. As the one-factor case is somewhat degenerate, we first look at the two-factor case for inspiration. We find it convenient to parameterize the quadratic term in such a way that we can identify one state variable as a "curve" factor and the other as a "volatility" factor (see Tezier [2005]). With that in mind, we set $d = 2$, $g_{1,2}(t) = g_{2,1}(t) = 0$, $h_1(t) = e^{-\varkappa t}$, $h_2(t) = 0$,

$$\gamma_{1,1}(t) = \eta\varpi h_1(t)^2, \quad \gamma_{1,2}(t) = \frac{\eta}{2}\overline{\varpi} h_1(t), \quad \gamma_{2,2}(t) = 0, \tag{12.78}$$

where $\overline{\varpi} = \sqrt{1 - \varpi^2}$. According to (12.65), the short rate is then given by

$$r(t) = (1 + \eta v(t))\, h_1(t) z_1(t) + a(t), \tag{12.79}$$

where

$$v(t) = \varpi h_1(t) z_1(t) + \overline{\varpi} z_2(t).$$

If $\eta = 0$, the expression for the short rate reduces to

$$r(t) = h_1(t) z_1(t) + a(t),$$

and the model then becomes a one-factor (linear) Gaussian,

$$dr(t) = \varkappa\,(\vartheta(t) - r(t))\, dt + e^{-\varkappa t} g_{1,1}(t)\, dW_1(t), \quad \vartheta(t) = a(t) + a'(t)/\varkappa.$$

Fittingly, we can identify $h_1(t)z_1(t)$ as a *curve factor*, i.e. as the factor that drives the state of the yield curve. If $\eta \neq 0$, the short rate is given by the curve factor times $1 + \eta v(t)$. As high values of $v(t)$ imply high volatility of $r(t)$, $\eta v(t)$ plays the role of[15] "stochastic volatility". Consequently, η may be interpreted as a volatility of volatility parameter. We notice that the volatility factor $v(t)$ is a linear combination of the curve factor $h_1(t)z_1(t)$ and a process $z_2(t)$ which is independent of the curve factor. The parameter ϖ therefore determines the correlation between the curve factor and the volatility factor.

As one would intuitively guess, the model outlined above is capable of producing volatility smiles that are similar to those of the stochastic volatility models we encountered in Chapter 8. Figure 12.3 shows a sample fit of the QG model to a market-implied volatility smile.

The parameterization (12.78) not only serves to identify one of the state variables as a curve factor and the other as a volatility factor, it also conceptually separates parameters that affect the volatility structure of the model ($h_1(t)$, $g_{1,1}(t)$), and those that affect the volatility smile (η, ϖ). Such separation is very convenient for building intuition for model dynamics and for the development of efficient European option approximations and practical calibration algorithms. Consequently, we seek to impose a similar structure as we build QG models of dimensions higher than two.

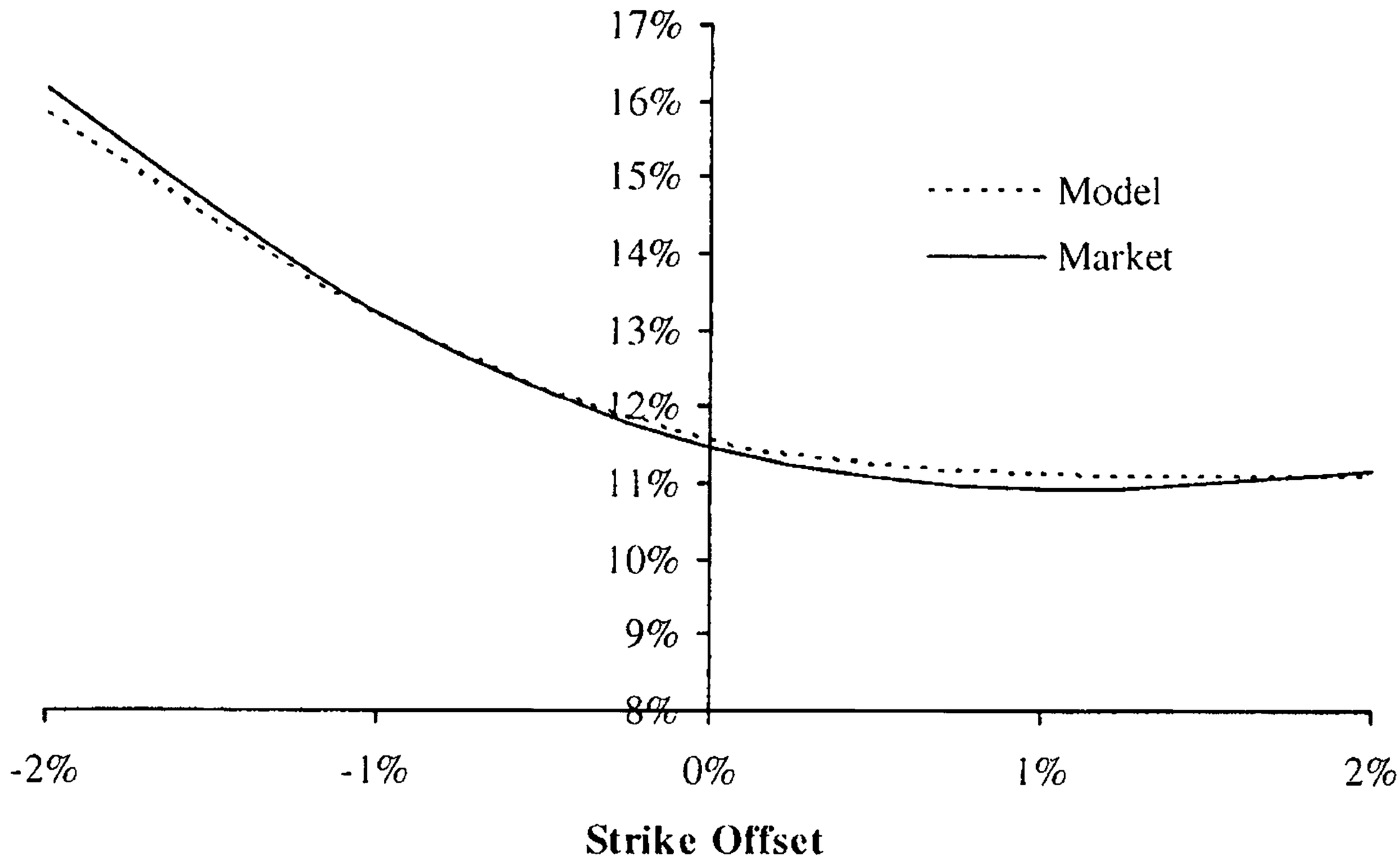
12.3.3.2 Quadratic Term

Given a budget of d curve factors and 1 volatility factor, we follow the example from the previous section and use a linear function of the d curve factors to define the yield curve dynamics, and the volatility factor to drive multiplicative scaling.

Let $z(t)$ be a $(d+1)$-dimensional column vector of factors, with the first d coordinates, denoted by $z_{1:d}(t)$, being curve factors and $z_{d+1}(t)$ being the

[15] In the language of Section 11.2.3, the stochastic volatility driver is here evidently of the *spanned* type. See also Section 12.3.6.

Fig. 12.3. Implied Volatility Smile



Notes: Fit of a four-factor quadratic Gaussian model to the volatility smile of 10y×10y swaptions as observed in the summer of 2007. The swaption strike ("Strike Offset") is set as an offset to the forward swap rate.

single volatility factor. Let the $(d+1) \times (d+1)$ matrix-valued function $g(t)$ in (12.66) be of the block form,

$$g(t) = \begin{pmatrix} g_{1:d}(t) & 0 \\ 0 & g_{d+1,d+1}(t) \end{pmatrix},$$

with $g_{1:d}(t)$ being a $d \times d$ diffusion matrix for curve factors, and $g_{d+1,d+1}(t)$ being a scalar diffusion coefficient for the volatility factor $z_{d+1}(t)$. We write (12.66) as

$$dz(t) = g(t)^\top \, dW(t),$$

where

$$W(t) = \begin{pmatrix} W_{1:d}(t) \\ W_{d+1}(t) \end{pmatrix}$$

is a $(d+1)$-dimensional Brownian motion in the risk-neutral measure Q. Notice that we assume that the volatility factor is independent of the curve factors. In (12.65), let the linear term $h(t)$, a $(d+1)$-dimensional column vector, have a last element 0 so that the volatility factor has no first-order effect on the short rate,

$$h(t) = \begin{pmatrix} h_{1:d}(t) \\ 0 \end{pmatrix}. \tag{12.80}$$

Finally, in (12.65) the quadratic term $\gamma(t)$ is specified to have the block form,

$$\gamma(t) = \begin{pmatrix} \gamma_{1:d}(t) & \gamma_{1:d,d+1}(t) \\ \gamma_{1:d,d+1}(t)^\top & 0 \end{pmatrix}, \tag{12.81}$$

where the $d \times d$ matrix $\gamma_{1:d}(t)$ is given by

$$\gamma_{1:d}(t) = \eta\varpi h_{1:d}(t) h_{1:d}(t)^\top,$$

and the $d \times 1$ vector $\gamma_{1:d,d+1}(t)$ is given by

$$\gamma_{1:d,d+1}(t) = \frac{1}{2}\eta\overline{\varpi} h_{1:d}(t), \quad \overline{\varpi} = \sqrt{1-\varpi^2},$$

where $\varpi \in [-1, 1]$.

Combining everything above, our model for the short rate is given by

$$r(t) = \left(1 + \eta\left(\varpi \times \left(h(t)^\top z(t)\right) + \overline{\varpi} \times z_{d+1}(t)\right)\right) \times \left(h(t)^\top z(t)\right) + a(t), \tag{12.82}$$

and we obtain a representation of the short rate as a curve factor times one plus a volatility process, similar to (12.79).

12.3.3.3 Linear Term

To understand the volatility structure of the QG model better, let us momentarily set $\eta = 0$ in (12.82). The model then reduces to a multi-factor (linear) Gaussian model,

$$r(t) = h(t)^\top z(t) + a(t), \tag{12.83}$$

for which we can use the tools and intuition we developed earlier in this chapter. In particular, the ideas of Section 12.1.7 could be fruitfully applied. Without repeating ourselves, we assume that d benchmark rates are specified, and the volatility structure is parameterized by their instantaneous volatilities $\lambda_i(t)$, $i = 1, \ldots, d$, and instantaneous correlations $\{\chi_{i,j}(t)\}$, $i, j = 1, \ldots, d$. Assuming, for concreteness, that the loadings vector $h(t)$ is specified through a series of d constant mean reversions,

$$h(t) = \left(e^{-\varkappa_1 t}, \ldots, e^{-\varkappa_d t}, 0\right)^\top, \tag{12.84}$$

we see that the "curve" part of the diffusion coefficient, i.e. the matrix $g_{1:d}(t)$, can be obtained by the algorithm from Section 12.1.7. Together with the quadratic form parameterization in (12.81), this completely specifies the model. Of course, it still remains to set the various model parameters in such a way that market prices for interest rate options are matched, a topic that we turn to next.

12.3.4 Swaption Pricing

12.3.4.1 State Vector Distribution Under the Annuity Measure

Adopting the notation employed in Section 12.1.6, we consider a swaption that fixes at $T_0 > 0$ and has fixed payments at $T_1 < \ldots < T_N$, $\tau_i = T_{i+1} - T_i$. We remind the reader that the relevant swap rate and the annuity are given by

$$S(t) = \frac{P(t,T_0) - P(t,T_N)}{A(t)}, \quad A(t) = \sum_{i=0}^{N-1} \tau_i P(t,T_{i+1}). \tag{12.85}$$

The corresponding annuity measure is the measure Q^A for which $A(t)$ is the numeraire. As exploited on numerous occasions already, in Q^A the swap rate $S(t)$ is a martingale and the swaption price may be obtained as a European option on $S(T_0)$.

The term distribution of the state vector on the fixing date T_0 of the swap rate, $z(T_0)$, is easily characterized.

Lemma 12.3.6. *The distribution of $z(T_0)$ in the annuity measure Q^A is a Gaussian mixture, with the density $\psi^A(z)$ of $z(T_0)$ given by*

$$\psi^A(z) = \sum_{i=0}^{N-1} w_i^A \phi\left(z; m^{T_{i+1}}(0,T_0,0), \nu^{T_{i+1}}(0,T_0,0)\right), \tag{12.86}$$

$$w_i^A = \frac{\tau_i P(0,T_{i+1})}{A(0)}, \quad i = 0,\ldots,N-1,$$

where $\phi(z;m,\nu)$ is a $(d+1)$-dimensional Gaussian density with mean m and covariance matrix ν. The mean $m^{T_{i+1}}(0,T_0,0)$ and the covariance matrix $\nu^{T_{i+1}}(0,T_0,0)$ are given in Proposition 12.3.4.

Proof. For an arbitrary scalar function $g(z)$ we have, from standard measure change arguments,

$$\begin{aligned}
\mathrm{E}^A\left(g(z(T_0))\right) &= \frac{P(0,T_0)}{A(0)} \mathrm{E}^T\left(g(z(T_0))A(T_0)\right) \\
&= \frac{P(0,T_0)}{A(0)} \sum_{i=0}^{N-1} \mathrm{E}^T\left(g(z(T_0))\tau_i P(T_0,T_{i+1})\right) \\
&= \sum_{i=0}^{N-1} \frac{\tau_i P(0,T_{i+1})}{A(0)} \mathrm{E}^{T_{i+1}}\left(g(z(T_0))\right).
\end{aligned}$$

The lemma follows directly from Proposition 12.3.4. □

It follows from Lemma 12.3.6 that the mean and the covariance matrix of $z(T_0)$ in the annuity measure Q^A are given by

$$\mathrm{E}^A(z(T_0)) = \sum_{i=0}^{N-1} w_i^A m^{T_{i+1}}(0, T_0, 0), \tag{12.87}$$

$$\mathrm{Var}^A(z(T_0)) = \sum_{i=0}^{N-1} w_i^A \left(\nu^{T_{i+1}}(0, T_0, 0) + m^{T_{i+1}}(0, T_0, 0)^2\right) - (\mathrm{E}^A(z(T_0)))^2.$$

We emphasize that $z(T_0)$ is not Gaussian under the annuity measure, although it is tempting to use the approximation

$$z(T_0) \overset{d}{\approx} \mathcal{N}\left(\mathrm{E}^A(z(T_0)), \mathrm{Var}^A(z(T_0))\right). \tag{12.88}$$

In practical uses of either (12.86) or (12.87), we need an efficient way to compute the moments $m^{T_{i+1}}(0, T_0, 0)$ and $\nu^{T_{i+1}}(0, T_0, 0)$ of $z(T_0)$. The results in Proposition 12.3.4 require an integration in the time domain for each $i = 0, \ldots, N-1$, and, with N being potentially large, the computational effort would therefore often be quite high. Fortunately, a much faster alternative is available. Once $m^{T_0}(0, T_0, 0)$ and $\nu^{T_0}(0, T_0, 0)$ are on hand — and they are always known as both are required for yield curve fitting via (12.69) — the moments of the factors observed on the same date, but under all other forward measures, can be calculated by a measure-change formula:

$$m^{T_{i+1}}(0, T_0, 0) = \mathrm{E}^{T_{i+1}}(z(T_0)) = \frac{\mathrm{E}^{T_0}\left(z(T_0) P(T_0, T_{i+1})\right)}{\mathrm{E}^{T_0}\left(P(T_0, T_{i+1})\right)}. \tag{12.89}$$

The expression on the right-hand side is obtained in closed form in Corollary 12.A.3 of Appendix 12.A, utilizing the fact that the discount bond is an exponential of a quadratic form of a Gaussian vector.

12.3.4.2 Exact Pricing of European Swaptions

The result of Lemma 12.3.6 that shows that the distribution of the state vector in the annuity measure is a mixture of Gaussian distributions leads us to one possible European swaption pricing method. To elaborate, let us define $S(T_0, z)$ to be the value of the swap rate $S(T_0)$ when $z(T_0) = z$. It follows from (12.86) that the value of a European swaption with strike c can be represented as

$$V_{\text{swaption}}(0) = A(0)\mathrm{E}^A\left(S(T_0, \mu_\xi + \sigma_\xi X) - c\right)^+, \tag{12.90}$$

where ξ is an integer-valued random variable with distribution

$$\mathrm{Q}(\xi = i) = w_i^A, \quad i = 0, \ldots, N-1,$$

X is a standard Gaussian $(d+1)$-dimensional random variable, and

$$\mu_i = m^{T_{i+1}}(0, T_0, 0), \quad \sigma_i = \sqrt{\nu^{T_{i+1}}(0, T_0, 0)}, \quad i = 0, \ldots, N-1.$$

Evaluation of (12.90) may, for instance, proceed by a simple one-step Monte Carlo simulation

$$\mathrm{E}^A\left(S(T_0)-c\right)^+ \approx \frac{1}{L}\sum_{l=1}^{L}\left(S\left(T_0,\mu_{\xi_l}+\sigma_{\xi_l}X_l\right)-c\right)^+,$$

where $\{\xi_l\}_{l=1}^L$ is an i.i.d. sample from the distribution of ξ, and $\{X_l\}_{l=1}^L$ is an i.i.d. sample from the standard $(d+1)$-dimensional Gaussian distribution.

While this scheme still requires a Monte Carlo simulation to compute a swaption value, we emphasize that it is simple and fast: one only needs to draw a sample of the variable ξ (which essentially defines what mean/variance to use) and one sample of a standard Gaussian vector, in order to sample *directly* the terminal distribution of the swap rate. When combined with quasi-random numbers as in Section 3.2.10.1, the method could be seen as a type of outright $(d+1)$-dimensional numerical integration.

12.3.4.3 Approximations for European Swaptions

As the short rate and all continuously compounded forward rates are quadratic forms of the state vector, it seems reasonable to assume that swap rates are approximately of this form as well. We can use this observation to develop analytically tractable approximations for swaptions. Let us define

$$h_S = \nabla^\top S\left(T_0,z^*\right), \quad \gamma_S = \frac{1}{2}\nabla^\top\nabla S\left(T_0,z^*\right), \tag{12.91}$$

where ∇ is the gradient operator, $\nabla = (\partial/\partial z_1,\ldots,\partial/\partial z_{d+1})$ (row vector). In essence, h_S is the first, and γ_S is the (half of the) second-order derivative of the swap rate function $S(T_0,\cdot)$ at a specific point z^*. Both are easily computed by a numerical finite difference algorithm. The expansion point z^* could be 0 or, for a slightly more accurate approximation,

$$z^* = \mathrm{E}^A(z(T_0))$$

as computed by (12.87). Applying Taylor's expansion,

$$S(T_0,z) \approx \left(z-z^*\right)^\top\gamma_S\left(z-z^*\right)+h_S^\top\left(z-z^*\right)+s\left(z^*\right).$$

To ensure that the forward swap rate is repriced correctly under this approximation, we adjust the constant term accordingly, and define the *quadratic approximation to the swap rate* by

$$\begin{aligned} S(T_0,z) &\approx S_q(T_0,z), \\ S_q(T_0,z) &= z^\top\gamma_S z + h_S^\top z - \mathrm{E}^A\left(z(T_0)^\top\gamma_S z(T_0)+h_S^\top z(T_0)\right)+S(0), \end{aligned} \tag{12.92}$$

where the required expected value is calculated in Corollary 12.A.1.

Under the quadratic approximation to the swap rate and Gaussian approximation to the distribution of $z(T_0)$ (see (12.88)), it becomes possible to price options on the swap rate using Fourier integration methods. For this, we need the moment-generating function

$$q(u) \triangleq \widetilde{\mathrm{E}}\left(e^{uS_q(T_0,z(T_0))}\right),$$

where $\widetilde{\mathrm{E}}$ is the expected value operator under the assumption that $z(T_0)$ is Gaussian. This expression is indeed available in closed form, thanks to Proposition 12.3.5:

$$q(u) = \exp\left(\Psi\left(uh_S, u\gamma_S; \mathrm{E}^A\left(z(T_0)\right), \mathrm{Var}^A\left(z(T_0)\right)\right)\right).$$

Given $q(u)$, we can compute the option price

$$\widetilde{\mathrm{E}}\left((S_q(T_0, z(T_0)) - c)^+\right)$$

by Theorem 8.4.3. A suitable control variate as in Theorem 8.4.4 is essential for improving numerical performance.

It should be noted here that the application of Fourier methods to swaption pricing does not hinge on the Gaussian approximation (12.88), as the true moment-generating function of $S_q(T_0, z(T_0))$ is readily available under Q^A. Indeed, from the mixing formula (12.86) we get

$$\begin{aligned}\mathrm{E}^A\left(e^{uS_q(T_0,z(T_0))}\right) &= \sum_{i=0}^{N-1} w_i^A \mathrm{E}^{T_{i+1}}\left(e^{uS_q(T_0,z(T_0))}\right)\\ &= \sum_{i=0}^{N-1} w_i^A \exp\left(\Psi\left(uh_S, u\gamma_S; m^{T_{i+1}}\left(0,T_0,0\right), \nu^{T_{i+1}}\left(0,T_0,0\right)\right)\right).\end{aligned}$$

We can use this formula in option pricing instead of the Gaussian approximation;; however, we find that the resulting increase in computational cost — for each value of u we now require N evaluations of the function Ψ, instead of just one — is rarely justified by the (slight) improvements of accuracy.

While we find the Fourier integration method to be robust and efficient[16], we can further explore the specifics of our parameterization to design even faster valuation algorithms. In particular, from (12.82) we notice that the quadratic form defining the short rate is not of full rank, as the short rate is a quadratic function of only two "aggregate" quantities, $h(t)^\top z(t)$ and $z_{d+1}(t)$. We can expect that this rank-2 structure is preserved, at least approximately, in swap rates. The linear term in the quadratic approximation for the swap rate, $h_S^\top z$, will be one of the two aggregate factors to use in re-parameterizing

[16] Contrary to some claims in the literature, see e.g. Boyarchenko and Levendorski [2007].

the quadratic part $z^\top \gamma_S z$. The other one, naturally, will be the stochastic volatility factor. In summary, we seek to approximate

$$z^\top \gamma_S z \approx \left(h_S^\top z, z_{d+1}\right) \widehat{\gamma}_S \left(h_S^\top z, z_{d+1}\right)^\top, \tag{12.93}$$

where $\widehat{\gamma}_S$ is a 2×2 (symmetric) matrix.

To formalize the idea outlined above, let us define a two-dimensional stochastic vector $\widehat{z}(T_0) = (\widehat{z}_1(T_0), \widehat{z}_2(T_0))^\top$ by

$$\widehat{z}_1(T_0) = h_S^\top z(T_0), \quad \widehat{z}_2(T_0) = z_{d+1}(T_0),$$

or, in matrix notation,

$$\begin{pmatrix} \widehat{z}_1(T_0) \\ \widehat{z}_2(T_0) \end{pmatrix} = Rz(T_0), \quad R = \begin{pmatrix} h_{S,1} & \dots & h_{S,d} & h_{S,d+1} \\ 0 & \dots & 0 & 1 \end{pmatrix}.$$

Then (12.93) can be re-written as

$$z^\top \gamma_S z \approx z^\top R^\top \widehat{\gamma}_S R z.$$

Formally, we set $\widehat{\gamma}_S$ to be a solution to the following minimization problem

$$\mathrm{Var}^A \left(z(T_0)^\top \left(\gamma_S - R^\top \widehat{\gamma}_S R\right) z(T_0)\right) \to \min; \tag{12.94}$$

we then call this $\widehat{\gamma}_S$ a *rank-2 quadratic approximation*. The problem (12.94) can be solved explicitly (if rather tediously) using Corollary 12.A.1, resulting in the following approximation to the value of an option on the swap rate.

Theorem 12.3.7. *Under the rank-2 quadratic approximation (12.94) to the swap rate defined by (12.85), and Gaussian approximation to the distribution of the state vector $z(T_0)$ under Q^A, the value of a European swaption with strike c is approximately given by*

$$V_{\text{swaption}}(0) \approx A(0) \int_{\mathbb{R}^2} \left(\widehat{z}^\top \widehat{\gamma}_S \widehat{z} + \widehat{z}_1 + \widehat{\alpha}_S + S(0) - c\right)^+ \phi\left(\widehat{z}; \widehat{m}, \widehat{\nu}\right) d\widehat{z}, \tag{12.95}$$

where $\widehat{z} = (\widehat{z}_1, \widehat{z}_2)^\top$ and $\phi(\widehat{z}; \widehat{m}, \widehat{\nu})$ a two-dimensional Gaussian density with mean $\widehat{m}$ and covariance matrix $\widehat{\nu}$. Also, $\widehat{\alpha}_S$ is defined by

$$\widehat{\alpha}_S = -\mathrm{E}^A \left(z(T_0)^\top R^\top \widehat{\gamma}_S R z(T_0) + h_S^\top z(T_0)\right), \tag{12.96}$$

with

$$\widehat{\gamma}_S = \left(2\widehat{m}\widehat{m}^\top + \widehat{\nu}\right)^{-1} R \left(2mm^\top + \nu\right) \gamma_S \nu R^\top \widehat{\nu}^{-1},$$

where

$$m = \mathrm{E}^A \left(z(T_0)\right), \quad \nu = \mathrm{Var}^A \left(z(T_0)\right), \quad \widehat{m} = Rm, \quad \widehat{\nu} = R\nu R^\top,$$

As in Section 12.1.6.1, the two-dimensional integral in (12.95) can be computed efficiently by conditioning on one of the integration variables, evaluating the resulting sub-expression in closed form, and performing the outer integration using, say, *Gauss-Hermite* quadrature (see Press et al. [1992]). We omit straightforward details. Table 12.1 demonstrates typical quality of various approximations. Data in the table represent 10y×10y swaption volatilities, computed from the same model settings as those used to construct Figure 12.3.

Strike	ATM-2%	ATM-1%	ATM	ATM+1%	ATM+2%
Model exact	15.84	13.17	11.54	11.12	11.09
Gauss approx	15.84	13.17	11.55	11.12	11.09
Gauss+Quadratic	15.89	13.17	11.51	11.06	11.00
Gauss+Quadratic+Rank 2	15.89	13.17	11.51	11.07	10.99

Table 12.1. Implied Black Volatilities in a Quadratic Gaussian Model. "Exact" is defined by (12.90). "Gauss approx" is defined by (12.88). "Quadratic" is defined by (12.92). "Rank 2" is defined by (12.95). Results for a 10y×10y swaption in %.

12.3.5 Calibration

While a number of viable approaches to calibration exist, we recommend organizing it as a multi-pass bootstrap algorithm, an approach that should be familiar to the reader by now (see e.g. Section 10.2.5.2). First, the parameters ϖ and η are fixed to the desired shape of volatility smile. Next, the correlation matrix of the benchmark rates $\{\chi_{i,j}(t)\}$ is parameterized by a convenient functional form (see the discussion in Section 14.3.2), generally to either match historical correlations of the relevant rates or to fit market-implied prices of CMS spread options. After that, the calibration problem is reduced to the problem of matching at-the-money swaption volatilities by manipulating the benchmark rate volatilities $\lambda_i(t)$, $i = 1, \ldots, d$ (the reader will recall from Sections 12.3.3.3 and 12.1.7 that they are used to construct the state vector diffusion matrix $g(t)$). Having d time-dependent volatilities allows us to calibrate to d swaption strips. While not strictly necessary, we find it convenient to choose the swaption strips to be of constant tenor, with the tenors matching those of the benchmark rates. Denoting $t_1 < \ldots < t_K$ to be the expiry dates of the swaptions in the calibration set, we break the calibration into K subproblems, where in the j-th sub-calibration we match the j-th row of the swaption matrix by tweaking $\lambda_i(t_j)$, $i = 1, \ldots, d$. In the linear case, i.e. when the quadratic term is zero, only one pass for $j = 1, \ldots, K$ is required, as swaption prices with expiry t_j depend on $\lambda(s)$ for $s \in [0, t_j]$ only. In the general quadratic case, this is no longer the case, and prices of swaptions with expiry t_j depend on $\lambda(s)$ for all s (for $s > t_j$

through bond reconstruction formulas). However, this "tail" dependence is minor, and we can still calibrate sequentially by performing multiple passes (typically two or three). For a more performant algorithm, we could use fast swaption approximations for initial pass(es), saving a more accurate one for the final pass. The specifics of such a multi-pass calibration should follow closely the ideas discussed in more details in the context of affine models in Section 10.2.5.

12.3.6 Spanned Stochastic Volatility

While we have used the term "stochastic volatility" throughout to describe our parameterization of the QG model, the model clearly does not involve true unspanned stochastic volatility, of the type defined in Section 11.2.3. In particular, the discount bond reconstitution formulas in Proposition 12.3.1 depend on the full vector of state variables $z(t)$. However, in parameterizing the model we were careful to assign zero weight to z_{d+1}, the "volatility" factor, in the linear part of the quadratic form for the short rate (see (12.80)), ensuring that discount bonds have rather small (second order) dependence on it. Hence, we expect the model to exhibit some traits of stochastic volatility models. Lending some credibility to this observation, Piterbarg [2009a] (and, with more details, Piterbarg [2008]) analyzes the dynamics of volatility smiles in two-factor quadratic Gaussian models and concludes that these models lie somewhere between local volatility and true stochastic volatility models (which we introduce in Chapter 13 below).

12.3.7 Numerical Methods

We round out our discussion of quadratic Gaussian models with a quick review of numerical methods available for derivatives pricing. The discussion is mercifully brief because the state variables in a quadratic Gaussian model follow the same process as the state variables in a linear Gaussian model, making the material of Section 12.1.8 directly applicable. This fortunate circumstance is, in fact, one of the key attractions of the quadratic Gaussian models, as we mentioned earlier. For instance, PDE methods for the quadratic Gaussian model carry over unchanged from the linear Gaussian case, as the state variables in both classes of models follow essentially identical processes. We refer to Section 12.1.9 for details.

As for Monte Carlo simulation, we can reuse results from Section 12.1.8, and emphasize that the state vector can be simulated at low cost and bias-free over a period of time of any length without adding any intermediate dates. As a result, the performance of the Monte Carlo method for the quadratic Gaussian model is on par with the linear Gaussian model, and far ahead of any alternative multi-factor model with volatility smile, such as the Libor market model (Chapter 14) or even the multi-factor quasi-Gaussian model (Chapter 13).

12.A Appendix: Quadratic Forms of Gaussian Vectors

First, we prove Proposition 12.3.5. We have,

$$\Psi(u, Q; m, \nu) = \frac{1}{(2\pi)^{K/2}\sqrt{\det(\nu)}} \times \int \exp\left(z^\top Q z + u^\top z\right) \exp\left(-\frac{1}{2}(z-m)^\top \nu^{-1}(z-m)\right) dz.$$

We have,

$$z^\top Q z + u^\top z = \left((z-m)^\top Q(z-m) + 2m^\top Q z - m^\top Q m + u^\top z\right),$$

so the integrand is equal to

$$\exp\left(-\frac{1}{2}(z-m)^\top (\nu^{-1} - 2Q)(z-m)\right) \exp\left(2m^\top Q z - m^\top Q m + u^\top z\right).$$

Define

$$\nu_Q \triangleq \left(\nu^{-1} - 2Q\right)^{-1} = \nu\left(I - 2Q\nu\right)^{-1}.$$

Let Q^{ν_Q} be a measure under which Z has mean m and variance ν_Q, and E^{ν_Q} the corresponding expected value operator. Then

$$\Psi(u, Q; m, \nu) = \exp\left(-m^\top Q m\right) \frac{\sqrt{\det(\nu_Q)}}{\sqrt{\det(\nu)}} \mathrm{E}^{\nu_Q}\left(\exp\left(2m^\top Q Z + u^\top Z\right)\right).$$

By the standard results for exponents of Gaussian linear forms (see e.g. Kotz et al. [2000]),

$$\begin{aligned}&\mathrm{E}^{\nu_Q}(\exp((2m^\top Q + u^\top)Z)) \\ &= \exp\left((2m^\top Q + u^\top)m + \frac{1}{2}(2m^\top Q + u^\top)\nu(I - 2Q\nu)^{-1}(2Qm + u)\right).\end{aligned}$$

Thus we get

$$\begin{aligned}\ln \mathrm{E}(\exp(Z^\top Q Z + u^\top Z)) = \frac{1}{2}(2m^\top Q + u^\top)\nu(I - 2Q\nu)^{-1}(2Qm + u) \\ + m^\top Q m + u^\top m + \frac{1}{2}\ln\det(\nu^{-1}\nu_Q).\end{aligned}$$

The proposition has been proven.

Once the moment-generating function is available, other characteristics of the distribution follow. For example, we can easily calculate the mean and the variance of a quadratic form of a Gaussian vector.

Corollary 12.A.1. *Let Z be a K-dimensional Gaussian vector with mean m and variance ν. Let Q be a symmetric $K \times K$ matrix and u a K-dimensional vector. Then*

$$\begin{aligned}
\mathrm{E}\left(Z^\top Q Z + u^\top Z\right) &= \left(m^\top Q m + u^\top m\right) + \mathrm{tr}\left(Q\nu\right), \\
\mathrm{Var}\left(Z^\top Q Z + u^\top Z\right) &= \left(2m^\top Q + u^\top\right)\nu\left(2Qm + u\right) + 2\mathrm{tr}\left(Q\nu Q\nu\right).
\end{aligned}$$

Proof. Clearly,

$$\begin{aligned}
\mathrm{E}\left(Z^\top Q Z + u^\top Z\right) &= \left.\frac{d}{d\epsilon}\Psi\left(\epsilon u, \epsilon Q; m, \nu\right)\right|_{\epsilon=0}, \\
\mathrm{Var}\left(Z^\top Q Z + u^\top Z\right) &= \left.\frac{d^2}{d\epsilon^2}\Psi\left(\epsilon u, \epsilon Q; m, \nu\right)\right|_{\epsilon=0}.
\end{aligned}$$

Recall

$$\begin{aligned}
\Psi\left(\epsilon u, \epsilon Q; m, \nu\right) = \frac{\epsilon^2}{2}&\left(2m^\top Q + u^\top\right)\left(\nu^{-1} - 2\epsilon Q\right)^{-1}\left(2Qm + u\right) \\
&+ \epsilon\left(m^\top Q m + u^\top m\right) - \frac{1}{2}\ln\det\left(I - 2\epsilon Q\nu\right).
\end{aligned}$$

By Jacobi's formula,

$$\frac{d}{d\epsilon}\det\left(I - 2\epsilon Q\nu\right) = -\det\left(I - 2\epsilon Q\nu\right)\mathrm{tr}\left(2\left(I - 2\epsilon Q\nu\right)^{-1} Q\nu\right),$$

so

$$\frac{d}{d\epsilon}\ln\det\left(I - 2\epsilon Q\nu\right) = -\mathrm{tr}\left(2\left(I - 2\epsilon Q\nu\right)^{-1} Q\nu\right).$$

Then

$$\begin{aligned}
\frac{d}{d\epsilon}\Psi\left(\epsilon u, \epsilon Q; m, \nu\right) = \epsilon&\left(2m^\top Q + u^\top\right)\left(\nu^{-1} - 2Q\epsilon\right)^{-1}\left(2Qm + u\right) \\
&+ \frac{1}{2}\epsilon^2 \times (\ldots) \\
&+ \left(m^\top Q m + u^\top m\right) + \mathrm{tr}\left(\left(I - 2\epsilon Q\nu\right)^{-1} Q\nu\right),
\end{aligned}$$

so that

$$\left.\frac{d}{d\epsilon}\Psi\left(\epsilon u, \epsilon Q; m, \nu\right)\right|_{\epsilon=0} = \left(m^\top Q m + u^\top m\right) + \mathrm{tr}\left(Q\nu\right).$$

Furthermore,

$$\begin{aligned}
\frac{d}{d\epsilon}\mathrm{tr}\left(\left(I - 2\epsilon Q\nu\right)^{-1} Q\nu\right) &= \mathrm{tr}\left(\frac{d}{d\epsilon}\left(\left(I - 2\epsilon Q\nu\right)^{-1} Q\nu\right)\right) \\
&= 2\mathrm{tr}\left(\left(I - 2\epsilon Q\nu\right)^{-1} Q\nu\left(I - 2\epsilon Q\nu\right)^{-1} Q\nu\right),
\end{aligned}$$

So,

$$\begin{aligned}\frac{d^2}{d\epsilon^2}\Psi\left(\epsilon u, \epsilon Q; m, \nu\right) = &\left(2m^\top Q + u^\top\right)\left(\nu^{-1} - 2\epsilon Q\right)^{-1}\left(2Qm + u\right)\\ &+ \epsilon \times (\ldots)\\ &+ 2\mathrm{tr}\left(\left(I - 2\epsilon Q\nu\right)^{-1} Q\nu \left(I - 2\epsilon Q\nu\right)^{-1} Q\nu\right),\end{aligned}$$

thus

$$\frac{d^2}{d\epsilon^2}\Psi\left(\epsilon u, \epsilon Q; m, \nu\right)\Big|_{\epsilon=0} = \left(2m^\top Q + u^\top\right)\nu\left(2Qm + u\right) + 2\mathrm{tr}\left(Q\nu Q\nu\right).$$

□

Interestingly, we can obtain covariances or, indeed, any cross-moments of multiple quadratic forms of the same vector Z using the same idea as in the previous corollary.

Corollary 12.A.2. *Let Z be a K-dimensional Gaussian vector with mean m and variance ν. Let Q_1, Q_2 be symmetric $K \times K$ matrices and u_1, u_2 be K-dimensional vectors. Then*

$$\begin{aligned}&\mathrm{E}\left(\left(Z^\top Q_1 Z + u_1^\top Z\right)^n \left(Z^\top Q_2 Z + u_2^\top Z\right)^m\right)\\ &\qquad = \frac{\partial^{n+m}}{\partial\epsilon_1^n \partial\epsilon_2^m}\exp\left(\Psi\left(\epsilon_1 u_1 + \epsilon_2 u_2, \epsilon_1 Q_1 + \epsilon_2 Q_2; m, \nu\right)\right)\Big|_{\epsilon_1=\epsilon_2=0}.\end{aligned}$$

In particular,

$$\begin{aligned}&\mathrm{Cov}\left(Z^\top Q_1 Z + u_1^\top Z, Z^\top Q_2 Z + u_2^\top Z\right)\\ &\qquad = \left(2m^\top Q_1 + u_1^\top\right)\nu\left(2Q_2 m + u_2\right) + 2\mathrm{tr}\left(Q_1\nu Q_2\nu\right).\end{aligned}$$

The next corollary helps with calculating moments of the state vector under different forward measures in quadratic Gaussian models, see (12.89).

Corollary 12.A.3. *Let Z be a K-dimensional Gaussian vector with mean m and variance ν. Let Q be a symmetric $K \times K$ matrix and u a K-dimensional vector. Denote*

$$\begin{aligned}\widehat{m} &= \frac{\mathrm{E}\left(Z\exp\left(-\left(Z^\top QZ + u^\top Z\right)\right)\right)}{\mathrm{E}\left(\exp\left(-\left(Z^\top QZ + u^\top Z\right)\right)\right)},\\ \widehat{\nu} &= \frac{\mathrm{E}\left(ZZ^\top\exp\left(-\left(Z^\top QZ + u^\top Z\right)\right)\right)}{\mathrm{E}\left(\exp\left(-\left(Z^\top QZ + u^\top Z\right)\right)\right)} - \widehat{m}\widehat{m}^\top.\end{aligned}$$

Then

$$\widehat{m} = m - \nu\left(I + 2Q\nu\right)^{-1}\left(2Qm + u\right), \tag{12.97}$$

$$\widehat{\nu} = \nu\left(I + 2Q\nu\right)^{-1}. \tag{12.98}$$

Proof. First, we note that

$$\widehat{m} = -\frac{d}{du}\Psi\left(-u, -Q; m, \nu\right),$$
$$\widehat{\nu} = -\frac{d}{dQ}\Psi\left(-u, -Q; m, \nu\right) - \widehat{m}\widehat{m}^{\top}.$$

From Proposition 12.3.5,

$$\Psi\left(-u, -Q; m, \nu\right) = \frac{1}{2}\left(2m^{\top}Q + u^{\top}\right)\left(\nu^{-1} + 2Q\right)^{-1}\left(2Qm + u\right)$$
$$- m^{\top}Qm - u^{\top}m - \frac{1}{2}\ln\det\left(I + 2Q\nu\right).$$

Then

$$\frac{d}{du}\left(\frac{1}{2}\left(2m^{\top}Q + u^{\top}\right)\nu\left(I + 2Q\nu\right)^{-1}\left(2Qm + u\right) - m^{\top}Qm - u^{\top}m\right)$$
$$= \nu\left(I + 2Q\nu\right)^{-1}\left(2Qm + u\right) - m,$$

and (12.97) follows.

The proof of (12.98) proceeds along similar lines, using the fact that for any matrix A

$$\frac{d}{dA}\det A = \left(\det A\right)\left(A^{-1}\right)^{\top}$$

and, in particular,

$$\frac{d}{dQ}\det\left(I + 2Q\nu\right) = 2\det\left(I + 2Q\nu\right)\nu\left(I + 2Q\nu\right)^{-1}.$$

□

13

The Quasi-Gaussian Model with Local and Stochastic Volatility

In this chapter we consider extensions to one- and multi-factor Gaussian short rate models (Chapters 10, 11 and 12) with local and stochastic volatility. The extensions come at additional computational cost, as extra state variables are required to preserve the Markovian structure of the model. Following the pioneering work of Jamshidian [1991b], we use the term *quasi-Gaussian*[1] for the models in this chapter; their development for practical applications was undertaken in Andreasen [2001], Andersen and Andreasen [2002] and Andreasen [2005], building on early work by Jamshidian [1991b], Babbs [1990], Cheyette [1991] and Ritchken and Sankarasubramanian [1995]. Low-dimensional versions of quasi-Gaussian models are, in our opinion, among the best — if not *the* best — low-factor short rate models, as they combine flexibility of volatility smile specification, relative ease of calibration, and efficient numerical implementation. Higher-dimensional quasi-Gaussian models, while not yet mainstream, provide an alluring alternative to the better-established Libor market models (see Chapter 14).

We start this chapter by developing a one-factor quasi-Gaussian model with a local volatility function. The problems of volatility and mean reversion calibration are given considerable attention, and are followed by a discussion of various numerical methods used for model implementation. A straightforward extension to stochastic volatility is presented next, followed by development of multi-factor quasi-Gaussian models.

13.1 One-Factor Quasi-Gaussian Model

13.1.1 Definition

Recall that any HJM model is defined by a volatility structure of instantaneous forward rates. In particular, for any "reasonable" random function $\sigma_f(t,T) = \sigma_f(t,T,\omega)$, the following SDE defines a valid HJM model,

[1] Also known as *pseudo-Gaussian* or *Cheyette* models.

$$df(t,T) = \sigma_f(t,T)\left(\left(\int_t^T \sigma_f(t,u)\,du\right)dt + dW(t)\right), \quad 0 \le t \le T < \infty. \tag{13.1}$$

Here $W(t)$ is a one-dimensional Brownian motion in the risk-neutral measure, and $\{f(t,T)\}_{T\ge t}$ is a collection of instantaneous forward rates.

It is shown in Section 4.5.2 that a one-factor Markovian Gaussian model is obtained by imposing a separability condition on the deterministic volatility structure of instantaneous forward rates, see (4.44). A general class of one-factor *quasi-Gaussian* (qG)[2] models is obtained by retaining the separability condition, but relaxing the deterministic requirement in a specific way. In particular, the component of the volatility structure that is a function of calendar time (the function g), is now allowed to be stochastic:

$$\sigma_f(t,T,\omega) = g(t,\omega)\,h(T). \tag{13.2}$$

In line with the notations of Section 10.1.2.2, we define

$$\begin{aligned} \varkappa(t) &= -\frac{h'(t)}{h(t)}, \\ G(t,T) &= \frac{\int_t^T h(s)\,ds}{h(t)}, \\ \sigma_r(t,\omega) &= \sigma_f(t,t,\omega) = g(t,\omega)\,h(t). \end{aligned} \tag{13.3}$$

The proof of Proposition 10.1.7 carries through unchanged even for stochastic $g(t,\omega)$, and we obtain the following result.

Proposition 13.1.1. *Consider a general one-factor qG model, i.e. the HJM model (13.1) with the separable volatility condition (13.2). Define stochastic processes $x(t)$ and $y(t)$ by*

$$\begin{aligned} dx(t) &= (y(t) - \varkappa(t)x(t))\,dt + \sigma_r(t,\omega)\,dW(t), \\ dy(t) &= \left(\sigma_r(t,\omega)^2 - 2\varkappa(t)y(t)\right)dt, \\ x(0) &= y(0) = 0. \end{aligned} \tag{13.4}$$

In the general qG model all zero-coupon discount bonds are deterministic functions of the processes $x(t)$ and $y(t)$,

$$P(t,T) = P(t,T,x(t),y(t)),$$

where

$$P(t,T,x,y) = \frac{P(0,T)}{P(0,t)}\exp\left(-G(t,T)\,x - \frac{1}{2}G(t,T)^2\,y\right), \tag{13.5}$$

[2] We use a small q in the abbreviation qG to avoid notational conflict with the quadratic Gaussian (QG) models of Chapter 12.

the instantaneous forward rates are given by

$$f(t,T) = f(0,T) + \frac{h(T)}{h(t)}\left(x(t) + G(t,T)\,y(t)\right), \tag{13.6}$$

and the short rate is

$$r(t) = f(t,t) = f(0,t) + x(t). \tag{13.7}$$

The proposition demonstrates that the evolution of the whole interest rate curve, as parameterized by either forward rates or discount bonds, in the model can be reduced to the evolution of just two *state variables* $x(t)$ and $y(t)$, with dynamics given by (13.4). Unlike many of the models in Chapter 11, the qG model has a closed-form bond reconstitution formula for arbitrary choices of $g(t,\omega)$; this tractability comes at the cost at requiring *two* state variables (x and y), even though the Brownian motion $W(t)$ is only one-dimensional. Observe that in general, the function $y(t)$ is *not* deterministic, except in the case of pure Gaussian dynamics, i.e. when $\sigma_r(t,\omega)$ is a deterministic function of t. However, even when it is not deterministic, $y(t)$ does not have the diffusion term; we call such processes *locally deterministic.*

The roles of the two state variables $x(t)$ and $y(t)$ in the qG model are rather different. The variable $x(t)$ constitutes the main yield curve driver, as evidenced in (13.7), whereas $y(t)$ is an auxiliary "convexity" variable required to uphold the no-arbitrage condition; in general, it is convenient to think of the model as having "one and a half" factors.

13.1.2 Local Volatility

A one-factor qG model with *local volatility* is obtained by requiring $g(\cdot)$ to be a deterministic, time-dependent function of the state variables,

$$g(t) = g\left(t, x(t), y(t)\right). \tag{13.8}$$

Then, the short rate volatility $\sigma_r(\cdot)$ is also a function of the state variables,

$$\sigma_r(t) = \sigma_r\left(t, x(t), y(t)\right) = g\left(t, x(t), y(t)\right) h(t), \tag{13.9}$$

and the dynamics of the state variables in the local volatility qG model are given by

$$dx(t) = \left(y(t) - \varkappa(t)x(t)\right)\,dt + \sigma_r\left(t, x(t), y(t)\right)\,dW(t), \tag{13.10}$$

$$dy(t) = \left(\sigma_r\left(t, x(t), y(t)\right)^2 - 2\varkappa(t)y(t)\right)\,dt. \tag{13.11}$$

Clearly, (13.10)–(13.11) define a two-dimensional Markovian process. As all zero-coupon discount bonds are functions of these two state variables by Proposition 13.1.1, the local volatility qG model is Markovian in two state variables.

For future use, we denote

$$\sigma_r^0(t) \triangleq \sigma_r(t,0,0). \tag{13.12}$$

If $\sigma_r(t,x,y)$ is independent of x, y, the model reduces to a purely Gaussian model with the deterministic short rate volatility $\sigma_r^0(t)$.

13.1.3 Swap Rate Dynamics

For the purpose of European swaption pricing in the qG model, we shall need to establish swap rate dynamics in an annuity measure. For this purpose, let us fix a tenor structure

$$0 < T_0 < T_1 < T_2 < \ldots < T_N,$$

with

$$\tau_n = T_{n+1} - T_n.$$

Consider a forward swap rate $S(t)$ with the first fixing T_0 and the last payment T_N (see Section 4.1.3), i.e.

$$S(t) \triangleq S_{0,N}(t) = \frac{P(t,T_0) - P(t,T_N)}{A(t)}, \tag{13.13}$$

$$A(t) \triangleq A_{0,N}(t) = \sum_{n=0}^{N-1} \tau_n P(t,T_{n+1}). \tag{13.14}$$

It follows from Proposition 13.1.1 that all zero-coupon bonds $P(t,\cdot)$ are deterministic functions of $x(t)$ and $y(t)$, and hence so is the forward swap rate; accordingly we define

$$S(t,x,y) \triangleq \frac{P(t,T_0,x,y) - P(t,T_N,x,y)}{\sum_{n=0}^{N-1} \tau_n P(t,T_{n+1},x,y)}. \tag{13.15}$$

The following proposition, a simple extension of the results from Section 10.1.3.2, determines the dynamics of the swap rate under its corresponding annuity measure, i.e. the measure Q^A for which $A(t)$ is a numeraire.

Proposition 13.1.2. *We have*

$$dS(t) = \left(\frac{\partial S}{\partial x}(t,x(t),y(t))\right) \sigma_r(t,x(t),y(t))\, dW^A(t), \tag{13.16}$$

where $W^A(t)$ is a Brownian motion in measure Q^A. Here

$$\frac{\partial S}{\partial x}(t,x,y) = -\frac{1}{A(t,x,y)}\left(P(t,T_0,x,y)\,G(t,T_0) - P(t,T_N,x,y)\,G(t,T_N)\right)$$
$$+ \frac{S(t,x,y)}{A(t,x,y)} \sum_{n=0}^{N-1} \tau_n P(t,T_{n+1},x,y)\,G(t,T_{n+1}). \tag{13.17}$$

Proof. By definition

$$S(t) = S\left(t, x(t), y(t)\right).$$

The statement of the proposition follows by applying Ito's lemma and dropping dt terms from the expression, as justified by the fact that the swap rate is a martingale in the annuity measure, per Lemma 4.2.4. □

13.1.4 Approximate Local Volatility Dynamics for Swap Rate

The SDE (13.16) shows that a swap rate follows a local volatility process where the volatility is a function of the short rate state x and the auxiliary variable y. Since the model is essentially one-factor, it is reasonable to assume that there is a strong linkage between a swap rate and the state of the short rate. Hence, it seems plausible that the dynamics for a swap rate could be written with a diffusion term that is just a function of the swap rate itself. Such a simplification would be convenient in many applications, as methods from Chapter 7 could be called upon to solve the resulting SDE or to price options. The following proposition proven by the methods of *Markovian projection* (see Appendix A) makes matters precise.

Lemma 13.1.3. *The values of all European options on the swap rate $S(t)$ in the model (13.16) are identical to values computed in a vanilla model with time-dependent local volatility function:*

$$dS(t) = \varphi\left(t, S(t)\right) dW^A(t), \tag{13.18}$$

where

$$\varphi\left(t, s\right)^2 = \mathrm{E}^A\left(\left(\frac{\partial S}{\partial x}\left(t, x(t), y(t)\right)\sigma_r\left(t, x(t), y(t)\right)\right)^2 \middle| S(t) = s\right). \tag{13.19}$$

While evaluating conditional expectations such as (13.19) is often rather difficult, here we are aided considerably by the essentially one-dimensional structure of the problem. If we assume that $y(t)$ is well approximated by a deterministic function $\overline{y}(t)$ — an approximation we shall use repeatedly in this chapter — then $S(t)$ would just be a deterministic function of $x(t)$ and time t. The opposite would also be true, i.e. $x(t)$ would be a deterministic function of $S(t)$, $x(t) = X(t, S(t))$. If this function were available, the evaluation of the conditional expected value in (13.19) would boil down to evaluating $\frac{\partial S}{\partial x}(t, x(t), y(t))\sigma_r(t, x(t), y(t))$ at $x(t) = X(t, s)$ and $y(t) = \overline{y}(t)$,

$$\varphi\left(t, s\right) \approx \frac{\partial S}{\partial x}\left(t, X\left(t, s\right), \overline{y}(t)\right)\sigma_r\left(t, X\left(t, s\right), \overline{y}(t)\right),$$

where $\partial S/\partial x$ is given by (13.17).

Before presenting various methods for approximating $\overline{y}$ and the function X, let us emphasize again that once the volatility function $\varphi(t, s)$ is determined, swaption values can be computed from (13.18) by methods developed for local volatility vanilla models in Chapter 7.

13.1.4.1 Simple Approximation

A simple approximation for the function φ is obtained by slightly extending the idea from Section 10.1.3.2. Setting

$$\overline{y}(t) = 0,$$

and applying a linear approximation

$$\begin{aligned} S(t,x,0) &\approx S(t,0,0) + \frac{\partial S}{\partial x}(t,0,0)\,x, \\ \frac{\partial S}{\partial x}(t,x,0) &\approx \frac{\partial S}{\partial x}(t,0,0), \end{aligned} \tag{13.20}$$

we obtain

$$x \approx \frac{S(t,x,0) - S(t,0,0)}{\partial S(t,0,0)/\partial x}. \tag{13.21}$$

Hence we arrive at the approximation

$$\varphi(t,s) \approx \frac{\partial S}{\partial x}(t,0,0)\,\sigma_r\left(t, \frac{s - S(t,0,0)}{\partial S(t,0,0)/\partial x}, 0\right). \tag{13.22}$$

13.1.4.2 Advanced Approximation

The approximation (13.22) is quite accurate provided that volatility is low or moderate. For an approximation with a greater range of applicability, we can consider improving (13.20) by using a higher-order expansion and by taking greater care in selecting the expansion point. Starting with the latter, we note that the conditional expectation in (13.19) is taken in the annuity measure, suggesting that the expected values of $x(t)$ and $y(t)$ under Q^A provide a good expansion point.

Proposition 13.1.4. *Let*

$$\overline{y}(t) = \mathrm{E}^A(y(t)).$$

Then, approximately,

$$\overline{y}(t) \approx h(t)^2 \int_0^t \sigma_r^0(s)^2 h(s)^{-2}\,ds, \quad t \in [0, T_0], \tag{13.23}$$

where, per (13.3),

$$h(t) = \exp\left(-\int_0^t \varkappa(u)\,du\right).$$

Proof. Recall the dynamics of $y(t)$ in the quasi-Gaussian model (Proposition 13.1.1),

$$dy(t) = \left(\sigma_r\left(t, x(t), y(t)\right)^2 - 2\varkappa(t)y(t)\right)\, dt, \quad y(0) = 0.$$

Taking expected values, we obtain

$$d\mathrm{E}^A\left(y(t)\right) = \left(\mathrm{E}^A\left(\sigma_r\left(t, x(t), y(t)\right)^2\right) - 2\varkappa(t)\mathrm{E}^A\left(y(t)\right)\right)\, dt, \tag{13.24}$$

subject to $\mathrm{E}^A(y(0)) = 0$. Approximating, for the purposes of this calculation,

$$\mathrm{E}^A\left(\sigma_r\left(t, x(t), y(t)\right)^2\right) \approx \sigma_r^0(t)^2,$$

the equation (13.24) yields

$$d\mathrm{E}^A\left(y(t)\right) \approx \left(\sigma_r^0(t)^2 - 2\varkappa(t)\mathrm{E}^A\left(y(t)\right)\right)\, dt, \quad \mathrm{E}^A\left(y(0)\right) = 0.$$

Solving this ODE leads to (13.23). □

As above, let $X(t,s)$ be the function inverse, in x, to $S(t,x,\overline{y}(t))$, i.e.

$$S\left(t, X\left(t,s\right), \overline{y}(t)\right) \equiv s, \tag{13.25}$$

and let $x_0(t)$ be given as the solution of

$$S\left(t, x_0(t), \overline{y}(t)\right) = S(0), \tag{13.26}$$

where $S(0)$ is the forward swap rate at time 0.

Remark 13.1.5. The function $S(t,x,\overline{y}(t))$ is known in closed form from (13.15) and is smooth and monotonic in x. As such, (13.26) can be solved for $x_0(t)$ in just a few iterations of the Newton algorithm. A good starting point for the search is $x = 0$.

It turns out that $x_0(t)$ in (13.26) is a good expansion point itself, and can also be used to calculate an even better one:

Lemma 13.1.6. *The function $x_0(t)$ is a first-order approximation to $\mathrm{E}^A(x(t))$. An approximation to second order is given by*

$$\overline{x}(t) = x_0(t) + \frac{\partial^2 X}{\partial s^2}\left(t, S(0)\right)\mathrm{Var}^A\left(S(t)\right), \quad t \in [0, T_0].$$

Proof. Expanding $X(t,s)$ around $s = S(0)$ to first order, we obtain

$$x(t) - x_0(t) = \left.\frac{\partial X}{\partial s}\left(t,s\right)\right|_{s=S(0)}\left(S(t) - S(0)\right) + O\left(\left(S(t) - S(0)\right)^2\right). \tag{13.27}$$

Taking expected values and using the fact that $S(t)$ is a martingale in measure Q^A, we get

$$\mathrm{E}^A\left(x(t)\right) - x_0(t) = O\left(\mathrm{E}^A\left(\left(S(t) - S(0)\right)^2\right)\right),$$

i.e. $x_0(t)$ is an approximation to $\mathrm{E}^A(x(t))$ to first order. The approximation $\overline{x}(t)$ is obtained by expanding (13.27) to second order,

$$\begin{aligned} x(t) - x_0(t) = & \frac{\partial X}{\partial s}\left(t, S(0)\right)\left(S(t) - S(0)\right) \\ & + \frac{1}{2}\frac{\partial^2 X}{\partial s^2}\left(t, S(0)\right)\left(S(t) - S(0)\right)^2 + O\left(\left(S(t) - S(0)\right)^3\right). \end{aligned}$$

Then

$$\mathrm{E}^A\left(x(t)\right) = x_0(t) + \frac{\partial^2 X}{\partial s^2}\left(t, S(0)\right)\mathrm{Var}^A\left(S(t)\right) + O\left(\mathrm{E}^A\left(\left(S(t) - S(0)\right)^3\right)\right).$$

□

Remark 13.1.7. As high precision is not required when calculating the variance $\mathrm{Var}^A(S(t))$ in Lemma 13.1.6, it can be evaluated by considering a simple Gaussian approximation to the dynamics of $S(t)$, i.e. using $x(t) = 0$, $y(t) = 0$ in (13.16),

$$dS(t) \approx \frac{\partial S}{\partial x}(t, 0, 0)\sigma_r^0(t)\, dW^A(t),$$

which would yield

$$\mathrm{Var}^A\left(S(t)\right) \approx \int_0^t \left(\frac{\partial S}{\partial x}(s, 0, 0)\sigma_r^0(s)\right)^2 ds.$$

The second derivative $\partial^2 X(t,s)/\partial s^2$ can be computed by differentiating the implicit definition (13.25) twice.

Having now established an expansion point, let us proceed to determine an approximation to $X(t,s)$ with higher accuracy than the linear one in (13.21). Empirically, it can be observed that $S(t,x)$ is closely approximated as a quadratic function of x across a wide range of the argument x. This suggests a second-order expansion of (13.25) around $\overline{x}(t)$, and approximating $X(t,s)$ with $\xi = \xi(t,s)$, the solution of the following quadratic equation in ξ:

$$\begin{aligned} S\left(t, \overline{x}(t), \overline{y}(t)\right) + \frac{\partial S}{\partial x}\left(t, \overline{x}(t), \overline{y}(t)\right)\left(\xi - \overline{x}(t)\right) \\ + \frac{1}{2}\frac{\partial^2 S}{\partial x^2}\left(t, \overline{x}(t), \overline{y}(t)\right)\left(\xi - \overline{x}(t)\right)^2 = s. \end{aligned} \tag{13.28}$$

With $S(t, \overline{x}(t), \overline{y}(t))$, $\partial S(t, \overline{x}(t), \overline{y}(t))/\partial x$ and $\partial^2 S(t, \overline{x}(t), \overline{y}(t))/\partial x^2$ precomputed, the evaluation of $\xi(t,s)$ for any s is essentially instantaneous, and we obtain the following efficient approximation for $\varphi(t,s)$.

Proposition 13.1.8. *An approximation to $\varphi(t,s)$ in (13.19) is given by*

$$\varphi(t,s) \approx \frac{\partial S}{\partial x}(t,\xi(t,s),\overline{y}(t))\,\sigma_r(t,\xi(t,s),\overline{y}(t)),$$

where $\xi(t,s)$ is the solution to the quadratic equation (13.28), with $\overline{x}(t)$ given by Lemma 13.1.6 and $\overline{y}(t)$ given by Proposition 13.1.4.

13.1.5 Linear Local Volatility

As demonstrated above, the local volatility qG model (13.9) has the flexibility to generate essentially arbitrary local volatility dynamics for swap rates. Using the results of Lemma 13.1.3 and Proposition 13.1.8, the function $\sigma_r(t,x,y)$ could therefore, in principle, be calibrated non-parametrically (see Dupire [1994] and the discussion in Section 7.1.3) to the implied volatilities of a collection of swaptions across all strikes. However, as explained in Section 7.1.3, we recommend the volatility function $\sigma_r(t,x,y)$ to be chosen from a parametric family of monotone, downward sloping functions of state variable(s). While power functions that give rise to models with CEV-type[3] dynamics could be used, as explained in Remark 7.2.14 linear functions provide a less-involved alternative capable of producing essentially the same range of volatility smiles as CEV models.

With the above in mind, let us consider the following short rate local volatility function

$$\sigma_r(t,x,y) = \lambda_r(t)\,(\alpha_r(t) + b_r(t)x)\,. \tag{13.29}$$

The scale function $\alpha_r(t)$ is redundant (as it can be absorbed in $\lambda_r(t)$) and may be set exogenously; Section 13.1.6 discusses a convenient choice. The functions $\lambda_r(t)$ (volatility) and $b_r(t)$ (skew) are calibrated to the market. Under (13.29), the local volatility of the swap rate S is given, approximately, by

$$\varphi(t,s) \approx \lambda_r(t)\frac{\partial S}{\partial x}(t,\xi(t,s),\overline{y}(t))\,(\alpha_r(t) + b_r(t)\xi(t,s))\,, \tag{13.30}$$

with $\xi(t,s)$ as in Proposition 13.1.8. As the local volatility of the short rate is linear in x, it seems reasonable that the local volatility of the swap rate would be well approximated by a linear function as well. To exploit this, let φ be given as in (13.30) and notice that

$$\begin{aligned}
\varphi(t,S(0)) &\approx \lambda_r(t)\frac{\partial S}{\partial x}(t,\xi(t,S(0)),\overline{y}(t))(\alpha_r(t) + b_r(t)\xi(t,S(0))),\\
\frac{\partial\varphi}{\partial s}(t,S(0)) &\approx \lambda_r(t)\frac{\partial S}{\partial x}(t,\xi(t,S(0)),\overline{y}(t))\frac{\partial\xi}{\partial s}(t,S(0))\\
&\quad\times\left[\frac{\frac{\partial^2 S}{\partial x^2}(t,\xi(t,S(0)),\overline{y}(t))}{\frac{\partial S}{\partial x}(t,\xi(t,S(0)),\overline{y}(t))}(\alpha_r(t) + b_r(t)\xi(t,S(0))) + b_r(t)\right].
\end{aligned}$$

[3] Constant Elasticity of Variance, see Section 7.2.

Clearly

$$\xi\left(t, S(0)\right) \approx \overline{x}(t),$$

and, with $\xi(t,\cdot)$ being an approximate inverse to $S(t,\cdot,\overline{y}(t))$,

$$\frac{\partial \xi}{\partial s}\left(t, S(0)\right) \approx \frac{1}{\frac{\partial S}{\partial x}\left(t, \overline{x}(t), \overline{y}(t)\right)}.$$

It follows that

$$\varphi\left(t, S(0)\right) \approx \lambda_r(t) \frac{\partial S}{\partial x}\left(t, \overline{x}(t), \overline{y}(t)\right)\left(\alpha_r(t) + b_r(t)\overline{x}(t)\right), \tag{13.31}$$

$$\frac{\partial \varphi}{\partial s}\left(t, S(0)\right) \approx \lambda_r(t)\left[\frac{\frac{\partial^2 S}{\partial x^2}\left(t, \overline{x}(t), \overline{y}(t)\right)}{\frac{\partial S}{\partial x}\left(t, \overline{x}(t), \overline{y}(t)\right)}\left(\alpha_r(t) + b_r(t)\overline{x}(t)\right) + b_r(t)\right]. \tag{13.32}$$

The following corollary to Proposition 13.1.8 holds.

Corollary 13.1.9. *Under the assumption of linear local short rate volatility (13.29) for the quasi-Gaussian model (13.10), the dynamics of the swap rate $S(t)$ are approximated by*

$$dS(t) \approx \lambda_S(t)\left(b_S(t)S(t) + \left(1 - b_S(t)\right)S(0)\right) dW^A(t), \tag{13.33}$$

where

$$\lambda_S(t) = \lambda_r(t)\frac{1}{S(0)}\frac{\partial S}{\partial x}\left(t, \overline{x}(t), \overline{y}(t)\right)\left(\alpha_r(t) + b_r(t)\overline{x}(t)\right), \tag{13.34}$$

$$b_S(t) = \frac{S(0)}{\left(\alpha_r(t) + b_r(t)\overline{x}(t)\right)}\frac{b_r(t)}{\frac{\partial S}{\partial x}\left(t, \overline{x}(t), \overline{y}(t)\right)} + \frac{S(0)\frac{\partial^2 S}{\partial x^2}\left(t, \overline{x}(t), \overline{y}(t)\right)}{\left(\frac{\partial S}{\partial x}\left(t, \overline{x}(t), \overline{y}(t)\right)\right)^2}. \tag{13.35}$$

Proof. Under a linear approximation to the local volatility function of the swap rate we have, with φ defined in (13.30),

$$dS(t) \approx \left(\varphi\left(t, S(0)\right) + \frac{\partial \varphi}{\partial s}\left(t, S(0)\right)\left(S(t) - S(0)\right)\right) dW^A(t)$$

which, after rearranging the terms, yields

$$dS(t) \approx \frac{\varphi\left(t, S(0)\right)}{S(0)}\left(S(0) + S(0)\frac{\partial \varphi\left(t, S(0)\right)/\partial s}{\varphi\left(t, S(0)\right)}\left(S(t) - S(0)\right)\right) dW^A(t).$$

Defining

$$\lambda_S(t) \triangleq \frac{\varphi\left(t, S(0)\right)}{S(0)}, \quad b_S(t) \triangleq S(0)\frac{\partial \varphi\left(t, S(0)\right)/\partial s}{\varphi\left(t, S(0)\right)},$$

the result follows from (13.31)–(13.32). □

We recognize (13.33) as a displaced log-normal SDE with time-dependent volatility $\lambda_S(t)$ and skew $b_S(t)$. Using averaging techniques from Section 7.6.2, we can convert it into a displaced log-normal SDE with time-constant parameters $\overline{\lambda}_S$ and $\overline{b}_S$, see Proposition 7.2.12. For convenience, we list the resulting swaption pricing formula below.

Proposition 13.1.10. *Consider a payer swaption with strike (i.e., coupon) c and expiry T_0 on the swap rate $S(t)$ defined in (13.13). In the quasi-Gaussian model (13.10) with linear short rate volatility (13.29), the swaption price can be approximated by the displaced log-normal option formula*

$$\begin{aligned} V_{\text{swaption}}(0) \approx A(0) \big[&\left(S(0) + S(0)\left(1-\overline{b}_S\right)/\overline{b}_S\right) \Phi(d_+) \\ &- \left(c + S(0)(1-\overline{b}_S)/\overline{b}_S\right) \Phi(d_-)\big], \end{aligned}$$

$$d_\pm = \frac{\ln\left(\frac{S(0)+S(0)\left(1-\overline{b}_S\right)/\overline{b}_S}{c+S(0)\left(1-\overline{b}_S\right)/\overline{b}_S}\right) \pm \frac{1}{2}\overline{b}_S^2\overline{\lambda}_S^2 T_0}{\overline{b}_S\overline{\lambda}_S\sqrt{T_0}},$$

where

$$\overline{\lambda}_S = \left(\frac{1}{T_0}\int_0^{T_0} \lambda_S(t)^2\,dt\right)^{1/2}, \tag{13.36}$$

$$\overline{b}_S = \int_0^{T_0} b_S(t) w_S(t)\,dt, \tag{13.37}$$

$$w_S(t) = \frac{\lambda_S(t)^2 \int_0^t \lambda_S(s)^2\,ds}{\int_0^{T_0}\left(\lambda_S(u)^2 \int_0^u \lambda_S(s)^2\,ds\right)\,du},$$

with $\lambda_S(t)$, $b_S(t)$ given by Corollary 13.1.9.

13.1.6 Linear Local Volatility for a Swaption Strip

The quasi-Gaussian model (13.10) is typically calibrated to a swaption strip (recall the definition in Section 10.1.4) on a maturity grid $0 = T_0 < \ldots < T_N$, i.e. a collection of $N-1$ swaptions with the n-th swaption expiring on T_n, $n = 1, \ldots, N-1$. Let us suppose a swaption strip is specified, and that the n-th swaption has an underlying swap with $\mu(n)$ periods. For each n, we denote the corresponding swap rate and the annuity by

$$S_n(t) \triangleq S_{n,\mu(n)}(t), \quad A_n(t) \triangleq A_{n,\mu(n)}(t), \quad n = 1, \ldots, N-1.$$

For the model with the linear volatility specification (13.29), Proposition 13.1.10 will, after proper adjustment of maturities (see footnote 6 in Chapter 10), allow us to value all swaptions of all strikes in the strip by using the

displaced log-normal model with the effective parameters computed from the local volatility function of the model.

As mentioned earlier, in the specification (13.29) the function $\alpha_r(t)$ is only included for convenient scaling. To find a good value for it, let us consider the relationship between the local swap rate skews $b_{S_n}(t)$ and the local short rate skew $b_r(t)$ in (13.35). Ignoring small terms,

$$b_{S_n}(t) \approx \frac{S_n(0)}{\alpha_r(t)} \frac{b_r(t)}{\partial S_n\left(t, \overline{x}(t), \overline{y}(t)\right)/\partial x}. \tag{13.38}$$

It is often convenient to parametrize the model in such a way that the values of model parameters (here $b_r(t)$) are roughly of the same order of magnitude as the output parameters (here $b_{S_n}(t)$). This allows one to quickly check whether model parameters are sensible, and may well lead to better numerical properties of the calibration algorithm. Based on (13.38), we elect to set $\alpha_r(t)$ equal to $S_n(0)$ (of course we need to account for different values of n), and rescale $b_r(t)$ to incorporate the term $\partial S_n/\partial x$ (again, for different n). In summary, we specialize the definition (13.29) to be

$$\sigma_r\left(t, x, y\right) = \sum_{n=1}^{N-1} \lambda_n \left(S_n(0) + b_n D_n x\right) 1_{\{t \in (T_{n-1}, T_n]\}}, \tag{13.39}$$

where the *skew scalings* D_n are given by

$$D_n = \frac{\partial S_n}{\partial x}\left(t, 0, 0\right).$$

This definition recognizes the fact that the behavior of $\lambda_r(t)$ and $b_r(t)$ between the knot dates $\{T_n\}$ is of no consequence, wherefore these functions can be taken to be piecewise constant,

$$\begin{aligned}
\lambda_r(t) &= \sum_{n=1}^{N-1} \lambda_n 1_{\{t \in (T_{n-1}, T_n]\}}, \\
\alpha_r(t) &= \sum_{n=1}^{N-1} S_n(0) 1_{\{t \in (T_{n-1}, T_n]\}}, \\
b_r(t) &= \sum_{n=1}^{N-1} b_n D_n 1_{\{t \in (T_{n-1}, T_n]\}}.
\end{aligned}$$

13.1.7 Volatility Calibration

Let us assume that a swaption strip is given, and the model is parameterized with the local volatility of the form (13.39). The model parameters λ_n and b_n, $n = 1, \ldots, N-1$, need to be determined by calibrating the model to

market prices of swaptions in the swaption strip. For now, let us suppose that the mean reversion function $\varkappa(t)$ in (13.10) is specified externally — we will return to its calibration later in the chapter.

As each swap rate has an approximately displaced log-normal distribution in the model, the calibration objective could be expressed as the problem of matching *displaced log-normal parameters*, as given by the model for each swap rate, to a similar set of market-implied parameters. We already saw a similar approach in Section 9.3.4 and recall that performing the calibration in model parameter space, rather than in the space of calibration instrument *values*, avoids the expense of invoking option pricing formulas within the calibration loop.

Accordingly, assume that a collection of market parameters, i.e. displaced log-normal volatilities and skews $(\widehat{\lambda}_{S_n}, \widehat{b}_{S_n})$, $n = 1, \ldots, N-1$, is given. In practice, these parameters are obtained by fitting a series of constant-parameter displaced log-normal vanilla models to the observed swaption volatility smiles at all expiries $T_1, \ldots, T_{N-1}$. A best-fit model calibration across swaption strikes is possible (at the expense of using a numerical optimizer), or we may simply set the volatility $\widehat{\lambda}_{S_n}$ to match at-the-money swaption volatilities, while the skew $\widehat{b}_{S_n}$ is fit to the slope of the volatility smile at-the-money, or to the volatility at some relevant non-ATM strike.

It is clear from the swaption pricing formula in Proposition 13.1.10 that the value of a swaption with expiry T_n depends on model parameters (λ_i, b_i) for $i = 1, \ldots, n$ only. Hence, the qG model can be calibrated by a bootstrap method, similarly to the pure Gaussian case from Section 10.1.4. In the bootstrap method, the equations (13.36)–(13.37) are solved sequentially for $n = 1, \ldots, N-1$, with the two equations on step n used to determine two unknown model parameters (λ_n, b_n). For example, the following algorithm could be used.

1. Set (λ_n, b_n), $n = 1, \ldots, N-1$, to some reasonable starting values, e.g. set λ_n's to (properly scaled) volatilities obtained by calibrating a pure Gaussian model as in Section 10.1.4, and $b_n = \widehat{b}_{S_n}$.
2. Set $n = 1$.
3. For given n, (λ_i^*, b_i^*) are known for $i = 1, \ldots, n-1$. Note we use a star to denote *calibrated* values of the model parameters.
4. Calculate $\overline{x}(t)$ (Lemma 13.1.6) and $\overline{y}(t)$ (Proposition 13.1.4) for $t \in [0, T_n]$ using (λ_i^*, b_i^*), $i = 1, \ldots, n-1$, and the initial guess for (λ_n, b_n) from Step 1. Note that $\overline{x}(t), \overline{y}(t)$ implicitly depend on n as their definition depends on the swap rate/annuity measure used.
5. Calculate $\lambda_{S_n}(t)$, $b_{S_n}(t)$ for $t \in [0, T_{n-1}]$ from (λ_i^*, b_i^*), $i = 1, \ldots, n-1$, using (13.34)–(13.35).
6. Make another guess for (λ_n, b_n).
7. Update $\lambda_{S_n}(t)$, $b_{S_n}(t)$ for $t \in (T_{n-1}, T_n]$ from (λ_i^*, b_i^*), $i = 1, \ldots, n-1$, using (13.34)–(13.35).
8. Calculate $\overline{\lambda}_{S_n}$, $\overline{b}_{S_n}$ using Proposition 13.1.10.

9. Compare $(\overline{\lambda}_{S_n}, \overline{b}_{S_n})$ to $(\widehat{\lambda}_{S_n}, \widehat{b}_{S_n})$. If not equal within given tolerance, go to Step 6. Otherwise, proceed to Step 10.
10. As we have reached acceptable convergence between $(\overline{\lambda}_{S_n}, \overline{b}_{S_n})$ and $(\widehat{\lambda}_{S_n}, \widehat{b}_{S_n})$, set the calibrated model parameter values to the latest trial values, $(\lambda_n^*, b_n^*) = (\lambda_n, b_n)$.
11. Update $n \to n+1$. If $n \leq N-1$ go to Step 3. Otherwise, conclude.

It may appear more accurate to make Step 4 a part of the calibration loop (for each n), with $\overline{x}(t)$, $\overline{y}(t)$, $t \in (T_{n-1}, T_n]$, and dependent quantities such as $\partial S_n(t, \overline{x}(t), \overline{y}(t))/\partial x$, etc. computed using the current guess for (λ_n, b_n) in each loop iteration (together with already-calibrated values (λ_i^*, b_i^*), $i = 1, \ldots, n-1$). While such extensions are relatively straightforward, our experience shows that the quality of the calibration is rarely improved enough to justify the additional complexity.

It is possible to add regularity terms to Step 9 of the algorithm above, to ensure that the resulting model parameters do not behave irregularly as functions of t. See Section 13.1.8.2 below for an example of this.

13.1.8 Mean Reversion Calibration

With the volatility calibration out of the way, let us discuss what to do with the remaining model parameter, the mean reversion function $\varkappa(t)$ in (13.10). We start with a short review of the effects of mean reversion.

13.1.8.1 Effects of Mean Reversion

Let us first consider a few simple examples as a way of building intuition about the effects of mean reversion on market values of various securities. For simplicity, we use continuously compounded rates as convenient proxies for Libor and swap rates, and consider a pure Gaussian model with constant volatility and mean reversion,

$$\sigma_r(t) \equiv \sigma_r, \quad \varkappa(t) \equiv \varkappa.$$

A continuously compounded forward yield over a period $[T, M]$, observed at time t, is given by[4]

$$F(t, T, M) = -\frac{1}{M-T} \ln \frac{P(t, M)}{P(t, T)}.$$

[4]We normally use $y(t, T, M)$ for continuously compounded forward yield, see Section 4.1.1, and $F(t, T, M)$ for a futures rate, see Section 4.1.2, but for the remainder of this chapter only we allow ourselves a slight notational inconsistency to avoid the possibility of confusion between y, the state variable, and y, the forward yield.

According to Proposition 10.1.7, the forward yield can be expressed as a function of the state variables and parameters of the model,

$$F(t,T,M) = \frac{G(t,M) - G(t,T)}{M-T} x(t) + \frac{1}{2} \frac{G(t,M)^2 - G(t,T)^2}{M-T} y(t),$$
$$G(t,u) = \frac{1 - e^{-\varkappa(u-t)}}{\varkappa}.$$

Recall that $y(t)$ is deterministic in the Gaussian case, so the standard deviation of $F(T,T,M)$ is equal to

$$\begin{aligned}\mathrm{Stdev}\left(F(T,T,M)\right) &= \frac{G(T,M) - G(T,T)}{M-T} \left(\mathrm{Var}\left(x(T)\right)\right)^{1/2} \\ &= \frac{1 - e^{-\varkappa(M-T)}}{\varkappa(M-T)} \left(\mathrm{Var}\left(x(T)\right)\right)^{1/2}.\end{aligned}$$

For two maturities M_1, M_2, $T \le M_1 \le M_2$, we therefore observe that

$$\frac{\mathrm{Stdev}\left(F(T,T,M_2)\right)}{\mathrm{Stdev}\left(F(T,T,M_1)\right)} = \left(\frac{1 - e^{-\varkappa(M_2-T)}}{\varkappa(M_2-T)}\right) \Big/ \left(\frac{1 - e^{-\varkappa(M_1-T)}}{\varkappa(M_1-T)}\right),$$

i.e. the ratio of standard deviations of two forward yields with the same expiry T but different tenors is independent of the volatility parameter σ_r, and solely determined by the mean reversion parameter $\varkappa$. Since standard deviations of forward yields can loosely be thought of as proxies for implied swaption volatilities, we observe that the mean reversion parameter changes the relative levels of implied volatilities of swaptions with the same expiry but different underlying swap tenors. Specifically, for a fixed level of volatility σ_r, an increase in mean reversion makes the volatility of a longer-tenor swaption decrease relative to the volatility of a shorter-tenor swaption, assuming both have the same expiry date.

With the mean reversion effect above in mind, consider (say) a caplet and a swaption with the same expiry, and imagine an experiment in which the mean reversion is changed, but the volatility σ_r is adjusted to keep the implied volatility of the caplet unchanged. It should be clear from the discussion above that as mean reversion increases, the swaption volatility will *decrease*. If, instead, the market volatility of the swaption is kept constant by adjusting σ_r for each level of mean reversion, then the caplet volatility will *increase* when the mean reversion increases. Such complementarity of effects of the mean reversion $\varkappa$ and volatility σ_r allows us, in principle, to set both in such a way that we match the market-implied volatilities of both the caplet and the swaption.

For an alternative look at the effect of mean reversion, let us consider two forward rates with different fixing dates, $F(T_1,T_1,M_1)$ and $F(T_2,T_2,M_2)$ with $T_1 \le T_2$. Observing that

$$x(t) = \sigma_r \int_0^t e^{-\varkappa(t-u)}\, dW(u),$$

it is easy to establish that

$$\begin{aligned}&\mathrm{Corr}\left(F\left(T_1, T_1, M_1\right), F\left(T_2, T_2, M_2\right)\right)\\ &\qquad = \mathrm{Corr}\left(x(T_1), x(T_2)\right) = e^{-\varkappa(T_2-T_1)}\left(\frac{1-e^{-2\varkappa T_1}}{1-e^{-2\varkappa T_2}}\right)^{1/2}.\end{aligned}$$

This correlation, which we can call *inter-temporal correlation* as the forward yields are observed at different times, depends only on $\varkappa$, i.e. on the auto-correlation properties of the process $x(t)$. At a level of $\varkappa = 0$, the inter-temporal correlation is $(T_1/T_2)^{1/2}$ and decreases to 0 as $\varkappa \to \infty$.

The dependence of inter-temporal correlation on mean reversion is potentially useful for calibration purposes. To see this, consider a hypothetical security, a basket option on a set of rates with different expiry dates, with a payoff

$$\max\left\{F\left(T_n, T_n, M_n\right),\ n = 1, \ldots, N-1\right\}.$$

It is well-known (and intuitively obvious) that prices of basket options are decreasing functions of correlation, hence we expect the price of the contract above to be increasing in the mean reversion $\varkappa$, *ceteris paribus*. If such a basket option were traded in the market, the mean reversion could in principle be implied from its price.

While basket options on forward yields are, of course, not traded outright, the example above is not as far-fetched as it might seem since a Bermudan swaption (see Section 5.12) gives the holder a right to exercise into one of several different swaps that, critically, are observed on different exercise dates. The Bermudan swaption is therefore conceptually similar to a basket option on a strip of swap rates, with each rate fixing on its own fixing date. The implications of this analogy, and the effect of mean reversion on inter-temporal correlations, will be exploited later in developing the local projection method for Bermudan swaptions (see Chapter 19).

13.1.8.2 Calibrating Mean Reversion to Volatility Ratios

We consider the qG model with the volatility function (13.39). In Section 13.1.7 we developed the volatility calibration algorithm for this model under the assumption that the mean reversion function $\varkappa(t)$ was already available. Since it is often beneficial to calibrate different model parameters separately, the calibration of mean reversion should then ideally not require the knowledge of the volatility $\sigma_r(t, x, y)$ of the model.

As presented earlier, the qG model volatility calibration involves a strip of swap rates $\{S_n(\cdot)\}_{n=1}^{N-1}$, with the rate $S_n(\cdot)$ fixing on T_n and having $\mu(n)$ periods; these swap rates define the volatility function $\sigma_r(t, x, y)$ in (13.39)

and serve as volatility calibration targets. As indicated in Section 13.1.8.1, the ratio of volatilities of two swaptions with the same expiry date is more-or-less independent of volatility, suggesting the usage of a *second* strip[5] of swap rates to define targets for mean reversion calibration as the ratios between these rates and the original ones.

We assume that a second strip of swap rates is given, with the n-th rate fixing on T_n and having $\nu(n)$ periods, where $\nu(n) \neq \mu(n)$. We use extended notations to distinguish the two strips, with

$$\left\{S_{n,\mu(n)}(\cdot)\right\}_{n=1}^{N-1}$$

used for the original strip and

$$\left\{S_{n,\nu(n)}(\cdot)\right\}_{n=1}^{N-1}$$

used for the additional one. The mean reversion is then calibrated to the pairwise ratios of implied volatilities of $S_{n,\nu(n)}(T_n)$ and $S_{n,\mu(n)}(T_n)$, $n = 1, \ldots, N-1$. The following result forms the basis of mean reversion calibration.

Proposition 13.1.11. *In the quasi-Gaussian model with local short rate volatility function (13.9), the ratio of variances of two swap rates fixing on T_n with m_1 and m_2 periods, respectively, is approximately given by either*

$$\frac{\mathrm{Var}\left(S_{n,m_1}(T_n)\right)}{\mathrm{Var}\left(S_{n,m_2}(T_n)\right)} \approx \frac{\int_0^{T_n}\left(\sigma_r^0(t)\frac{\partial S_{n,m_1}}{\partial x}(t,0,0)\right)^2 dt}{\int_0^{T_n}\left(\sigma_r^0(t)\frac{\partial S_{n,m_2}}{\partial x}(t,0,0)\right)^2 dt}, \tag{13.40}$$

where $\sigma_r^0(t)$ is defined in (13.12), or

$$\frac{\mathrm{Var}\left(S_{n,m_1}(T_n)\right)}{\mathrm{Var}\left(S_{n,m_2}(T_n)\right)} \approx \frac{\int_0^{T_n}\left(\frac{\partial S_{n,m_1}}{\partial x}(t,0,0)\right)^2 dt}{\int_0^{T_n}\left(\frac{\partial S_{n,m_2}}{\partial x}(t,0,0)\right)^2 dt}. \tag{13.41}$$

Proof. The first formula is obtained by using $x(t) = 0$, $y(t) = 0$ in (13.16); see also Remark 13.1.7. The second one follows from the first under the approximation of the time-dependent volatility $\sigma_r^0(t)$ by a constant,

$$\sigma_r^0(t) \approx \sigma_r^0(0). \tag{13.42}$$

□

[5]Note that it is also possible to calibrate mean reversion to a whole *collection* of European swaptions (i.e. more than two) sharing the same expiry; the motivation for such calibration and an outline of the algorithm are presented in Section 19.4.4.

Remark 13.1.12. In practice, using the simpler approximation (13.41) instead of (13.40) does not reduce the accuracy of mean reversion calibration much. However, if the more accurate formula (13.40) is preferred, one can use an estimate of $\sigma_r^0(t)$ obtained from, for example, a pre-calibration of a pure Gaussian model.

Proposition 13.1.11 suggests an algorithm for mean reversion calibration. Recalling formulas (13.17), (13.3) we notice that the ratios

$$\left\{ \frac{\mathrm{Var}\left(S_{n,\nu(n)}(T_n)\right)}{\mathrm{Var}\left(S_{n,\mu(n)}(T_n)\right)} \right\}_{n=1}^{N-1}, \tag{13.43}$$

as computed in (13.41) depend on the mean reversion parameter $\varkappa(t)$ only, allowing us to set up an optimization problem in which $\varkappa(t)$ is chosen to match the ratios (13.43) to their market-implied values.

For a possible calibration algorithm, suppose that we discretize $\varkappa(t)$ on the time grid,

$$\varkappa(t) = \sum_{n=1}^{N-1} \varkappa_n \times 1_{\{t \in (T_{n-1}, T_n]\}} + \varkappa_N \times 1_{\{t \in (T_{N-1}, \infty)\}}.$$

It is generally not advisable to find these mean reversions only by best-fitting to variance ratios, as the function $\varkappa(t)$ would likely end up being quite irregular. To prevent this, we suggest the inclusion of a regularization term penalizing non-stationary behavior. While there are multiple ways of doing this, one simple approach would be to set optimal mean reversion levels $\{\varkappa_n^*\}_{n=1}^N$ as the solution to the following optimization problem,

$$\begin{aligned} \{\varkappa_n^*\} = \mathrm{argmin} \Bigg\{ \sum_{n=1}^{N-1} \Bigg(& \frac{\mathrm{Var}\left(S_{n,\nu(n)}(T_n)\right)}{\mathrm{Var}\left(S_{n,\mu(n)}(T_n)\right)} (\{\varkappa_n\}) \\ & - \frac{\widehat{\mathrm{Var}}\left(S_{n,\nu(n)}(T_n)\right)}{\widehat{\mathrm{Var}}\left(S_{n,\mu(n)}(T_n)\right)} \Bigg)^2 + w \sum_{n=1}^{N-1} (\varkappa_{n+1} - \varkappa_n)^2 \Bigg\}, \end{aligned} \tag{13.44}$$

where

$$\frac{\mathrm{Var}\left(S_{n,\nu(n)}(T_n)\right)}{\mathrm{Var}\left(S_{n,\mu(n)}(T_n)\right)} (\{\varkappa_n\}) = \frac{\int_0^{T_n} \left(\frac{\partial S_{n,\nu(n)}}{\partial x}(t,0,0) \right)^2 dt}{\int_0^{T_n} \left(\frac{\partial S_{n,\mu(n)}}{\partial x}(t,0,0) \right)^2 dt}.$$

Here $w > 0$ is a user-specified regularization weight, and the $\widehat{\mathrm{Var}}(S_{n,m})$'s are market-implied variances of swap rates. This minimization problem is easily solved by standard non-linear optimization methods, to be discussed in some detail later (in particular in Section 14.5).

At this point the reader might, of course, wonder whether target variances of swap rates, as required for the mean reversion calibration, could indeed be observed in the market. The answer is yes: in general, a market-implied variance of a swap rate can be calculated directly from values of options on the swap rate (i.e. swaptions) across a collection of strikes, as discussed in detail in Chapter 16. In the most commonly used linear local short rate volatility (13.29) case, we may simplify matters further. Indeed, in this model the market volatility parameters of swap rates are the displaced log-normal volatilities $\widehat{\lambda}_{S_{n,m}}$ (see (13.33) and Section 13.1.7), in which case the market-implied variance of a swap rate can be approximated by

$$\widehat{\mathrm{Var}}\left(S_{n,m}(T_n)\right) \approx \left(S_{n,m}(0)\widehat{\lambda}_{S_{n,m}}\right)^2 T_n. \tag{13.45}$$

Hence, in this case the mean reversion calibration targets could be specified directly in terms of market-implied displaced log-normal volatilities. In any case, when paired up with the volatility calibration procedure from Section 13.1.7, numerical solution of the optimization problem (13.44) ultimately allows us to calibrate the model to market-implied values of two separate swaption strips.

13.1.8.3 Calibrating Mean Reversion to Inter-Temporal Correlations

In the previous section, we took advantage of the observation from Section 13.1.8.1 that a ratio of standard deviations of two rates with the same expiry in a one-factor constant-volatility Gaussian model is independent of volatility. In this section, we develop a different calibration method for mean reversion, based on our earlier observation that inter-temporal correlations between forward yields are also nearly independent of volatility.

We focus on correlations (for the rest of this section, we omit the "inter-temporal" qualifier for brevity) of the original strip of swap rates $\{S_n(\cdot)\}_{n=1}^{N-1}$. While a different swap rate strip could be used, the practical importance of such a generalization is limited.

Proposition 13.1.13. *Let $\sigma_r^0(t)$ be given by (13.12). The correlation between two rates $S_{n_1}(T_{n_1})$ and $S_{n_2}(T_{n_2})$, $n_1 \leq n_2$, in the quasi-Gaussian model with the general local volatility function (13.9) can be approximated by either*

$$\begin{aligned}
&\operatorname{Corr}\left(S_{n_1}\left(T_{n_1}\right), S_{n_2}\left(T_{n_2}\right)\right) \\
&\quad \approx \int_0^{T_{n_1}} \left(\frac{\partial S_{n_1}}{\partial x}(t,0,0)\right)\left(\frac{\partial S_{n_2}}{\partial x}(t,0,0)\right) \sigma_r^0(t)^2\, dt \\
&\qquad \times \left(\int_0^{T_{n_1}} \left(\frac{\partial S_{n_1}}{\partial x}(t,0,0)\right)^2 \sigma_r^0(t)^2\, dt\right)^{-1/2} \\
&\qquad \times \left(\int_0^{T_{n_2}} \left(\frac{\partial S_{n_2}}{\partial x}(t,0,0)\right)^2 \sigma_r^0(t)^2\, dt\right)^{-1/2},
\end{aligned} \tag{13.46}$$

or

$$\begin{aligned}
&\operatorname{Corr}\left(S_{n_1}\left(T_{n_1}\right), S_{n_2}\left(T_{n_2}\right)\right) \approx \int_0^{T_{n_1}} \left(\frac{\partial S_{n_1}}{\partial x}(t,0,0)\right)\left(\frac{\partial S_{n_2}}{\partial x}(t,0,0)\right) dt \\
&\times \left(\int_0^{T_{n_1}} \left(\frac{\partial S_{n_1}}{\partial x}(t,0,0)\right)^2 dt\right)^{-1/2} \left(\int_0^{T_{n_2}} \left(\frac{\partial S_{n_2}}{\partial x}(t,0,0)\right)^2 dt\right)^{-1/2}.
\end{aligned} \tag{13.47}$$

Proof. By Proposition 13.1.2,

$$dS_{n_i}(t) = \frac{\partial S_{n_i}}{\partial x}(t, x(t), y(t))\, \sigma_r\left(t, x(t), y(t)\right)\, dW^{A_{n_i}}(t)$$

in $\mathrm{Q}^{A_{n_i}}$, $i = 1, 2$. In the risk-neutral measure, the SDE for $S_{n_i}(t)$ will have a stochastic drift; however, for the purposes of calculating the correlations in question, we ignore drift contributions. Thus, in the risk-neutral measure Q we have approximately that

$$\begin{aligned}
&\mathrm{E}\left(S_{n_1}\left(T_{n_1}\right) S_{n_2}\left(T_{n_2}\right)\right) - S_{n_1}(0) S_{n_2}(0) \\
&\quad = \mathrm{E}\left(S_{n_1}\left(T_{n_1}\right) S_{n_2}\left(T_{n_1}\right)\right) - S_{n_1}(0) S_{n_2}(0) \\
&\quad = \mathrm{E} \int_0^{T_{n_1}} \frac{\partial S_{n_1}}{\partial x}(t, x(t), y(t)) \frac{\partial S_{n_2}}{\partial x}(t, x(t), y(t)) \sigma_r\left(t, x(t), y(t)\right)^2 dt.
\end{aligned}$$

Using (13.12) and approximating all functions of state variables with their values at $x = y = 0$, we obtain

$$\begin{aligned}
&\mathrm{E}\left(S_{n_1}\left(T_{n_1}\right) S_{n_2}\left(T_{n_2}\right)\right) - S_{n_1}(0) S_{n_2}(0) \\
&\qquad \approx \int_0^{T_{n_1}} \left(\frac{\partial S_{n_1}}{\partial x}(t,0,0)\right)\left(\frac{\partial S_{n_2}}{\partial x}(t,0,0)\right) \sigma_r^0(t)^2\, dt.
\end{aligned}$$

Similar formulas hold for $\mathrm{E}(S_{n_i}(T_{n_i})^2)$, $i = 1, 2$, and (13.46) follows. The result (13.47) follows from (13.46) by approximating the deterministic volatility function $\sigma_r^0(t)$ with a constant,

$$\sigma_r^0(t) \approx \sigma_r^0(0).$$

□

If a "market-implied" correlation matrix

$$\widehat{\chi}_{n_1,n_2} = \mathrm{Corr}\left(S_{n_1}\left(T_{n_1}\right), S_{n_2}\left(T_{n_2}\right)\right), \quad 1 \leq n_1 \leq n_2 < N-1,$$

is somehow known — extracted, for example, from Bermudan swaption prices that, per discussion in Section 13.1.8.1, depend strongly on such correlations — then Proposition 13.1.13 can be used to calibrate the mean reversion function $\varkappa(t)$ to this matrix. The formula (13.47) is independent of the volatility term $\sigma_r(t,x,y)$, and the mean reversion calibration can precede volatility calibration. Alternatively, the more accurate formula (13.46) could be used with the volatility obtained from a pre-calibration of a pure Gaussian model.

13.1.8.4 Final Comments on Mean Reversion Calibration

The reader has undoubtedly noticed that the formulas for mean reversion calibration were derived using rather crude approximations. Refinements are certainly possible, but achieving a high level of accuracy for mean reversion calibration was never our objective. Indeed, while it is possible to execute a global calibration to vanilla option in which the mean reversion and the volatility function are calibrated together using numerical valuation methods (such as the PDE method, see later in the chapter), we believe that in setting mean reversion, market information should be a rough guide rather than a "hard" calibration target. For this reason, we recommend using mean reversion calibration to match ratios of volatilities or inter-temporal correlations in an approximate sense only; subsequently, we can apply much more precise volatility calibration to recover our main targets, namely the implied volatilities of the primary swaption strip.

Calibrating a model with relatively limited set of parameters to market inevitably leads to time-dependent parameters and, subsequently, questions about the stationarity of the resulting volatility structure. In most applications, a completely time-stationary model would yield a clearly unacceptable fit to market data, and a certain degree of non-stationarity is unavoidable. However, as far as mean reversion calibration is concerned, we often advocate using a *constant* mean reversion function $\varkappa(t) \equiv \varkappa$ in the calibration routines developed previously. For example, we can use a constant mean reversion $\varkappa$ to roughly match the volatility ratios of caplets and swaptions, and then calibrate a time-dependent volatility function to match the volatilities of swaptions in the swaption strip exactly. To encourage even more time stationarity in the model, it is also possible to set (through an optimizer) the mean reversion $\varkappa$ in such a way that the volatility parameters of the calibrated model (i.e. the λ_n's) are as close to constant as possible.

Let us note that there is one instance of the application of the model where we recommend time-dependent mean reversion, namely when the

model is used as part of the *local projection method*.. The method is developed in detail later in the book (see e.g. Sections 18.4, 19.2, 20.1.3, 20.2.1), but it can loosely be described as using a "small" model, such as the qG model, as a local or instrument-specific proxy for a "big" model, such as a Libor market model (see Chapter 14). In this case the dynamics of the volatility structure are defined by the "big" and, hopefully, realistic model, and the local model is effectively just a mechanism to reduce numerical complexity of valuation.

Finally, we note that it is, of course, also possible to use the qG model without explicitly calibrating the mean reversion parameter. For example, it could be exposed as an "exotic risk" parameter and set exogenously by a trader to reflect his estimation of the market prices of Bermudan swaptions or other exotic securities. Such practice is, we believe, quite common.

13.1.9 Numerical Methods

13.1.9.1 Direct Integration

We start the discussion of numerical methods for pricing derivatives in the quasi-Gaussian model by deriving an approximation to the density of the state variables $x(T)$ and $y(T)$, $T > 0$. Our approximation is constructed to be suitable for small T; we find that it has good accuracy for T around 1–2 years or less, depending on the level of volatility. While usable for valuing European-style derivatives by direct integration — a method generally preferable to PDE or Monte Carlo methods when available — the real utility of having a probability density comes in improving the accuracy of the PDE method, as described in Sections 2.8.2 and 2.8.3.

Consider a contract with a payoff $V(x(T), y(T))$ at time T. Its value at time 0 is given by

$$V_0 = P(0,T)\,\mathrm{E}^T\left(V(x(T), y(T))\right),$$

where E^T (as always) denotes expectation in the T-forward measure Q^T. Accordingly, we seek the density of $x(T), y(T)$ in the T-forward measure. As we focus on short times to maturity, it suffices to replace $y(T)$ with a deterministic approximation $\overline{y}(T)$ from Proposition 13.1.4. The local volatility (13.9) and mean reversion $\varkappa(t)$ are, as a rule, piecewise constant in time; thus, in the small-T regime, they can be assumed to be independent of t. It is also safe to ignore the dependence of the volatility function on y. With this in mind, we define

$$\begin{aligned} \varkappa &= \varkappa(0), \\ \upsilon(x) &= \sigma_r(0, x, 0)\,/\sigma_r^0(0). \end{aligned} \tag{13.48}$$

The following result holds.

Theorem 13.1.14. *Let us define $\pi(x)$ by the ODE*

$$\upsilon(x)\left(\pi'(x)\right)^2 + 2\varkappa G_2(T)\pi(x)\left(x/\pi(x)\right)' - 1 = 0, \quad x \in \mathbb{R}, \quad \pi(0) = 0,$$

where $G_2(T)$ is defined by (see (13.3) for the definition of $h(t)$)

$$G_2(T) = \int_0^T h(s)^2\, ds, \tag{13.49}$$

and the prime denotes differentiation with respect to x. Set

$$\varpi(x) = x/\pi(x),$$

and let $\Psi(T,x)$ be the CDF of $x(T)$ in the T-forward measure. Then

$$\Psi(T,x) \approx \Phi\left(\frac{x}{\sigma_r^0(0)\varpi(x)\sqrt{G_2(T)}}\right) + \sigma_r^0(0)\sqrt{G_2(T)}\varpi'(x)\phi\left(-\frac{x}{\sigma_r^0(0)\varpi(x)\sqrt{G_2(T)}}\right),$$

where $\Phi(z)$, $\phi(z)$ are the standard Gaussian CDF and PDF, respectively.

Proof. While lengthy and somewhat technical, the proof is instructive as it shows a general approach to deriving short-time densities for local volatility term structure models. Full details are shown in Appendix 13.A of this chapter. □

Remark 13.1.15. If $\sigma_r(0,x,0) = \text{const}$ (pure Gaussian case), then $\upsilon(x) = 1$, and $\pi(x) = x$ is a solution to the ODE. Therefore, $\varpi(x) = 1$, and we recover the Gaussian CDF of $x(T)$ as expected. The function $\varpi(x) = x/\pi(x)$ measures the deviation of the model from the Gaussian case; $\sigma_r^0(0)\varpi(x)$ can be thought of as an "effective term volatility" at x.

Equipped with Theorem 13.1.14, we can recover an approximation for the density

$$\psi(T,x) = \frac{\partial \Psi}{\partial x}(T,x),$$

which allows us to value (short-dated) derivative contracts by numerical integration. Specifically, the value of a contract with a payoff $V(x(T), y(T))$ at time T in the quasi-Gaussian local volatility model is approximately equal to

$$V_0 \approx P(0,T)\int_{-\infty}^{\infty} V\left(x, \overline{y}(T)\right)\psi(T,x)\, dx,$$

where $\overline{y}(T)$ is given in Proposition 13.1.4. Again, the primary use of the results in Theorem 13.1.14 is to improve on the finite difference method, using the results of Sections 2.8.2 and 2.8.3.

13.1.9.2 Finite Difference Methods

We consider the model in its general local volatility form (13.10),

$$\begin{aligned} dx(t) &= (y(t) - \varkappa(t)x(t))\, dt + \sigma_r\left(t, x(t), y(t)\right) dW(t), \\ dy(t) &= \left(\sigma_r\left(t, x(t), y(t)\right)^2 - 2\varkappa(t)y(t)\right) dt. \end{aligned}$$

Using methods from Chapter 2, a PDE for the value of a security as a function of x, y can be easily derived from these SDEs. We find it more convenient, however, to transform the variables first to replace $y(t)$ with a locally deterministic process $u(t)$ that is drift-free on average.

Recall the definition (13.12), and define the deterministic function $\overline{y}(t)$ by (as in Proposition 13.1.4)

$$d\overline{y}(t) = \left(\sigma_r^0(t)^2 - 2\varkappa(t)\overline{y}(t)\right) dt, \quad \overline{y}(0) = 0,$$

or

$$\overline{y}(t) = h(t)^2 \int_0^t \sigma_r^0(s)^2 h(s)^{-2} ds.$$

We define a new, normalized auxiliary variable $u(t)$ by

$$u(t) = y(t) - \overline{y}(t), \tag{13.50}$$

so that the state process $(x(t), u(t))$ satisfies

$$dx(t) = (u(t) + \overline{y}(t) - \varkappa(t)x(t))\, dt + \sigma_r\left(t, x(t), u(t) + \overline{y}(t)\right) dW(t), \tag{13.51}$$

$$du(t) = \left(\left(\sigma_r\left(t, x(t), u(t) + \overline{y}(t)\right)^2 - \sigma_r^0(t)^2\right) - 2\varkappa(t)u(t)\right) dt, \tag{13.52}$$

subject to $x(0) = u(0) = 0$. Values of zero-coupon discount bonds can easily be re-expressed in terms of the new state variables,

$$P(t, T, x, u) = \frac{P(0,T)}{P(0,t)} \exp\left(-G(t,T)\, x - \frac{1}{2} G(t,T)^2\, u - \frac{1}{2} G(t,T)^2\, \overline{y}(t)\right).$$

The new parameterization of the qG model reduces nicely to the (Gaussian) case of deterministic volatility, in the following sense. If $\sigma_r(t, x, y)$ is independent of x and y, then

$$\sigma_r(t, x, y) \equiv \sigma_r^0(t), \tag{13.53}$$

and the SDE for the state variable $u(t)$ becomes

$$du(t) = -2\varkappa(t)u(t)\, dt, \quad u(0) = 0,$$

with the unique solution

$$u(t) \equiv 0. \tag{13.54}$$

Thus, in the pure Gaussian case, the system of SDEs (13.51) reduces to a single SDE, in line with the way a one-factor Gaussian model was developed in Section 10.1.2.2.

Aesthetic reasons aside, the change of variable from $y(t)$ to $u(t)$ improves the numerical properties of a discretization of the PDE. Specifically, the variable y (or u) does not have a diffusion term so, at least in the y direction, the PDE is convection-dominated, in the sense described in Section 2.6. Removing most of the drift outside of the time-stepping scheme alleviates some of the numerical issues associated with such PDEs[6].

The PDE associated with the dynamics (13.51) is derived in the standard way. Let $V(t,x,u)$ be the value, at time t, of a derivative with a payoff $V(x(T),u(T))$ at time T, given that $x(t)=x$, $u(t)=u$,

$$V(t,x,u) = \mathrm{E}\left(e^{-\int_t^T r(s)\,ds} V(x(T),u(T)) \middle| x(t)=x, u(t)=u\right).$$

Then the function $V(t,x,u)$ satisfies the PDE

$$\begin{aligned} \frac{\partial V}{\partial t}(t,x,u) + (\mathcal{L}V)(t,x,u) &= (f(0,t)+x)\,V(t,x,u), \quad 0 \le t < T, \\ V(T,x,u) &= V(x,u), \end{aligned} \tag{13.55}$$

where

$$\mathcal{L} = \mathcal{L}_x + \mathcal{L}_u,$$

and

$$\begin{aligned} \mathcal{L}_x &= (u+\overline{y}(t)-\varkappa(t)x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma_r\left(t,x,u+\overline{y}(t)\right)^2 \frac{\partial^2}{\partial x^2}, \\ \mathcal{L}_u &= \left(\left(\sigma_r\left(t,x,u+\overline{y}(t)\right)^2 - \sigma_r^0(t)^2\right) - 2\varkappa(t)u\right)\frac{\partial}{\partial u}. \end{aligned}$$

The PDE (13.55) has two space dimensions and no mixed derivatives, and can be solved numerically by the Douglas-Rachford ADI method outlined in Section 2.10. The fact that $u(t)$ has no diffusion term does not complicate the situation much; even without upwinding (which we nevertheless recommend), we have observed no noticeable deterioration in the stability or accuracy of the scheme. There is some evidence that a 5-point discretization in the u direction may improve precision slightly, as may semi-Lagrangian schemes (see e.g. Chen and Forsyth [2007]); as standard methods produce adequate results, we consider such improvements optional and leave them to the reader to explore.

[6] Removing deterministic drift components from variables before discretizing a PDE is a useful trick for any model.

The simple nature of the process for $u(t)$ typically allows one to use a rather coarse discretization in the u direction. For instance, a typical setting might involve $n_t = 100$ time steps, $n_x = 150$ steps in x direction, and $n_u = 10$ steps in the u direction. With these settings, most instruments are priced to basis point precision. It should be noted that n_u can be chosen to reflect the degree to which the model deviates from the pure Gaussian case. To elaborate, consider a volatility term of the form (13.29). If $b = 0$ the model is Gaussian, in which case only *one* discretization point ($u_0 = 0$) is required, as is clear from (13.54). As b is increased, the model becomes increasingly non-Gaussian, and an ever larger number of points in u direction are required to maintain adequate precision. In other words, a practical scheme would set u as an increasing function of the skew parameter b.

The choice of the domain for $(x(t), u(t))$ follows standard prescriptions from Chapter 2. Boundaries in the x dimension are most easily obtained under the Gaussian approximation to the short rate state dynamics

$$dx(t) \approx (\overline{y}(t) - \varkappa(t)x(t))\, dt + \sigma_r^0(t)\, dW(t), \quad x(0) = 0, \tag{13.56}$$

using the formula (10.31) for the size of the grid while calculating $\mathrm{E}(x(T))$, $\mathrm{Var}(x(T))$ from (13.56) (see also footnote 7 in Chapter 10). For a slightly more refined approach, we note that with the linear local volatility (13.29) (or under a linear approximation in x to the general volatility function $\sigma_r(t, x, y)$) the distribution of $x(T)$ is closer to a displaced log-normal than to a Gaussian; we can then set the boundaries by moment-matching a displaced log-normal variable to the distribution of $x(T)$ and using appropriate quantiles. We leave it to the reader to fill in missing details.

A slightly more interesting question is how to dimension the grid in the u direction. As $u(t)$ does not have its own diffusion term, the randomness in $u(T)$ comes from the stochasticity of the drift in (13.52) which, to first order, is driven by $x(t)$. To extract this dependence, we apply a linear approximation to the volatility function $\sigma_r(t, x, y)$,

$$\sigma_r\left(t, x, u + \overline{y}(t)\right)^2 - \sigma_r^0(t)^2 \approx 2\sigma_r^0(t)\left(\partial\sigma_r\left(t, 0, 0\right)/\partial x\right) x,$$

in the SDE (13.52), which allows us to integrate (13.52) in an approximate sense,

$$u(T) \approx 2h(T)^2 \int_0^T x(t)\sigma_r^0(t)\left(\partial\sigma_r\left(t, 0, 0\right)/\partial x\right) h(t)^{-2}\, dt.$$

Then, under the Gaussian approximation (13.56) to the dynamics of $x(t)$, we see that $u(T)$ is also Gaussian with an easily calculated variance, which allows us to set the boundaries in the u direction using an analog to the formula (10.31).

Once the grid boundaries are determined, appropriate spatial boundary conditions need to be specified. This is straightforward, with the ideas from

Section 10.1.5 easily transferable to the quasi-Gaussian model. For instance, a reasonable boundary specification is to assume linearity in V as a function of u at the u-boundaries (as in Section 2.2.2), while using the PDE itself (as in Section 10.1.5.2) to establish the boundary conditions at the x-boundaries.

13.1.9.3 Monte Carlo Simulation

Application of the Monte Carlo method to the quasi-Gaussian model is straightforward and can follow general guidelines from Chapter 3. As with the PDE method, there may be some accuracy gains associated with using the state variable $u(t)$ in (13.50) instead of $y(t)$.

Consider the problem of computing the value of a derivative with the payoff $V(x(T), u(T))$ at time T. Let

$$0 = t_0 < t_1 < \ldots < t_N = T, \quad \Delta_n = t_n - t_{n-1},$$

be the discretization of the time domain. By applying a standard Euler discretization (see Section 3.2.3) to (13.51), the following stepping scheme is obtained,

$$\begin{aligned}
\widehat{x}_n &= \widehat{x}_{n-1} + \left(\widehat{u}_{n-1} + \overline{y}\left(t_{n-1}\right) - \varkappa\left(t_{n-1}\right)\widehat{x}_{n-1}\right)\Delta_n \\
&\quad + \sigma_r\left(t_{n-1}, \widehat{x}_{n-1}, \widehat{u}_{n-1} + \overline{y}\left(t_{n-1}\right)\right) Z_n \sqrt{\Delta_n}, \\
\widehat{u}_n &= \widehat{u}_{n-1} + \Delta_n \\
&\quad \times \left(\sigma_r\left(t_{n-1}, \widehat{x}_{n-1}, \widehat{u}_{n-1} + \overline{y}\left(t_{n-1}\right)\right)^2 - \sigma_r^0\left(t_{n-1}\right)^2 - 2\varkappa\left(t_{n-1}\right)\widehat{u}_{n-1}\right),
\end{aligned}$$

where $\widehat{x}_0 = \widehat{u}_0 = 0$, $\{Z_n\}_{n=1}^N$ is a collection of i.i.d. standard Gaussian random variables, and $\{\widehat{x}_n, \widehat{u}_n\}_{n=0}^N$ is an approximation to $\{x(t_n), u(t_n)\}_{n=0}^N$. Of course, more advanced discretization schemes are possible, as explained in Chapter 3. Some of the ideas of Section 10.1.6 are also applicable here, including the observation that the model can be simulated under either the terminal measure or the spot measure to avoid the bias involved in time-discretizing the continuously compounded money market account $e^{\int_0^T r(s)\,ds}$ in the risk-neutral measure.

13.1.9.4 Single-State Approximations

The extra state variable, and the resultant requirement of a *two*-dimensional PDE scheme (see (13.55)) for an essentially one-factor model, is the price one has to pay for the flexibility of volatility specification in the qG model. This price is relatively modest in practice, but does make the model slightly slower than a classical one-factor short rate model, and also makes it somewhat more challenging to use as a building block for more complicated models, such as equity or FX-linked interest rate hybrid models. In this section we

briefly outline a few ideas for reducing the dimensionality of the model to one state only.

A very simple idea that can be traced to Hagan and Woodward [1999a] (in the multi-dimensional setting) is to force the $y(t)$ variable to be deterministic in the SDE for $x(t)$:

$$dx(t) = (y(t) - \varkappa(t)x(t))\, dt + \sigma_r(t, x(t))\, dW(t), \quad x(0) = 0,$$

where now $y(t)$ is deterministic. Then, using generic machinery of Section 11.3, we can fit $y(t)$ to the initial yield curve via forward induction. Of course, tractability of bond reconstruction formulas is then largely lost, although much of the intuition behind the model is retained and we still maintain separate control over the at-the-money volatility structure (via mean reversion) and the volatility smile (via the local volatility function). Also, the bond reconstruction formulas from Proposition 13.1.1 can be considered as approximations in the new model, and potentially could be used to speed up volatility calibration.

A straight deterministic approximation is a rather blunt tool and would most likely not deliver the level of accuracy we require. A more refined approach for replacing the stochastic variable $y(t)$ in the qG model involves using its *projection* on the variable $x(t)$, as proposed in Kramin [2008]. To develop this idea in a bit more detail, consider a qG model with local volatility (13.10). Focusing first on calculating the following one-dimensional risk-neutral expectation

$$\mathrm{E}\left(V(x(T))\right),$$

with the payoff $V(x)$ is a function of $x(T)$ only, the ideas behind Markovian projection (Gyöngy's theorem, see Theorem A.1.1) give us the following exact result.

Proposition 13.1.16. *The undiscounted expected value of a payoff $V(x(T))$ in the model (13.10) is equal to*

$$\mathrm{E}\left(V(x(T))\right) = \mathrm{E}\left(V(\widetilde{x}(T))\right), \tag{13.57}$$

where the process $\widetilde{x}(t)$ satisfies

$$d\widetilde{x}(t) = (\widetilde{y}(t, \widetilde{x}(t)) - \varkappa(t)\widetilde{x}(t))\, dt + \widetilde{\sigma}_r(t, \widetilde{x}(t))\, dW(t), \tag{13.58}$$

with

$$\widetilde{y}(t, x) = \mathrm{E}\left(y(t) | x(t) = x\right), \quad \widetilde{\sigma}_r(t, x)^2 = \mathrm{E}\left(\sigma_r(t, x(t), y(t))^2 \middle| x(t) = x\right). \tag{13.59}$$

The equality (13.57) does not hold for the more realistic, and useful, case of calculating *discounted* expected values of a payoffs that depend on *both* x and y. Nor is there much theoretical justification for using this

projection when calculating expected values of payoffs that depend on values of the state variables at *multiple times*. Nevertheless, there is some empirical evidence that the approximations work reasonably well in practice. With that in mind, let us define a generic one-state approximate quasi-Gaussian local volatility model by

$$\begin{aligned} dx(t) &= \left(\widetilde{y}\left(t, x(t)\right) - \varkappa(t)x(t)\right) dt + \widetilde{\sigma}_r\left(t, x(t)\right) dW(t), \qquad (13.60)\\ P(t,T) &= \frac{P(0,T)}{P(0,t)} \exp\left(-G(t,T)\, x(t) - \frac{1}{2} G(t,T)^2\, \widetilde{y}\left(t, x(t)\right)\right),\\ r(t) &= f(0,t) + x(t), \end{aligned}$$

with $\widetilde{y}(t,x)$, $\widetilde{\sigma}_r(t,x)$ given by (13.59). Then, the value $V(t,x)$ of a given security at time t in state x satisfies the following one-dimensional PDE

$$\frac{\partial V}{\partial t} + \left(\widetilde{y}\left(t,x\right) - \varkappa(t)x\right) \frac{\partial V}{\partial x} + \frac{1}{2}\widetilde{\sigma}_r\left(t,x\right)^2 \frac{\partial^2 V}{\partial x^2} = \left(f\left(0,t\right) + x\right) V. \quad (13.61)$$

Assuming that we can evaluate all terms efficiently, solving the one-dimensional PDE (13.61) is typically quicker than solving the PDE (13.55) for the real model. There is little, if any, benefit in applying the approximation to the Monte Carlo method.

Needless to say, the PDE (13.61) should only be considered for problems inside the domain of applicability of the approximation (13.60). In general, we would expect the approximation to work reasonably well for low to moderate volatilities and maturities (up to, say, 20 years), and deteriorate for longer maturities and/or for large volatilities. Kramin [2008] reports good results across a wide maturity spectrum.

To effectively use (13.60), we need to compute/approximate $\widetilde{y}(t,x)$ and $\widetilde{\sigma}_r(t,x)$. Typically the volatility term $\sigma_r(t,x,y)$ either does not depend on y at all (see e.g. (13.29)), or depends on y in a close-to-linear fashion, due to the low variance of $y(t)$ compared to $x(t)$. In both cases, the following simple approximation

$$\widetilde{\sigma}_r\left(t,x\right) = \sigma_r\left(t,x,\widetilde{y}\left(t,x\right)\right)$$

appears to be justified.

To calculate $\widetilde{y}(t,x)$, we recall the definition of $y(t)$,

$$y(t) = h(t)^2 \int_0^t \sigma_r\left(s, x(s), y(s)\right)^2 h(s)^{-2}\, ds.$$

Conditioning on $x(t)$ and replacing $y(s)$ with $\overline{y}(s)$ in the argument of σ_r, where $\overline{y}(\cdot)$ is defined in (13.23), we obtain

$$\widetilde{y}\left(t,x\right) \approx h(t)^2 \int_0^t \mathrm{E}\left(\sigma_r\left(s, x(s), \overline{y}(s)\right)^2 \middle| x(t) = x\right) h(s)^{-2}\, ds.$$

Invoking approximate linearity of $\sigma_r(s,x,y)^2$ in x we obtain

$$\widetilde{y}(t,x) \approx h(t)^2 \int_0^t \sigma_r\left(s, \mathrm{E}\left(x(s)|\, x(t)=x\right), \overline{y}(s)\right)^2 h(s)^{-2}\, ds.$$

Under the Gaussian approximation

$$\mathrm{E}\left(x(s)|\, x(t)=x\right) \approx \frac{\mathrm{Var}\left(x(s)\right)}{\mathrm{Var}\left(x(t)\right)} x \approx \frac{h(s)^2 \int_0^s \sigma_r^0(u)^2 h(u)^{-2} du}{h(t)^2 \int_0^t \sigma_r^0(u)^2 h(u)^{-2} du} x = \frac{\overline{y}(s)}{\overline{y}(t)} x,$$

so we obtain

$$\widetilde{y}(t,x) \approx h(t)^2 \int_0^t \sigma_r\left(s, \left(\overline{y}(s)/\overline{y}(t)\right) x, \overline{y}(s)\right)^2 h(s)^{-2}\, ds. \tag{13.62}$$

A direct application of (13.62) is rather costly: with n_t discretized points in t direction and n_x points in x direction, the cost of computing $\widetilde{y}(\cdot,\cdot)$ for all t, x on the grid is $O(n_t^2 n_x)$, i.e. higher than for solving the PDE (13.61) itself. One remedy for this issue is to approximate $\sigma_r(s,x,\overline{y}(s))^2$ by a first- or second-order polynomial in x for each s, to obtain a polynomial approximation to $\widetilde{y}(s,x)$ with the coefficients computed at $O(n_t)$ cost. Alternatively, we can derive a recursive update equation for $\widetilde{y}(t,x)$ with the additional advantage that it does not rely on approximate linearity of $\sigma_r(s,x,y)^2$ in x (unlike (13.62)), a condition that, although generally desirable as we pointed out before, is not necessarily satisfied in all applications. We recall the equation (13.11) satisfied by $y(t)$, and discretize it for the time step $[t_n, t_{n+1}]$. Using short-hand notations $x_n = x(t_n)$, etc., we obtain

$$y_{n+1} = y_n + \left(\sigma_r\left(t_{n+1}, x_{n+1}, \overline{y}_{n+1}\right)^2 - 2\varkappa(t) y_n\right) \Delta_n, \quad , \Delta_n = t_{n+1} - t_n,$$

where $\sigma_r(t,x,\overline{y})^2$ is evaluated at the right point of the interval for reasons that will be clear momentarily. Next, conditioning on x_n, x_{n+1}, we obtain

$$\begin{aligned}\mathrm{E}\left(y_{n+1}|\, x_n, x_{n+1}\right) &= \mathrm{E}\left(y_n|\, x_n, x_{n+1}\right) \\ &\quad + \left(\sigma_r\left(t_{n+1}, x_{n+1}, \overline{y}_{n+1}\right)^2 - 2\varkappa(t)\mathrm{E}\left(y_n|\, x_n, x_{n+1}\right)\right) \Delta_n.\end{aligned}$$

By the Markov property

$$\mathrm{E}\left(y_n|\, x_n, x_{n+1}\right) = \mathrm{E}\left(y_n|\, x_n\right),$$

so

$$\begin{aligned}\mathrm{E}\left(y_{n+1}|\, x_n, x_{n+1}\right) &= \mathrm{E}\left(y_n|\, x_n\right) \\ &\quad + \left(\sigma_r\left(t_{n+1}, x_{n+1}, \overline{y}_{n+1}\right)^2 - 2\varkappa(t)\mathrm{E}\left(y_n|\, x_n\right)\right) \Delta_n,\end{aligned}$$

which gives us a way to obtain $\mathrm{E}(y_{n+1}|x_n, x_{n+1})$ from $\mathrm{E}(y_n|x_n)$. To get at the quantity that we want, namely $\mathrm{E}(y_{n+1}|x_{n+1})$, we average over x_n,

$$\begin{aligned}
&\mathrm{E}\left(y_{n+1}|\, x_{n+1}\right) \\
&\quad = \int \mathrm{E}\left(y_{n+1}|\, x_n = x, x_{n+1}\right) \mathrm{Q}\left(x_n \in dx|\, x_{n+1}\right) \\
&\quad = \int \Big(\mathrm{E}\left(y_n|\, x_n = x\right) \\
&\qquad + \Big(\sigma_r\left(t_{n+1}, x_{n+1}, \overline{y}_{n+1}\right)^2 - 2\varkappa(t)\mathrm{E}\left(y_n|\, x_n = x\right)\Big)\, \Delta_n\Big) \\
&\qquad \times \mathrm{Q}\left(x_n \in dx|\, x_{n+1}\right)
\end{aligned}$$

and, after rearranging some terms, we obtain a recursive formula for $\mathrm{E}(y_{n+1}|x_{n+1})$,

$$\begin{aligned}
\mathrm{E}\left(y_{n+1}|\, x_{n+1}\right) = (1 - 2\varkappa(t)\Delta_n) \int \mathrm{E}\left(y_n|\, x_n = x\right) \mathrm{Q}\left(x_n \in dx|\, x_{n+1}\right) \\
+ \sigma_r\left(t_{n+1}, x_{n+1}, \overline{y}_{n+1}\right)^2 \Delta_n.
\end{aligned}$$

For small Δ_n, the density

$$\mathrm{Q}\left(x_n \in dx|\, x_{n+1}\right)$$

is approximately Gaussian, and the required integral can be quickly computed numerically with just a few terms, giving us an algorithm of numerical complexity $O(n_t n_x)$.

Finally, we point out that the model (13.60) could be made exactly arbitrage-free by introducing a time-dependent deterministic component in its drift that is fit numerically to the initial yield curve, in line with the discussion at the beginning of this section — but of course at the cost of losing analytical tractability.

13.2 One-Factor Quasi-Gaussian Model with Stochastic Volatility

The most general one-factor quasi-Gaussian model specification allows for the short rate volatility to be a stochastic process, see (13.2). While so far we only considered the case of deterministic dependence of the volatility on state variables of the model, we now proceed to generalize the setup to include stochastic volatility.

13.2.1 Definition

Introduction of a stochastic variance process (see Chapter 8) in the specification of $g(\cdot)$ in (13.2) leads to a *stochastic volatility quasi-Gaussian* model. In particular, defining $z(t)$ to be the familiar CIR process,

$$dz(t) = \theta\left(z_0 - z(t)\right)\, dt + \eta(t)\sqrt{z(t)}\, dZ(t), \quad \langle dZ(t), dW(t)\rangle = 0,$$

we obtain a stochastic volatility qG model by specifying the volatility structure of the form

$$g\left(t,\omega\right) = \sqrt{z(t)}g\left(t, x(t), y(t)\right), \tag{13.63}$$

where $g(t,x,y)$ is a function of t, x, y only. With the standard definition

$$\sigma_r\left(t,x,y\right) = g\left(t,x,y\right)h(t),$$

the model is defined by the collection of SDEs

$$\begin{aligned} dx(t) &= \left(y(t) - \varkappa(t)x(t)\right)\, dt + \sqrt{z(t)}\sigma_r\left(t, x(t), y(t)\right)\, dW(t), \\ dy(t) &= \left(z(t)\sigma_r\left(t, x(t), y(t)\right)^2 - 2\varkappa(t)y(t)\right)\, dt, \\ dz(t) &= \theta\left(z_0 - z(t)\right)\, dt + \eta(t)\sqrt{z(t)}\, dZ(t), \end{aligned} \tag{13.64}$$

subject to

$$x(0) = y(0) = 0, \quad z(0) = z_0 = 1, \quad \langle dZ(t), dW(t)\rangle = 0.$$

When specifying the local volatility function, it was natural to use piecewise constant functions for various parameters (see (13.39)), and we do the same with the volatility of variance,

$$\eta(t) = \sum_{n=1}^{N-1} \eta_n 1_{\{t\in(T_{n-1},T_n]\}}. \tag{13.65}$$

The bond reconstitution formulas in the model (13.64) are the same as for the local volatility case; as follows from Proposition 13.1.1, they are the same for *any* quasi-Gaussian model. In particular, the zero-coupon discount bond formulas do not depend on the stochastic volatility process $z(t)$, and thus the model is a "true" stochastic volatility model, i.e. its stochastic volatility is unspanned and cannot be hedged by discount bonds. We remind the reader of the discussion of this topic in Section 11.2.3 and note that the model (13.64) has the lowest possible number of state variables — three — for an unspanned stochastic volatility term structure model, see Collin-Dufresne and Goldstein [2002b].

We note in passing that the assumption of zero correlation $\langle dZ(t), dW(t)\rangle = 0$ is a technical restriction helpful for developing efficient calibration formulas. It does not, however, restrict the range of available volatility smiles in the model, as the local volatility term can be used to control the slope of the smile. See also our discussion in Section 13.2.5.

13.2.2 Swap Rate Dynamics

Many results obtained in the local volatility case extend naturally to incorporate stochastic volatility, including Proposition 13.1.2, Lemma 13.1.3 and Proposition 13.1.8. The following analog to Corollary 13.1.9 is particularly useful for calibration.

Proposition 13.2.1. *Under the assumption of linear local short rate volatility (13.29), the dynamics of a swap rate $S(t)$ in the stochastic volatility quasi-Gaussian model (13.64) are given approximately by*

$$\begin{aligned} dS(t) &= \sqrt{z(t)}\lambda_S(t)\left(b_S(t)S(t) + (1 - b_S(t))\, S(0)\right)\, dW^A(t), \\ dz(t) &= \theta\left(z_0 - z(t)\right)\, dt + \eta(t)\sqrt{z(t)}\, dZ(t), \\ z(0) &= z_0 = 1, \quad \langle dZ(t), dW^A(t)\rangle = 0, \end{aligned}$$

where $\lambda_S(t)$ and $b_S(t)$ are given by (13.34)–(13.35).

The dynamics in Proposition 13.2.1 are easily recognized to be those of a time-dependent SV model, see Chapter 9. As time averaging methods are available for stochastic volatility models (see Section 9.3), the following proposition, an analog to Proposition 13.1.10, should not come as a surprise. See also Theorem 9.3.1, Corollary 9.3.5, and Theorem 9.3.6.

Proposition 13.2.2. *In the setting of the stochastic volatility quasi-Gaussian model (13.64) with the linear local volatility (13.29), consider a T-maturity swaption on a swap rate $S(T)$. For the purpose of European option pricing, the dynamics of $S(t)$ in its annuity measure can be approximated by the following time-homogeneous stochastic volatility model,*

$$\begin{aligned} dS(t) &= \sqrt{z(t)}\,\overline{\lambda}_S\left(\overline{b}_S S(t) + \left(1 - \overline{b}_S(t)\right) S(0)\right)\, dW^A(t), \\ dz(t) &= \theta\left(z_0 - z(t)\right)\, dt + \overline{\eta}_S\sqrt{z(t)}\, dZ(t), \end{aligned}$$

where

- *The effective volatility of variance $\overline{\eta}_S$ is given by*

$$\overline{\eta}_S^2 = \frac{\int_0^T \eta(t)^2 \rho_S(t)\, dt}{\int_0^T \rho_S(t)\, dt}, \tag{13.66}$$

  with the weight function $\rho_S(t)$ given by

$$\rho_S(r) = \int_r^T \int_s^T \lambda_S(t)^2 \lambda_S(s)^2 e^{-\theta(t-s)} e^{-2\theta(s-r)}\, dt\, ds.$$

- *The effective skew* $\overline{b}_S$ *is given by*

$$\overline{b}_S = \int_0^T b_S(t) w_S(t)\, dt, \tag{13.67}$$

with the weight function $w_S(t)$ *given by*

$$w_S(t) = \frac{v_S(t)^2 \lambda_S(t)^2}{\int_0^T v_S(u)^2 \lambda_S(u)^2\, du},$$

$$v_S(t)^2 = z_0^2 \int_0^t \lambda_S(s)^2\, ds + z_0 e^{-\theta t} \int_0^t \lambda_S(s)^2 e^{-\theta s} \int_0^s \eta(u)^2 e^{2\theta u} du\, ds.$$

- *The effective volatility* $\overline{\lambda}_S$ *is given by the solution to the equation*

$$\Psi_{\overline{z}}\left(\frac{h''(\zeta_S)}{h'(\zeta_S)}\overline{\lambda}_S^2, 0; T\right) = \Psi_{\overline{z\lambda^2}}\left(\frac{h''(\zeta_S)}{h'(\zeta_S)}, 0; T\right), \tag{13.68}$$

where (see Theorem 9.3.1)

$$\zeta_S = z_0 \int_0^T \lambda_S(t)^2\, dt,$$

$$\Psi_{\overline{z\lambda^2}}(v, 0; T) = \mathrm{E}\left(\exp\left(v \int_0^T \lambda_S(t)^2 z(t)\, dt\right)\right),$$

$$\Psi_{\overline{z}}(v, 0; T) = \mathrm{E}\left(\exp\left(v \int_0^T z(t)\, dt\right)\right),$$

$$h(x) = \frac{S(0)}{\overline{b}_S}\left(2\Phi\left(\overline{b}_S\sqrt{x}/2\right) - 1\right).$$

The functions $\lambda_S(t)$, $b_S(t)$ *are given by (13.34)–(13.35).*

13.2.3 Volatility Calibration

With a swaption strip $\{S_n(\cdot)\}_{n=1}^{N-1}$ given as in Sections 13.1.6 and 13.1.7, the volatility calibration algorithm can proceed along the same principles as in Section 13.1.7, where swap rate distributions in the stochastic volatility qG model can be found from the constant-parameter displaced SV SDEs in Proposition 13.2.2. As before, we assume that a collection of market parameters $(\widehat{\lambda}_{S_n}, \widehat{b}_{S_n}, \widehat{\eta}_{S_n})$, $n = 1, \ldots, N-1$, is given; in practice, these parameters may be obtained by fitting a vanilla SV model to swaptions of a given expiry/tenor across strikes, a procedure described in more detail in Section 16.1.4.

While it is easy to modify the algorithm of Section 13.1.7 to introduce one more variable to solve for (η_n) in Step 6 for each n, the calibration algorithm

is typically more stable if we first solve for the volatility of variance function $\eta(t)$, i.e. find $\{\eta_n^*\}$ for all n, and then follow the algorithm from Section 13.1.7 to solve recursively in n for (λ_n^*, b_n^*), using slightly modified formulas from Proposition 13.1.10 as given in Proposition 13.2.2. For completeness, we repeat the algorithm with the necessary modifications.

1. Set (λ_n, b_n), $n = 1, \ldots, N-1$, to some reasonable starting values, e.g. set λ_n's to (properly scaled) volatilities obtained by calibrating a pure Gaussian model as in Section 10.1.4, and $b_n = \widehat{b}_{S_n}$.
2. Solve for η_n^* for $n = 1, \ldots, N-1$, using (13.66). For weights $\rho_{S_n}(t)$, use $\lambda_{S_n}(t)$ as computed from the first guess for λ_n's (using (13.34)) obtained on Step 1.
3. Set $n = 1$.
4. For given n, (λ_i^*, b_i^*) are known for $i = 1, \ldots, n-1$. Note we use a star to denote *calibrated* values of the model parameters.
5. Calculate $\overline{x}(t)$ (Lemma 13.1.6) and $\overline{y}(t)$ (Proposition 13.1.4) for $t \in [0, T_n]$ using (λ_i^*, b_i^*), $i = 1, \ldots, n-1$, and the initial guess for (λ_n, b_n) from Step 1. Note that $\overline{x}(t), \overline{y}(t)$ implicitly depend on n as their definition depends on the swap rate/annuity measure used.
6. Calculate $\lambda_{S_n}(t)$, $b_{S_n}(t)$ for $t \in [0, T_{n-1}]$ from (λ_i^*, b_i^*), $i = 1, \ldots, n-1$, using (13.34)–(13.35).
7. Make another guess for (λ_n, b_n).
8. Update $\lambda_{S_n}(t)$, $b_{S_n}(t)$ for $t \in (T_{n-1}, T_n]$ from (λ_i^*, b_i^*), $i = 1, \ldots, n-1$, using (13.34)–(13.35).
9. Calculate $\overline{\lambda}_{S_n}$, $\overline{b}_{S_n}$ using Proposition 13.2.2.
10. Compare $(\overline{\lambda}_{S_n}, \overline{b}_{S_n})$ to $(\widehat{\lambda}_{S_n}, \widehat{b}_{S_n})$. If not equal within given tolerance, go to Step 7. Otherwise, proceed to Step 11.
11. As we have reached acceptable convergence between $(\overline{\lambda}_{S_n}, \overline{b}_{S_n})$ and $(\widehat{\lambda}_{S_n}, \widehat{b}_{S_n})$, set the calibrated model parameter values to the latest trial values, $(\lambda_n^*, b_n^*) = (\lambda_n, b_n)$.
12. Update $n \to n+1$. If $n \leq N-1$ go to Step 4. Otherwise, conclude.

As in Section 13.1.7, Step 5 can be performed inside of the calibration loop with some positive impact on the quality of calibration at a (moderate) cost of extra complexity.

13.2.4 Mean Reversion Calibration

The stochastic variance scaling $z(t)$ has some impact on the inter-temporal correlations or volatility ratios of swaptions of different maturities, an effect we discuss further in Section 20.2.4. Still, as mean reversion calibration is not meant to be overly precise, we can continue to use formulas developed in Section 13.1.8 that, in the context of a stochastic volatility qG model, imply that we roundly ignore any such effects.

While on the topic of mean reversions, let us not forget the parameter θ, the mean reversion of the variance process. The parameter θ fundamentally determines the decay of volatility smile curvature as a function of option maturity T; a good setting of this parameter will help keep the parameter $\eta(t)$ from depending excessively on calendar time t. In general, we inherit the same θ as used for the vanilla SV model calibration to European swaptions, a subject we discuss in depth in Section 16.1.4.

13.2.5 Non-Zero Correlation

A question sometimes arises whether it is too restrictive to assume (as we do) that the correlation between the Brownian motions driving the curve factor and the stochastic volatility is zero. Empirical evidence from all major fixed income markets generally suggests that correlations between interest rates and their (short-dated) volatilities are small; see e.g. the analysis in Chen and Scott [2001]. Moreover, the assumption of zero correlation has little impact on the range of volatility smiles that the model can produce, as the skew term in the local volatility can produce the necessary tilting of the smile. Nor does it affect hedging implications of the model as long as minimum variance hedging is employed; see our discussion in Section 8.9. In our view, the zero-correlation constraint has little consequence in practice, but brings substantial technical benefits, particularly the ability to shift pricing numeraires without affecting the form of the stochastic variance process. If desired, non-zero correlation can still be accommodated, as most numerical schemes — including averaging formulas — are easy to adapt, as we briefly explain in Remark 9.3.7. One should be mindful, however, of the fact that under non-zero correlation, measure changes introduce a rate-dependent term in the drift of the stochastic variance (see Proposition 8.3.9), which requires additional considerations when deriving approximations for European swaptions, say. If interested, the reader can attack the problem along the same lines as in Chapter 15, where the case of non-zero correlation is considered in a context of a different model (the Libor market model).

13.2.6 PDE and Monte Carlo Methods

The PDE method of Section 13.1.9.2 can be extended to cover the stochastic volatility case, using the techniques of Section 9.4 for the stochastic volatility part. The resulting PDE will involve three spatial dimensions, which can be handled by the Craig-Sneyd scheme. The same is true for the Monte Carlo method, where a combination of ideas from Sections 13.1.9.3 and 9.5 cover most practical issues.

13.3 Multi-Factor Quasi-Gaussian Model

13.3.1 General Multi-Factor Model

Multi-factor quasi-Gaussian models combine the flexibility of volatility specification of multi-factor models (see Chapter 12) with the ability to generate volatility skews and smiles. Practical multi-factor quasi-Gaussian models are relatively new (see Andreasen [2005]), but could provide a compelling alternative to the Libor market model in Chapter 14[7], the current *de-facto* market standard for multi-factor models.

Following the steps that lead to the one-factor quasi-Gaussian model, the multi-factor qG model is obtained by imposing a separability condition on the volatility structure of a multi-factor HJM model. Specifically, let us consider the forward rate process

$$df(t,T) = \sigma_f(t,T,\omega)^\top \left(\left(\int_t^T \sigma_f(t,u,\omega)\, du \right) dt + dW(t) \right), \tag{13.69}$$

where $\sigma_f(t,T,\omega)$ is a d-dimensional stochastic process, and $W(t)$ a d-dimensional Brownian motion in the risk-neutral measure. Let us assume that $\sigma_f(t,T,\omega)$ is separable, in the sense that it can be written as

$$\sigma_f(t,T,\omega) = g(t,\omega)h(T), \tag{13.70}$$

where $g(t,\omega)$ is a $d \times d$ stochastic matrix-valued process, and $h(t)$ is a d-dimensional deterministic vector-valued function of time. Then we define

$$H(t) = \operatorname{diag}(h(t)) = \begin{pmatrix} h_1(t) & 0 & \ddots & 0 \\ 0 & h_2(t) & \ddots & \ddots \\ \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & 0 & h_d(t) \end{pmatrix}.$$

Let us assume further that $h_i(t) \neq 0$, $i = 1, \ldots, d$, for all t, whereby $H(t)$ is then invertible, and so we can define a diagonal $d \times d$ matrix $\varkappa(t)$ by

$$\varkappa(t) = -\frac{dH(t)}{dt} H(t)^{-1} \tag{13.71}$$

(this is the same as in (12.7)–(12.8)). Moreover, let us define

$$G(t,T) = \int_t^T H(u)H(t)^{-1}\mathbf{1}\, du, \quad \sigma_r(t,\omega) = g(t,\omega)H(t),$$

where we use the notation $\mathbf{1} = (1,1,\ldots,1)^\top$ from Section 12.1.1.1.

[7]For readers not fully familiar with Libor market models, we recommend reading Chapters 14 and 15 before proceeding with this section.

Proposition 13.3.1. *Consider a general multi-factor quasi-Gaussian model, i.e. an HJM model (13.69) with the separable volatility condition (13.70). Define stochastic processes $x(t)$, $y(t)$ by*

$$\begin{aligned} dx(t) &= (y(t)\mathbf{1} - \varkappa(t)x(t))\,dt + \sigma_r(t,\omega)^\top dW(t), \\ dy(t) &= \left(\sigma_r(t,\omega)^\top \sigma_r(t,\omega) - \varkappa(t)y(t) - y(t)\varkappa(t)\right) dt, \end{aligned} \tag{13.72}$$

where $x(t) \in \mathbb{R}^d$, $y(t) \in \mathbb{R}^{d\times d}$, and $x(0) = 0$, $y(0) = 0$. Then, zero-coupon discount bonds are given by

$$P(t,T) = P(t,T,x(t),y(t)),$$

with

$$P(t,T,x,y) = \frac{P(0,T)}{P(0,t)} \exp\left(-G(t,T)^\top x - \frac{1}{2}G(t,T)^\top y G(t,T)\right).$$

In addition, the instantaneous forward rates are given by

$$f(t,T) = f(0,T) + \mathbf{1}^\top H(T)H(t)^{-1}\left(x(t) + y(t)G(t,T)\right), \tag{13.73}$$

with the short rate

$$r(t) = f(0,t) + \mathbf{1}^\top x(t).$$

Proof. Follows closely that of Proposition 12.1.2. □

13.3.2 Local and Stochastic Volatility Parameterization

While a pure local volatility specification of the multi-factor qG model is certainly possible, for brevity let us proceed directly to a more general setting where we have both local and stochastic volatility. Following Section 13.2, we start by specifying a one-dimensional process $z(t)$ by

$$dz(t) = \theta\left(z_0 - z(t)\right) dt + \eta(t)\sqrt{z(t)}\,dZ(t), \quad z(0) = z_0 = 1, \tag{13.74}$$

with $\langle dZ(t), dW(t)\rangle = 0$. Inspired by the one-dimensional case, we would like to specify a model with the volatility structure of the type

$$\sigma_r(t,\omega)^\top = \sqrt{z(t)}\sigma_x(t,x(t),y(t))^\top, \tag{13.75}$$

where $\sigma_x(t,x,y)$ is a multi-dimensional local volatility function responsible for inducing the skews in volatility smiles of swaptions. However, it is not entirely obvious how to parametrize $\sigma_x(t,x,y)$ sensibly, as the volatility function is not only responsible for skews but also for the general volatility structure of the model, including volatilities and correlations of all the rates.

Fortunately, the ideas of Section 12.1.7 could be fruitfully recycled and extended here, as suggested by Andreasen [2005] (see also Cheyette [1991]).

Recalling the definition of benchmark rates from Section 12.1.7, we take d benchmark tenors $\delta_1 < \ldots < \delta_d$, and define d "rolling" benchmark rates $f_i(t) = f(t, t+\delta_i)$, $i = 1, \ldots, d$. Ideally, it would be convenient if the qG model specification was such that the dynamics of the benchmark rates $f_i(t)$, $i = 1, \ldots, d$, were of the familiar form

$$df_i(t) = \sqrt{z(t)}\lambda_i^f(t)\left(\alpha_i^f(t) + b_i^f(t)f_i(t)\right)dU_i(t) + O(dt), \quad i = 1, \ldots, d, \tag{13.76}$$

where $\{U_i(t)\}_{i=1}^d$ is a d-dimensional vector of Brownian motions with the correlation matrix $X^f(t) = \{\chi_{i,j}(t)\}$. The following proposition shows how model parameters need to be set for these dynamics to hold.

Proposition 13.3.2. *Let us define the $d \times d$ matrix-valued process H^f by*

$$H^f(t) = \begin{pmatrix} h(t+\delta_1)^\top \\ \vdots \\ h(t+\delta_d)^\top \end{pmatrix},$$

and σ^f by

$$\begin{aligned}&\sigma^f(t, f(t)) \\ &\quad = \operatorname{diag}\left((\lambda_1^f(t)(\alpha_1^f(t) + b_1^f(t)f_1(t)), \ldots, \lambda_d^f(t)(\alpha_d^f(t) + b_d^f(t)f_d(t)))^\top\right),\end{aligned}$$

where $f(t) = (f_1(t), \ldots, f_d(t))^\top$. Also, let $D^f(t)$ be specified by $X^f(t) = D^f(t)^\top D^f(t)$. In (13.72), let us set

$$\begin{aligned}\sigma_r(t, \omega)^\top &= \sqrt{z(t)}\sigma_x(t, x(t), y(t))^\top, \\ \sigma_x(t, x(t), y(t))^\top &= H(t)H^f(t)^{-1}\sigma^f(t, f(t))D^f(t)^\top,\end{aligned} \tag{13.77}$$

where σ_x is a function of $x(t)$, $y(t)$ because f_i's are, see (13.73). Then the qG model in Proposition 13.3.1 is consistent with the benchmark rate dynamics of equation (13.76).

Proof. From (13.73) and using the definition of the vector $f(t)$, we obtain

$$f(t) = f(0) + H^f(t)H(t)^{-1}(x(t) + y(t)G(t, T)).$$

Then, with the help of (13.72),

$$\begin{aligned}df(t) &= O(dt) + H^f(t)H(t)^{-1}\,dx(t) \\ &= O(dt) + H^f(t)H(t)^{-1}\sigma_r(t, \omega)^\top\,dW(t).\end{aligned}$$

Using (13.77), we obtain

$$\begin{aligned} df(t) &= O(dt) + \sqrt{z(t)} H^f(t) H(t)^{-1} H(t) H^f(t)^{-1} \sigma^f(t, f(t)) D^f(t)^\top \, dW(t) \\ &= O(dt) + \sqrt{z(t)} \sigma^f(t, f(t)) D^f(t)^\top \, dW(t). \end{aligned}$$

The statement of the proposition follows once we set $dU(t) = D^f(t)^\top \, dW(t)$. □

With the parameterization outlined in Proposition 13.3.2, the benchmark rates follow "SV-like" dynamics (13.76), and we can reasonably expect Libor and swap rates to follow similar dynamics, at least approximately. Besides reducing the generic qG model to a familiar class of dynamics, the parameterization in Proposition 13.3.2 also achieves a clear distinction between the effects of the various model parameters: we now have volatility parameters $\{\lambda_i^f(t)\}$, rate correlation parameters $\{\chi_{i,j}(t)\}$, skew parameters $\{b_i^f(t)\}$, and the volatility of variance $\eta(t)$.

As was the case for the model in Section 12.1.7, the qG model above has enough flexibility to calibrate to d swaption strips, if we assume that the mean reversions $\varkappa(t)$ and the correlation matrix $X^f(t)$ are specified prior to volatility calibration. If the swaption strips are of constant-tenor type — a sensible choice for $d > 1$ — then it is natural to set the tenors of the swaptions we want to use in calibration equal to the benchmark tenors δ_i, $i = 1, \ldots, d$. With the stochastic volatility parameterization (13.74), (13.77), we can calibrate

1. At-the-money volatilities for d swaption strips.
2. Slopes of the volatility smiles for d swaption strips.
3. Curvature of the smile for one swaption strip.

The last point is not as restrictive as it may appear, since the curvatures of the smiles for swaptions of the same expiry but different tenors tend to be fairly similar.

To parametrize the model in a suitable way, let us choose a tenor structure $0 = T_0 < \ldots < T_N$ and denote the swap rates/annuities for the i-th swaption strip, $i = 1, \ldots, d$, in the calibration set by

$$S_{n,\mu_i(n)}(t, x(t), y(t)), \quad A_{n,\mu_i(n)}(t, x(t), y(t)), \quad n = 1, \ldots, N-1.$$

Extending (13.39), it is natural to define, for $i = 1, \ldots, d$,

$$\lambda_i^f(t) = \sum_{n=1}^{N-1} \lambda_{i,n} 1_{\{t \in (T_{n-1}, T_n]\}}, \quad \alpha_i^f(t) = \sum_{n=1}^{N-1} S_{n,\mu_i(n)}(0) 1_{\{t \in (T_{n-1}, T_n]\}},$$

$$b_i^f(t) = \sum_{n=1}^{N-1} b_{i,n} D_{i,n} 1_{\{t \in (T_{n-1}, T_n]\}}, \quad \eta(t) = \sum_{n=1}^{N-1} \eta_n 1_{\{t \in (T_{n-1}, T_n]\}},$$

where the skew scalings $D_{i,n}$ are given by an approximate (as we ignore ∂y terms) derivative of $S_{n,\mu_i(n)}$ "in the direction" of f_i,

$$D_{i,n} = \mathbf{1}^\top H(t+\delta_i)H(t)^{-1}\left(\nabla S_{n,\mu_i(n)}\right),$$

with

$$\nabla S \triangleq (\partial S/\partial x_1, \ldots, \partial S/\partial x_d).$$

In summary, the volatility smile parameters in the model are the $(2d+1) \times (N-1)$ parameters $\{\lambda_{i,n}\}$, $\{b_{i,n}\}$ and $\{\eta_n\}$.

13.3.3 Swap Rate Dynamics and Approximations

The swap rate dynamics in the multi-factor qG model can be derived and simplified by techniques similar to those we applied earlier in the one-factor case. First, we establish a multi-dimensional counterpart to Proposition 13.1.2, i.e. the exact dynamics of a given swap rate $S(t)$ in the annuity measure corresponding to its annuity $A(t)$, see (13.13)–(13.14).

Proposition 13.3.3. *In a multi-factor stochastic volatility quasi-Gaussian model with volatility parameterization (13.75), the dynamics of a swap rate $S(t)$ defined by (13.13) are given by*

$$dS(t) = \sqrt{z(t)}\left((\nabla S)\,\sigma_x^\top \sigma_x\,(\nabla S)^\top\right)^{1/2} dU^A(t), \tag{13.78}$$

where all functions are understood to be evaluated at $(t, x(t), y(t))$, and $U^A(t)$ is a one-dimensional Brownian motion in the annuity measure Q^A.

Proof. By standard arguments of using Ito's lemma on $S(t)$ and a martingale property of $S(t)$. □

Using the Markovian projection method (see Appendix A), the dynamics of (13.78) can be approximated, for the purposes of pricing European options, with

$$dS(t) = \sqrt{z(t)}\varphi\left(t, S(t)\right) dU^A(t), \tag{13.79}$$

where

$$\varphi\left(t,s\right)^2 = \mathrm{E}^A\left((\nabla S)\,c_x\,(\nabla S)^\top \middle| S(t) = s\right),$$

and we have denoted

$$c_x = c_x\left(t,x,y\right) = \sigma_x\left(t,x,y\right)^\top \sigma_x\left(t,x,y\right). \tag{13.80}$$

We expect the local volatility term in (13.79) to mostly control the slope of the volatility smile, hence we look for a linear approximation to $\varphi(t,s)$ in s. First, we need to choose the point for expansion. One can use $x(t) = 0$, $y(t) = 0$ as a decent choice; however, using $\mathrm{E}^A(x(t))$, $\mathrm{E}^A(y(t))$ or approximations thereof is, as always, preferable.

Proposition 13.3.4. *Let*

$$\sigma_x^0(t) = \sigma_x(t, 0, 0). \tag{13.81}$$

Then

$$\mathrm{E}^A(y(t)) \approx \overline{y}(t) \triangleq H(t)\left(\int_0^t H(s)^{-1}\sigma_x^0(s)^\top \sigma_x^0(s) H(s)^{-1}\, ds\right) H(t).$$

An approximation $\overline{x}(t)$ to $\mathrm{E}^A(x(t))$ is given by

$$\overline{x}(t) = H(t)\left(\int_0^t H(s)^{-1}\left(\overline{y}(s)\mathbf{1} - \sigma_x^0(s)^\top \sigma_x^0(s) G_A(s)\right) ds\right), \quad t \in [0, T_1],$$

where (recall that $S(t) = S_{0,N}(t)$, $A(t) = A_{0,N}(t)$)

$$G_A(s) = \frac{1}{A(0)} \sum_{n=0}^{N-1} \tau_n P(0, T_{n+1}) G(s, T_{n+1}).$$

Proof. The result for $\mathrm{E}^A(y(t))$ follows after approximating $\sigma_r(t, \omega)$ in (13.72) by $\sigma_x^0(t)$ and then proceeding as in the proof of Proposition 13.1.4 (see also Section 12.1.1.1).

Using the same approximation and replacing $y(t)$ with $\overline{y}(t)$, we obtain from (13.72) the following SDE for $x_g(t)$, a Gaussian approximation to $x(t)$,

$$dx_g(t) = (\overline{y}(t)\mathbf{1} - \varkappa(t) x_g(t))\, dt + \sigma_x^0(t)^\top dW(t).$$

For a given $T > 0$, in the T-forward measure,

$$dx_g(t) = \left(\overline{y}(t)\mathbf{1} - \sigma_x^0(t)^\top \sigma_x^0(t) G(t, T) - \varkappa(t) x_g(t)\right) dt + \sigma_x^0(t)^\top dW^T(t),$$

where dW^T is a driftless Brownian motion; hence

$$\mathrm{E}^T(x_g(t)) = H(t)\left(\int_0^t H(s)^{-1}\left(\overline{y}(s)\mathbf{1} - \sigma_x^0(s)^\top \sigma_x^0(s) G(s, T)\right) ds\right).$$

Then,

$$\begin{aligned}
\mathrm{E}^A(x(t)) &\approx \mathrm{E}^A(x_g(t)) \\
&= \frac{1}{A(0)} \mathrm{E}\left(\beta(t)^{-1} A(t) x_g(t)\right) \\
&= \frac{1}{A(0)} \sum_{n=0}^{N-1} \tau_n \mathrm{E}\left(x_g(t)\beta(t)^{-1} P(t, T_{n+1})\right) \\
&= \sum_{n=0}^{N-1} \frac{\tau_n P(0, T_{n+1})}{A(0)} \mathrm{E}^{T_{n+1}}(x_g(t)) \\
&= \sum_{n=0}^{N-1} \frac{\tau_n P(0, T_{n+1})}{A(0)} H(t) \\
&\quad \times \int_0^t H(s)^{-1}\left(\overline{y}(s)\mathbf{1} - \sigma_x^0(s)^\top \sigma_x^0(s) G(s, T_{n+1})\right) ds,
\end{aligned}$$

and the result follows. □

In preparation for our next result, let us define $S_g(t)$ to be the Gaussian approximation to $S(t)$, i.e. a process with the dynamics given by (13.78) where all functions are evaluated at $(t, 0, 0)$ and there is no stochastic volatility,

$$dS_g(t) = \left((\nabla S) \left(\sigma_x^0\right)^\top \sigma_x^0 (\nabla S)^\top \right)^{1/2} \Big|_{(t,0,0)} dU^A(t).$$

The dynamics of $S_g(t)$ can alternatively be represented using a multi-dimensional Brownian motion,

$$dS_g(t) = \nabla S\,(t, 0, 0) \left(\sigma_x^0\right)^\top dW^A(t). \tag{13.82}$$

Theorem 13.3.5. *Let $\overline{x}(t)$ and $\overline{y}(t)$ be as given in Proposition 13.3.4, and let $c_x(t)$ be defined as in (13.80). For pricing European swaptions, the dynamics of the swap rate $S(t)$ (defined by (13.13)) in the multi-factor quasi-Gaussian model with the volatility structure given by (13.77) can be approximated by the following time-dependent Heston dynamics,*

$$dS(t) \approx \sqrt{z(t)}\lambda_S(t) \left(b_S(t)S(t) + (1 - b_S(t))\, S(0)\right) dU^A(t),$$

where

$$\lambda_S(t) = \frac{1}{S(0)} \left((\nabla S)\, c_x\, (\nabla S)^\top \right)^{1/2} (t, \overline{x}(t), \overline{y}(t)),$$

$$b_S(t) = S(0) \left(\frac{1}{2} \frac{(\nabla S)\, d_x\, (\nabla S)^\top}{(\nabla S) \left(c_x\, (\nabla S)^\top (\nabla S)\, c_x \right) (\nabla S)^\top} + \frac{(\nabla S) \left(c_x \left(\nabla^2 S\right) c_x \right) (\nabla S)^\top}{(\nabla S) \left(c_x\, (\nabla S)^\top (\nabla S)\, c_x \right) (\nabla S)^\top} \right) \Bigg|_{(t, \overline{x}(t), \overline{y}(t))}.$$

Here we have denoted

$$d_x = \sum_{l=1}^{d} \left(c_x\, (\nabla S)^\top \right)_l \left(\frac{\partial \sigma_x^\top}{\partial x_l} \sigma_x + \sigma_x^\top \frac{\partial \sigma_x}{\partial x_l} \right),$$

$$\frac{\partial \sigma_x^\top}{\partial x_l} = H(t) H^f(t)^{-1} \mathrm{diag} \left(\lambda_1^f(t) b_1^f(t) \frac{h_l\,(t + \delta_1)}{h_l(t)}, \ldots \right.$$
$$\left. \ldots, \lambda_d^f(t) b_d^f(t) \frac{h_l\,(t + \delta_d)}{h_l(t)} \right) D^f(t)^\top, \tag{13.83}$$

and

$$\nabla^2 S \triangleq \left\{ \partial^2 S / \partial x_i \partial x_j \right\}_{i,j=1}^{d}.$$

Proof. In (13.79), consider the conditional expected value

$$\mathrm{E}^A\left((\nabla S)\, c_x\, (\nabla S)^\top \middle| S(t) = s \right).$$

Expanding the integrand $(\nabla S) c_x (\nabla S)^\top$ around $(t, \overline{x}(t), \overline{y}(t))$ to first order in x, we obtain

$$\begin{aligned}\left((\nabla S)\, c_x\, (\nabla S)^\top \right)(t, x(t), y(t)) \approx & \left((\nabla S)\, c_x\, (\nabla S)^\top \right)(t, \overline{x}(t), \overline{y}(t)) \\ & + \nabla \left((\nabla S)\, c_x\, (\nabla S)^\top \right)\Big|_{(t,\overline{x}(t),\overline{y}(t))} (x(t) - \overline{x}(t)).\end{aligned}$$

Then

$$\begin{aligned}\mathrm{E}^A\left((\nabla S)\, c_x\, (\nabla S)^\top \middle| S(t) = s \right) \approx & \left((\nabla S)\, c_x\, (\nabla S)^\top \right)(t, \overline{x}(t), \overline{y}(t)) \\ & + \nabla \left((\nabla S)\, c_x\, (\nabla S)^\top \right)\Big|_{(t,\overline{x}(t),\overline{y}(t))} \mathrm{E}^A\left(x(t) - \overline{x}(t) \middle| S(t) = s \right).\end{aligned}$$

Using a Gaussian approximation for the conditional expected value,

$$\begin{aligned}\mathrm{E}^A\left(x(t) - \overline{x}(t) \middle| S(t) = s \right) &\approx \mathrm{E}^A\left(x_g(t) - \overline{x}(t) \middle| S_g(t) = s \right) \\ &= \frac{\mathrm{Cov}\left(S_g(t), x_g(t)\right)}{\mathrm{Var}\left(S_g(t)\right)} (s - S(0)) \\ &= \frac{(\sigma_x^0)^\top \sigma_x^0 (\nabla S\,(t,0,0))^\top}{(\nabla S(t,0,0))(\sigma_x^0)^\top \sigma_x^0 (\nabla S(t,0,0))^\top} (s - S(0)) \\ &\approx \frac{c_x\, (\nabla S)^\top}{(\nabla S)\, c_x\, (\nabla S)^\top}\Bigg|_{(t,\overline{x}(t),\overline{y}(t))} (s - S(0)).\end{aligned}$$

For any l, $l = 1, \ldots, d$, we have

$$\begin{aligned}&\frac{\partial}{\partial x_l} \left((\nabla S)\, c_x\, (\nabla S)^\top \right) \\ &\quad = \left((\nabla^2 S)_l \right)^\top c_x (\nabla S)^\top + (\nabla S) \left(\frac{\partial c_x}{\partial x_l} \right) (\nabla S)^\top + (\nabla S) c_x (\nabla^2 S)_l \\ &\quad = (\nabla S) \left(\frac{\partial \sigma_x^\top}{\partial x_l} \sigma_x + \sigma_x^\top \frac{\partial \sigma_x}{\partial x_l} \right) (\nabla S)^\top + 2 (\nabla S) c_x (\nabla^2 S)_l,\end{aligned}$$

where $(\nabla^2 S)_l$ is the l-th column of the matrix $\nabla^2 S$. Thus

$$\begin{aligned}&\nabla \left((\nabla S)\, c_x\, (\nabla S)^\top \right) c_x (\nabla S)^\top \\ &\qquad = (\nabla S)\, d_x\, (\nabla S)^\top + 2\, (\nabla S) \left(c_x \left(\nabla^2 S \right) c_x \right) (\nabla S)^\top,\end{aligned}$$

where the matrix d_x is given in the statement of the theorem, and

$$\nabla\left((\nabla S)\, c_x\, (\nabla S)^\top\right)\Big|_{(t,\overline{x}(t),\overline{y}(t))} \mathrm{E}^A\left(x(t)-\overline{x}(t)\middle| S(t)=s\right)$$
$$\approx\left(\frac{(\nabla S)\, d_x\, (\nabla S)^\top}{(\nabla S)\, c_x\, (\nabla S)^\top}\right.$$
$$\left.+\,2\,\frac{(\nabla S)\left(c_x\left(\nabla^2 S\right) c_x\right)(\nabla S)^\top}{(\nabla S)\, c_x\, (\nabla S)^\top}\right)\Bigg|_{(t,\overline{x}(t),\overline{y}(t))}(s-S(0))\,.$$

Thus, we obtain (all terms on the right-hand side evaluated at $(t,\overline{x}(t),\overline{y}(t))$)

$$\varphi(t,s)^2\approx(\nabla S)\, c_x\, (\nabla S)^\top$$
$$+\left(\frac{(\nabla S)\, d_x\, (\nabla S)^\top}{(\nabla S)\, c_x\, (\nabla S)^\top}+2\frac{(\nabla S)\left(c_x\left(\nabla^2 S\right) c_x\right)(\nabla S)^\top}{(\nabla S)\, c_x\, (\nabla S)^\top}\right)(s-S(0))\,,$$

so that, linearizing around $s\approx S(0)$,

$$\varphi(t,s)\approx\left((\nabla S)\, c_x\, (\nabla S)^\top\right)^{1/2}+\frac{1}{2\left((\nabla S)\, c_x\, (\nabla S)^\top\right)^{1/2}}$$
$$\times\left(\frac{(\nabla S)\, d_x\, (\nabla S)^\top}{(\nabla S)\, c_x\, (\nabla S)^\top}+2\frac{(\nabla S)\left(c_x\left(\nabla^2 S\right) c_x\right)(\nabla S)^\top}{(\nabla S)\, c_x\, (\nabla S)^\top}\right)(s-S(0))\,.$$

Setting

$$\lambda_S(t)=\frac{\varphi(t,S(0))}{S(0)},\quad b_S(t)=S(0)\frac{\frac{\partial\varphi}{\partial s}(t,S(0))}{\varphi(t,S(0))},$$

the main statement of the theorem follows. Finally, from the definition of $\sigma_x(t,x,y)$ (see (13.77)),

$$\frac{\partial\sigma_x^\top}{\partial x_l}=H(t)H^f(t)^{-1}\mathrm{diag}\left(\lambda_1^f(t)b_1^f(t)\frac{\partial f_1(t)}{\partial x_l},\ldots,\lambda_d^f(t)b_d^f(t)\frac{\partial f_d(t)}{\partial x_l}\right)D^f(t)^\top$$

and the expression (13.83) follows from

$$\frac{\partial f_i(t)}{\partial x_l}=h_l\left(t+\delta_i\right)/h_l(t).$$

□

Remark 13.3.6. Using averaging techniques, the time-dependent Heston dynamics could be easily translated into time-independent ones. The derivation and the result essentially mimic those for the one-dimensional case, see Proposition 13.2.2.

13.3.4 Volatility Calibration

As explained in Section 13.3.2, the model (13.72), (13.74), (13.77) has enough degrees of freedom to calibrate to the smiles of d swaption strips if the vector function $h(t)$, as well as the time-dependent correlation matrix of benchmark rates $X^f(t)$, are specified exogenously. For multi-factor quasi-Gaussian models, the strips are usually taken to be constant-tenor strips with swap tenors matching benchmark rate tenors. With $d = 4$ or 5 factors being a typical choice of dimensionality, a calibration to 4 or 5 swaption strips essentially recovers the whole universe of swaption volatilities, so there is little need to choose calibration targets in a product-specific way.

As with the one-factor stochastic volatility qG model, we favor splitting the calibration into two main steps. First, we calibrate the volatility of volatility curve $\eta(t)$ to the market-implied curvatures of the smiles or, better yet, to average curvatures of volatility smiles across swap tenors, as we only have the flexibility of making $\eta(t)$ time-specific, not tenor-specific. After that, the main calibration is performed, matching the overall levels and slopes of volatility smiles of d swaption strips. We omit the details as they follow closely the algorithm of Section 13.2.3, with the only difference being that on each time step, the calibration problem involves d swaptions and not 1. Since all formulas are closed-form, the calibration is essentially instantaneous.

13.3.5 Mean Reversions, Correlations, and Numerical Schemes

In the multi-factor context, the time-dependent "loadings" vector $h(t)$ essentially defines the interpolation rule, i.e. how the volatilities and correlations of non-benchmark rates are obtained from those of the benchmark rates. We advocate choosing d fixed values of mean reversions and using them for all cases — note that they should all be different, since the inverse of matrix $H_f(t)$ is required to exist. For example, a reasonable choice is to span the interval $[0, 1]$ with mean reversions while always including the point 0, i.e. set

$$\varkappa(t) = \operatorname{diag}\left((0.015, 0.15, 0.30, 1.20)^\top\right)$$

for a 4-dimensional model, corresponding to benchmark tenors

$$\{\delta_1, \ldots, \delta_4\} = \{6m, 2y, 10y, 30y\}.$$

However, in principle at least, the mean reversions can be calibrated as well, giving us additional d strips to calibrate to. Formulas for mean reversion calibration could be derived in the same way as for the one-dimensional case, see Section 13.1.8.

Additionally, the correlation matrix between benchmark rates, $X^f(t)$, could in principle be used in calibration, particularly when valuing products with strong correlation sensitivity. In this case, to capture market-implied correlation information in the model, one sometimes chooses to best-fit

market-observed prices of CMS spread options by tweaking benchmark rate correlations. We discuss this in more details in the context of LM models in Section 14.5.9. For now we just note that spread option values exhibit a *correlation smile*, i.e. the dependence of implied correlation on the strike of the spread option (see Section 17.4.2), so the choice of the strike to calibrate to should be carefully considered.

Finally, some brief words on numerical implementation. For $d > 1$, PDE methods quickly become impractical — even for the simple case of $d = 2$, there are 3 auxiliary (y) variables to take care of, pushing the dimension of the PDE to 5 which is prohibitively expensive in virtually all applications. However, by using tricks such as "freezing" or projecting some of the auxiliary variables, as in Section 13.1.9.4, a PDE scheme for $d = 2$ or $d = 3$ could possibly be made viable.

With Monte Carlo methods, the usual considerations apply, and no special tricks beyond those for a one-dimensional stochastic volatility model are required.

13.A Appendix: Density Approximation

We prove Theorem 13.1.14 in a number of steps. Denoting for brevity

$$\sigma = \sigma^0(0),$$

and using the notations of Section 13.1.9.1, we can write down the approximate risk-neutral model dynamics as

$$dx(t) = (\overline{y}(t) - \varkappa x(t))\, dt + \sigma \upsilon\,(x(t))\, dW(t), \tag{13.84}$$

$$d\overline{y}(t) = \left(\sigma^2 - 2\varkappa \overline{y}(t)\right)\, dt. \tag{13.85}$$

13.A.1 Simplified Forward Measure Dynamics

As a first step we simplify the dynamics of the state process.

Proposition 13.A.1. *For small time T, the distribution of $x(T)$ in the T-forward measure can be approximated by the distribution of $\widetilde{x}(T)$, with the dynamics of $\widetilde{x}(t)$ given by*

$$d\widetilde{x}(t) = -\varkappa \widetilde{x}(t)\, dt + \sigma \upsilon\,(\widetilde{x}(t))\, dW^T(t), \tag{13.86}$$

where $W^T(t)$ is a Brownian motion in the T-forward measure, and $\upsilon(x)$ is the same as in (13.84) (and defined by (13.48)).

Remark 13.A.2. Note that the statement is only about approximating the marginal distribution of $x(T)$ with $\widetilde{x}(T)$, not the dynamics of $x(\cdot)$ with $\widetilde{x}(\cdot)$ in the T-forward measure.

Proof. The process

$$dW^T(t) = dW(t) + \sigma \upsilon\left(x(t)\right) G\left(t, T\right)\, dt$$

is a driftless Brownian motion in the T-forward measure, hence

$$dx(t) = \left(\overline{y}(t) - \sigma^2 \upsilon\left(x(t)\right)^2 G\left(t, T\right) - \varkappa x(t)\right)\, dt + \sigma \upsilon\left(x(t)\right)\, dW^T(t).$$

Also,

$$\overline{y}(t) = \sigma^2 \int_0^t e^{-2\varkappa(t-s)} ds = \sigma^2 G_2(t),$$

using notation introduced in (13.3). Thus

$$x(T) = \sigma^2 \int_0^T e^{-\varkappa(T-t)} \left(G_2(t) - \upsilon\left(x(t)\right)^2 G\left(t, T\right)\right)\, dt + \sigma \int_0^T e^{-\varkappa(T-t)} \upsilon\left(x(t)\right)\, dW^T(t).$$

On the other hand, the instantaneous forward rate $f(t, T)$ is a martingale in the T-forward measure,

$$\mathrm{E}^T\left(f\left(T, T\right)\right) = f\left(0, T\right),$$

which implies that

$$\mathrm{E}^T\left(x(T)\right) = 0.$$

Therefore

$$\mathrm{E}^T\left(\int_0^T e^{-\varkappa(T-t)} \left(G_2(t) - \upsilon\left(x(t)\right)^2 G\left(t, T\right)\right)\, dt\right) = 0,$$

where the equality is only approximate as we replaced $y(t)$ with $\overline{y}(t)$, but is as accurate as the approximation of $y(t)$ with $\overline{y}(t)$. We replace the equality with a somewhat stronger condition

$$\int_0^T e^{-\varkappa(T-t)} \left(G_2(t) - \upsilon\left(x(t)\right)^2 G\left(t, T\right)\right)\, dt = 0,$$

leading to

$$x(T) = \sigma \int_0^T e^{-\varkappa(T-t)} \upsilon\left(x(t)\right)\, dW^T(t),$$

which is equivalent to (13.86). □

13.A.2 Effective Volatility

From now on we consider the distribution of $x(T)$ to be given by (13.86), and we drop the tilde to simplify the notations. The (undiscounted) value of a call option on $x(t)$ with strike k is denoted by

$$c(t,k) = \mathrm{E}^T (x(t) - k)^+ . \tag{13.87}$$

By the Bachelier formula (Remark 7.2.9), this function is known explicitly for $\upsilon(x) \equiv 1$ (since (13.86) is a Gaussian SDE then), and we denote it c_0:

$$c_0(t,k,\sigma) = (x_0 - k)\,\Phi\left(\frac{x_0 - k}{\sigma\sqrt{G_2(t)}}\right) + \sigma\sqrt{G_2(t)}\phi\left(\frac{x_0 - k}{\sigma\sqrt{G_2(t)}}\right),$$

where $\phi(z)$ is the standard Gaussian PDF, $\Phi(z)$ is the standard Gaussian CDF, $x_0 = x(0) = 0$, and $G_2(t)$ is defined in (13.49). Using this expression as the base case, we look for the approximate value of $c(t,k)$, for $x(t)$ governed by (13.86), of the form (compare to the methods of Section 7.5)

$$c(t,k) = c_0(t,k,\sigma\varpi(k)) . \tag{13.88}$$

Here $\varpi(k)$ has the meaning of the *effective term volatility*. For notational convenience, we also define

$$\zeta(t,k) = \frac{x_0 - k}{\sigma\varpi(k)\sqrt{G_2(t)}}. \tag{13.89}$$

Then

$$c(t,k) = (x_0 - k)\,\Phi(\zeta(t,k)) + \sigma\varpi(k)\sqrt{G_2(t)}\phi(\zeta(t,k)) . \tag{13.90}$$

In the next few sections, we obtain an expression for $\zeta(t,k)$ in the small-σ limit. To do so, we, firstly, derive a PDE for $c(t,k)$ defined by (13.87). Then, we substitute the expression (13.90) into the PDE to derive an equation on $\zeta(t,k)$. We drop the terms of order $O(\sigma^2)$ and smaller, and solve the simplified equation for $\zeta(t,k)$. Finally, we obtain a CDF and a PDF of $x(t)$ by differentiating $c(t,k)$ in strike.

13.A.3 The Forward Equation for Call Options

In this section we derive a PDE for $c(t,k)$ in the variables t (time to expiry) and k (strike), just as we did in Proposition 7.4.2 for vanilla local volatility models. Let $\psi(t,k)$ be the density of $x(t)$ and $\Psi(t,k)$ its CDF. We use subscripts to denote partial derivatives, and primes to denote derivatives of functions of a single variable.

Proposition 13.A.3. *The function $c(t,k)$ defined by (13.87) for $x(t)$ following (13.86) satisfies*

$$c_t(t,k) = -\varkappa\left[c(t,k) - kc_k(t,k)\right] + \frac{\sigma^2}{2}\upsilon(k)^2 c_{kk}(t,k) \tag{13.91}$$

with the initial condition

$$c(0,k) = \delta(k),$$

where $\delta(k)$ is the Dirac delta function at 0.

Proof. Follows that of Proposition 7.4.2 closely, with the use of the following identity

$$\begin{aligned}\mathrm{E}\left(x(t)1_{\{x(t)>k\}}\right) &= \mathrm{E}\left((x(t)-k)\,1_{\{x(t)>k\}}\right) + k\mathrm{E}\left(1_{\{x(t)>k\}}\right)\\ &= c(t,k) - kc_k(t,k).\end{aligned}$$

□

13.A.4 Asymptotic Expansion

Lemma 13.A.4. *The following holds for $c(t,k)$ as defined by (13.90),*

$$\frac{c_t}{\phi(\zeta)} = \sigma\varpi(k)\left(\sqrt{G_2(t)}\right)', \tag{13.92}$$

$$\frac{c - kc_k}{\phi(\zeta)} = (\varpi(k) - k\varpi'(k))\,\sigma\sqrt{G_2(t)}, \tag{13.93}$$

$$\frac{c_{kk}}{\phi(\zeta)} = -\zeta_k + \sigma\sqrt{G_2(t)}\left(\varpi''(k) - \varpi'(k)\zeta_k\zeta\right), \tag{13.94}$$

where ϖ, ζ are defined by (13.88), (13.89).

Proof. Follows by straightforward differentiation of $c(t,k)$ defined by (13.90) with the help of the identities

$$\phi'(z) = -z\phi(z), \quad \phi''(z) = \left(z^2 - 1\right)\phi(z). \tag{13.95}$$

□

Proposition 13.A.5. *If the function $\varpi(k)$ is such that*

$$\pi(k) \triangleq \frac{k}{\varpi(k)}$$

satisfies the ODE

$$\upsilon(k)^2\left(\pi'(k)\right)^2 + 2\varkappa G_2(t)\pi(k)\left(\frac{k}{\pi(k)}\right)' - 1 = 0, \quad k \in \mathbb{R},$$

with the boundary condition

$$\pi(0) = 0,$$

then

$$c(t,k) = c_0\left(t,k,\sigma\varpi(k)\right) + O(\sigma^2),$$

i.e. $c_0(t,k,\sigma\varpi(k))$ *is an approximation to* $c(t,k)$ *(given in (13.90)) to the first order in* σ, $\sigma \to 0$.

Proof. Substituting (13.92)–(13.94) into the PDE (13.91) and keeping only the terms of order $O(\sigma^2)$ we obtain,

$$\sigma\varpi(k)\frac{d}{dt}\sqrt{G_2(t)} = -\varkappa\left(\varpi(k) - k\varpi'(k)\right)\sigma\sqrt{G_2(t)} + \frac{\sigma^2}{2}\upsilon(k)^2\left(-\zeta'(k) - \sigma\sqrt{G_2(t)}\varpi'(k)\zeta'(k)\zeta(k)\right).$$

Dividing by σ and using the fact that

$$\sigma\varpi(k)\sqrt{G_2(t)}\zeta(k) = -k, \tag{13.96}$$

we obtain

$$\varpi(k)\frac{d}{dt}\sqrt{G_2(t)} = -\varkappa\left(\varpi(k) - k\varpi'(k)\right)\sqrt{G_2(t)} + \frac{\sigma}{2}\upsilon(k)^2\left(\left(\frac{k\varpi'(k)}{\varpi(k)} - 1\right)\zeta'(k)\right). \tag{13.97}$$

By definition of $\pi(k)$, with the help of (13.96), and using the definition of $G_2(t)$,

$$\zeta'(k) = -\frac{1}{\sigma\sqrt{G_2(t)}}\pi'(k), \quad \frac{k\varpi'(k)}{\varpi(k)} - 1 = -\pi'(k)\varpi(k),$$

$$\frac{d}{dt}\sqrt{G_2(t)} = \frac{dG_2(t)/dt}{2\sqrt{G_2(t)}}, \quad \frac{d}{dt}G_2(t) = e^{-2\varkappa t} = 1 - 2\varkappa G_2(t),$$

which, substituted into (13.97), gives us, after some simplifications,

$$2\varkappa G_2(t)\frac{k\varpi'(k)}{\varpi(k)} + \upsilon(k)^2\left(\pi'(k)\right)^2 = 1.$$

In addition,

$$\frac{k\varpi'(k)}{\varpi(k)} = \pi(k)\left(\frac{k}{\pi(k)}\right)'$$

so, finally,

$$\upsilon(k)^2\left(\pi'(k)\right)^2 + 2\varkappa G_2(t)\pi(k)\left(\frac{k}{\pi(k)}\right)' - 1 = 0.$$

To obtain the boundary conditions on $\pi(k)$, we recall that $\varpi(k) = k/\pi(k)$ and, as $\varpi(k)$ has to be bounded at $k = 0$, we must have $\pi(k) = 0$. □

13.A.5 Proof of Theorem 13.1.14

The statement of the theorem follows by using Proposition 13.A.3 to simplify the model dynamics to (13.86), and then differentiating $c_0(T, x, \sigma\varpi(x))$ with $\varpi(x)$ from Proposition 13.A.5 once with respect to x to obtain the approximate CDF of $x(T)$,

$$\Psi(T, x) = \frac{\partial c_0(T, x, \sigma\varpi(x))}{\partial x} + 1.$$

We omit tedious but straightforward details.

14

The Libor Market Model I

Many of the models considered so far describe the evolution of the yield curve in terms of a small set of Markov state variables. While proper calibration procedures allow for successful application of such models to the pricing and hedging of a surprising variety of securities, many exotic derivatives require richer dynamics than what is possible with low-dimensional Markov models. For instance, exotic derivatives may be strongly sensitive to the joint evolution of multiple points of the yield curve, necessitating the usage of several driving Brownian motions. Also, most exotic derivatives may not be related in any obvious way to vanilla European options, making it hard to confidently identify a small, representative set of vanilla securities to which a low-dimensional Markovian model can feasibly be calibrated. What is required in such situations is a model sufficiently rich to capture the full correlation structure across the entire yield curve, and to allow for volatility calibration to a large enough set of European options that the volatility characteristics of most exotic securities can be considered "spanned" by the calibration. Candidates for such a model include the multi-factor short rate models in Chapter 12 and the multi-factor quasi-Gaussian models in Section 13.3. In this chapter, we shall cover an alternative approach to the construction of multi-factor interest rate models, the so-called *Libor market (LM)* model framework. Originally developed in Brace et al. [1997], Jamshidian [1997], and Miltersen et al. [1997], the LM model class enjoys significant popularity with practitioners and is in many ways easier to grasp than, say, the multi-factor quasi-Gaussian models in Chapter 13.

This chapter develops the basic LM model and provides a series of extensions to the original log-normal framework in Brace et al. [1997] and Miltersen et al. [1997] in order to better capture observed volatility smiles. To facilitate calibration of the model, efficient techniques for the pricing of European securities are developed. We provide a detailed discussion of the modeling of forward rate correlations which, along with the pricing formulas for caps and swaptions, serves as the basis for most of the calibration

strategies that we proceed to examine. Many of these strategies are generic in nature and apply to multi-factor models other than the LM class, including the models discussed in Chapters 12 and 13. We wrap up the chapter with a careful discussion of schemes for Monte Carlo simulation of LM models. A number of advanced topics in LM modeling is postponed to Chapter 15.

14.1 Introduction and Setup

14.1.1 Motivation and Historical Notes

Chapter 4 introduced the HJM framework which, in its most general form, involves multiple driving Brownian motions and an infinite set of state variables (namely the set of instantaneous forward rates). As argued earlier, the HJM framework contains any arbitrage-free interest rate model adapted to a finite set of Brownian motions. Working directly with instantaneous forward rates is, however, not particularly attractive in applications, for a variety of reasons. First, instantaneous forward rates are never quoted in the market, nor do they figure directly in the payoff definition of any traded derivative contract. As discussed in Chapter 5, realistic securities (swaps, caps, futures, etc.) instead involve simply compounded (Libor) rates, effectively representing *integrals* of instantaneous forward rates. The disconnect between market observables and model primitives often makes development of market-consistent pricing expression for simple derivatives cumbersome. Second, an infinite set of instantaneous forward rates can generally[1] not be represented exactly on a computer, but will require discretization into a finite set. Third, prescribing the form of the volatility function of instantaneous forward rates is subject to a number of technical complications, requiring sub-linear growth to prevent explosions in the forward rate dynamics, which precludes the formulation of a log-normal forward rate model (see Sandmann and Sondermann [1997] and the discussion in Sections 4.5.3 and 11.1.3).

As discovered in Brace et al. [1997], Jamshidian [1997], and Miltersen et al. [1997], the three complications above can all be addressed simultaneously by simply formulating the model in terms of a non-overlapping set of simply compounded Libor rates. Not only do we then conveniently work with a finite set of directly observable rates that can be represented on a computer but, as we shall show, an explosion-free log-normal forward rate model also becomes possible. Despite the change to simply compounded rates, we should emphasize that the Libor market model will still be a special case of an HJM model, albeit one where we only indirectly specify the volatility function of the instantaneous forward rates.

[1] As we have seen in earlier chapters, for special choices of the forward rate volatility we can sometimes identify a finite-dimensional Markovian representation of the forward curve that eliminates the need to store the entire curve. This is not possible in general, however.

14.1.2 Tenor Structure

The starting point for our development of the LM model is a fixed tenor structure

$$0 = T_0 < T_1 < \ldots < T_N. \tag{14.1}$$

The intervals $\tau_n = T_{n+1} - T_n$, $n = 0, \ldots, N-1$, would typically be set to be either 3 or 6 months, corresponding to the accrual period associated with observable Libor rates. Rather than keeping track of an entire yield curve, at any point in time t we are (for now; but see Section 15.1) focused only on a finite set of zero-coupon bonds $P(t, T_n)$ for the set of n's for which $T_N \geq T_n > t$; notice that this set shrinks as t moves forward, becoming empty when $t > T_N$. To formalize this "roll-off" of zero-coupon bonds in the tenor structure as time progresses, it is often useful to work with an *index function* $q(t)$, defined by the relation

$$T_{q(t)-1} \leq t < T_{q(t)}. \tag{14.2}$$

We think of $q(t)$ as representing the tenor structure index of the shortest-dated discount bond still alive.

On the fixed tenor structure, we proceed to define Libor forward rates according to the relation (see (4.2))

$$L(t, T_n, T_{n+1}) = L_n(t) = \frac{1}{\tau_n}\left(\frac{P(t, T_n)}{P(t, T_{n+1})} - 1\right), \quad N-1 \geq n \geq q(t).$$

We note that when considering a given forward Libor rate $L_n(t)$, we always assume $n \geq q(t)$ unless stated otherwise. For any $T_n > t$,

$$P(t, T_n) = P(t, T_{q(t)}) \prod_{i=q(t)}^{n-1} (1 + L_i(t)\tau_i)^{-1}. \tag{14.3}$$

Notice that knowledge of $L_n(t)$ for all $n \geq q(t)$ is generally *not* sufficient to reconstruct discount bond prices on the entire (remaining) tenor structure; the front "stub" discount bond price $P(t, T_{q(t)})$ must also be known.

14.2 LM Dynamics and Measures

14.2.1 Setting

In the Libor market model, the set of Libor forward rates $L_{q(t)}(t), L_{q(t)+1}(t), \ldots, L_{N-1}(t)$ constitutes the set of state variables for which we wish to specify dynamics. As a first step, we pick a probability measure P and assume that those dynamics originate from an m-dimensional Brownian motion $W(t)$, in the sense that all Libor rates are measurable with

respect to the filtration generated by $W(t)$. Further assuming that the Libor rates are square integrable, it follows from the martingale representation theorem that, for all $n \geq q(t)$,

$$dL_n(t) = \sigma_n(t)^\top \left(\mu_n(t)\, dt + dW(t)\right), \tag{14.4}$$

where μ_n and σ_n are m-dimensional processes, respectively, both adapted to the filtration generated by $W(t)$. From the diffusion invariance principle (see Section 1.5) we know that $\sigma_n(t)$ is measure invariant, whereas $\mu_n(t)$ is not.

As it turns out, for a given choice of $\sigma_n(t)$ in the specification (14.4), it is quite straightforward to work out explicitly the form of $\mu_n(t)$ in various martingale measures of practical interest. We turn to this shortly, but let us first stress that (14.4) allows us to use a *different* volatility function σ_n for each of the forward rates $L_n(t)$, $n = q(t), \ldots, N-1$, in the tenor structure. This obviously gives us tremendous flexibility in specifying the volatility structure of the forward curve evolution, but in practice will require us to impose quite a bit of additional structure on the model to ensure realism and to avoid an excess of parameters. We shall return to this topic later in this chapter.

14.2.2 Probability Measures

As shown in Lemma 4.2.3, $L_n(t)$ must be a martingale in the T_{n+1}-forward measure $\mathrm{Q}^{T_{n+1}}$, such that, from (14.4),

$$dL_n(t) = \sigma_n(t)^\top\, dW^{n+1}(t), \tag{14.5}$$

where $W^{n+1}(t) \triangleq W^{T_{n+1}}(t)$ is an m-dimensional Brownian motion in $\mathrm{Q}^{T_{n+1}}$. It is to be emphasized that only *one* specific Libor forward rate — namely L_n — is a martingale in the T_{n+1}-forward measure. To establish dynamics in other probability measures, the following proposition is useful.

Proposition 14.2.1. *Let $L_n(t)$ satisfy (14.5). In measure Q^{T_n} the process for $L_n(t)$ is*

$$dL_n(t) = \sigma_n(t)^\top \left(\frac{\tau_n \sigma_n(t)}{1+\tau_n L_n(t)} dt + dW^n(t)\right),$$

where $W^n(t)$ is an m-dimensional Brownian motion in measure Q^{T_n}.

Proof. From Theorem 1.4.2 we know that the density $\varsigma(t)$ relating the measures $\mathrm{Q}^{T_{n+1}}$ and Q^{T_n} is given by

$$\begin{aligned}\varsigma(t) &= \mathrm{E}_t^{T_{n+1}}\left(\frac{d\mathrm{Q}^{T_n}}{d\mathrm{Q}^{T_{n+1}}}\right)\\ &= \frac{P(t,T_n)/P(0,T_n)}{P(t,T_{n+1})/P(0,T_{n+1})} = (1+\tau_n L_n(t))\,\frac{P(0,T_{n+1})}{P(0,T_n)}.\end{aligned}$$

Clearly, then,

$$d\varsigma(t) = \frac{P(0,T_{n+1})}{P(0,T_n)}\tau_n\, dL_n(t) = \frac{P(0,T_{n+1})}{P(0,T_n)}\tau_n\sigma_n(t)^\top\, dW^{n+1}(t),$$

or

$$d\varsigma(t)/\varsigma(t) = \frac{\tau_n\sigma_n(t)^\top\, dW^{n+1}(t)}{1+\tau_n L_n(t)}.$$

From the Girsanov theorem (Theorem 1.5.1), it follows that the process

$$dW^n(t) = dW^{n+1}(t) - \frac{\tau_n\sigma_n(t)}{1+\tau_n L_n(t)}\, dt \tag{14.6}$$

is a Brownian motion in Q^{T_n}. The proposition then follows directly from (14.5). □

To gain some further intuition for the important result in Proposition 14.2.1, let us derive it in less formal fashion. For this, consider the forward discount bond $P(t,T_n,T_{n+1}) = P(t,T_{n+1})/P(t,T_n) = (1+\tau_n L_n(t))^{-1}$. An application of Ito's lemma to $P(t,T_n,T_{n+1})$, with the help of (14.5), shows that

$$\begin{aligned}
&dP(t,T_n,T_{n+1})\\
&\quad= \tau_n^2\,(1+\tau_n L_n(t))^{-3}\,\sigma_n(t)^\top\sigma_n(t)\,dt\\
&\qquad- \tau_n\,(1+\tau_n L_n(t))^{-2}\,\sigma_n(t)^\top\, dW^{n+1}(t)\\
&\quad= \tau_n\,(1+\tau_n L_n(t))^{-2}\,\sigma_n(t)^\top\left\{\tau_n\,(1+\tau_n L_n(t))^{-1}\,\sigma_n(t)\,dt - dW^{n+1}(t)\right\}.
\end{aligned}$$

As $P(t,T_n,T_{n+1})$ must be a martingale in the Q^{T_n}-measure, it follows that

$$-dW^n(t) = \tau_n\,(1+\tau_n L_n(t))^{-1}\,\sigma_n(t)\,dt - dW^{n+1}(t)$$

is a Brownian motion in Q^{T_n}, consistent with the result in Proposition 14.2.1.

While Proposition 14.2.1 only relates "neighboring" measures $\mathrm{Q}^{T_{n+1}}$ and Q^{T_n}, it is straightforward to use the proposition iteratively to find the drift of L_n in any of the probability measures discussed in Section 4.2. Let us give some examples.

Lemma 14.2.2. *Let $L_n(t)$ satisfy (14.5). Under the* terminal measure Q^{T_N} *the process for $L_n(t)$ is*

$$dL_n(t) = \sigma_n(t)^\top\left(-\sum_{j=n+1}^{N-1}\frac{\tau_j\sigma_j(t)}{1+\tau_j L_j(t)}dt + dW^N(t)\right),$$

where $W^N(t)$ is an m-dimensional Brownian motion in measure Q^{T_N}.

Proof. From (14.6) we know that

$$\begin{aligned} dW^N(t) &= dW^{N-1}(t) + \frac{\tau_{N-1}\sigma_{N-1}(t)}{1+\tau_{N-1}L_{N-1}(t)}\,dt \\ &= dW^{N-2}(t) + \frac{\tau_{N-2}\sigma_{N-2}(t)}{1+\tau_{N-2}L_{N-2}(t)}\,dt + \frac{\tau_{N-1}\sigma_{N-1}(t)}{1+\tau_{N-1}L_{N-1}(t)}\,dt. \end{aligned}$$

Continuing this iteration down to W^{n+1}, we get

$$dW^N(t) = dW^{n+1}(t) + \sum_{j=n+1}^{N-1} \frac{\tau_j\sigma_j(t)}{1+\tau_j L_j(t)}\,dt.$$

The lemma now follows from (14.5). □

Lemma 14.2.3. *Let $L_n(t)$ satisfy (14.5). Under the* spot measure Q^B *(see Section 4.2.3) the process for $L_n(t)$ is*

$$dL_n(t) = \sigma_n(t)^\top \left(\sum_{j=q(t)}^{n} \frac{\tau_j\sigma_j(t)}{1+\tau_j L_j(t)} dt + dW^B(t) \right), \tag{14.7}$$

where $W^B(t)$ is an m-dimensional Brownian motion in measure Q^B.

Proof. Recall from Section 4.2.3 that the spot measure is characterized by a rolling or "jumping" numeraire

$$B(t) = P\left(t, T_{q(t)}\right) \prod_{n=0}^{q(t)-1} \left(1+\tau_n L_n(T_n)\right). \tag{14.8}$$

At any time t, the random part of the numeraire is the discount bond $P(t, T_{q(t)})$, so effectively we need to establish dynamics in the measure $\mathrm{Q}^{T_{q(t)}}$. Applying the iteration idea shown in the proof of Lemma 14.2.2, we get

$$dW^{n+1}(t) = dW^{q(t)}(t) + \sum_{j=q(t)}^{n} \frac{\tau_j\sigma_j(t)}{1+\tau_j L_j(t)}\,dt,$$

as stated. □

The spot and terminal measures are, by far, the most commonly used probability measures in practice. Occasionally, however, it may be beneficial to use one of the hybrid measures discussed earlier in this book, for instance if one wishes to enforce that a particular Libor rate $L_n(t)$ be a martingale. As shown in Section 4.2.4, we could pick as a numeraire the asset price process

$$\widetilde{P}_{n+1}(t) = \begin{cases} P(t, T_{n+1}), & t \le T_{n+1}, \\ B(t)/B(T_{n+1}), & t > T_{n+1}, \end{cases} \tag{14.9}$$

where $B(t)$ is the spot numeraire (14.8). Using the same technique as in the proofs of Lemmas 14.2.2 and 14.2.3, it is easily seen that when $i \geq n$, then

$$dL_i(t) = \begin{cases} \sigma_i(t)^\top \left(\sum_{j=n+1}^{i} \frac{\tau_j \sigma_j(t)}{1+\tau_j L_j(t)} dt + d\widetilde{W}^{n+1}(t) \right), & t \leq T_{n+1}, \\ \sigma_i(t)^\top \left(\sum_{j=q(t)}^{i} \frac{\tau_j \sigma_j(t)}{1+\tau_j L_j(t)} dt + d\widetilde{W}^{n+1}(t) \right), & t > T_{n+1}, \end{cases}$$

where $\widetilde{W}^{n+1}(t)$ is a Brownian motion in the measure induced by the numeraire $\widetilde{P}_{n+1}(t)$. Note in particular that $L_n(t)$ is a martingale as desired, and that we have defined a numeraire which — unlike $P(t, T_{n+1})$ — will be alive at any time t.

We should note that an equally valid definition of a hybrid measure will replace (14.9) with the asset process

$$\overline{P}_{n+1}(t) = \begin{cases} B(t), & t \leq T_{n+1}, \\ B(T_{n+1})P(t, T_N)/P(T_{n+1}, T_N), & t > T_{n+1}. \end{cases} \tag{14.10}$$

This type of numeraire process is often useful in discretization of the LM model for simulation purposes; see Section 14.6.1.2 for details.

14.2.3 Link to HJM Analysis

As discussed earlier, the LM model is a special case of the general HJM class of diffusive interest rate models. To explore this relationship a bit further, we recall that HJM models generally have risk-neutral dynamics of the form

$$df(t,T) = \sigma_f(t,T)^\top \int_t^T \sigma_f(t,u)\, du\, dt + \sigma_f(t,T)^\top dW(t),$$

where $f(t,T)$ is the time t instantaneous forward rate to time T and $\sigma_f(t,T)$ is the instantaneous forward rate volatility function. From the results in Chapter 4, it follows that dynamics for the forward bond $P(t, T_n, T_{n+1})$ are of the form

$$\frac{dP(t, T_n, T_{n+1})}{P(t, T_n, T_{n+1})} = O(dt) - \left(\sigma_P(t, T_{n+1})^\top - \sigma_P(t, T_n)^\top\right) dW(t),$$

where $O(dt)$ is a drift term and

$$\sigma_P(t,T) = \int_t^T \sigma_f(t,u)\, du.$$

By definition $L_n(t) = \tau_n^{-1}(P(t, T_n, T_{n+1})^{-1} - 1)$, so that

$$dL_n(t) = O(dt) + \tau_n^{-1}(1 + \tau_n L_n(t)) \int_{T_n}^{T_{n+1}} \sigma_f(t,u)^\top du\, dW(t).$$

By the diffusion invariance principle, it follows from (14.5) that the LM model volatility $\sigma_n(t)$ is related to the HJM instantaneous forward volatility function $\sigma_f(t,T)$ by

$$\sigma_n(t) = \tau_n^{-1}(1 + \tau_n L_n(t)) \int_{T_n}^{T_{n+1}} \sigma_f(t,u)\, du. \tag{14.11}$$

Note that, as expected, $\sigma_n(t) \to \sigma_f(t,T_n)$ as $\tau_n \to 0$.

It should be obvious from (14.11) that a complete specification of $\sigma_f(t,T)$ uniquely determines the LM volatility $\sigma_n(t)$ for all t and all n. On the other hand, specification of $\sigma_n(t)$ for all t and all n does *not* allow us to imply a unique HJM forward volatility function $\sigma_f(t,T)$ — all we are specifying is essentially a strip of contiguous integrals of this function in the T-direction. This is hardly surprising, inasmuch as the LM model only concerns itself with a finite set of discretely compounded forward rates and cannot be expected to uniquely characterize the behaviors of instantaneous forward rates and their volatilities. Along the same lines, we note that the LM model does not uniquely specify the behavior of the short rate $r(t) = f(t,t)$; as a consequence, the rolling money market account $\beta(t)$ and the risk-neutral measure are not natural constructions in the LM model[2]. Section 15.3 discusses these issues in more detail.

14.2.4 Separable Deterministic Volatility Function

So far, our discussion of the LM model has been generic, with little structure imposed on the $N-1$ volatility functions $\sigma_n(t)$, $n = 1,2,\ldots,N-1$. To build a workable model, however, we need to be more specific about our choice of $\sigma_n(t)$. A common prescription of $\sigma_n(t)$ takes the form

$$\sigma_n(t) = \lambda_n(t)\varphi\left(L_n(t)\right), \tag{14.12}$$

where $\lambda_n(t)$ is a bounded vector-valued deterministic function and $\varphi : \mathbb{R} \to \mathbb{R}$ is a time-homogeneous local volatility function. This specification is conceptually very similar to the local volatility models in Chapter 7, although here $\sigma_n(\cdot)$ is vector-valued and the model involves joint dynamics of multiple state variables (the $N-1$ Libor forward rates).

At this point, the reader may reasonably ask whether the choice (14.12) in fact leads to a system of SDEs for the various Libor forward rates that is "reasonable", in the sense of existence and uniqueness of solutions. While we here shall not pay much attention to such technical regularity issues, it should be obvious that not all functions φ can be allowed. One relevant result is given below.

[2] In fact, as discussed in Jamshidian [1997], one does not need to assume that a short rate process exists when constructing an LM model.

Proposition 14.2.4. *Assume that (14.12) holds with $\varphi(0) = 0$ and that $L_n(0) \geq 0$ for all n. Also assume that φ is locally Lipschitz continuous and satisfies the growth condition*

$$\varphi(x)^2 \leq C\left(1 + x^2\right), \quad x > 0,$$

where C is some positive constant. Then non-explosive, pathwise unique solutions of the no-arbitrage SDEs for $L_n(t)$, $q(t) \leq n \leq N-1$, exist under all measures Q^{T_i}, $q(t) \leq i \leq N$. If $L_n(0) > 0$, then $L_n(t)$ stays positive at all t.

Proof. (Sketch) Due to the recursive relationship between measures, it suffices to consider the system of SDEs (14.7) under the spot measure Q^B:

$$dL_n(t) = \varphi\left(L_n(t)\right)\lambda_n(t)^\top\left(\mu_n(t)\,dt + dW^B(t)\right), \tag{14.13}$$

$$\mu_n(t) = \sum_{j=q(t)}^{n} \frac{\tau_j \varphi\left(L_j(t)\right)\lambda_j(t)}{1 + \tau_j L_j(t)}. \tag{14.14}$$

Under our assumptions, it is easy to see that each term in the sum for μ_n is locally Lipschitz continuous and bounded. The growth condition on φ in turn ensures that the product $\varphi(L_n(t))\lambda_n(t)^\top\mu_n(t)$ is also locally Lipschitz continuous and, due to the boundedness of μ_n, satisfies a linear growth condition. Existence and uniqueness now follow from Theorem 1.6.1. The result that 0 is a non-accessible boundary for the forward rates if started above 0 follows from standard speed-scale boundary classification results; see Andersen and Andreasen [2000b] for the details. □

Some standard parameterizations of φ are shown in Table 14.1. Of those, only the log-normal specification and the LCEV specification directly satisfy the criteria in Proposition 14.2.4. The CEV specification violates Lipschitz continuity at $x = 0$, and as a result uniqueness of the SDE fails. As shown in Andersen and Andreasen [2000b], we restore uniqueness by specifying that forward rates are *absorbed* at the origin (see also Section 7.2.3). As for the displaced log-normal specification $\varphi(x) = ax + b$, we here violate the assumption that $\varphi(0) = 0$, and as a result we cannot always guarantee that forward rates stay positive. Also, to prevent explosion of the forward rate drifts, we need to impose additional restrictions to prevent terms of the form $1 + \tau_n L_n(t)$ (in the denominator) from becoming zero. As displaced log-normal models are of considerable practical importance, we list the relevant restrictions in Lemma 14.2.5 below.

Lemma 14.2.5. *Consider a local volatility Libor market model with local volatility function $\varphi(x) = bx + a$, where $b > 0$ and $a \neq 0$. Assume that $bL_n(0) + a > 0$ and $a/b < \tau_n^{-1}$ for all $n = 1, 2, \ldots, N-1$. Then non-explosive, pathwise unique solutions of the no-arbitrage SDEs for $L_n(t)$, $q(t) \leq n \leq N-1$, exist under all measures Q^{T_i}, $q(t) \leq i \leq N$. All $L_n(t)$ are bounded from below by $-a/b$.*

Name	$\varphi(x)$
Log-normal	x
CEV	$x^p, \quad 0 < p < 1$
LCEV	$x \min(\varepsilon^{p-1}, x^{p-1}), \quad 0 < p < 1, \varepsilon > 0$
Displaced log-normal	$bx + a, \quad b > 0,\ a \neq 0$

Table 14.1. Common DVF Specifications

Proof. Define $H_n(t) = bL_n(t) + a$. By Ito's lemma, we have

$$dH_n(t) = b\,dL_n(t) = bH_n(t)\lambda_n(t)^\top \left(\mu_n(t)\,dt + dW^B(t)\right),$$
$$\mu_n(t) = \sum_{j=q(t)}^{n} \frac{\tau_j H_j(t)\lambda_j(t)}{1 + \tau_j\left(H_j(t) - a\right)/b}.$$

From the assumptions of the lemma, we have $H_n(0) > 0$ for all n, allowing us to apply the result of Proposition 14.2.4 to $H_n(t)$, provided that we can guarantee that $\mu_j(t)$ is bounded for all positive H_j, $j = q(t), \ldots, n$. This follows from $1 - \tau_j a/b > 0$ or $a/b < \tau_j^{-1}$. □

We emphasize that the requirement $a/b < \tau_n^{-1}$ implies that only in the limit of $\tau_j \to 0$ — where the discrete forward Libor rates become instantaneous forward rates — will a pure Gaussian LM model specification ($b = 0$) be meaningful; such a model was outlined in Section 4.5.1. On the flip-side, according to Proposition 14.2.4, a finite-sized value of τ_j ensures that a well-behaved log-normal forward rate model exists, something that we saw earlier (Section 11.1.3) was *not* the case for models based on instantaneous forward rates. The existence of log-normal forward rate dynamics in the LM setting was, in fact, a major driving force behind the development and popularization of the LM framework, and all early examples of LM models (see Brace et al. [1997], Jamshidian [1997], and Miltersen et al. [1997]) were exclusively log-normal.

We recall from earlier chapters that it is often convenient to specify displaced log-normal models as $\varphi(L_n(t)) = (1-b)L_n(0) + bL_n(t)$, in which case the constant a in Lemma 14.2.5 is different from one Libor rate to the next. In this case, we must require

$$(1-b)/b < (L_n(0)\tau_n)^{-1}, \quad n = 1, \ldots, N-1.$$

As $L_n(0)\tau_n$ is typically in the magnitude of a few percent, the regularity requirement on b in (14.2.4) is not particularly restrictive.

14.2.5 Stochastic Volatility

As discussed earlier in this book, to ensure that the evolution of the volatility smile is reasonably stationary, it is best if the skew function φ in (14.14)

is (close to) monotonic in its argument. Typically we are interested in specifications where $\varphi(x)/x$ is downward-sloping, to establish the standard behavior of interest rate implied volatilities tending to increase as interest rates decline. In reality, however, markets often exhibit non-monotonic volatility smiles or "smirks" with high-struck options trading at implied volatilities above the at-the-money levels. An increasingly popular mechanism to capture such behavior in LM models is through the introduction of stochastic volatility. We have already encountered stochastic volatility models in Chapters 8, 9 and, in the context of term structure models, in Sections 13.2 and 13.3; we now discuss how to extend the notion of stochastic volatility models to the simultaneous modeling of multiple Libor forward rates.

As our starting point, we take the process (14.14), preferably equipped with a φ that generates either a flat or monotonically downward-sloping volatility skew, but allow the term on the Brownian motion to be scaled by a stochastic process. Specifically, we introduce a mean-reverting scalar process $z(t)$, with dynamics of the form

$$dz(t) = \theta\left(z_0 - z(t)\right)dt + \eta\psi\left(z(t)\right)dZ(t), \quad z(0) = z_0, \tag{14.15}$$

where θ, z_0, and η are positive constants, Z is a Brownian motion under the spot measure Q^B, and $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is a well-behaved function. We impose that (14.15) will not generate negative values of $z(t)$, which requires $\psi(0) = 0$. We will interpret the process in (14.15) as the (scaled) variance process for our forward rate diffusions, in the sense that the square root of $z(t)$ will be used as a stochastic, multiplicative scaling of the diffusion term in (14.14). That is, our forward rate processes in Q^B are, for all $n \geq q(t)$,

$$dL_n(t) = \sqrt{z(t)}\varphi\left(L_n(t)\right)\lambda_n(t)^\top\left(\sqrt{z(t)}\mu_n(t)\,dt + dW^B(t)\right), \tag{14.16}$$

$$\mu_n(t) = \sum_{j=q(t)}^{n} \frac{\tau_j\varphi\left(L_j(t)\right)\lambda_j(t)}{1+\tau_j L_j(t)},$$

where $z(t)$ satisfies (14.15). This construction naturally follows the specification of vanilla stochastic volatility models in Chapter 8, and the specification of stochastic volatility quasi-Gaussian models in Chapter 13. As we discussed previously, it is often natural to scale the process for $z(t)$ such that $z(0) = z_0 = 1$.

Let us make two important comments about (14.16). First, we emphasize that a single common factor $\sqrt{z(t)}$ simultaneously scales all forward rate volatilities; movements in volatilities are therefore perfectly correlated across the various forward rates. In effect, our model corresponds only to the first principal component of the movements of the instantaneous forward rate volatilities. This is a common assumption that provides good balance between realism and parsimony, and we concentrate mostly on this case — although we do relax it later in the book, in Chapter 15. Second, we note

that the clean form of the z-process (14.15) in the measure Q^B generally does not carry over to other probability measures, as we would expect from Proposition 8.3.9. To state the relevant result, let $\langle Z(t), W(t)\rangle$ denote the vector of quadratic covariations between $Z(t)$ and the m components of $W(t)$ (recall the definition of covariation in Remark 1.1.7). We then have

Lemma 14.2.6. *Let dynamics for $z(t)$ in the measure Q^B be as in (14.15). Then the SDE for $z(t)$ in measure $\mathrm{Q}^{T_{n+1}}$, $n \geq q(t) - 1$, is*

$$\begin{aligned} dz(t) &= \theta\left(z_0 - z(t)\right) dt + \eta\psi\left(z(t)\right) \\ &\qquad \times \left(-\sqrt{z(t)}\mu_n(t)^\top \left\langle dZ(t), dW^B(t)\right\rangle + dZ^{n+1}(t)\right), \end{aligned} \tag{14.17}$$

where $\mu_n(t)$ is given in (14.16) and $Z^{n+1}(t)$ is a Brownian motion in measure $\mathrm{Q}^{T_{n+1}}$.

Proof. From earlier results, we have

$$dW^{n+1}(t) = \sqrt{z(t)}\mu_n(t)\, dt + dW^B(t).$$

Let us introduce the m-dimensional vector

$$a(t) = \left\langle dZ(t), dW^B(t)\right\rangle / dt,$$

so that we can write

$$dZ(t) = a(t)^\top\, dW^B(t) + \sqrt{1 - \|a(t)\|^2}\, d\widetilde{W}(t),$$

where $\widetilde{W}(t)$ is a scalar Brownian motion independent of $W^B(t)$. In the measure $\mathrm{Q}^{T_{n+1}}$, we then have

$$\begin{aligned} dZ(t) &= a(t)^\top \left(dW^{n+1}(t) - \sqrt{z(t)}\mu_n(t)\, dt\right) + \sqrt{1 - \|a(t)\|^2}\, d\widetilde{W}(t) \\ &= dZ^{n+1}(t) - a(t)^\top \sqrt{z(t)}\mu_n(t)\, dt, \end{aligned}$$

and the result follows. □

The process (14.17) is awkward to deal with, due to presence of the drift term $\mu_n(t)^\top \langle dZ(t), dW^B(t)\rangle$ which will, in general, depend on the state of the Libor forward rates at time t. For tractability, on the other hand, we would like for the z-process to only depend on $z(t)$ itself. To achieve this, and to generally simplify measure shifts in the model, we make the following assumption[3] about (14.15)–(14.16):

Assumption 14.2.7. *The Brownian motion $Z(t)$ of the variance process $z(t)$ is independent of the vector-valued Brownian motion $W^B(t)$.*

[3] We briefly return to the general case in Section 15.6.

We have already encountered the same assumption in the context of stochastic volatility quasi-Gaussian models, see Section 13.2.1, where we also discussed the implications of such a restriction.

The diffusion coefficient of the variance process, the function ψ, is traditionally chosen to be of power form, $\psi(x) = x^\alpha, \alpha > 0$. While it probably makes sense to keep the function monotonic, the power specification is likely a nod to tradition rather than anything else. Nevertheless, some particular choices lead to analytically tractable specifications, as we saw in Chapter 8; for that reason, $\alpha = 1/2$ (the Heston model) is popular.

Remark 14.2.8. Going forward we shall often use the stochastic volatility model in this section as a benchmark for theoretical and numerical work. As the stochastic volatility model reduces to the local volatility model in Section 14.2.4 when $z(t)$ is constant, all results for the stochastic volatility model will carry over to the DVF setting.

14.2.6 Time-Dependence in Model Parameters

In the models we outlined in Sections 14.2.4 and 14.2.5, the main role of the vector-valued function of time $\lambda_n(t)$ was to establish a term structure "spine" of at-the-money option volatilities. To build volatility smiles around this spine, we further introduced a universal skew-function φ, possibly combined with a stochastic volatility scale $z(t)$ with time-independent process parameters. In practice, this typically gives us a handful of free parameters with which we can attempt to match the market-observed volatility smiles for various cap and swaption tenors. As it turns out, a surprisingly good fit to market skew data can, in fact, often be achieved with the models of Sections 14.2.4 and 14.2.5. For a truly precise fit to volatility skews across all maturities and swaption tenors it may, however, be necessary to allow for time-dependence in both the process parameters for $z(t)$ and, more importantly, the skew function φ. The resulting model is conceptually similar to the model in Section 14.2.5, but involves a number of technical intricacies that draw heavily on the material presented in Chapter 9. To avoid cluttering this first chapter on LM models with technical detail, we postpone the treatment of time-inhomogeneous φ and z-process parameters to Chapter 15.

14.3 Correlation

In one-factor models for interest rates — such as the ones presented in Chapters 10 and 11 — all points on the forward curve always move in the same direction. While this type of forward curve move indeed is the most commonly observed type of shift to the curve, "rotational steepenings" and

the formation of "humps" may also take place, as may other more complex types of curve changes. The empirical presence of such non-trivial curve movements is an indication of the fact that various points on the forward curve do not move co-monotonically with each other, i.e. they are imperfectly correlated. A key characteristic of the LM model is the consistent use of vector-valued Brownian motion drivers, of dimension m, which gives us control over the instantaneous correlation between various points on the forward curve.

Proposition 14.3.1. *The correlation between forward rate increments $dL_k(t)$ and $dL_j(t)$ in the SV model (14.16) is*

$$\mathrm{Corr}\left(dL_k(t), dL_j(t)\right) = \frac{\lambda_k(t)^\top \lambda_j(t)}{\|\lambda_k(t)\| \, \|\lambda_j(t)\|}.$$

Proof. Using the covariance notation of Remark 1.1.7, we have, for any j and k,

$$d\langle L_k(t), L_j(t)\rangle = z(t)\varphi\left(L_k(t)\right)\varphi\left(L_j(t)\right)\lambda_k(t)^\top \lambda_j(t)\, dt.$$

Using this in the definition of the correlation,

$$\mathrm{Corr}\left(dL_k(t), dL_j(t)\right) = \frac{\langle dL_k(t), dL_j(t)\rangle}{\sqrt{\langle dL_k(t)\rangle\langle dL_j(t)\rangle}},$$

which gives the result of the proposition. □

A trivial corollary of Proposition 14.3.1 is the fact that $\mathrm{Corr}(dL_k(t), dL_j(t)) = 1$ always when $m = 1$, i.e. when we only have one Brownian motion. As we add more Brownian motions, our ability to capture increasingly complicated correlation structures progressively improves (in a sense that we shall examine further shortly), but at a cost of increasing the model complexity and, ultimately, computational effort. To make rational decisions about the choice of model dimension m, let us turn to the empirical data.

14.3.1 Empirical Principal Components Analysis

For some fixed value of τ (e.g. 0.25 or 0.5), let us define "sliding" forward rates[4] $l(t, x)$ with tenor x as

$$l(t, x) = L(t, t + x, t + x + \tau).$$

[4] The use of sliding forward rates, i.e. forward rates with a fixed time **to** maturity rather than a fixed time **of** maturity, is often known as the *Musiela parameterization*.

For a given set of tenors $x_1, \ldots, x_{N_x}$ and a given set of calendar times $t_0, t_1, \ldots, t_{N_t}$, we can use market observations[5] to set up the $N_x \times N_t$ observation matrix O with elements

$$O_{i,j} = \frac{l(t_j, x_i) - l(t_{j-1}, x_i)}{\sqrt{t_j - t_{j-1}}}, \quad i = 1, \ldots, N_x, \; j = 1, \ldots, N_t.$$

Notice the normalization with $\sqrt{t_j - t_{j-1}}$ which annualizes the variance of the observed forward rate increments. Also note that we use absolute increments in forward rates here. This is arbitrary — we could have used, say, relative increases as well, if we felt that rates were more log-normal than Gaussian. For small sampling periods, the precise choice is of little importance.

Assuming time-homogeneity and ignoring small drift terms, the data collected above will imply a sample $N_x \times N_x$ variance-covariance matrix equal to

$$C = \frac{OO^\top}{N_t}. \tag{14.18}$$

For our LM model to conform to empirical data, we need to use a sufficiently high number m of Brownian motions to closely replicate this variance-covariance matrix. A formal analysis of what value of m will suffice can proceed with the tools of principal components analysis (PCA), as established in Section 3.1.3.

14.3.1.1 Example: USD Forward Rates

To give a concrete example of a PCA run, we set $N_x = 9$ and use tenors of $\{x_1, \ldots, x_9\} = \{0.5, 1, 2, 3, 5, 7, 10, 15, 20\}$ years. We fix $\tau = 0.5$ (i.e., all forward rates are 6 months discrete rates) and use 4 years of weekly data from the USD market, spanning January 2003 to January 2007, for a total of $N_t = 203$ curve observations. The eigenvalues of the matrix C in (14.18) are listed in Table 14.2, along with the percentage of variance that is explained by using only the first m principal components.

m	1	2	3	4	5	6	7	8	9
Eigenvalue	7.0	0.94	0.29	0.064	0.053	0.029	0.016	0.0091	0.0070
% Variance	83.3	94.5	97.9	98.7	99.3	99.6	99.8	99.9	100

Table 14.2. PCA for USD Rates. All eigenvalues are scaled up by 10^4.

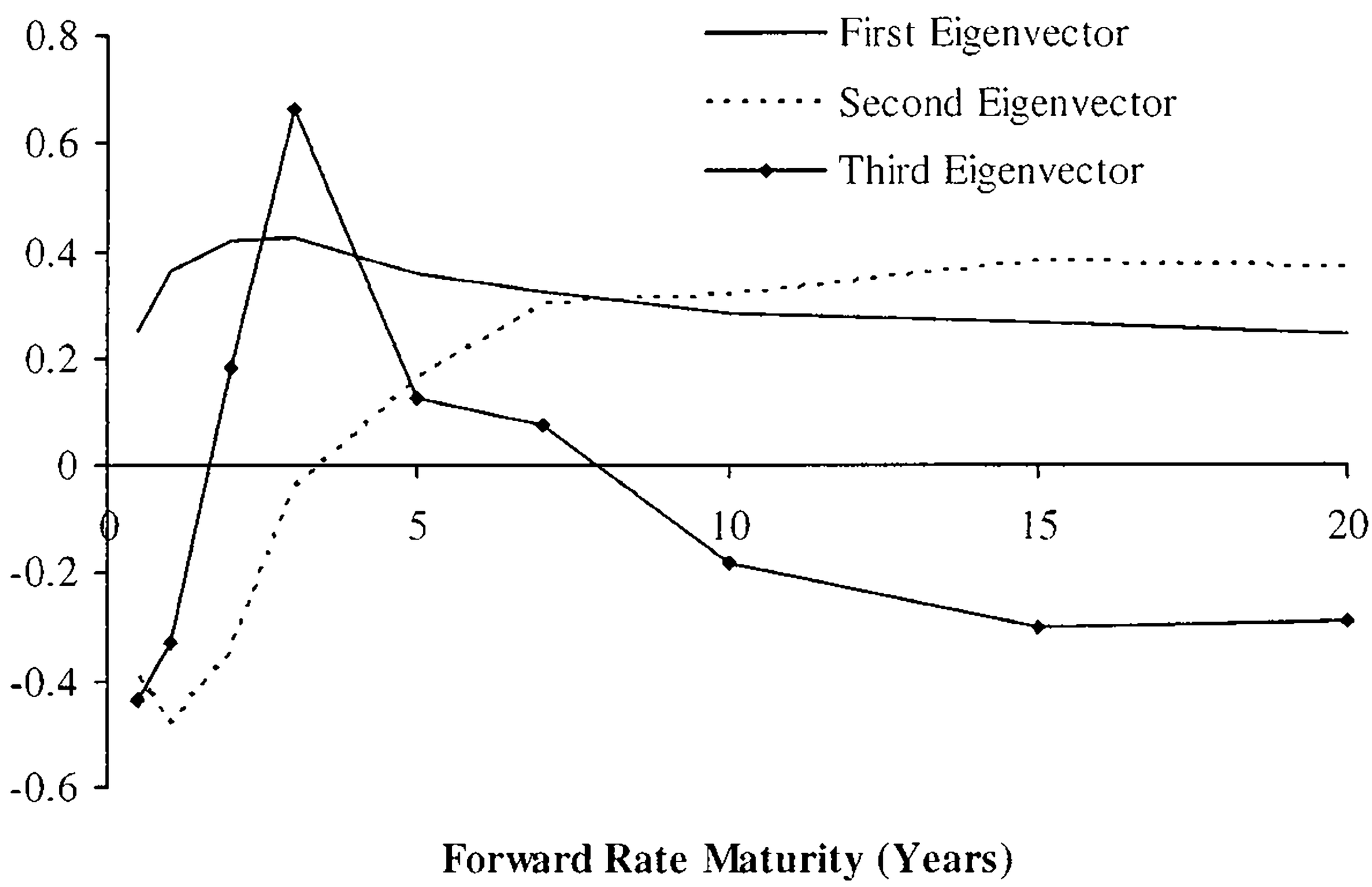
As we see from the table, the first principal component explains about 83% of the observed variance, and the first three principal components together

[5]For each date in the time grid t_j we construct the forward curve from market observable swaps, futures, and deposits, using the techniques from Chapter 6.

explain nearly 98%. This pattern carries over to most major currencies, and in many applications we would consequently expect that using $m = 3$ or $m = 4$ Brownian motions in a LM model would adequately capture the empirical covariation of the points on the forward curve. An exception to this rule-of-thumb occurs when a particular derivative security depends strongly on the correlation between forward rates with tenors that are close to each other; in this case, as we shall see in Section 14.3.4, a high number of principal components is required to provide for sufficient decoupling of nearby forward rates.

The eigenvectors corresponding to the largest three eigenvectors in Table 14.2 are shown in the Figure 14.1; the figure gives us a reasonable idea about what the (suitably scaled) first three elements of the $\lambda_k(t)$ vectors should look like as functions of $T_k - t$. Loosely speaking, the first principal component can be interpreted as a near-parallel shift of the forward curve, whereas the second and third principal components correspond to forward curve twists and bends, respectively.

Fig. 14.1. Eigenvectors



Notes: Eigenvectors for the largest three eigenvalues in Table 14.2.

14.3.2 Correlation Estimation and Smoothing

Empirical estimates for forward rate correlations can proceed along the lines of Section 14.3.1. Specifically, if we introduce the diagonal matrix

$$c \triangleq \begin{pmatrix} \sqrt{C_{1,1}} & 0 & \ddots & 0 \\ 0 & \sqrt{C_{2,2}} & \ddots & \ddots \\ \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & 0 & \sqrt{C_{N_x,N_x}} \end{pmatrix},$$

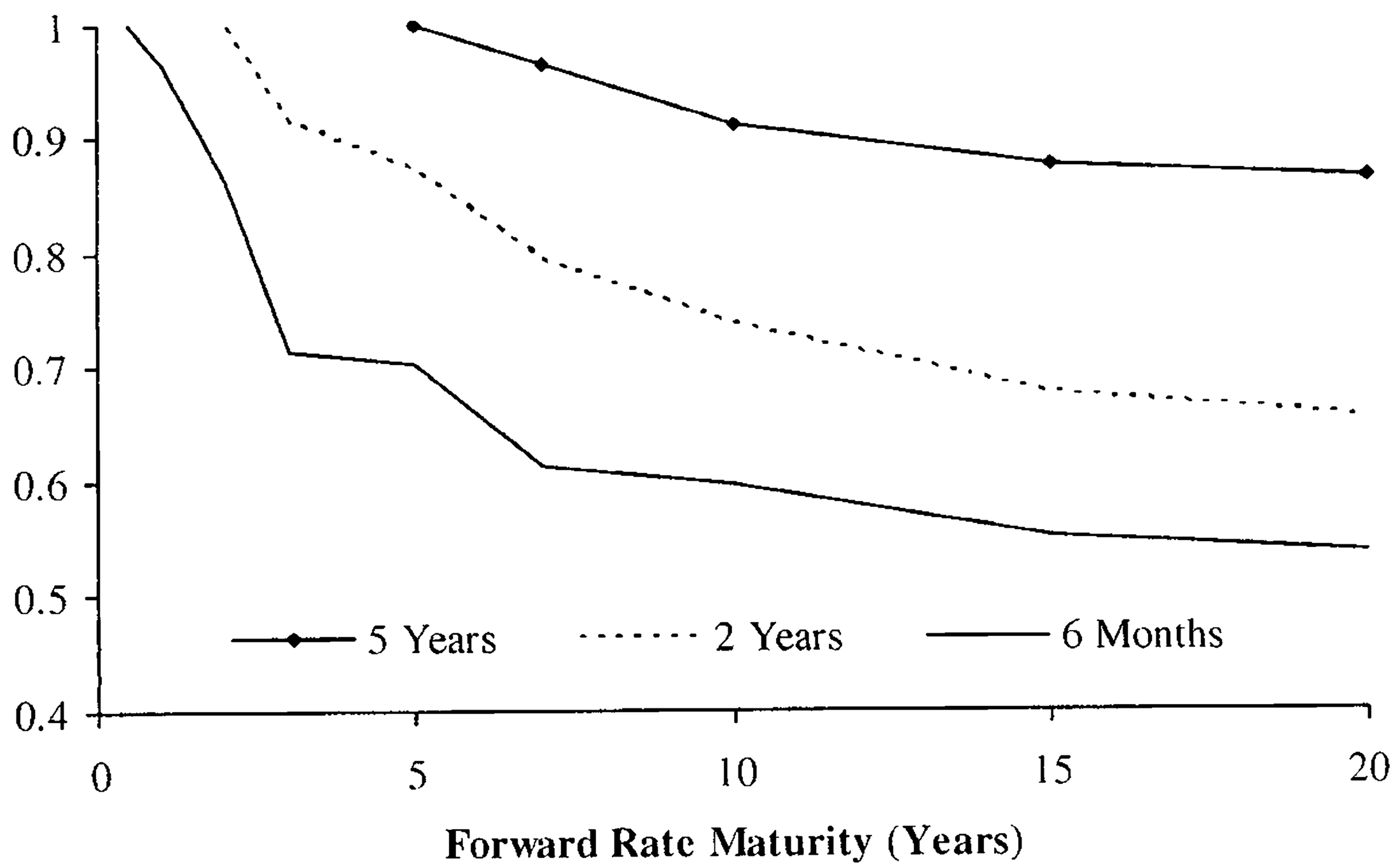
then the empirical $N_x \times N_x$ forward rate correlation matrix R becomes

$$R = c^{-1} C c^{-1}.$$

Element $R_{i,j}$ of R provides a sample estimate of the instantaneous correlation between increments in $l(t, x_i)$ and $l(t, x_i)$, under the assumption that this correlation is time-homogeneous.

The matrix R is often relatively noisy, partially as a reflection of the fact that correlations are well-known to be quite variable over time, and partially as a reflection of the fact that the empirical correlation estimator has rather poor sample properties with large confidence bounds (see Johnson et al. [1995] for details). Nevertheless, several stylistic facts can be gleaned from the data, as demonstrated in Figure 14.2 where we have graphed a few slices of the correlation matrix for the USD data in Section 14.3.1.1.

Fig. 14.2. Forward Rate Correlations



Notes: For each of three fixed forward rate maturities (6 months, 2 years, and 5 years), the figure shows the correlation between the fixed forward rate and forward rates with other maturities (as indicated on the x-axis of the graph).

To make a few qualitative observations about Figure 14.2, we notice that correlations between forward rates $l(\cdot, x_k)$ and $l(\cdot, x_j)$ generally decline in $|x_k - x_j|$; this decline appears near-exponential for x_k and x_j close to each other, but with a near-flat asymptote for large $|x_k - x_j|$. It appears that the rate of the correlation decay and the level of the asymptote depend not only on $|x_k - x_j|$, but also on $\min(x_k, x_j)$. Specifically, the decay rate decreases with $\min(x_k, x_j)$, and the asymptote level increases with $\min(x_k, x_j)$.

In practice, unaltered empirical correlation matrices are typically too noisy for comfort, and might contain non-intuitive entries (e.g., correlation between a 10 year forward and a 2 year forward might come out higher than between a 10 year forward and a 5 year forward). As such, it is common practice in multi-factor yield curve modeling to work with simple parametric forms; this not only smoothes the correlation matrix, but also conveniently reduces the effective parameter dimension of the correlation matrix object, from $N_x(N_x - 1)/2$ distinct matrix elements to the number of parameters in the parametric form.

Several candidate parametric forms for the correlation have been proposed in the literature, see Schoenmakers and Coffey [2000], Jong et al. [2001], and Rebonato [2002], among many others. Rather than list all of these, we instead focus on a few reasonable forms that we have designed to encompass most or all of the empirical facts listed above. Our first parametric form is as follows:

$$\mathrm{Corr}\left(dL_k(t), dL_j(t)\right) = q_1\left(T_k - t, T_j - t\right),$$

where

$$q_1(x, y) = \rho_\infty + (1 - \rho_\infty)\exp\left(-a\left(\min(x, y)\right)|y - x|\right), \tag{14.19}$$
$$a(z) = a_\infty + (a_0 - a_\infty)e^{-\kappa z},$$

subject to $0 \le \rho_\infty \le 1$, $a_0, a_\infty, \kappa \ge 0$. Fundamentally, $q_1(x, y)$ exhibits correlation decay at a rate of a as $|y - x|$ is increased, with the decay rate a itself being an exponential function of $\min(x, y)$. We would always expect to have $a_0 \ge a_\infty$, in which case

$$\frac{\partial q_1(x, y)}{\partial x} = (1 - \rho_\infty)e^{-a(x)(y-x)}\left[a(x) + (y - x)\kappa(a_0 - a_\infty)e^{-\kappa x}\right], \quad x < y,$$

is non-negative, as one would expect.

Variations on (14.19) are abundant in the literature — the case $a_0 = a_\infty$ is particularly popular — and q_1 generally has sufficient degrees of freedom to provide a reasonable fit to empirical data. One immediate issue, however, is a lack of control of the asymptotic correlation level at $|x - y| \to \infty$ which, as we argued above, is typically not independent of x and y. As the empirical data suggests that ρ_∞ tends to increase with $\min(x, y)$, we could introduce yet another decaying function

$$\rho_\infty(z) = b_\infty + (b_0 - b_\infty)e^{-\alpha z}, \tag{14.20}$$

and extend q_1 to the "triple-decaying" form

$$q_2(x,y) = \rho_\infty(\min(x,y)) + (1-\rho_\infty(\min(x,y)))\exp(-a(\min(x,$$

with $a(z)$ given in (14.19), and where $0 \le b_0, b_\infty \le 1, \alpha \ge 0$. Empirical data suggests that normally $b_0 \le b_\infty$, in which case we have

$$\begin{aligned}\frac{\partial q_2(x,y)}{\partial x} &= -\alpha(b_0 - b_\infty)e^{-\alpha x}\left(1 - e^{-a(x)(y-x)}\right)\\ &+ (1-\rho_\infty(x))\, e^{-a(x)(y-x)}\left[a(x) + (y-x)\kappa(a_0 - a_\infty)e^{-\kappa x}\right], \quad x <\end{aligned}$$

which remains non-negative if $b_0 \le b_\infty$ and $a_0 \ge a_\infty$.

In a typical application, the four parameters of q_1 and the six parameters of q_2 are found by least-squares optimization against an empirical correlation matrix. Any standard optimization algorithm, such as the Levenberg-Marquardt algorithm in Press et al. [1992], can be used for this purpose. Some parameters are here subject to simple box-style constraints (e.g. $\rho_\infty \in [0,1]$), which poses no particular problems for most commercial optimizers. In any case, we can always use functional mappings to rewrite our optimization problem in terms of variables with unbounded domains. For instance, for the form q_1, we can set

$$\rho_\infty = \frac{1}{2} + \frac{\arctan(u)}{\pi}, \quad u \in (-\infty, \infty),$$

and optimize on the variable u instead of ρ_∞. Note that we sometimes may wish to optimize correlation parameters against more market-driven targets than empirical correlation matrices, an idea that we shall investigate further in Section 14.5.9.

14.3.2.1 Example: Fit to USD Data

Let R be the 9×9 empirical correlation matrix generated from the data in Section 14.3.1.1, and let $R_2(\xi)$, $\xi \triangleq (a_0, a_\infty, \kappa, b_0, b_\infty, \alpha)^\top$, be the 9×9 correlation matrix generated from the form q_2, when using the 9 specific forward tenors in 14.3.1.1. To determine the optimal parameter vector ξ^*, we minimize an unweighted Frobenius (least-squares) matrix norm, subject to a non-negativity constraint

$$\xi^* = \underset{\xi}{\operatorname{argmin}}\left(\operatorname{tr}\left((R - R_2(\xi))\,(R - R_2(\xi))^\top\right)\right), \text{ subject to } \xi \ge 0.$$

The resulting fit is summarized in Table 14.3; Figure 14.3 in Section 14.3.4.1 contains a 3D plot of the correlation matrix $R_2(\xi^*)$.

The value of the Frobenius norm at ξ^* is 0.070, which translates into an average absolute correlation error (excluding diagonal elements) of around

a_0	a_∞	κ	b_0	b_∞	α
0.312	0.157	0.264	0.490	0.946	0.325

Table 14.3. Best-Fit Parameters for q_2 in USD Market

2%. If we use the four parameter form q_1 instead of q_2 in the optimization exercise, the Frobenius norm at the optimum increases to 0.164. As we would expect from Figure 14.2, allowing correlation asymptotes to increase in tenors thus adds significant explanatory power to the parametric form.

14.3.3 Negative Eigenvalues

While some functional forms are designed to always return valid correlation matrices (the function in Schoenmakers and Coffey [2000] being one such example), many popular forms — including our q_1 and q_2 above — can, when stressed, generate matrices R that fail to be positive definite. While this rarely happens in real applications, it is not inconceivable that on occasion one or more eigenvalues of R may turn out to be negative, requiring us to somehow "repair" the matrix. A similar problem can also arise due to rounding errors when working with large empirical correlation matrices.

Formally, when faced with an R matrix that is not positive definite, we would ideally like to replace it with a modified matrix R^* which i) is a valid correlation matrix; and ii) is as close as possible to R, in the sense of some matrix norm. The problem of locating R^* then involves computing the norm

$$\{\|R - X\| : X \text{ is a correlation matrix}\}$$

and setting R^* equal to the matrix X that minimizes this distance. If $\|\cdot\|$ is a weighted Frobenius norm, good numerical algorithms for the computation of R^* have recently emerged, see Higham [2002] for a review and a clean approach.

If the negative eigenvalues are small in absolute magnitude (which is often the case in practice), it is often reasonable to abandon a full-blown optimization algorithm in favor of a more heuristic approach where we simply raise all offending negative eigenvalues to some positive cut-off value. To present one obvious algorithm, let us start by writing

$$R = E\Lambda E^\top,$$

where Λ is a diagonal matrix of eigenvalues, and E is a matrix with the eigenvectors of R in its columns. Let Λ^* be the diagonal matrix with all-positive entries

$$\Lambda^*_{i,i} = \max(\epsilon, \Lambda_{i,i}), \quad i = 1, \ldots, N_x,$$

for some small cut-off value $\epsilon > 0$. Then set

$$C^* = E\Lambda^* E^\top,$$

which we interpret as a *covariance* matrix, i.e. of the form

$$C^* = c^* R^* c^*,$$

where c^* is a diagonal matrix with elements $c^*_{i,i} = \sqrt{C^*_{i,i}}$ and R^* is the valid, positive definite correlation matrix we seek. R^* is then computed as

$$R^* = (c^*)^{-1} C^* (c^*)^{-1}. \tag{14.22}$$

We emphasize that R^* as defined in (14.22) will have 1's in its diagonal, whereas C^* will not. Both C^* and R^* are, by construction, positive definite.

14.3.4 Correlation PCA

We now turn to a problem that arises in certain important applications, such as the calibration procedure we shall discuss in Section 14.5. Consider a p-dimensional Gaussian variable Y, where all elements of Y have zero mean and unit variance. Let Y have a positive definite correlation matrix R, given by

$$R = \mathrm{E}\left(YY^\top\right).$$

Consider now writing, as an approximation,

$$Y \approx DX, \tag{14.23}$$

where X is an m-dimensional vector of independent standard Gaussian variables, $m < p$, and D is a $(p \times m)$-dimensional matrix. We wish to strictly enforce that DX remains a vector of variables with zero means and unit variances, thereby ensuring that the matrix $DD^\top$ has the interpretation of a valid correlation matrix. In particular, we require that $DD^\top$ has ones on its diagonal.

Let $v(D)$ be the p-dimensional vector of the diagonal elements of $DD^\top$, i.e. $v_i = (DD^\top)_{i,i}$, $i = 1, \ldots, p$. Working as before with an unweighted[6] Frobenius norm, we set

$$h(D; R) = \mathrm{tr}\left(\left(R - DD^\top\right)\left(R - DD^\top\right)^\top\right), \tag{14.24}$$

and define the optimal choice of D, denoted D^*, as

$$D^* = \underset{D}{\mathrm{argmin}}\, h(D; R), \quad \text{subject to } v(D) = \mathbf{1}, \tag{14.25}$$

where $\mathbf{1}$ is a p-dimensional vector of 1's.

[6] The introduction of user-specified weights into this norm is a straightforward extension.

Proposition 14.3.2. *Let μ be a p-dimensional vector, and let D_μ be given as the* unconstrained *optimum*

$$D_\mu = \underset{D}{\operatorname{argmin}}\, h(D; R + \operatorname{diag}(\mu)),$$

with h given in (14.24). Define D^* as in (14.25) and let μ^* be the solution to

$$v(D_\mu) - \mathbf{1} = \mathbf{0}.$$

Then $D^* = D_{\mu^*}$.

Proof. We only provide a sketch of the proof; for more details, see Zhang and Wu [2003]. First, we introduce the Lagrangian

$$\Im\,(D, \mu) = h(D; R) - 2\mu^\top\,(v(D) - \mathbf{1})\,.$$

(The factor 2 on $\mu^\top$ simplifies results.) Standard matrix calculus shows that

$$\frac{dh(D;R)}{dD} = \left\{\frac{dh(D;R)}{dD_{i,j}}\right\} = -4RD + 4DD^\top D.$$

We can use this result to compute the derivative of the Lagrangian with respect to D, which in turn yields the following condition for an optimum

$$-\,(R + \operatorname{diag}(\mu))\,D + DD^\top D = \mathbf{0}, \tag{14.26}$$

where we still must enforce the condition $v(D) = \mathbf{1}$.. Equation (14.26) identifies the optimum as minimizing the (unconstrained) optimization norm $h(D; R + \operatorname{diag}(\mu))$. □

Remark 14.3.3. For any fixed value of μ, D_μ can be computed easily by standard PCA methods provided we interpret $R + \operatorname{diag}(\mu)$ as the target covariance matrix.

With Proposition 14.3.2, determination of D^* is reduced to solving the p-dimensional root-search problem $v(D_\mu) - \mathbf{1} = \mathbf{0}$ for μ. Many standard methods will suffice; for instance, one can use straightforward secant search methods such as the Broyden algorithm on p. 389 of Press et al. [1992].

As is the case for ordinary PCA approximations of covariance matrices, the "correlation PCA" algorithm outlined so far will return a correlation matrix approximation $D^*(D^*)^\top$ that has reduced rank (from p down to m), a consequence of the PCA steps taken in estimating D_μ.

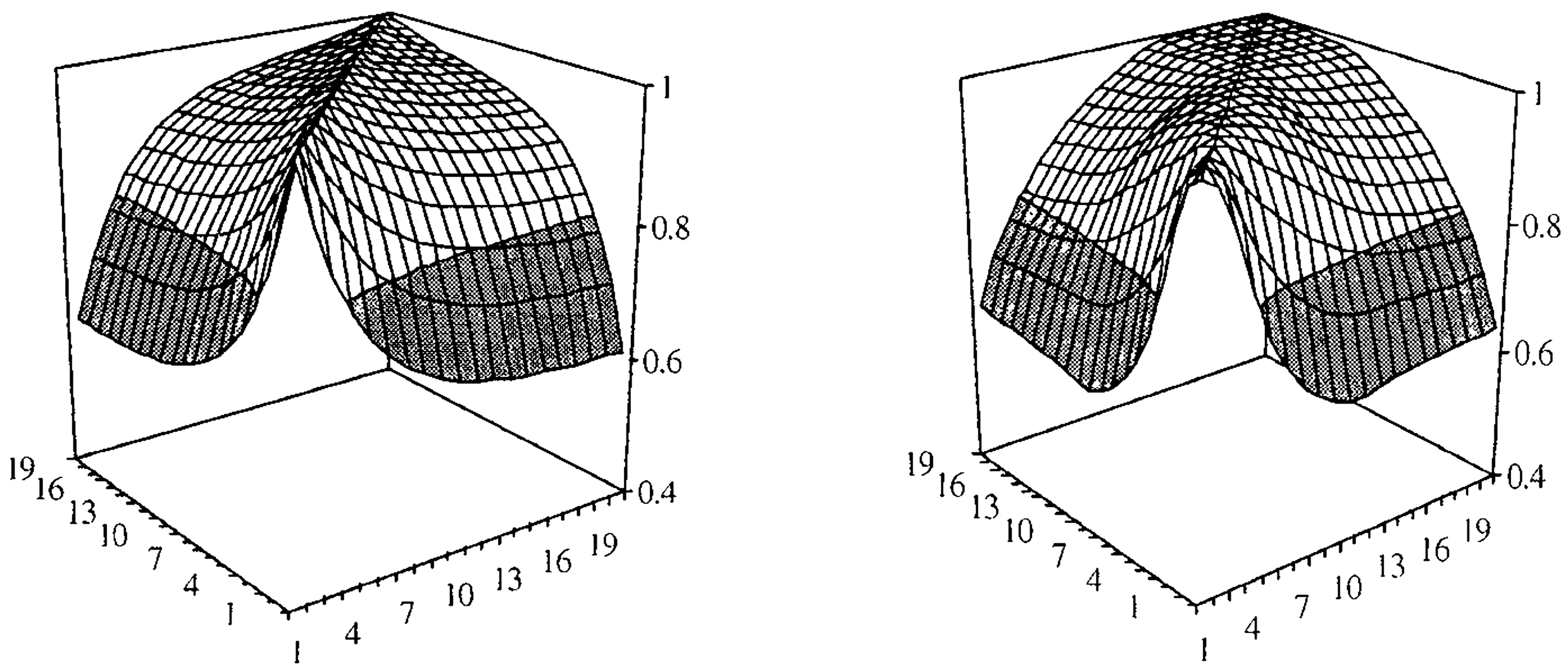
Computation of optimal rank-reduced correlation approximations is a relatively well-understood problem, and we should note the existence of several recent alternatives to the basic algorithm we outlined above. A survey can be found in Pietersz and Groenen [2004] where an algorithm

based on *majorization* is also developed[7]. A recent paper by Li and Qi [2009] not covered in the survey of Pietersz and Groenen [2004] develops an approach based on representing the problem as a non-convex semi-definite programming problem. The authors show how a numerical algorithm that is highly efficient for large scale problems can be constructed. We should also note that certain heuristic (and non-optimal) methods have appeared in the literature, some of which are closely related to the simple algorithm we outlined in Section 14.3.3 for repair of correlation matrices. We briefly outline one such approach below (in Section 14.3.4.2), but first we consider a numerical example.

14.3.4.1 Example: USD Data

We here consider performing a correlation PC analysis on the correlation matrix R generated from our best-fit form q_2 in Section 14.3.2.1. The 3D plots in Figure 14.3 below show the correlation fit we get with a rank-3 correlation matrix.

Fig. 14.3. Forward Rate Correlation Matrix in USD



Notes: The left-hand panel shows the correlation matrix R for form q_2 calibrated to USD data. The right-hand panel shows the best-fitting rank-3 correlation matrix, computed by the algorithm in Proposition 14.3.2. In both graphs, the x- and y-axes represent the Libor forward rate maturities in years.

Looking at Figure 14.3, the effect of rank reduction is, loosely, that the exponential decay of our original matrix R away from the diagonal has been

[7]In our experience, the majorization method in Pietersz and Groenen [2004] is faster than the method in Proposition 14.3.2 but less robust, at least for large and irregular correlation matrices.

replaced with a "sigmoid" shape (to paraphrase Riccardo Rebonato) that is substantially too high close to the matrix diagonal. As the rank of the approximating correlation matrix is increased, the sigmoid shape is — often rather slowly — pulled towards the exponential shape of the full-rank data. Intuitively, we should not be surprised at this result: with the rank m being a low number, we effectively only incorporate smooth, large-scale curve movements (e.g. parallel shifts and twists) into our statistical model, and there is no mechanism to "pull apart" Libor forward rates with maturities close to each other.

Analysis of this difference — rather than the simple PCA considerations of Section 14.3.1 — often forms the basis for deciding how many factors m to use in the model, especially for pricing derivatives with strong correlation dependence. For the reader's guidance, we find that $m = 5$ to 10 suffices to recover the full-rank correlation shape in most cases.

14.3.4.2 Poor Man's Correlation PCA

For the case where the $p \times p$ correlation matrix R is well-represented by a rank-m representation of the form (14.23), it may sometimes be sufficiently accurate to compute the loading matrix D by a simpler algorithm based on *standard* PCA applied directly to the correlation matrix. Specifically, suppose that we as a first step compute

$$R_m = E_m \Lambda_m E_m^\top,$$

where Λ_m is an $m \times m$ diagonal matrix of the m largest eigenvalues of R, and E_m is a $p \times m$ matrix of eigenvectors corresponding to these eigenvalues. While the error $R_m - R$ minimizes a least-squares norm, R_m itself is obviously not a valid approximation to the correlation matrix R as no steps were taken to ensure that R_m has a unit diagonal. A simple way to accomplish this borrows the ideas of Section 14.3.3 and writes

$$R \approx r_m^{-1} R_m r_m^{-1}, \tag{14.27}$$

where r_m is a diagonal matrix with elements $(r_m)_{i,i} = \sqrt{(R_m)_{i,i}}$, $i = 1, \ldots, p$. We note that this approximation sets the matrix D in (14.23) to

$$D = r_m^{-1} E_m \sqrt{\Lambda_m}.$$

It is clear that the difference between the "poor man's" PCA result (14.27) and the optimal result in Proposition 14.3.2 will generally be small if R_m is close to having a unit diagonal, as the heuristic step taken in (14.27) will then have little effect. For large, complex correlation matrices, however, the optimal approximation in Proposition 14.3.2 will often be quite different from (14.27) unless m is quite large.

14.4 Pricing of European Options

The previous section laid the foundation for calibrating an LM model to empirical forward curve correlation data, a topic that we shall return to in more detail in Section 14.5. Besides correlation calibration, however, we need to ensure that the forward rate *variances* implied by the LM model are in line with market data. In most applications — and certainly in all those that involve pricing and hedging of traded derivatives — this translates into a requirement that the vectors $\lambda_k(t)$ of the model are such that it will successfully reproduce the prices of liquid plain-vanilla derivatives, i.e. swaptions and caps. A condition for practical uses of the LM model is thus that we can find pricing formulas for vanilla options that are fast enough to be embedded into an iterative calibration algorithm.

14.4.1 Caplets

Deriving formulas for caplets is generally straightforward in the LM model, a consequence of the fact that Libor rates — which figure directly in the payout formulas for caps — are the main primitives of the LM model itself. Indeed, the word "market" in the term "Libor market model" originates from the ease with which the model can accommodate market-pricing of caplets by the Black formula.

As our starting point here, we use the generalized version of the LM model with skews and stochastic volatility; see (14.15) and (14.16). Other, simpler models, are special cases of this framework, and the fundamental caplet pricing methodology will carry over to these cases in a transparent manner. We consider the price of a c-strike caplet $V_{\text{caplet}}(\cdot)$ maturing at time T_n and settling at time T_{n+1}. That is,

$$V_{\text{caplet}}\left(T_{n+1}\right) = \tau_n \left(L_n(T_n) - c\right)^+ .$$

For the purpose of pricing the caplet, the m-dimensional Brownian motion $W^{n+1}(t)$ can here be reduced to one dimension, as shown in the following result.

Proposition 14.4.1. *Assume that the forward rate dynamics in the spot measure are as in (14.15)–(14.16), and that Assumption 14.2.7 holds. Then*

$$V_{\text{caplet}}(0) = P(0, T_{n+1})\tau_n \mathrm{E}^{T_{n+1}}\left(\left(L_n(T_n) - c\right)^+\right),$$

where

$$\begin{aligned} dL_n(t) &= \sqrt{z(t)}\varphi\left(L_n(t)\right) \|\lambda_n(t)\| \, dY^{n+1}(t), \\ dz(t) &= \theta\left(z_0 - z(t)\right) dt + \eta\psi\left(z(t)\right) dZ(t), \end{aligned} \tag{14.28}$$

and $Y^{n+1}(t)$ *and* $Z(t)$ *are independent* scalar *Brownian motions in measure* $Q^{T_{n+1}}$. *Specifically,* $Y^{n+1}(t)$ *is given by*

$$Y^{n+1}(t) = \int_0^t \frac{\lambda_n(s)^\top}{\|\lambda_n(s)\|} dW^{n+1}(s).$$

Proof. $Y^{n+1}(t)$ is clearly Gaussian, with mean 0 and variance $\sqrt{t}$, identifying $Y^{n+1}(t)$ as a Brownian motion such that $\|\lambda_n(t)\| dY^{n+1}(t) = \lambda_n(t)^\top dW^{n+1}(t)$. The remainder of the proposition follows from the martingale property of $L_n(t)$ in $Q^{T_{n+1}}$, combined with the assumed independence of the forward rates and the process for $z(t)$. □

While rather obvious, Proposition 14.4.1 is useful as it demonstrates that caplet pricing reduces to evaluation of an expectation of $(L_n(T_n) - c)^+$, where the process for $L_n(t)$ is now identical to the types of scalar stochastic volatility diffusions covered in detail in Chapters 8 and 9; the pricing of caplets can therefore be accomplished with the formulas listed in these chapters. In the same way, when dealing with LM models of the simpler local volatility type, we compute caplet prices directly from formulas in Chapter 7.

14.4.2 Swaptions

Whereas pricing of caplets is, by design, convenient in LM models, swaption pricing requires a bit more work and generally will involve some amount of approximation if a quick algorithm is required. In this section, we will outline one such approximation which normally has sufficient accuracy for calibration applications. A more accurate (but also more complicated) approach can be found in Section 15.2.

First, let us recall some notations. Let $V_{\text{swaption}}(t)$ denote the time t value of a payer swaption that matures at time $T_j \geq t$, with the underlying security being a fixed-for-floating swap making payments at times $T_{j+1}, \ldots, T_k$, where $j < k \leq N$. We define an annuity factor for this swap as (see (4.8))

$$A(t) \triangleq A_{j,k-j}(t) = \sum_{n=j}^{k-1} P(t, T_{n+1})\tau_n, \quad \tau_n = T_{n+1} - T_n. \tag{14.29}$$

Assuming that the swap underlying the swaption pays a fixed coupon of c against Libor, the payout of V_{swaption} at time T_j is (see Section 4.1.3)

$$V_{\text{swaption}}(T_j) = A(T_j)\,(S(T_j) - c)^+,$$

where we have defined a par forward swap rate (see (4.10))

$$S(t) \triangleq S_{j,k-j}(t) = \frac{P(t, T_j) - P(t, T_k)}{A(t)}.$$

Assume, as in Section 14.4.1, that we are working in the setting of a stochastic volatility LM model, of the type defined in Section 14.2.5; the procedure we shall now outline will carry over to simpler models unchanged.

Proposition 14.4.2. *Assume that the forward rate dynamics in the spot measure are as in (14.15)–(14.16). Let* Q^A *be the measure induced by using* $A(t)$ *in (14.29) as a numeraire, and let* $W^A(t)$ *be an* m*-dimensional Brownian motion in* Q^A*. Then, in measure* Q^A*,*

$$dS(t) = \sqrt{z(t)}\varphi\left(S(t)\right)\sum_{n=j}^{k-1} w_n(t)\lambda_n(t)^\top\, dW^A(t), \tag{14.30}$$

where the stochastic weights are

$$\begin{aligned} w_n(t) &= \frac{\varphi\left(L_n(t)\right)}{\varphi\left(S(t)\right)} \times \frac{\partial S(t)}{\partial L_n(t)} = \frac{\varphi\left(L_n(t)\right)}{\varphi\left(S(t)\right)} \times \frac{S(t)\tau_n}{1+\tau_n L_n(t)} \\ &\quad \times \left[\frac{P(t,T_k)}{P(t,T_j)-P(t,T_k)} + \frac{\sum_{i=n}^{k-1}\tau_i P(t,T_{i+1})}{A(t)}\right]. \end{aligned} \tag{14.31}$$

Proof. It follows from Lemma 4.2.4 that $S(t)$ is a martingale in measure Q^A, hence we know that the drift of the process for $S(t)$ must be zero in this measure. From its definition, $S(t)$ is a function of $L_j(t), L_{j+1}(t), \ldots, L_{k-1}(t)$, and an application of Ito's lemma shows that

$$dS(t) = \sum_{n=j}^{k-1}\sqrt{z(t)}\varphi\left(L_n(t)\right)\frac{\partial S(t)}{\partial L_n(t)}\lambda_n(t)^\top dW^A(t).$$

Evaluating the partial derivative proves the proposition. □

It should be immediately obvious that the dynamics of the par rate in (14.30) are too complicated to allow for analytical treatment. The main culprit are the random weights $w_n(t)$ in (14.31) which depend on the entire forward curve in a complex manner. All is not lost, however, as one would intuitively expect that $S(t)$ is well-approximated by a weighted sum of its "component" forward rates $L_j(t), L_{j+1}(t), \ldots, L_{k-1}(t)$, with weights varying little over time. In other words, we expect that, for each n, $\partial S(t)/\partial L_n(t)$ is a near-constant quantity.

Consider now the ratio $\varphi(L_n(t))/\varphi(S(t))$ which multiplies $\partial S(t)/\partial L_n(t)$ in (14.31). For forward curves that are reasonably flat and forward curve movements that are predominantly parallel (which is consistent with our earlier discussion in Section 14.3.1.1), it is often reasonable to assume that the ratio is close to constant. This assumption obviously hinges on the precise form of φ, but tends to hold well for many of the functions that we would consider using in practice. To provide some loose motivation for this statement, consider first the extreme case where $\varphi(x) = \text{const}$ (i.e. the

model is Gaussian) in which case the ratio $\varphi(L_n(t))/\varphi(S(t))$ is constant, by definition. Second, let us consider the log-normal case where $\varphi(x) = x$. In this case, a parallel shift h of the forward curve at time t would move the ratio to

$$\frac{L_n(t)+h}{S(t)+h} = \frac{L_n(t)}{S(t)} + h\frac{S(t)-L_n(t)}{S(t)^2} + O(h^2),$$

which is small if the forward curve slope (and thereby $S(t) - L_n(t)$) is small. As the φ's that we use in practical applications are mostly meant to produce skews that lie somewhere between log-normal and Gaussian ones, assuming that $\varphi(L_n(t))/\varphi(S(t))$ is constant thus appears reasonable.

The discussion above leads to the following approximation, where we "freeze" the weights $w_n(t)$ at their time 0 values.

Proposition 14.4.3. *The time 0 price of the swaption is given by*

$$V_{\text{swaption}}(0) = A(0)\mathrm{E}^A\left((S(T_j)-c)^+\right). \tag{14.32}$$

Let $w_n(t)$ be as in Proposition 14.4.2 and set

$$\lambda_S(t) = \sum_{n=j}^{k-1} w_n(0)\lambda_n(t),$$

The swap rate dynamics in Proposition 14.4.2 can be approximated as

$$\begin{aligned} dS(t) &\approx \sqrt{z(t)}\varphi\left(S(t)\right)\|\lambda_S(t)\|\, dY^A(t), \\ dz(t) &= \theta\left(z_0 - z(t)\right)dt + \eta\psi\left(z(t)\right)dZ(t), \end{aligned} \tag{14.33}$$

where $Y^A(t)$ and $Z(t)$ are independent scalar *Brownian motions in measure Q^A, and*

$$\|\lambda_S(t)\|\, dY^A(t) = \sum_{n=j}^{k-1} w_n(0)\lambda_n(t)^\top dW^A(t).$$

Proof. Equation (14.32) follows from standard properties of Q^A. The remainder of the proposition is proven the same way as Proposition 14.4.1, after approximating $w_n(t) \approx w_n(0)$. □

We emphasize that the scalar term $\|\lambda_S(t)\|$ is purely deterministic, whereby the dynamics of $S(t)$ in the annuity measure have precisely the same form as the Libor rate SDE in Proposition 14.4.1. Therefore, computation of the Q^A-expectation in (14.32) can lean directly on the analytical results we established for scalar stochastic volatility processes in Chapter 8 and, for simpler DVF-type LM models, in Chapter 7. We review relevant results and apply them to LM models in Chapter 15; here, to give an example, we merely list a representative result for a displaced log-normal local volatility LM model.

Proposition 14.4.4. *Let each rate $L_n(t)$ follow a displaced log-normal process in its own forward measure,*

$$dL_n(t) = (bL_n(t) + (1-b)L_n(0))\lambda_n(t)^\top dW^{n+1}(t), \quad n = 1, \ldots, N-1.$$

Then the time 0 price of the swaption is approximated by

$$V_{\text{swaption}}(0) \approx A(0)c_{\text{B}}(0, S(0)/b; T_j, c - S(0) + S(0)/b; b\overline{\lambda}_S),$$

where $c_{\text{B}}(t, S; T, K; \sigma)$ is the Black call option formula with volatility σ, see Remark 7.2.8, and the term swap rate volatility $\overline{\lambda}_S$ is given by

$$\overline{\lambda}_S = \left(\frac{1}{T_j} \int_0^{T_j} \|\lambda_S(t)\|^2 \, dt \right)^{1/2},$$

with $\lambda_S(t)$ defined in Proposition 14.4.3.

Proof. By Proposition 14.4.3, the approximate dynamics of $S(t)$ are given by

$$dS(t) \approx (bS(t) + (1-b)S(0)) \, \|\lambda_S(t)\| \, dY^A(t).$$

The result then follows from Proposition 7.2.12. □

While we do not document the performance of the approximation (14.33) in detail here, many tests are available in the literature; see e.g. Andersen and Andreasen [2000b], Glasserman and Merener [2001], and Rebonato [2002]. Suffice to say that the approximation above is virtually always accurate enough for the calibration purposes for which it is designed, particularly if we restrict ourselves to pricing swaptions with strikes close to the forward swap rate. As mentioned earlier, should further precision be desired, one can turn to the more sophisticated swaption pricing approximations that we discuss in Chapter 15. Finally, we should note the existence of models where no approximations are required to price swaptions; these so-called *swap market models* are reviewed in Section 15.4.

14.4.3 Spread Options

When calibrating LM models to market data, the standard approach is to fix the correlation structure in the model to match empirical forward rate correlations. It is, however, tempting to consider whether one alternatively could imply the correlation structure directly from traded market data, thereby avoiding the need for "backward-looking" empirical data altogether. As it turns out, the dependence of swaptions and caps on the correlation structure is, not surprisingly, typically too indirect to allow one to simultaneously back out correlations and volatilities from the prices of these types of instruments alone. To overcome this, one can consider amending the set of calibration instruments with securities that have stronger sensitivity to forward rate

correlations. A good choice would here be to use *yield curve spread options*, a type of security that we encountered earlier in Section 5.13.3. Spread options are natural candidates, not only because their prices are highly sensitive to correlation, but also because they are relatively liquid and not too difficult to value in an LM model setting.

14.4.3.1 Term Correlation

Let $S_1(t) = S_{j_1,k_1-j_1}(t)$ and $S_2(t) = S_{j_2,k_2-j_2}(t)$ be two forward swap rates, and assume that we work with a stochastic volatility LM model of type (14.15)–(14.16). Following the result of Proposition 14.4.3, for $i = 1, 2$ we have, to good approximation,

$$dS_i(t) \approx O(dt) + \sqrt{z(t)}\varphi\left(S_i(t)\right)\lambda_{S_i}(t)^\top dW^B(t),$$
$$\lambda_{S_i}(t) \triangleq \sum_{n=j_i}^{k_i-1} w_{S_i,n}(0)\lambda_n(t),$$

where $W^B(t)$ is a vector-valued Brownian motion in the spot measure, and we use an extended notation $w_{S_i,n}$ to emphasize which swap rate a given weight relates to. Notice the presence of drift terms, of order $O(dt)$. The quadratic variation and covariation of $S_1(t)$ and $S_2(t)$ satisfy

$$d\langle S_1(t), S_2(t)\rangle = z(t)\varphi\left(S_1(t)\right)\varphi\left(S_2(t)\right)\lambda_{S_1}(t)^\top\lambda_{S_2}(t)\,dt,$$
$$d\langle S_i(t)\rangle = z(t)\varphi\left(S_i(t)\right)^2\left\|\lambda_{S_i}(t)\right\|^2\,dt, \quad i = 1, 2,$$

and the instantaneous correlation is

$$\mathrm{Corr}(dS_1(t), dS_2(t)) = \frac{\lambda_{S_1}(t)^\top\lambda_{S_2}(t)}{\left\|\lambda_{S_1}(t)\right\|\left\|\lambda_{S_2}(t)\right\|}. \tag{14.34}$$

Instead of the instantaneous correlation, in many applications we are normally more interested in an estimate for *term correlation* $\rho_{\text{term}}(T', T)$ of S_1 and S_2 on some finite interval $[T', T]$. Formally, we define this time 0 measurable quantity as

$$\rho_{\text{term}}(T', T) \triangleq \mathrm{Corr}\left(S_1(T) - S_1(T'), S_2(T) - S_2(T')\right).$$

Ignoring drift terms and freezing the swap rates at their time 0 forward levels, to decent approximation we can write

$$\rho_{\text{term}}\left(T', T\right) \approx \frac{\varphi\left(S_1(0)\right)\varphi\left(S_2(0)\right)\int_{T'}^{T}\mathrm{E}^B\left(z(t)\right)\lambda_{S_1}(t)^\top\lambda_{S_2}(t)\,dt}{\varphi\left(S_1(0)\right)\varphi\left(S_2(0)\right)\prod_{i=1}^{2}\sqrt{\int_{T'}^{T}\mathrm{E}^B\left(z(t)\right)\left\|\lambda_{S_i}(t)\right\|^2\,dt}}$$
$$= \frac{\int_{T'}^{T}\lambda_{S_1}(t)^\top\lambda_{S_2}(t)\,dt}{\sqrt{\int_{T'}^{T}\left\|\lambda_{S_1}(t)\right\|^2\,dt}\sqrt{\int_{T'}^{T}\left\|\lambda_{S_2}(t)\right\|^2\,dt}}, \tag{14.35}$$

where in the second equality we have used that fact that the parameterization (14.15) implies that, for all $t \geq 0$,

$$\mathrm{E}^B\left(z(t)\right) = z_0.$$

14.4.3.2 Spread Option Pricing

Consider a spread option paying at time $T \leq \min(T_{j_1}, T_{j_2})$

$$V_{\text{spread}}(T) = \left(S_1(T) - S_2(T) - K\right)^+,$$

such that

$$V_{\text{spread}}(0) = P(0,T)\mathrm{E}^T\left(\left(S_1(T) - S_2(T) - K\right)^+\right),$$

where, as always, E^T denotes expectations in measure Q^T. An accurate (analytic) evaluation of this expected value is somewhat involved, and we postpone it until Chapter 17. Here, as a preview, we consider a cruder approach which may, in fact, be adequate for calibration purposes. We assume that the spread

$$\varepsilon(T) = S_1(T) - S_2(T)$$

is a Gaussian variable with mean

$$\mathrm{E}^T\left(\varepsilon(T)\right) = \mathrm{E}^T\left(S_1(T)\right) - \mathrm{E}^T\left(S_2(T)\right).$$

In a pinch, the mean of $\varepsilon(T)$ can be approximated as $S_1(0) - S_2(0)$, which assumes that the drift terms of $S_1(t)$ and $S_2(t)$ in the T-forward measure are approximately identical. For a better approximation, see Chapter 16. As for the variance of $\varepsilon(T)$, it can be approximated in several different ways, but one approach simply writes

$$\begin{aligned}\mathrm{Var}^T\left(\varepsilon(T)\right) \approx & \sum_{i=1}^{2} \varphi\left(S_i(0)\right)^2 z_0 \int_0^T \left\|\lambda_{S_i}(t)\right\|^2 dt \\ & - 2\rho_{\text{term}}(0,T) z_0 \prod_{i=1}^{2} \varphi\left(S_i(0)\right) \left(\int_0^T \left\|\lambda_{S_i}(t)\right\|^2 dt\right)^{1/2}. \end{aligned} \tag{14.36}$$

With these approximations, the Bachelier formula (7.16) yields

$$V_{\text{spread}}(0) = P(0,T)\sqrt{\mathrm{Var}^T\left(\varepsilon(T)\right)}\left(d\Phi(d) + \phi(d)\right), \quad d = \frac{\mathrm{E}^T\left(\varepsilon(T)\right) - K}{\sqrt{\mathrm{Var}^T\left(\varepsilon(T)\right)}}. \tag{14.37}$$

14.5 Calibration

14.5.1 Basic Principles

Suppose that we have fixed the tenor structure, have decided upon the number of factors m to be used, and have selected the basic form (e.g. DVF or SV) of the LM model that we are interested in deploying. Suppose also, for now, that any skew functions and stochastic volatility dynamics have been exogenously specified by the user. To complete our model specification, what then remains unanswered is the fundamental question of how to establish the set of m-dimensional deterministic volatility vectors $\lambda_k(\cdot)$, $k = 1, 2, \ldots, N-1$, that together determine the overall correlation and volatility structure of forward rates in the model.

As evidenced by the large number of different calibration approaches proposed in the literature, there are no precise rules for calibration of LM models. Still, certain common steps are nearly always invoked:

- Prescribe the basic form of $\|\lambda_k(t)\|$, either through direct parametric assumptions, or by introduction of discrete time- and tenor-grids.
- Use correlation information to establish a map from $\|\lambda_k(t)\|$ to $\lambda_k(t)$.
- Choose the set of observable securities against which to calibrate the model.
- Establish the norm to be used for calibration.
- Recover $\lambda_k(t)$ by norm optimization.

In the next few sections, we will discuss these steps in sequence. In doing so, our primary aim is to expose a particular calibration methodology that we personally prefer for most applications, rather than give equal mention to all possible approaches that have appeared in the literature. We note up front that our discussion is tilted towards applications that ultimately involve pricing and hedging of exotic Libor securities (see e.g. Chapters 18 and 19).

14.5.2 Parameterization of $\|\lambda_k(t)\|$

For convenience, let us write

$$\lambda_k(t) = h(t, T_k - t), \quad \|\lambda_k(t)\| = g\,(t, T_k - t)\,, \tag{14.38}$$

for some functions $h : \mathbb{R}_+^2 \to \mathbb{R}^m$ and $g : \mathbb{R}_+^2 \to \mathbb{R}_+$ to be determined. We focus on g in this section, and start by noting that many ad-hoc parametric forms for this function have been proposed in the literature. A representative example is the following 4-parameter specification, due to Rebonato [1998]:

$$g\,(t, x) = g(x) = (a + bx)e^{-cx} + d, \quad a, b, c, d \in \mathbb{R}_+. \tag{14.39}$$

We notice that this specification is *time-stationary* in the sense that $\|\lambda_k(t)\|$ does not depend on calendar time t, but only on the remaining time to maturity $(T_k - t)$ of the forward rate in question. While attractive from a perspective of smoothness of model volatilities, assumptions of perfect time stationarity will generally not allow for a sufficiently accurate fit to market prices. To address this, some authors have proposed "separable" extensions of the type

$$g(t, x) = g_1(t)g_2(x), \tag{14.40}$$

where g_1 and g_2 are to be specified separately. See Brace et al. [1997] for an early approach along these lines.

For the applications we have in mind, relying on separability or parametric forms is ultimately too inflexible, and we seek a more general approach. For this, let us introduce a rectangular grid of times and tenors $\{t_i\} \times \{x_j\}$, $i = 1, \ldots, N_t$, $j = 1, \ldots, N_x$; and an $(N_t \times N_x)$-dimensional matrix G. The elements $G_{i,j}$ will be interpreted as

$$g(t_i, x_j) = G_{i,j}. \tag{14.41}$$

When dimensioning the grid $\{t_i\} \times \{x_j\}$, we would normally[8] require that $t_1 + x_{N_x} \geq T_N$, to ensure that all forward rates on the Libor forward curve are covered by the table; beyond this, there need not be any particular relationship between the grid and the chosen tenor structure, although we find it convenient to ensure that $t_i + x_j \in \{T_n\}$ as long as $t_i + x_j \leq T_N$ — a convention we adopt from now on. Note that the bottom right-hand corner of the grid contains Libor maturities beyond that of our tenor structure and is effectively redundant.

A few further comments on the grid-based approach above are in order. First, we notice that both time-stationary and separable specifications along the lines of (14.39) and (14.40) can be emulated closely as special cases of the grid-based approach. For instance, the parametric specification (14.39) would give rise to a matrix G where

$$G_{i,j} = (a + bx_j)e^{-cx_j} + d,$$

i.e. all rows would be perfectly identical. We also point out that free parameters to be determined here equate all non-superfluous elements in G. In practice N_t and N_x would often both be around 10–15, so even after accounting for the fact that the bottom-right corner of G is redundant, the total number of free parameters to be determined is potentially quite large. To avoid overfitting, additional regularity conditions must be imposed — an important point to which we return in Section 14.5.6.

[8] An alternative would be to rely on extrapolation.

14.5.3 Interpolation on the Whole Grid

Suppose that we have somehow managed to construct the matrix G in (14.41), i.e. we have uncovered $\|\lambda_k(t)\| = g(t, T_k - t)$ for the values of t and $x = T_k - t$ on the grid $\{t_i\} \times \{x_j\}$. The next step is to construct $\|\lambda_k(t)\|$ for *all* values of t and k, $k = 1, \ldots, N-1$.

It is common[9] to assume that for each k, the function $\|\lambda_k(t)\|$ is piecewise constant in t, with discontinuities at T_n, $n = 1, \ldots, k-1$,

$$\|\lambda_k(t)\| = \sum_{n=1}^{k} 1_{\{T_{n-1} \le t < T_n\}} \|\lambda_{n,k}\| = \sum_{n=1}^{k} 1_{\{q(t)=n\}} \|\lambda_{n,k}\| . \tag{14.42}$$

In this case, we are left with constructing the matrix $\|\lambda_{n,k}\|$ from G, for all $1 \le n \le k \le N-1$. This is essentially a problem of two-dimensional interpolation (and, perhaps, extrapolation if the $\{t_i\} \times \{x_j\}$ grid does not cover the whole tenor structure). Simple, robust schemes such as separate t- and x-interpolation of low order seem to perform well, whereas high-order interpolation (cubic or beyond) may lead to undesirable effects during risk calculations. In practice, one would normally use either piecewise constant or piecewise linear interpolation.

Suppose, for concreteness, that linear interpolation in both dimensions of G is chosen. Then for each n, k $(1 \le n \le k \le N-1)$ we have the following scheme

$$\|\lambda_{n,k}\| = w_{++} G_{i,j} + w_{+-} G_{i,j-1} + w_{-+} G_{i-1,j} + w_{--} G_{i-1,j-1}, \tag{14.43}$$

where, denoting $\tau_{n,k} = T_k - T_{n-1}$, we have

$$i = \min\{a : t_a \ge T_{n-1}\}, \quad j = \min\{b : x_b \ge \tau_{n,k}\},$$

$$w_{++} = \frac{(T_{n-1} - t_{i-1})(\tau_{n,k} - x_{j-1})}{(t_i - t_{i-1})(x_j - x_{j-1})}, \quad w_{+-} = \frac{(T_{n-1} - t_{i-1})(x_j - \tau_{n,k})}{(t_i - t_{i-1})(x_j - x_{j-1})},$$

$$w_{-+} = \frac{(t_i - T_{n-1})(\tau_{n,k} - x_{j-1})}{(t_i - t_{i-1})(x_j - x_{j-1})}, \quad w_{--} = \frac{(t_i - T_{n-1})(x_j - \tau_{n,k})}{(t_i - t_{i-1})(x_j - x_{j-1})}.$$

Apart from the order of interpolation, we can also choose which type of volatilities we want to interpolate. To explain, let us recall from Chapters 7 and 8 that we often normalize the local volatility function in such a way that $\varphi(L_n(0)) \approx L_n(0)$. Then, $\|\lambda_k(\cdot)\|$'s have the dimensionality of log-normal, or percentage, volatilities, and (14.43) defines interpolation in *log-normal* Libor volatilities. This is not the only choice, and using volatilities that are scaled differently in the interpolation will sometimes lead to smoother and more robust results, as was the case during much of the 2007–2009 financial

[9] A more refined approach, especially for low values of time-to-maturity, is advisable for some applications where the fine structure of short-term volatilities is important. See the discussion in Remark 15.1.1.

crisis. To demonstrate the basic idea, let us fix p, $0 \leq p \leq 1$. Then we can replace (14.43) with

$$L_k(0)^{1-p}\,\|\lambda_{n,k}\| = w_{++}L_{n(i,j)}(0)^{1-p}G_{i,j} + w_{+-}L_{n(i,j-1)}(0)^{1-p}G_{i,j-1}$$
$$+\, w_{-+}L_{n(i-1,j)}(0)^{1-p}G_{i-1,j} + w_{--}L_{n(i-1,j-1)}(0)^{1-p}G_{i-1,j-1}, \quad (14.44)$$

where the indexing function $n(i,j)$ is defined by $T_{n(i,j)} = t_i + x_j$. For $p = 0$, this can be interpreted as interpolation in Gaussian volatilities (see Remark 7.2.9). For arbitrary p, the formula (14.44) specifies interpolation in "CEV" volatilities.

Finally, note that even if we use linear interpolation between the knot points (either in t or x or both), it is normally better to use *constant* extrapolation before the initial t_1 and x_1 and after the final t_{N_t} and x_{N_x}.

14.5.4 Construction of $\lambda_k(t)$ from $\|\lambda_k(t)\|$

Suppose the values of volatility norm $\|\lambda_{n,k}\|$ are known on the full grid $1 \leq n \leq k \leq N-1$. For each T_n, the components of the m-dimensional $\lambda_k(T_n)$ vectors may now be obtained from instantaneous Libor rate volatilities $\|\lambda_{n,k}\|$ for $k \geq n$, and instantaneous correlations of Libor rates fixing on or after T_n. The procedure is similar in spirit to the one we employed previously for parameterizing multi-factor Gaussian short rate models in Section 12.1.7. So, with the calendar time fixed at some value T_n, we introduce an $(N-n) \times (N-n)$ instantaneous correlation matrix $R(T_n)$, with elements

$$(R(T_n))_{i,j} = \mathrm{Corr}\left(dL_i\left(T_n-\right), dL_j\left(T_n-\right)\right), \quad i,j = n,\ldots,N-1.$$

The correlation matrix would, in many applications, be computed from an estimated parametric form, such as those covered in Section 14.3.2. Furthermore, we define a diagonal volatility matrix $c(T_n)$ with elements $\|\lambda_{n,n}\|, \|\lambda_{n,n+1}\|, \ldots, \|\lambda_{n,N-1}\|$ along its diagonal. That is,

$$(c(T_n))_{j,j} = \|\lambda_{n,n+j-1}\|, \quad j = 1,\ldots,N-n,$$

with all other elements set to zero. Given $R(T_n)$ and $c(T_n)$, an instantaneous covariance matrix[10] $C(T_n)$ for forward rates on the grid can now be computed as

$$C(T_n) = c(T_n)R(T_n)c(T_n). \quad (14.45)$$

Let us define $H(T_n)$ to be an $(N-n) \times m$ matrix composed by stacking each dimension of $h(T_n, T_{n+j-1} - T_n)$ (see 14.38) side by side, with j running on the grid:

[10]Earlier results show that the true instantaneous covariance matrix for forward rates may involve DVF- or SV-type scales on the elements of c. For the purposes of calibration of λ_k, we omit these scales.

$$(H(T_n))_{j,i} = h_i\left(T_n, T_{n+j-1} - T_n\right), \quad j = 1, \ldots, N-n, \quad i = 1, \ldots, m.$$

Then, it follows that we should have

$$C(T_n) = H(T_n)H(T_n)^\top. \tag{14.46}$$

Equations (14.45) and (14.46) specify two different representations of the covariance matrix, and we want them to be identical, i.e.

$$H(T_n)H(T_n)^\top = c(T_n)R(T_n)c(T_n), \tag{14.47}$$

which gives us a way to construct the $H(T_n)$ matrix, and thereby the vectors $h(T_n, T_{n+j-1})$ for all values of n, j on the full grid $1 \leq n \leq N-1$, $1 \leq j \leq N-n$. Assuming, as before, piecewise constant interpolation of $\lambda_k(t)$ for t between knot dates $\{T_i\}$, the full set of factor volatilities $\lambda_k(t)$ can be constructed for all t and T_k.

As written, equation (14.47) will normally *not* have a solution as the left-hand side is rank-deficient, whereas the right-hand side will typically have full rank. To get around this, we can proceed to apply PCA methodology, in several different ways. We discuss two methods below, but first quickly note that for n close to N (in particular for $N-n < m$), the equation (14.47) will have *too many* solutions. A pragmatic approach here is to zero out the last few (namely, $m-(N-n)$) columns of the matrix $H(T_n)$ before solving the equation, in effect "forbidding" Brownian motions with high index affecting remaining Libor rates. We trust the reader can fill in the details of this scheme, and will ignore this slight complication going forward.

14.5.4.1 Covariance PCA

In this approach, we apply PCA decomposition to the entire right-hand side of (14.47), writing

$$c(T_n)R(T_n)c(T_n) \approx e_m(T_n)\Lambda_m(T_n)e_m(T_n)^\top,$$

where $\Lambda_m(T_n)$ is an $m \times m$ diagonal matrix of the m largest eigenvalues of $c(T_n)R(T_n)c(T_n)$, and $e_m(T_n)$ is an $(N-n) \times m$ matrix of eigenvectors corresponding to these eigenvalues. Inserting this result into (14.47) leads to

$$H(T_n) = e_m(T_n)\sqrt{\Lambda_m(T_n)}. \tag{14.48}$$

As discussed in Chapter 3, this approximation is optimal in the sense of minimizing the Frobenius norm of the covariance matrix errors.

14.5.4.2 Correlation PCA

An attractive alternative to the approach in Section 14.5.4.1 uses the correlation PCA decomposition discussed in Section 14.3.4. Here we write

$$R(T_n) = D(T_n)D(T_n)^\top, \tag{14.49}$$

for an $(N-n) \times m$ matrix D found by the techniques in Section 14.3.4. Inserting this into (14.47) yields

$$H(T_n) = c(T_n)D(T_n). \tag{14.50}$$

In computing the matrix D, we would normally use the result from Proposition 14.3.2, which would minimize the Frobenius norm on correlation matrix errors.

14.5.4.3 Discussion and Recommendation

Several papers in the literature focus on the method in Section 14.5.4.1 (e.g. Sidenius [2000], and Pedersen [1998]), but we nevertheless strongly prefer the approach in Section 14.5.4.2 for calibration applications. Although performing the PCA decomposition (as in Proposition 14.3.2) of a correlation matrix is technically more difficult than the same operation on a covariance matrix, the correlation PCA is independent of the c matrix and as such will not have to be updated when we update guesses for the G matrix (on which c depends) in a calibration search loop. When the correlation matrix $R(T_n)$ originates from a parametric form independent of calendar time (which we recommend), the matrix D in (14.49) will, in fact, need estimation only once per tenor date[11] T_n, at a minimal computational overhead cost. In comparison, the covariance PCA operation will have to be computed at each T_n every time G gets updated in the calibration loop. We also notice that $D(T_n)D(T_n)^\top$ having a unit diagonal will automatically ensure that the total forward rate volatility will be preserved if m is changed; this is *not* the case for covariance PCA, where the total volatility of forward rates will normally increase as m is increased, *ceteris paribus*.

If the complexity of the optimal PCA algorithm in Proposition 14.3.2 of Section 14.3.4 is deemed too egregious, the simplified approach of Section 14.3.4.2 could be used instead. It shares the performance advantages of the "true" correlation PCA as it only needs to be run once outside the calibration loop, but its theoretical deficiencies suggest that its use should, in most circumstances, be limited to the case where the correlations are themselves calibrated, rather than exogenously specified by the user. We return to the concept of correlation calibration in Section 14.5.9.

14.5.5 Choice of Calibration Instruments

In a standard LM model calibration, we choose a set of swaptions and caps (and perhaps Eurodollar options) with market-observable prices; these prices

[11] Since the matrix D shrinks in T_n, we need to repeat the PCA analysis at each tenor date. Alternatively, but suboptimally, we can do the PCA analysis only once at time 0, pruning the results as needed for other values of T_n.

serve as calibration targets for our model. The problem of determining precisely which caps and swaptions should be included in the calibration is a difficult and contentious one, with several opposing schools of thought represented in the literature. We shall spend this section[12] outlining the major arguments offered in the literature as well as our own opinion on the subject. Before commencing on this, we emphasize that the calibration algorithm we develop in this book accommodates arbitrary sets of calibration instruments and as such will work with any selection philosophy.

One school of thought — the *fully calibrated* or *global* approach — advocates calibrating an LM model to a large set of available interest options, including both caps and swaptions in the calibration set. When using grid-based calibration, this camp would typically recommend using at-the-money swaptions with maturities and tenors chosen to coincide with each point in the grid. That is, if T_s is the maturity of a swaption and T_e is the end date of its underlying swap, then we would let T_s take on all values in the time grid $\{t_i\}$, while at the same time letting $T_e - T_s$ progress through all values[13] of the tenor grid $\{x_j\}$. On top of this, one would often add at-the-money caps at expiries ranging from $T = t_1$ to $T = t_{N_t}$.

The primary advantage of the fully calibrated approach is that a large number of liquid volatility instruments are consistently priced within the model. This, in turn, gives us some confidence that the vanilla option market is appropriately "spanned" and that the calibrated model can be used on a diverse set of exotic securities. In vega hedging (see Section 8.9.1 for definition and Chapter 26 for much more on vega hedging in LM models) of an exotic derivative, one will undoubtedly turn to swaptions and caps, so mispricing these securities in the model would be highly problematic.

Another school of thought — *the parsimonious* or *local* approach — judiciously chooses a small subset of caps and swaptions in the market, and puts significant emphasis on specification of smooth and realistic term structures of forward rate volatilities. Typically this will involve imposing strong time-homogeneity assumptions, or observed statistical relationships, on the $\lambda_k(\cdot)$ vectors. The driving philosophy behind the parsimonious approach (besides the desire for calibration speed) is the observation that, fundamentally, the price of a security in a model is equal to the model-predicted cost of hedging the security over its lifetime. Hedging profits in the future as specified by the model are, in turn, directly related to the forward rate volatility structures that the model predicts for the future. For these model-predicted hedging profits to have any semblance to the actual realized hedging profits, the dynamics of the volatility structure in the model should be a reasonable

[12]We also revisit the subject in the context of callable Libor exotics in Section 18.1.

[13]One would here limit T_s to be no larger than T_N, so the total number of swaptions would be less that $N_t \cdot N_x$. See our discussion of redundant grid entries in Section 14.5.2.

estimate of the actual dynamics. In many cases, however, our best estimate of future volatility structures might be today's volatility structures (or those we have estimated historically), suggesting that the evolution of volatility should be as close to being time-homogeneous as possible. This can be accomplished, for instance, by using time-homogeneous mappings such as (14.39) or similar.

The strong points of the parsimonious approach are, of course, weak ones of the fully calibrated approach. Forward rate volatilities produced by the fully calibrated model can easily exhibit excessively non-stationary behavior, impairing the performance of dynamic hedging. On the other hand, the inevitable mispricings of certain swaptions and/or caps in the parsimonious approach are troublesome. In a pragmatic view of a model as a (sophisticated, hopefully) interpolator that computes prices of complex instruments from prices of simple ones, mispricing of simple instruments obviously does not inspire confidence in the prices returned for complex instruments. As discussed, the parsimonious approach involves an attempt to identify a small enough set of "relevant" swaptions and caps that even a time-homogeneous model with a low number of free parameters can fit reasonably well, but it can often be very hard to judge which swaption and cap volatilities are important for a particular exotic security. In that sense, a fully calibrated model is more universally applicable, as the need to perform trade-specific identification of a calibration set is greatly reduced. Notice also that the risk profile of a given security may change greatly over time as market rates move around, potentially necessitating the use of *different* calibration instruments over time. Changing the calibration instrument set will obviously trigger a discontinuity in the hedge strategy, which is not ideal.

It is easy to imagine taking both approaches to the extremes to generate results that would convincingly demonstrate the perils of using either of them. To avoid such pitfalls we recommend looking for an equilibrium between the two. While we overall favor the fully calibrated approach, it is clear that, at the very least, it should be supplemented by an explicit mechanism to balance price precision versus regularity (e.g. smoothness and time-homogeneity) of the forward rate volatility functions. In addition, one should always perform rigorous checks of the effects of calibration assumptions on pricing and hedging results produced by the model. These checks should cover, at a minimum, result variations due to changes in

- Number of factors used (m).
- Relative importance of recovering all cap/swaption prices vs. time-homogeneity of the resulting volatility structure.
- Correlation assumptions.

A final question deserves a brief mention: should one calibrate to either swaptions or caps, or should one calibrate to both simultaneously? Followers of the parsimonious approach will often argue that there is a persistent

basis between cap and swaption markets, and any attempt to calibrate to both markets simultaneously is bound to distort the model dynamics. Instead, it is argued, one should only calibrate to one of the two markets, based on an analysis of whether the security to be priced is more cap- or swaption-like. Presumably this analysis would involve judging whether either caps or swaptions will provide better vega hedges for the security in question. The drawback of this approach is obvious: many complicated interest rates securities depend on the evolution of both Libor rates as well as swap rates and will simultaneously embed "cap-like" and "swaption-like" features.

To avoid discarding potentially valuable information from either swaption or cap markets, we generally recommend that both markets be considered in the calibration of the LM model. However, we do not necessarily advocate that both types of instruments receive equal weighting in the calibration objective function; rather, the user should be allowed some mechanism to affect the relative importance of the two markets. We return to this idea in the next section.

14.5.6 Calibration Objective Function

As discussed above, several issues should be considered in the choice of a calibration norm, including the smoothness and time-stationarity of the $\lambda_k(\cdot)$ functions; the precision to which the model can replicate the chosen set of calibration instruments; and the relative weighting of caps and swaptions. To formally state a calibration norm that will properly encompass these requirements, assume that we have chosen calibration targets that include N_S swaptions, $V_{\text{swaption},1}$, $V_{\text{swaption},2}$, $\ldots$, V_{swaption,N_S}, and N_C caps, $V_{\text{cap},1}$, $V_{\text{cap},2}$, $\ldots$, V_{cap,N_C}. Strategies for selecting these instruments were discussed in the previous section. We let $\widehat{V}$ denote their quoted market prices and, adopting the grid-based framework from Section 14.5.2, we let $\overline{V}(G)$ denote their model-generated prices as functions of the volatility grid G defined in Section 14.5.2. We introduce a calibration objective function $\mathcal{I}$ as

$$
\begin{aligned}
\mathcal{I}(G) = {} & \frac{w_S}{N_S}\sum_{i=1}^{N_S}\left(\overline{V}_{\text{swaption},i}(G) - \widehat{V}_{\text{swaption},i}\right)^2 \\
& + \frac{w_C}{N_C}\sum_{i=1}^{N_C}\left(\overline{V}_{\text{cap},i}(G) - \widehat{V}_{\text{cap},i}\right)^2 \\
& + \frac{w_{\partial t}}{N_x N_t}\sum_{i=1}^{N_t}\sum_{j=1}^{N_x}\left(\frac{\partial G_{i,j}}{\partial t_i}\right)^2 + \frac{w_{\partial x}}{N_x N_t}\sum_{i=1}^{N_t}\sum_{j=1}^{N_x}\left(\frac{\partial G_{i,j}}{\partial x_j}\right)^2 \\
& + \frac{w_{\partial t^2}}{N_x N_t}\sum_{i=1}^{N_t}\sum_{j=1}^{N_x}\left(\frac{\partial^2 G_{i,j}}{\partial t_i^2}\right)^2 + \frac{w_{\partial x^2}}{N_x N_t}\sum_{i=1}^{N_t}\sum_{j=1}^{N_x}\left(\frac{\partial^2 G_{i,j}}{\partial x_j^2}\right)^2, \qquad (14.51)
\end{aligned}
$$

where $w_S, w_C, w_{\partial t}, w_{\partial x}, w_{\partial t^2}, w_{\partial x^2} \in \mathbb{R}_+$ are exogenously specified weights. In (14.51) the various derivatives of the elements in the table G are, in practice, to be interpreted as discrete difference coefficients on neighboring table elements — see (14.52) below for an example definition[14].

As we have defined it, $\mathcal{I}(G)$ is a weighted sum of i) the mean-squared swaption price error; ii) the mean-squared cap price error; iii) the mean-squared average of the derivatives of G with respect to calendar time; iv) the mean-squared average of the second derivatives of G with respect to calendar time; v) the mean-squared average of the derivatives of G with respect to forward rate tenor; and vi) the mean-squared average of the second derivatives of G with respect to forward rate tenor. The terms in i) and ii) obviously measure how well the model is capable of reproducing the supplied market prices, whereas the remaining four terms are all related to regularity. The term iii) measures the degree of volatility term structure time homogeneity and penalizes volatility functions that vary too much over calendar time. The term iv) measures the smoothness of the calendar time evolution of volatilities and penalizes deviations from linear evolution (a straight line being perfectly smooth). Terms v) and vi) are similar to iii) and iv) and measure constancy and smoothness in the tenor direction. In (14.51), the six weights $w_S, w_C, w_{\partial t}, w_{\partial x}, w_{\partial t^2}, w_{\partial x^2}$ determine the trade-off between volatility smoothness and price accuracy, and are normally to be supplied by the user based on his or her preferences. In typical applications, the most important regularity terms are those scaled by the weights $w_{\partial t}$ and $w_{\partial x^2}$ which together determine the degree of time homogeneity and tenor smoothness in the resulting model.

We should note that there are multiple ways to specify smoothness criteria, with (14.51) being one of many. For example, as we generalized the basic log-normal interpolation scheme (14.43) to allow for interpolation in "CEV" volatilities in (14.44), we can adjust the definition of smoothness to be in terms of compatible quantities. In particular, instead of using

$$\frac{w_{\partial x}}{N_x N_t} \sum_{i=1}^{N_t} \sum_{j=2}^{N_x} \left(\frac{G_{i,j} - G_{i,j-1}}{x_j - x_{j-1}} \right)^2 \tag{14.52}$$

as implicit in (14.51) for the tenor-smoothness term, we could use

$$\frac{w_{\partial x}}{N_x N_t} \sum_{i=1}^{N_t} \sum_{j=2}^{N_x} \left(\frac{L_{n(i,j)}(0)^{1-p} G_{i,j} - L_{n(i,j-1)}(0)^{1-p} G_{i,j-1}}{x_j - x_{j-1}} \right)^2 \tag{14.53}$$

for some p, $0 \leq p \leq 1$. The case of $p = 0$ would then correspond to smoothing basis-point Libor volatilities rather than log-normal Libor volatilities.

As written, the terms of the calibration norm that measure precision in cap and swaption pricing involve mean-squared errors directly on prices.

[14] Depending on how table boundary elements are treated, notice that the range for i and j may not always be as stated in (14.51).

In practice, however, the error function is often applied to some transform of outright prices, e.g. implied volatilities. For an SV-type LM model, for instance, we could institute a pre-processing step where the market price of each swaption $\widehat{V}_{\text{swaption},i}$ would be converted into a constant implied volatility $\widehat{\lambda}_{S_i}$, in such a way that the scalar SDE for the swap rate S_i underlying the swaption $V_{\text{swaption},i}$,

$$dS_i(t) = \sqrt{z(t)}\widehat{\lambda}_{S_i}\varphi\left(S_i(t)\right)dY_i(t),$$

would reproduce the observed swaption market price. Denoting by $\overline{\lambda}_{S_i}(G)$ the corresponding model volatility of the swap rate S_i (as given by, for example, Proposition 14.4.4) and repeating this exercise for all caps and swaptions in the calibration set, we obtain an alternative calibration norm definition where the cap and swaption terms in (14.51) are modified as follows:

$$\mathcal{I}(G) = \frac{w_S}{N_S}\sum_{i=1}^{N_S}\left(\overline{\lambda}_{S_i}(G) - \widehat{\lambda}_{S_i}\right)^2 + \frac{w_C}{N_C}\sum_{i=1}^{N_C}\left(\overline{\lambda}_{C_i}(G) - \widehat{\lambda}_{C_i}\right)^2 + \cdots. \quad (14.54)$$

The advantage of working with implied volatilities in the precision norm is two-fold. First, the relative scaling of individual swaptions and caps is more natural; when working directly with prices, high-value (long-dated) trades would tend to be overweighed relative to low-value (short-dated) trades[15]. Second, in many models computation of the implied volatility terms $\overline{\lambda}_{S_i}$ and $\overline{\lambda}_{C_i}$ can often be done by simple time integration of (combinations of) $\lambda_k(\cdot)$ (see e.g. Proposition 14.4.4) avoiding the need to apply a possibly time-consuming option pricing formula to compute the prices $\overline{V}_{\text{swaption},i}$ and $\overline{V}_{\text{cap},i}$. Considerable attention to this particular issue was paid in Section 9.3 (for SV models) and Section 7.6.2 (for DVF models), and we review relevant results and apply them to LM models in Chapter 15.

The quality-of-fit objective can be expressed in terms of *scaled* volatilities, which sometimes improves performance. Following the ideas developed for interpolation (14.44) and smoothing (14.53), we could express the fit objective as

$$\mathcal{I}(G) = \frac{w_S}{N_S}\sum_{i=1}^{N_S}\left(S_i(0)^{1-p}\left(\overline{\lambda}_{S_i}(G) - \widehat{\lambda}_{S_i}\right)\right)^2 + \cdots,$$

for a given p, $0 \leq p \leq 1$. Taking this idea further we note that a more refined structure of mean-squared weights in the definition of calibration norm is possible. For instance, rather than weighting all swaptions equally with the term w_S, one could use different weights for each swaption in the calibration set. Similarly, by using node-specific weights on the derivatives of the entries in G one may, say, express the view that time homogeneity is more important for large t than for small t.

[15] Another approach to producing more equitable scaling involves using relative (=percentage) price errors, rather than absolute price errors.

14.5.7 Sample Calibration Algorithm

At this point, we are ready to state our full grid-based calibration algorithm. We assume that a tenor structure and a time/tenor grid $\{t_i\} \times \{x_j\}$ have been selected, as have the number of Brownian motions (m), a correlation matrix R, and the set of calibration swaptions and caps. In addition, the user must select the weights in the calibration norm $\mathcal{I}$ in (14.51) or (14.54). Starting from some guess for G, we then run the following iterative algorithm:

1. Given G, interpolate using (14.43) or (14.44) to obtain the full norm volatility grid $\|\lambda_{n,k}\|$ for all Libor indices $k = 1, \ldots, N-1$ and all expiry indices $n = 1, \ldots, k$.
2. For each $n = 1, \ldots, N-1$, compute the matrix $H(T_n)$, and ultimately volatility loadings $\lambda_k(T_n)$, from $\|\lambda_{n,k}\|$, $k \geq n$, by PCA methodology, using either (14.48) or (14.50).
3. Given $\lambda_k(\cdot)$ for all $k = 1, \ldots, N-1$, use the formulas in Sections 14.4.1 and 14.4.2 to compute model prices for all swaptions and caps in the calibration set.
4. Establish the value of $\mathcal{I}(G)$ by direct computation of either (14.51) or (14.54).
5. Update G and repeat Steps 1–4 until $\mathcal{I}(G)$ is minimized.

Step 5 in the above algorithm calls for the use of a robust high-dimensional numerical optimizer. Good results can, in our experience, be achieved with several algorithms, including the Spellucci algorithm[16], the Levenberg-Marquardt algorithm, and the downhill simplex method (the last two can be found in Press et al. [1992]). These, and many alternative algorithms, are available in standard numerical packages, such as IMSL[17] and NAG[18]. On a standard PC, a well-implemented calibration algorithm should generally complete in about 10 seconds from a cold start (i.e. where we do not have a good initial guess for G) for, say, a 40 year model with quarterly Libor rolls.

14.5.8 Speed-Up Through Sub-Problem Splitting

An LM model calibration problem involves a substantial number of free input variables to optimize over, namely all elements of the matrix G. In a typical setup, the number of such variables may range from a few dozen to a few hundred. As the number of terms, or "targets", in the calibration norm is of the same order of magnitude, we are dealing with a fairly sizable optimization problem. While modern optimization algorithms implemented on modern hardware can successfully handle the full-blown problem, it is still

[16] donlp2 SQP/ECQP algorithm, available on www.mathematik.tu-darmstadt.de:8080/ags/ag8/Mitglieder/spellucci_de.html.

[17] www.imsl.com.

[18] www.nag.com.

of interest to examine whether there are ways of to improve computational efficiency. For instance, if we could split the optimization problem into a sequence of smaller sub-problems solved separately and sequentially, the performance of the algorithm would typically improve. Indeed, imagine for illustrative purposes that we have an optimization problem with $m = m_1 m_2$ variables and computational complexity of the order[19] $O(m^2) = O(m_1^2 m_2^2)$. However, if we could find the solution by sequentially solving m_1 problems of m_2 variables each, then the computational cost would be $m_1 O(m_2^2)$, yielding savings of the order $O(m_1)$.

Our ability to split the problem into sub-problems typically relies on exploring its particular structure, i.e. the relationship between input variables and targets. If, for example, target 1 depends on variable 1 but not — or only mildly — on other variables, then it makes sense to find the optimal value for the variable 1 by optimizing for target 1 while keeping other variables constant, and so on. Fortunately, the LM model optimization problem presents good opportunities for this type of analysis. First, recall that the main calibration targets for the problem are the differences in market and model prices (or implied volatilities) of caps and swaptions. Let us consider a swaption with expiry T_j and final payment date T_n; let i be such that $T_j = t_i$. Then, as follows from the swaption approximation formula (14.33), the model volatility for this swaption depends on $\lambda_k(t)$'s for $t \in [0, t_i]$ and for $k = j, \ldots, n-1$. Hence, the part of the calibration norm associated with the price fit of the swaption will depend on the first i rows of the matrix G *only*. This observation suggests splitting the calibration problem into a collection of "row by row" calibration problems.

To simplify notations, we assume that the set of fit targets consists of *all* swaptions with expiries t_i and tenors x_l, $i = 1, \ldots, N_t$, $l = 1, \ldots, N_x$. In a row-by-row calibration algorithm, the first row of the matrix G is calibrated to all N_x swaptions with expiry t_1, then the second row of G is calibrated to the swaptions with expiry t_2, and so on.

As we emphasized earlier, having regularity terms in the calibration norm is important to ensure a smooth solution. Fortunately, regularity terms can generally be organized in the same row-by-row format as the precision terms. For instance, the regularity terms in the tenor direction naturally group into row-specific collections. As for the terms controlling the regularity of the matrix G in calendar time t, when optimizing on time slice t_i, we would only include in the norm the terms that involve rows of G with an index less than or equal to i. We trust that the reader can see how to arrange this, and omit straightforward details.

[19] As many of the algorithms we have in mind compute, at the very least, the sensitivity of each calibration target to each input variable, the computational complexity is at least $O(m^2)$; if the order of complexity is higher, the case for problem splitting is even more compelling.

Computational savings from the row by row scheme could be substantial — for a 40 year model with quarterly Libor rolls, a well-tuned algorithm should converge in less than a second or two. There are, however, certain drawbacks associated with problem splitting. In particular, as the calibration proceeds from one row to the next, the optimizer does not have the flexibility to adjust previous rows of the matrix G to the current row of swaption volatilities. This may result in a tell-tale "ringing" pattern of the Libor volatilities in the time direction, as the optimizer attempts to match each row of price targets through excessively large moves in the elements of G, in alternating up and down directions. Judicious application of regularity terms in the optimization norm can, however, help control this behavior, and overall the row-by-row scheme performs well. We recommend it as the default method for most applications, but note that sometimes a combination of full-blown and row-by-row calibration is the best choice.

Returning to the row-by-row calibration idea, one can try to take it further and split the calibration to an ever-finer level, eventually fitting each individual price target — a given caplet or swaption volatility, say — separately, by moving just a *single* element of the matrix G. This should seemingly work because the (t_i, x_{l+1})-swaption volatility depends on the same elements of matrix G as the (t_i, x_l)-swaption volatility *plus* $G_{i,l+1}$. (This is not entirely true due to some grid interpolation effects, but the general idea is correct). So, in principle, $G_{i,l+1}$ can be found by just solving a quadratic equation, i.e. in closed form. For full details we refer the reader to Brigo and Mercurio [2001] where this *bootstrap*, or *cascade*, algorithm is described in detail. While this may appear to be a strong contender for practical LM calibration — full calibration is performed by just solving $N_t N_x$ quadratic equations — the scheme generally does not work for practically-sized problems. The cascade calibration suffers strongly from the ringing problem discussed above, and the quadratic equations typically fail to have a solution for swaption targets with just a few years of total maturity (i.e. from today to the final payment date). While it is possible to include regularity terms that preserve the closed-form nature of the solution, the ringing problem is difficult to remedy and calibration to long-dated options is rarely feasible. We find this to be true, even if one applies ad-hoc remediation methods proposed by various authors (see e.g. Brigo and Morini [2006]).

We should note that bootstrap calibration does have certain uses. For instance, one could use full-blown (or row by row) optimization to fundamentally calibrate G, and then use some version of bootstrap calibration to examine the effect of making small perturbations to input prices, e.g. when computing vegas. We discuss this idea in Chapter 26.

14.5.9 Correlation Calibration to Spread Options

In the calibration algorithm in Section 14.5.7, the matrix R was specified exogenously and would typically originate from an empirical analysis similar

to that in Section 14.3.2. As we discussed in Section 14.4.3, an alternative approach attempts to imply R directly from market data for spread options. Less is known about the robustness of calibrations based on this approach, but this shall not prevent us from listing a possible algorithm.

First, to make the problem tractable, we assume that the matrix R is time-homogeneous and specified as some parametric function of a low-dimension parameter-vector ξ,

$$R = R(\xi).$$

Possible parameterizations include those listed in Section 14.3.2. We treat ξ as an unknown vector, to be determined in the calibration procedure along with the elements of the volatility matrix G. For this, we introduce a set of market-observable spread option prices $\widehat{V}_{\text{spread},1}, \widehat{V}_{\text{spread},2}, \ldots, \widehat{V}_{\text{spread},N_{SP}}$, their corresponding model-based prices $\overline{V}_{\text{spread},1}(G,\xi), \overline{V}_{\text{spread},2}(G,\xi), \ldots, \overline{V}_{\text{spread},N_{SP}}(G,\xi)$, and update the norm $\mathcal{I}$ in (14.51) (or (14.54)) to the norm $\mathcal{I}^*(G,\xi)$, where[20]

$$\mathcal{I}^*(G,\xi) = \mathcal{I}(G,\xi) + \frac{w_{SP}}{N_{SP}} \sum_{i=1}^{N_{SP}} \left(\overline{V}_{\text{spread},i}(G,\xi) - \widehat{V}_{\text{spread},i} \right)^2 . \tag{14.55}$$

The algorithm in Section 14.5.7 proceeds as before with a few obvious changes; we list the full algorithm here for completeness.

1. Given G, interpolate using (14.43) or (14.44) to obtain the full norm volatility grid $\|\lambda_{n,k}\|$ for all Libor indices $k = 1, \ldots, N-1$ and all expiry indices $n = 1, \ldots, k$.
2. Given ξ, compute $R = R(\xi)$.
3. For each $n = 1, \ldots, N-1$ and using $R(\xi)$, compute the matrix H, and ultimately volatility loadings $\lambda_k(T_n)$, from $\|\lambda_{n,k}\|$, $k \geq n$, by PCA methodology, using either (14.48) or (14.50).
4. Given $\lambda_k(\cdot)$ for all $k = 1, \ldots, N-1$, use the formulas in Sections 14.4.1, 14.4.2 and 14.4.3 to compute model prices for all swaptions, caps and spread options in the calibration set.
5. Establish the value of $\mathcal{I}^*(G,\xi)$ by direct computation of (14.55).
6. Update G and ξ and repeat Steps 1–5 until $\mathcal{I}(G,\xi)$ is minimized.

When using a correlation PCA algorithm in Step 3, in practice one may find that it is most efficient to use the "poor man's" approach in Section 14.3.4.2, rather than the slower expression listed in Proposition 14.3.2. Indeed, as long as the spread option prices ultimately are well-matched, we can be confident that our model has a reasonable correlation structure, irrespective of which PCA technique was used.

As was the case for our basic algorithm, let us note that it may be useful to transform spread option prices into implied volatilities or, even better,

[20] Note that our original norm $\mathcal{I}$ now also is a function of ξ, since cap and swaption prices depend on the correlation matrix R.

into implied term correlations[21] when evaluating the mean-squared error. For spread options, a definition of implied term correlation can be extracted from the simple Gaussian spread approach in Section 14.4.3, equations (14.36) and (14.37) or, for more accurate formulas, using the results of Chapter 17 and in particular Sections 17.4.2 and 17.9.1.

Finally, we should note that the optimization problem embedded in the algorithm above can be quite challenging to solve in practice. To stabilize the numerical solution, it may be beneficial to employ a split calibration approach, where we first freeze correlation parameters ξ and then optimize G over the parts of the calibration norm that do not involve spread options. Then we freeze G at its optimum and optimize ξ over the parts of the calibration norm that do not involve caps and swaptions. This alternating volatility- and correlation-calibration is then repeated iteratively until (hopefully) convergence. A similar idea can be employed when calibrating models to a volatility smile; see Section 15.2.3 for LM model applications and Section 16.2.3 for applications to vanilla models.

14.5.10 Volatility Skew Calibration

The calibration algorithm we have discussed so far will normally take at-the-money options as calibration targets when establishing the $\lambda_k(t)$ functions. Establishing the volatility smile away from at-the-money strikes must be done in a separate step, through specification of a DVF skew function φ and, possibly, a stochastic volatility process $z(t)$. For the time-stationary specifications of these two mechanisms that we considered in Section 14.2.5, best-fitting to the volatility skew can be done relatively easily — in fact, it is probably often best to leave the parameters[22] of the skew function φ as a free parameter for trader's input. We study the problem of volatility skew calibration for LM models in more detail in Chapter 15.

14.6 Monte Carlo Simulation

Once an LM model has been calibrated to market data, we can proceed to use the parameterized model for the pricing and risk management of non-vanilla options. In virtually all cases, pricing of such options will involve numerical methods. As the LM model involves a very large number of Markov state

[21] By representing spread options through implied term correlations, the information extracted from spread options is more "orthogonal" to that extracted from caps and swaptions, something that can help improve the numerical properties of the calibration algorithm, particularly if split calibration approach is used.

[22] Assuming that there are only a few parameters that define the shape of the function. We generally recommend using simple skew functions that can be described by a single-parameter family, such as linear or power functions.

variables — namely the full number of Libor forward rates on the yield curve plus any additional variables used to model stochastic volatility — finite difference methods are rarely applicable (but see the brief discussion in Section 15.3 for a special case), and we nearly always have to rely on Monte Carlo methods. As we discussed in Chapter 3, the main idea of Monte Carlo pricing is straightforward: i) simulate independent paths of the collection of Libor rates through time; ii) for each path, sum the numeraire-deflated values of all cash flows generated by the specific interest rate dependent security at hand; iii) repeat i)-ii) many times and form the average. Proper execution of step i) is obviously key to this algorithm, and begs an answer to the following question: given a probability measure and the state of the Libor forward curve at time t, how do we move the entire Libor curve (and the numeraire) forward to time $t + \Delta$, $\Delta > 0$, in a manner that is consistent with the LM model dynamics? We address this question here.

14.6.1 Euler-Type Schemes

Assume that we stand at time t, and have knowledge of forward Libor rates maturing at all dates in the tenor structure after time t. We wish to devise a scheme to advance time to $t + \Delta$ and construct a sample of $L_{q(t+\Delta)}(t + \Delta), \ldots, L_{N-1}(t + \Delta)$. Notice that $q(t + \Delta)$ may or may not exceed $q(t)$; if it does, some of the front-end forward rates expire and "drop off" the curve as we move to $t + \Delta$.

For concreteness, assume for now that we work in the spot measure Q^B in which case Lemma 14.2.3 tells us that general LM model dynamics are of the form

$$dL_n(t) = \sigma_n(t)^\top \left(\mu_n(t)\, dt + dW^B(t)\right), \quad \mu_n(t) = \sum_{j=q(t)}^{n} \frac{\tau_j \sigma_j(t)}{1 + \tau_j L_j(t)}, \tag{14.56}$$

where the $\sigma_n(t)$ are adapted vector-valued volatility functions and $W^B(t)$ is an m-dimensional Brownian motion in measure Q^B. The simplest way of drawing an approximate sample $\widehat{L}_n(t + \Delta)$ for $L_n(t + \Delta)$ would be to apply a first-order Euler-type scheme. Applying results from Section 3.2.3, Euler (14.57) and log-Euler (14.58) schemes for (14.56) are, for $n = q(t + \Delta), \ldots, N - 1$,

$$\widehat{L}_n(t + \Delta) = \widehat{L}_n(t) + \sigma_n(t)^\top \left(\mu_n(t)\Delta + \sqrt{\Delta} Z\right), \tag{14.57}$$

$$\widehat{L}_n(t + \Delta) = \widehat{L}_n(t) \exp\left\{ \frac{\sigma_n(t)^\top}{\widehat{L}_n(t)} \left(\left(\mu_n(t) - \frac{1}{2} \frac{\sigma_n(t)}{\widehat{L}_n(t)} \right) \Delta + \sqrt{\Delta} Z \right) \right\}, \tag{14.58}$$

where Z is a vector of m independent $\mathcal{N}(0,1)$ Gaussian draws[23]. For specifications of $\sigma_n(t)^\top$ that are close to proportional in $L_n(t)$ (e.g. the log-normal LM model), we would expect the log-Euler scheme (14.58) to produce lower biases than the Euler scheme (14.57). As discussed in Chapter 3, the log-Euler scheme will keep forward rates positive, whereas the Euler scheme will not.

As shown, both schemes (14.57), (14.58) advance time only by a single time step, but creation of a full path of forward curve evolution through time is merely a matter of repeated application of the single-period stepping schemes on a (possibly non-equidistant) time line $t_0, t_1, \ldots$. When working in the spot measure, it is preferable to have the tenor structure dates $T_1, T_2, \ldots, T_{N-1}$ among the simulation dates, in order to keep track of the spot numeraire $B(\cdot)$ without having to resort to extrapolations. In fact, it is common in practice to set $t_i = T_i$, which, unless accrual periods τ_i are unusually long or volatilities unusually high, will normally produce an acceptable discretization error for many types of LM models. See e.g. Andersen and Andreasen [2000b] and Glasserman and Zhao [2000] for some numerical investigations of the Euler bias.

Remark 14.6.1. When t coincides with a date in the tenor structure, $t = T_k$, say, $q(t)$ will equal T_{k+1} due to our definition of q being right-continuous. As a result, when stepping forward from time $t = T_k$, $\widehat{L}_k(T_k)$ will *not* be included in the computation of the drifts μ_n, $n \geq k+1$. As it turns out, this convention reduces discretization bias, a result that makes sense when we consider that the contribution from $L_k(t)$ to the drifts drops to zero at time $T_k + dt$ in a continuous-time setting.

While Euler-type schemes such as (14.57) and (14.58) are not very sophisticated and, as we recall from Chapter 3, result in rather slow convergence of the discretization bias ($O(\Delta)$), these schemes are appealing in their straightforwardness and universal applicability. Further, they serve to highlight the basic structure of an LM simulation and the computational effort in advancing the forward curve.

14.6.1.1 Analysis of Computational Effort

Focusing on the straight Euler scheme (14.57), a bit of contemplation reveals that the computational effort involved in advancing L_n is dominated by the computation of $\mu_n(\cdot)$ which, in a direct implementation of (14.56), involves

$$m \cdot (n - q(t) + 1) = O(mn)$$

[23] In addition to these time-stepping schemes for the forward rates, it may be necessary to simultaneously evolve stochastic volatility variables if one works with models such as those in Section 14.2.5.

operations for a given value of n. To advance all $N - q(t+\Delta)$ forward rates. it follows that the computational effort is $O(mN^2)$ for a single time step. Assuming that our simulation time line coincides with the tenor structure dates, generation of a full path of forward curve scenarios from time 0 to time T_{N-1} will thus require a total computational effort of $O(mN^3)$. As N is often big (e.g., a 25 year curve of quarterly forward rates will have $N = 100$), a naive application of the Euler scheme will often require considerable computing resources.

As should be rather obvious, however, the computational order of $O(mN^3)$ is easy to improve on, as there is no need to spend $O(mN)$ operations on the computation of each μ_n. Instead, we can invoke the recursive relationship

$$\mu_n(t) = \mu_{n-1}(t) + \frac{\tau_n \sigma_n(t)}{1 + \tau_n \widehat{L}_n(t)}, \tag{14.59}$$

which allows us to compute all μ_n, $n = q(t+\Delta), \ldots, N-1$, by an $O(mN)$-step iteration starting from

$$\mu_{q(t+\Delta)}(t) = \sum_{j=q(t)}^{q(t+\Delta)} \frac{\tau_j \sigma_j(t)}{1 + \tau_j \widehat{L}_j(t)}.$$

In total, the computational effort of advancing the full curve one time step will be $O(mN)$, and the cost of taking N such time steps will be $O(mN^2)$ — and not $O(mN^3)$.

We summarize this result in a lemma.

Lemma 14.6.2. *Assume that we wish to simulate the entire Libor forward curve on a time line that contains the dates in the tenor structure and has $O(N)$ points. The computational effort of Euler-type schemes — such as (14.57) and (14.58) — is $O(mN^2)$.*

Remark 14.6.3. The results of the lemma can be verified to hold for any of the probability measures we examined in Section 14.2.2.

We note that when simulating in other measures, the starting point of the iteration for μ_n will be measure-dependent. For instance, in the terminal measure,

$$\mu_n(t) = -\sum_{j=n+1}^{N-1} \frac{\tau_j \sigma_j(t)}{1 + \tau_j \widehat{L}_j(t)}$$

and the equation (14.59) still holds. Now, however, the iteration starts at

$$\mu_{N-1}(t) = 0,$$

and proceeds *backwards* through $\mu_{N-2}, \mu_{N-3}, \ldots, \mu_{q(t+\Delta)}$. We leave it to the reader to carry out the analysis for other probability measures.

14.6.1.2 Long Time Steps

Most exotic interest rate derivatives involve revolving cash flows paid on a tightly spaced schedule (e.g. quarterly). As our simulation time line should always include dates on which cash flows take place, the average time spacing used in path generation will thus normally, by necessity, be quite small. In certain cases, however, there may be large gaps between cash flow dates, e.g. when a security is forward-starting or has an initial lock-out period. When simulating across large gaps, we may always choose to sub-divide the gap into smaller time steps, thereby retaining a tightly spaced simulation time line. To save computational time, however, it is often tempting to cover large gaps in a small number of coarse time steps, in order to lower overall computation effort. Whether such coarse stepping is possible is, in large part, a question of how well we can keep the discretization bias under control as we increase the time step, something that is quite dependent on the magnitude of volatility and the particular formulation of the LM model under consideration. Section 14.6.2 below deals with this question and offers strategies to improve on the basic Euler scheme. Here, we instead consider the pure mechanics of taking large time steps, i.e. steps that skip past several dates in the tenor structure.

Assume that we stand at the j-th date in the tenor structure, $t = T_j$, and wish to simulate the forward curve to time T_k, $k > j+1$, in a single step. As noted earlier, the mere notion of skipping over dates in the tenor structure makes usage of the spot measure Q^B inconvenient, as the numeraire $B(T_k)$ cannot be constructed without knowledge of the realizations of $L_{j+1}(T_{j+1}), L_{j+2}(T_{j+2}), \ldots, L_{k-1}(T_{k-1})$; in turn, numeraire-deflation of cash flows is not possible and derivatives cannot be priced. Circumventing this issue, however, is merely a matter of changing the numeraire from $B(t)$ to the price of an asset that involves no roll-over in the interval $[T_j, T_k]$. One such asset price is $P(t, T_N)$, the choice of which corresponds to running our simulated paths in the terminal measure. In particular, we recognize that

$$P(T_k, T_N) = \prod_{n=k}^{N-1} \frac{1}{1 + \tau_n L_n(T_k)}, \tag{14.60}$$

which depends only on the state of the forward curve at time T_k. Another valid numeraire asset would be $\overline{P}_j(t)$, as defined in Section 14.2.2:

$$\overline{P}_j(t) = \begin{cases} B(t), & t \leq T_j, \\ B(T_j)P(t, T_N)/P(T_j, T_N), & t > T_j. \end{cases}$$

The numeraire $\overline{P}_j(T_k)$ can always be computed without knowledge of $L_{j+1}(T_{j+1}), \ldots, L_{k-1}(T_{k-1})$, as long as $B(T_j)$ is known[24]. In the measure induced by this asset, the LM model dynamics are

[24] This precludes the existence of other large gaps in the simulation time line prior to time T_j. When using a hybrid measure such as $\overline{P}_j$, we would need to

$$dL_n(t) = \begin{cases} \sigma_n(t)^\top \left(-\sum_{l=n+1}^{N-1} \frac{\tau_l \sigma_l(t)}{1+\tau_l L_l(t)} dt + d\overline{W}^j(t) \right), & t > T_j, \\ \sigma_n(t)^\top \left(\sum_{l=q(t)}^{n} \frac{\tau_l \sigma_l(t)}{1+\tau_l L_l(t)} dt + d\overline{W}^j(t) \right), & t \leq T_j. \end{cases}$$

14.6.1.3 Notes on the Choice of Numeraire

Given our discussion above, the terminal measure may strike the reader as an obvious first choice for simulating the LM model — after all, simulations in the terminal measure will never fail to be meaningful, irrespective of the coarseness of the simulation time line. Other issues, however, come in play here as well. For instance, updating the numeraire $P(t, T_N)$ from one time step to the next is generally a more elaborate operation that updating the spot numeraire $B(t)$: the former requires multiplying together $O(N)$ terms (see (14.60)), whereas the latter only involves multiplying $B(t)$ at the previous time step with a single discount bond price. Also, the statistical sample properties of price estimators in the terminal measure may be inferior to those in the spot measure, in the sense that the Monte Carlo noise is larger in the terminal measure. Glasserman and Zhao [2000] list empirical results indicating that this is, indeed, often the case for many common interest rate derivatives. A formal analysis of this observation is complex, but we can justify it by considering the pricing of a very simple derivative security, namely a discount bond maturing at some arbitrary time T_k in the tenor structure. In the spot measure, we would estimate the price of this security by forming the sample average of random variables

$$P(T_k, T_k)/B(T_k) = B(T_k)^{-1} = \frac{1}{\prod_{n=0}^{k-1} (1 + \tau_n L_n(T_n))}, \tag{14.61}$$

whereas in the terminal measure we would form the sample average of random variables

$$P(T_k, T_k)/P(T_k, T_N) = P(T_k, T_N)^{-1} = \prod_{n=k}^{N-1} (1 + \tau_n L_n(T_k)). \tag{14.62}$$

Assuming that Libor rates stay positive, the important thing to notice is that the right-hand side of (14.61) is bounded from above by 1, whereas the right-hand side of (14.62) can grow arbitrarily large. For moderate to high Libor rate volatilities, we would thus intuitively expect price estimators based on (14.62) to have higher sample error.

As discussed in Section 14.6.1.2, sometimes it is mechanically inconvenient to simulate in the spot measure, due to a desire to take large time steps. In these cases, usage of a hybrid numeraire $\overline{P}(t)$ that switches from $B(t)$ to $P(t, T_N)$ at the latest possible date may be a useful strategy.

---

position T_j at the start of the first simulation time step that spans multiple dates in the tenor structure.

14.6.2 Other Simulation Schemes

When simulating on a reasonably tight time schedule, the accuracy of the Euler or log-Euler schemes is adequate for most applications. However, as discussed above, we may occasionally be interested in using coarse time steps in some parts of the path generation algorithm, requiring us to pay more attention to the discretization scheme. Generic techniques for these purposes were introduced in detail in Chapter 3; we proceed to discuss a few of these in the context of LM models. We also consider the case where special-purpose schemes happen to exist for the discretization of the stochastic integral in the forward rate dynamics.

14.6.2.1 Special-Purpose Schemes with Drift Predictor-Corrector

In integrated form, the general LM dynamics in (14.56) become

$$\begin{aligned} L_n(t+\Delta) &= L_n(t) + \int_t^{t+\Delta} \sigma_n(u)^\top \mu_n(u)\, du + \int_t^{t+\Delta} \sigma_n(u)^\top dW^B(u) \\ &\triangleq L_n(t) + D_n(t, t+\Delta) + M_n(t, t+\Delta), \end{aligned} \tag{14.63}$$

where $M_n(t, t+\Delta)$ is a zero-mean martingale increment and $D_n(t, t+\Delta)$ is the increment of a predictable process. In many cases of practical interest, high-performance special-purpose schemes exist for simulation of $M_n(t, t+\Delta)$. This, for instance, is the case for the SV-LM model specification (Section 14.2.5), as discussed in detail in Section 9.5. In such cases, we obviously will choose to generate $M_n(t, t+\Delta)$ from the special-purpose scheme, and it thus suffices to focus on the term $D_n(t, t+\Delta)$. A simple approach is to use Euler stepping:

$$\widehat{L}_n(t+\Delta) = \widehat{L}_n(t) + \sigma_n(t)^\top \mu_n(t)\, \Delta + \widehat{M}_n(t, t+\Delta), \tag{14.64}$$

where $\widehat{M}_n(t, t+\Delta)$ is generated by a special-purpose scheme.

The drift adjustments in (14.64) are explicit in nature, as they are based only on the forward curve at time t. To incorporate information from time $t+\Delta$, we can use the *predictor-corrector* scheme from Section 3.2.5, which for (14.64) will take the two-step form

$$\overline{L}_n(t+\Delta) = \widehat{L}_n(t) + \sigma_n\left(t, \widehat{\mathbf{L}}(t)\right)^\top \mu_n\left(t, \widehat{\mathbf{L}}(t)\right) \Delta + \widehat{M}_n(t, t+\Delta), \tag{14.65}$$

$$\begin{aligned} \widehat{L}_n(t+\Delta) = \widehat{L}_n(t) &+ \theta_{\mathrm{PC}} \sigma_n\left(t, \widehat{\mathbf{L}}(t)\right)^\top \mu_n\left(t, \widehat{\mathbf{L}}(t)\right) \Delta \\ &+ (1-\theta_{\mathrm{PC}}) \sigma_n\left(t+\Delta, \overline{\mathbf{L}}(t+\Delta)\right)^\top \mu_n\left(t+\Delta, \overline{\mathbf{L}}(t+\Delta)\right) \Delta \\ &+ \widehat{M}_n(t, t+\Delta), \end{aligned} \tag{14.66}$$

where θ_{PC} is a parameter in $[0,1]$ that determines the amount of implicitness we want in our scheme ($\theta_{\mathrm{PC}} = 1$: fully explicit; $\theta_{\mathrm{PC}} = 0$: fully implicit). In practice, we would nearly always go for the balanced choice of $\theta_{\mathrm{PC}} = 1/2$. In (14.65)–(14.66), $\mathbf{L}$ denotes the vector of all Libor rates, $\mathbf{L}(t) = (L_1(t), \ldots, L_{N-1}(t))^\top$ (with the convention that $L_i(t) \equiv L_i(T_i)$ for $i < q(t)$), and $\widehat{\mathbf{L}}$, $\overline{\mathbf{L}}$ defined accordingly. In particular, the short-hand notation $\mu_n(t, \widehat{\mathbf{L}}(t))$ is used to indicate that μ_n (and σ_n) may depend on the state of the entire forward curve at time t.

The technique above is based on a standard (additive) Euler scheme. If one is more inclined to use a multiplicative scheme in the vein of (14.58), we may replace the explicit scheme (14.64) with

$$\widehat{L}_n(t+\Delta) = \widehat{L}_n(t) \exp\left\{ \frac{\sigma_n(t)^\top}{\widehat{L}_n(t)} \mu_n(t)\Delta \right\} \widehat{M}_n(t, t+\Delta), \tag{14.67}$$

where $\widehat{M}_n(t, t+\Delta)$ now has been redefined to be a unit-mean positive random variable, often a discretized multiplicative increment of an exponential martingale. The construction of a predictor-corrector extension of (14.67) follows closely the steps above, and is left for the reader.

While the weak convergence order of simulation schemes may not be affected by predictor-corrector schemes (Section 3.2.5), experiments show that (14.65)–(14.66) often will reduce the bias significantly relative to a fully explicit Euler scheme. Some results for the simple log-normal LM model can be found in Hunter et al. [2001] and Rebonato [2002]. As the computational effort of applying the predictor step is not insignificant, the speed-accuracy trade-off must be evaluated on a case-by-case basis. Section 14.6.2.3 below discusses a possible modification of the predictor-corrector scheme to improve efficiency.

14.6.2.2 Euler Scheme with Predictor-Corrector

In simulating the term $M_n(t, t+\Delta)$ in the predictor-corrector scheme above, we can always use the Euler scheme, i.e. in (14.64) we set

$$\widehat{M}_n(t, t+\Delta) = \sigma_n(t)^\top \sqrt{\Delta} Z,$$

where Z is an m-dimensional vector of standard Gaussian draws. As we recall from Chapter 3, however, it may also be useful to apply the predictor-corrector principle to the martingale part of the forward rate evolution itself, although this would involve the evaluation of derivatives of the LM volatility term with respect to the forward Libor rates; see Chapter 3 for details.

14.6.2.3 Lagging Predictor-Corrector Scheme

Drift calculations, as was pointed out earlier, are the most computationally expensive part of any Monte Carlo scheme for a Libor market model. The

predictor-corrector scheme of (14.65)–(14.66) requires *two* calculations of the drift and is thus considerably more expensive than the standard Euler scheme. We often prefer to use a "lagging" modified predictor-corrector scheme which, as it turns out, allows us to realize most of the benefits of the predictor-corrector scheme, while keeping computational costs comparable to the standard Euler scheme.

Recall the definition of the drift of the n-th Libor rate under the spot measure,

$$\mu_n(t) = \sum_{j=q(t)}^{n} \frac{\tau_j \sigma_j(t)}{1 + \tau_j L_j(t)}.$$

Note that the drift depends on the values of the Libor rates of indices less than or equal to n. Let us split the contributions coming from Libor rates with an index strictly less than n, and the n-th Libor rate,

$$\mu_n(t) = \sum_{j=q(t)}^{n-1} \frac{\tau_j \sigma_j(t)}{1 + \tau_j L_j(t)} + \frac{\tau_n \sigma_n(t)}{1 + \tau_n L_n(t)}.$$

Denoting $t' = t + \Delta$, we observe that if we simulate the Libor rates in the order of increasing index, then by the time we need to simulate $L_n(t')$, we have already simulated $L_j(t')$, $j = q(t), \ldots, n-1$. Hence, it is natural to use the predictor-corrector technique for the part of the drift that depends on Libor rates maturing strictly before T_n, while treating the part of the drift depending on the n-th Libor rate explicitly. This idea leads to the following scheme (compare to (14.64) or (14.65)–(14.66) with $\theta_{\mathrm{PC}} = 1/2$),

$$\begin{aligned}
&\widehat{L}_n(t') = \widehat{L}_n(t) + \sigma_n(t)^{\top} \\
&\times \left(\frac{1}{2} \sum_{j=q(t)}^{n-1} \left(\frac{\tau_j \sigma_j(t)}{1 + \tau_j \widehat{L}_j(t)} + \frac{\tau_j \sigma_j(t')}{1 + \tau_j \widehat{L}_j(t')} \right) + \frac{\tau_n \sigma_n(t)}{1 + \tau_n \widehat{L}_n(t)} \right) \Delta + \widehat{M}_n(t, t').
\end{aligned} \tag{14.68}$$

Importantly, the drifts required for this scheme also satisfy a recursive relationship, allowing for an efficient update. Defining

$$\widehat{\alpha}_n(t') = \sum_{j=q(t)}^{n} \left(\frac{\tau_j \sigma_j(t)}{1 + \tau_j \widehat{L}_j(t)} + \frac{\tau_j \sigma_j(t')}{1 + \tau_j \widehat{L}_j(t')} \right),$$

we see that, clearly,

$$\widehat{\alpha}_n(t') = \widehat{\alpha}_{n-1}(t') + \frac{\tau_n \sigma_n(t)}{1 + \tau_n \widehat{L}_n(t)} + \frac{\tau_n \sigma_n(t')}{1 + \tau_n \widehat{L}_n(t')},$$

and (14.68) can be rewritten as

$$\widehat{L}_n(t') = \widehat{L}_n(t) + \sigma_n(t)^\top \left(\frac{1}{2}\widehat{\alpha}_{n-1}(t') + \frac{\tau_n \sigma_n(t)}{1 + \tau_n \widehat{L}_n(t)} \right) \Delta + \widehat{M}_n(t, t'). \tag{14.69}$$

The scheme above can easily be applied to other probability measures. In fact, since in the terminal measure the drift

$$\mu_n(t) = -\sum_{j=n+1}^{N-1} \frac{\tau_j \sigma_j(t)}{1 + \tau_j L_j(t)},$$

does not depend on $L_n(t)$ in the first place, no "lag" is required in this measure. Indeed, we simply redefine

$$\widehat{\alpha}_n(t') = -\sum_{j=n+1}^{N-1} \left(\frac{\tau_j \sigma_j(t)}{1 + \tau_j \widehat{L}_j(t)} + \frac{\tau_j \sigma_j(t')}{1 + \tau_j \widehat{L}_j(t')} \right)$$

and, starting from $n = N - 1$ and working backwards, use the scheme

$$\widehat{L}_n(t') = \widehat{L}_n(t) + \sigma_n(t)^\top \frac{1}{2}\widehat{\alpha}_n(t')\Delta + \widehat{M}_n(t, t'). \tag{14.70}$$

Notice that $\widehat{\alpha}_n$ now satisfies the recursion

$$\widehat{\alpha}_{n-1}(t') = \widehat{\alpha}_n(t') - \frac{\tau_n \sigma_n(t)}{1 + \tau_n \widehat{L}_n(t)} - \frac{\tau_n \sigma_n(t')}{1 + \tau_n \widehat{L}_n(t')},$$

to be started at $\widehat{\alpha}_{N-1}(t) = 0$.

The modifications of (14.69) and (14.70) to accommodate log-Euler stepping are trivial and left to the reader to explore. The lagging predictor-corrector scheme in the spot Libor measure has, as far as we know, not appeared in the literature, and its theoretical properties are not well-known (although the terminal measure version was studied in Joshi and Stacey [2008]). Still, its practical performance is very good and we do not hesitate recommending it as the default choice for many applications.

14.6.2.4 Further Refinements of Drift Estimation

For large time steps, it may be useful to explicitly integrate the time-dependent parts of the drift, rather than rely on pure Euler-type approximations. Focusing on, say, (14.63), assume that we can write

$$\sigma_n(u)^\top \mu_n(u) \approx g\left(u, \mathbf{L}(t)\right), \quad u \geq t, \tag{14.71}$$

for a function g that depends on time as well as the state of the forward rates frozen at time t. Then,

$$D_n(t, t + \Delta) = \int_t^{t+\Delta} \sigma_n(u)^\top \mu_n(u)\, du \approx \int_t^{t+\Delta} g\left(u, \mathbf{L}(t)\right) du. \tag{14.72}$$

As g evolves deterministically for $u > t$, the integral on the right-hand size can be evaluated either analytically (if g is simple enough) or by numerical quadrature. If doing the integral numerically, a decision must be made on the spacing of the integration grid. For volatility functions that are piecewise flat on the tenor-structure — which is a common assumption in model calibration — it is natural to align the grid with dates in the tenor structure.

To give an example, consider the DVF LM model, where we get (in the terminal measure, for a change)

$$\begin{aligned}\sigma_n(u)^\top \mu_n(u) &= -\varphi\left(L_n(u)\right)\lambda_n(u)^\top \sum_{j=n+1}^{N-1} \frac{\tau_j \lambda_j(u)\varphi\left(L_j(u)\right)}{1+\tau_j L_j(u)} \\ &\approx -\varphi\left(L_n(t)\right)\lambda_n(u)^\top \sum_{j=n+1}^{N-1} \frac{\tau_j \lambda_j(u)\varphi\left(L_j(t)\right)}{1+\tau_j L_j(t)}, \quad u \geq t,\end{aligned}$$

which is of the form (14.71). For stochastic volatility models we might, say, additionally assume that the process $z(t)$ would stay on its expected path, i.e. $z(u) \approx \mathrm{E}_t^N(z(u))$ which can often be computed in closed form for models of interest. For instance, for the SV model in (14.15) we have

$$\mathrm{E}_t^N\left(z(u)\right) = z_0 + (z(t) - z_0)e^{-\theta(u-t)}.$$

The approach in (14.72) easily combines with predictor-corrector logic, i.e. we could write

$$\begin{aligned}D_n(t, t+\Delta) \approx \theta_{\mathrm{PC}} &\int_t^{t+\Delta} g\left(u, \mathbf{L}(t)\right) du \\ &+ (1-\theta_{\mathrm{PC}}) \int_t^{t+\Delta} g\left(u, \overline{\mathbf{L}}(t+\Delta)\right) du, \quad (14.73)\end{aligned}$$

where $\overline{L}_i(t+\Delta)$ has been found in a predictor step using (14.72) in (14.64). The "lagged" schemes in Section 14.6.2.3 work equally well. Formula (14.72) also applies to exponential-type schemes such as (14.67), with or without predictor-corrector adjustment; we leave details to the reader.

14.6.2.5 Brownian-Bridge Schemes and Other Ideas

As a variation on the predictor-corrector scheme, we could attempt a further refinement of taking into account variance of the Libor curve between the sampling dates t and $t+\Delta$. Schemes attempting to do so by application of *Brownian bridge techniques*[25] were proposed in Andersen [2000b] and Pietersz et al. [2004], among others. While performance of these schemes is

[25] See Section 3.2.9 for an introduction to the Brownian bridge, albeit for a somewhat different application.

mixed — tests in Joshi and Stacey [2008] show rather unimpressive results in comparison to simpler predictor-corrector schemes — the basic idea is sufficiently simple and instructive to merit a brief mention. In a nutshell, the Brownian bridge approach aims to replace in (14.72) all forward rates $\mathbf{L}(t)$ with the expectation of $\mathbf{L}(u)$, *conditional* upon the forward rates ending up at $\overline{\mathbf{L}}(t+\Delta)$, where $\overline{\mathbf{L}}(t+\Delta)$ is generated in a predictor step. Under simplifying assumptions on the dynamics of $L_n(t)$, a closed-form expression is possible for this expectation.

Proposition 14.6.4. *Assume that*

$$dL_n(t) \approx \sigma_n(t)^\top \, dW(t),$$

where $\sigma_n(t)$ *is* deterministic *and* $W(t)$ *is an* m*-dimensional Brownian motion in some probability measure* P. *Let*

$$v_n(t,T) = \int_t^T \|\sigma_n(s)\|^2 \, ds, \quad T \geq t.$$

Then, for $u \in [t, t+\Delta]$,

$$\mathrm{E}\left(L_n(u)|L_n(t), L_n(t+\Delta)\right) = L_n(t) + \frac{v_n(t,u)}{v_n(t,t+\Delta)}\left(L_n(t+\Delta) - L_n(t)\right).$$

Proof. We first state a very useful general result for multi-variate Gaussian variables.

Lemma 14.6.5. *Let* $X = (X_1, X_2)^\top$ *be a partitioned vector of Gaussian variables, where* X_1 *and* X_2 *are themselves vectors. Assume that the covariance matrix between* X_i *and* X_j *is* $\Sigma_{i,j}$ *such that the total covariance matrix of* X *is*

$$\Sigma = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{pmatrix}$$

(where, of course, $\Sigma_{2,1} = \Sigma_{1,2}^\top$*). Let the vector means of* X_i *be* μ_i, $i = 1, 2$, *and assume that* $\Sigma_{2,2}$ *is invertible. Then* $X_1|X_2 = x$ *is Gaussian:*

$$(X_1|X_2 = x) \sim \mathcal{N}\left(\mu_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(x - \mu_2), \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}\right).$$

In Lemma 14.6.5, now set $X_1 = L_n(u) - L_n(t)$ and $X_2 = L_n(t+\Delta) - L_n(t)$. Note that $\mu_1 = \mu_2 = 0$ and

$$\Sigma_{1,2} = \Sigma_{2,1} = \Sigma_{1,1} = v_n(t,u), \quad \Sigma_{2,2} = v_n(t, t+\Delta).$$

The result of Proposition 14.6.4 follows. □

We can use the result of Proposition 14.6.4 in place of the ordinary corrector step. For instance, in (14.73) we write

$$D_n(t, t+\Delta) \approx \int_t^{t+\Delta} g\left(u, \mathbf{m}(u)\right) du,$$

where, for $\mathbf{m}(u) = (m_1(u), \ldots, m_{N-1}(u))$,

$$m_i(u) = \mathrm{E}^B\left(L_i(u) | L_i(t), \overline{L}_i(t+\Delta)\right)$$

is computed according to Proposition 14.6.4 once $\overline{L}_i(t+\Delta)$ has been sampled in a predictor step.

In some cases, it may be more appropriate to assume that L_n is roughly log-normal, in which case Proposition 14.6.4 must be altered slightly.

Lemma 14.6.6. *Assume that*

$$dL_n(t)/L_n(t) \approx \sigma_n(t)^\top dW(t),$$

where $\sigma_n(t)$ *is* deterministic *and* $W(t)$ *is an* m*-dimensional Brownian motion in some probability measure* P. *Then, for* $u \in [t, t+\Delta]$,

$$\mathrm{E}\left(L_n(u) | L_n(t), L_n(t+\Delta)\right) = L_n(t) \left(\frac{L_n(t+\Delta)}{L_n(t)}\right)^{v_n(t,u)/v_n(t,t+\Delta)} \times \exp\left(\frac{v_n(t,u)\left(v_n(t,t+\Delta) - v_n(t,u)\right)}{2v_n(t,t+\Delta)}\right),$$

where $v_n(t,T)$ *is given in Proposition 14.6.4.*

Proof. Apply Lemma 14.6.5 to $X_1 = \ln L_n(u) - \ln L_n(t)$ and $X_2 = \ln L_n(t+\Delta) - \ln L_n(t)$. To translate back to find the conditional mean of e^{X_1}, one may use the fact that $\mathrm{E}(e^{a+bY}) = e^{a+b^2/2}$ if Y is Gaussian $\mathcal{N}(0,1)$. □

Joshi and Stacey [2008] investigate a number of other possible discretization schemes for the drift term in the LM model, including ones that attempt to incorporate information about the correlation between various forward rates. In general, many of these schemes will result in some improvement of the discretization error, but at the cost of more computational complexity and effort. All things considered, we hesitate to recommend any of these methods (and this goes for the Brownian bridge scheme above) for general-purpose use, as the bias produced by simpler methods is often adequate. If not, it may, in fact, often be the case that we can insert a few extra simulation dates inside large gaps to bring down the bias, yet still spend less computational time than we would if using a more complex method of bridging the gap in a single step. Finally, we should note that most authors (including Joshi and Stacey [2008]) exclusively examine simple log-normal models where the martingale component (M_n in the notation of Section 14.6.2.1) can be simulated completely bias-free. When using more realistic models, this will not always be the case, in which case high-precision simulation of the drift term D_n will likely be a waste of time.

14.6.2.6 High-Order Schemes

Even with predictor-corrector adjustment, all Euler-type discretization schemes are limited to a convergence order of Δ. To raise this, one possibility is to consider higher-order schemes, such as the Milstein scheme and similar Taylor-based approaches; see Section 3.2.6 for details. Many high-order schemes unfortunately become quite cumbersome to deal with for the type of high-dimensional vector-SDE that arises in the context of LM models and, possibly as a consequence of this, there are currently very few empirical results in the literature to lean on. One exception is Brotherton-Ratcliffe (Brotherton-Ratcliffe [1997]) where a Milstein scheme was developed for the basic log-normal LM model with piecewise flat volatilities. The efficacy of this, and similar high-order schemes, in the context of the generalized LM model would obviously depend strongly on the particular choice of model formulation.

A simple alternative to classical Taylor-based high-order schemes involves Richardson extrapolation based on prices found by simulating on two separate time lines, one coarser than the other (see Section 3.2.7 for details). Andersen and Andreasen [2000b] list some results for this scheme, the efficacy of which seems to be rather modest.

14.6.3 Martingale Discretization

Consider again the hybrid measure induced by the numeraire $\widetilde{P}_{n+1}$, defined in Section 14.2.2. As discussed, one effect of using this measure is to render the process for the n-th forward Libor rate $L_n(t)$ a martingale. When time-discretizing the LM model using, say, an Euler scheme, the martingale property of $L_n(t)$ is automatically preserved, ensuring that the expectation of the discretized approximation $\widehat{L}_n(t)$ will have expectation $L_n(0)$, with no discretization bias. Also, when using Monte Carlo to estimate the price of the zero-coupon bond maturing at time T_{n+1}, we get

$$P(0, T_{n+1}) = \widetilde{P}_{n+1}(0)\mathrm{E}^{n+1}(1),$$

which will (obviously) be estimated bias-free as well.

As the discussion above highlights, it is possible to select a measure such that a particular zero-coupon bond and a particular FRA will be priced bias-free[26] by Monte Carlo simulation, even when using a simple Euler scheme. While we are obviously rarely interested in pricing zero-coupon bonds by Monte Carlo methods, this observation can nevertheless occasionally help guide the choice of simulation measure, particularly if, say, a security can be argued to depend primarily on a single forward rate (e.g. caplet-like securities). In practice, matters are rarely this clear-cut, and one wonders

[26] But not error-free, of course — there will still be a statistical zero-mean error on the simulation results. See Section 14.6.4 below.

whether perhaps simulation schemes exist that will simultaneously price all zero-coupon bonds $P(t,T_1)$, $P(t,T_2),\ldots,P(t,T_N)$ bias-free. It should be obvious that this cannot be accomplished by a simple measure-shift, but will require a more fundamental change in simulation strategy.

14.6.3.1 Deflated Bond Price Discretization

Fundamentally, we are interested in a simulation scheme that by construction will ensure that all numeraire-deflated bond prices are martingales. The easiest way to accomplish this is to follow a suggestion offered by Glasserman and Zhao [2000]: instead of discretizing the dynamics for Libor rates directly, simply discretize the deflated bond prices themselves. To demonstrate, let us consider the spot measure, and define

$$U(t,T_{n+1}) = \frac{P(t,T_{n+1})}{B(t)}. \tag{14.74}$$

Lemma 14.6.7. *Let dynamics in the spot measure* Q^B *be as in Lemma 14.2.3. The dynamics for deflated zero-coupon bond prices (14.74) are given by*

$$\frac{dU(t,T_{n+1})}{U(t,T_{n+1})} = -\sum_{j=q(t)}^{n} \tau_j \frac{U(t,T_{j+1})}{U(t,T_j)} \sigma_j(t)^\top dW^B(t), \quad n = q(t),\ldots,N-1. \tag{14.75}$$

Proof. We note that, by definition,

$$U(t,T_{n+1}) = \frac{P(t,T_{q(t)}) \prod_{j=q(t)}^{n} \frac{1}{1+\tau_j L_j(t)}}{P(t,T_{q(t)}) B(T_{q(t)-1})} = \frac{\prod_{j=q(t)}^{n} \frac{1}{1+\tau_j L_j(t)}}{B(T_{q(t)-1})},$$

where $B(T_{q(t)-1})$ is non-random at time t. We have that $U(t,T_{n+1})$ must, by construction, be a martingale in Q^B. An application of Ito's lemma to the diffusion term of U gives

$$dU(t,T_{n+1}) = -U(t,T_{n+1}) \sum_{j=q(t)}^{n} \frac{\tau_j \sigma_j(t)^\top}{1+\tau_j L_j(t)} dW^B(t),$$

and the lemma follows once we note that

$$\frac{U(t,T_{j+1})}{U(t,T_j)} = \frac{1}{1+\tau_j L_j(t)}.$$

□

Discretization schemes for (14.75) that preserve the martingale property are easy to construct. For instance, we could use the log-Euler scheme

$$\widehat{U}(t+\Delta, T_{n+1}) = \widehat{U}(t, T_{n+1}) \exp\left(-\frac{1}{2}\|\gamma_{n+1}(t)\|^2 \Delta + \gamma_{n+1}(t)^\top Z\sqrt{\Delta}\right). \tag{14.76}$$

where, as before, Z is an m-dimensional standard Gaussian random variable and

$$\gamma_{n+1}(t) \triangleq -\sum_{j=q(t)}^{n} \tau_j \frac{\widehat{U}(t, T_{j+1})}{\widehat{U}(t, T_j)} \sigma_j(t). \tag{14.77}$$

We have several remarks to the log-Euler scheme (14.76). First, for models where interest rates cannot become negative, $U(t, T_{n+1})/U(t, T_n) = P(t, T_{n+1})/P(t, T_n)$ cannot exceed 1 in a continuous-time model, so it might be advantageous to replace (14.77) with

$$\gamma_{n+1}(t) \triangleq -\sum_{j=q(t)}^{n} \tau_j \min\left(\frac{\widehat{U}(t, T_{j+1})}{\widehat{U}(t, T_j)}, 1\right) \sigma_j(t),$$

as recommended in Glasserman and Zhao [2000]. Second, for computational efficiency we should rely on iterative updating,

$$\gamma_{n+1}(t) = \gamma_n(t) - \tau_n \min\left(\frac{\widehat{U}(t, T_{n+1})}{\widehat{U}(t, T_n)}, 1\right) \sigma_n(t),$$

using the same arguments as those presented in Section 14.6.1.1. Third, once $\widehat{U}(t+\Delta, T_n)$ has been drawn for all possible n, we can reconstitute the Libor curve from the relation

$$\widehat{L}_n(t+\Delta) = \frac{\widehat{U}(t+\Delta, T_n) - \widehat{U}(t+\Delta, T_{n+1})}{\tau_n \widehat{U}(t+\Delta, T_{n+1})}, \quad n = q(t+\Delta), \ldots, N-1. \tag{14.78}$$

For completeness, we note that dynamics of the deflated bond prices in the terminal measure Q^{T_N} can easily be derived to be

$$\frac{dU(t, T_{n+1})}{U(t, T_{n+1})} = \sum_{j=n+1}^{N-1} \tau_j \frac{U(t, T_{j+1})}{U(t, T_j)} \sigma_j(t)^\top dW^N(t), \tag{14.79}$$

where we must now (re-)define $U(t, T_n)$ as

$$U(t, T_n) = P(t, T_n)/P(t, T_N).$$

Equation (14.79) can form the basis of a discretization scheme in much the same manner as above.

14.6.3.2 Comments and Alternatives

The discretization scheme presented above will preserve the martingale property of all deflated bonds maturing in the tenor structure, and in this

sense can be considered arbitrage-free. The resulting lack of bias on bond prices, however, does not necessarily translate into a lack of bias on any other derivative security price, e.g. a caplet or a swaption. In particular, we notice that nothing in the scheme above will ensure that bond price moments of any order other than one will be simulated accurately.

The extent of the bias induced by the scheme in Section 14.6.3.1 is specific to the security and model under consideration. For instance, using a log-Euler scheme for deflated bonds might work well in an LM model with rates that are approximately Gaussian, but might work less well in a model where rates are approximately log-normal. If results are disappointing, we can replace (14.76) with another discretization of (14.75) (see Chapter 3 for many examples), or we can try to discretize a quantity other than the deflated bonds $U(t, T_n)$. The latter idea is pursued in Glasserman and Zhao [2000], where several suggestions for discretization variables are considered. For instance, one can consider the differences

$$U(t, T_n) - U(t, T_{n+1}) \tag{14.80}$$

which are martingales since the U's are. As follows from (14.78), discretizing $U(t, T_n) - U(t, T_{n+1})$ is, in a sense, close to discretizing $L_n(t)$ itself which may be advantageous. Joshi and Stacey [2008] contains some tests of discretization schemes based on (14.80), but, again, only in a log-normal setting. Additional tests in Beveridge et al. [2008] (in a displaced log-normal case) find (14.80) inferior to the standard predictor-corrector scheme and demonstrate that the scheme can lead to negative (path realizations of) bond prices.

14.6.4 Variance Reduction

We recall from the discussion in Chapter 3 that the errors involved in Monte Carlo pricing of derivatives can be split into two sources: the statistical Monte Carlo error (the standard error), and a bias unique to the discretization scheme employed. So far, our discussion has centered exclusively on the latter of these two types of errors and we now wish to provide some observations about the former. We should note, however, that it is difficult to provide generic prescription for variance reduction techniques in the LM model, as most truly efficient schemes tend be quite specific to the product being priced. We shall offer several such product-specific variance reduction schemes in later chapters, and here limit ourselves to rather brief suggestions.

We recall that Chapter 3 discussed three types of variance reduction techniques: i) antithetic sampling; ii) control variates; and iii) importance sampling. All have potential uses in simulation of LM models.

14.6.4.1 Antithetic Sampling

Application of antithetic sampling to LM modeling is straightforward. Using the Euler scheme as an example, each forward rate sample path generated

from the relation

$$\widehat{L}_n(t+\Delta) = \widehat{L}_n(t) + \sigma_n(t)^\top \left(\mu_n(t)\Delta + \sqrt{\Delta}Z\right)$$

is simply accompanied by a "reflected" sample path computed by flipping the vector-valued Gaussian variable Z around the origin, i.e.

$$\widehat{L}_n^{(a)}(t+\Delta) = \widehat{L}_n^{(a)}(t) + \sigma_n(t)^\top \left(\mu_n^{(a)}(t)\Delta - \sqrt{\Delta}Z\right).$$

The reflection of Z is performed at each time step, with both paths having identical starting points, $\widehat{L}_n^{(a)}(0) = \widehat{L}_n(0) = L_n(0)$. Using antithetic variates thus doubles the number of sample paths that will be generated from a fixed budget of random number draws. In practice, the variance reduction associated with antithetic variates is often relatively modest.

14.6.4.2 Control Variates

As discussed in Chapter 3, the basic (product-based) control variate method involves determining a set of securities (control variates) that i) have payouts close to that of the instrument we are trying to price; and ii) have known expected values in the probability measure in which we simulate. Obvious control variates in the LM model include (portfolios of) zero-coupon bonds and caplets. Due to discretization errors in generation of sample paths, we should note, however, that the sample means of zero-coupon bonds and caplets will deviate from their true continuous-time means with amounts that depend on the time step and the discretization scheme employed. This error will nominally cause a violation of condition ii) — we are generally able only to compute in closed-form the continuous-time expected values — but the effect is often benign and will theoretically be of the same order[27] as the weak convergence order of the discretization scheme employed. Swaptions can also be included in the control variate set, although additional care must be taken here due to the presence of hard-to-quantify approximation errors in the formulas in Section 14.4.2. See Jensen and Svenstrup [2003] for an example of using swaptions as control variates for Bermudan swaptions.

An alternative interpretation of the control variate idea involves pricing a particular instrument using, in effect, two different LM models, one of which allows for an efficient computation of the instrument price, and one of which is the true model we are interested in applying. We shall return to this in Chapter 25.

Finally, the dynamic control variate method, based on the idea that an (approximate) self-financed hedging strategy could be a good proxy for the value of a security, is available for LM models as well. The method was developed in Section 3.4.3.2.

[27] Suppose that we estimate $\mathrm{E}(X) \approx \mathrm{E}(X' + Y' - \mu_Y)$, where $\mu_Y = \mathrm{E}(Y') + O(\Delta^p)$ and $\mathrm{E}(X') = \mathrm{E}(X) + O(\Delta^p)$. Then clearly also $\mathrm{E}(X' + Y' - \mu_Y) = \mathrm{E}(X) + O(\Delta^p)$.

14.6.4.3 Importance Sampling

Importance sampling techniques have so far found relatively limited use in the simulation of LM models, although Capriotti [2007] demonstrates that least-squares importance sampling (see Section 3.4.4.4) gives good results when pricing simple European options (caps and swaptions) in a three-factor log-normal LM model. As the variance reduction efficiency of importance sampling depends strongly on the payout, it is, however, unclear to what extent the results in Capriotti [2007] carry over to more complex security payouts (and models, for that matter).

Probably the most fruitful application of importance sampling in LM modeling is in the pricing of securities with a knock-out barrier. The basic idea is here that sample paths are generated conditional on a barrier not being breached, ensuring that all paths survive to maturity; this conditioning step induces a change of measure. We will expose the details of this technique in Chapter 20, where we discuss the pricing of the TARN product introduced in Section 5.15.2.

15

The Libor Market Model II

For the sake of cohesion, our discussion of LM models in Chapter 14 silently skipped over a number of practical issues. Chief amongst these is the problem of how to construct an entire (continuous) discount curve from knowledge of only a finite set of simply compounded Libor rates. This surprisingly subtle issue shall be discussed in this chapter, along with a select set of other advanced pricing and calibration topics in LM modeling. For instance, we provide a number of extensions to the stochastic volatility setup of Chapter 14 and also show how to construct swaption pricing formulas more accurate and more general than those in Chapter 14. We cover the generic problem of evolving separate discount and forward curves, and also include brief discussions of the so-called *swap market models* and of LM models with "near-Markov" structure. The latter topic shall be taken up again in Chapter 25.

15.1 Interpolation

The simulation schemes that we developed so far (see Section 14.6) allow us to obtain at any time t a vector of forward Libor rates on a pre-specified tenor structure. As should be obvious, and as pointed out previously in Sections 14.1.2 and 14.2.3, this information is not sufficient to recover the full interest rate yield curve at time t. At the very least, to be able to compute $P(t, T_n)$ for all n (see (14.3)), we need to additionally establish the *front stub* discount factor $P(t, T_{q(t)})$. In addition, as many actual security payoffs dictate that we calculate $P(t, T)$ for an arbitrary T, the *back stub* (forward) discount factor $P(t, T, T_{q(T)}) = P(t, T_{q(T)})/P(t, T)$ will also be required, since

$$P(t, T) = P\left(t, T_{q(t)}\right) \times \left(\prod_{i=q(t)}^{q(T)-1} \left(1 + \tau_i L_i(t)\right)^{-1}\right) / P\left(t, T, T_{q(T)}\right).$$

Both the front and back stubs cannot, in general, be expressed as a function of Libor rates on a fixed tenor structure.

There are a number of approaches that could be employed to obtain the front and back stub in a simulation. We start with the back stub as it is somewhat easier to handle.

15.1.1 Back Stub, Simple Interpolation

Let us fix the discount factor maturity time T, and set $m = q(T)$ such that

$$T_{m-1} \leq T < T_m.$$

Observe that as $T \to T_{m-1}$ the back stub $P(t,T,T_m) = P(t,T_m)/P(t,T)$ converges to $P(t,T_{m-1},T_m)$, a discount factor that can be calculated from Libor rates as $P(t,T_{m-1},T_m) = (1 + L_{m-1}(t)\tau_{m-1})^{-1}$. At the other limit, when $T \to T_m$, the back stub converges to 1. Hence, it seems reasonable to approximate $P(t,T,T_m)$ by interpolating between these two known extremes. This idea gives rise to a number of plausible schemes that we now proceed to describe.

A particularly simple idea is to apply linear interpolation directly to bond prices, resulting in the scheme

$$P(t,T,T_m) = \frac{T - T_{m-1}}{T_m - T_{m-1}} + \frac{T_m - T}{T_m - T_{m-1}} P(t,T_{m-1},T_m). \tag{15.1}$$

Using $P(t,T_{m-1},T) = P(t,T_{m-1},T_m)/P(t,T,T_m)$ as the interpolation variable instead, another linear interpolation scheme arises:

$$P(t,T_{m-1},T) = \frac{T_m - T}{T_m - T_{m-1}} + \frac{T - T_{m-1}}{T_m - T_{m-1}} P(t,T_{m-1},T_m),$$

or

$$P(t,T,T_m) = \frac{P(t,T_{m-1},T_m)}{\frac{T_m - T}{T_m - T_{m-1}} + \frac{T - T_{m-1}}{T_m - T_{m-1}} P(t,T_{m-1},T_m)}. \tag{15.2}$$

Alternatively, we can apply piecewise constant interpolation to continuously compounded instantaneous forward rates $f(t,u)$ for $u \in [T_{m-1},T_m]$, yielding

$$-\frac{1}{T_m - T} \ln P(t,T,T_m) = -\frac{1}{T_m - T_{m-1}} \ln P(t,T_{m-1},T_m),$$

or, in terms of forward bond prices,

$$P(t,T,T_m) = P(t,T_{m-1},T_m)^{\frac{T_m - T}{T_m - T_{m-1}}}. \tag{15.3}$$

Yet another interpolation scheme is obtained by constant interpolation of simply compounded rates,

$$\frac{1}{T_m - T}\left(\frac{1}{P(t,T,T_m)} - 1\right) = \frac{1}{T_m - T_{m-1}}\left(\frac{1}{P(t,T_{m-1},T_m)} - 1\right),$$

resulting in the scheme

$$P(t,T,T_m) = \left(\frac{T - T_{m-1}}{T_m - T_{m-1}} + \frac{T_m - T}{T_m - T_{m-1}}\frac{1}{P(t,T_{m-1},T_m)}\right)^{-1}. \quad (15.4)$$

While the interpolation schemes (15.1), (15.2), (15.3), and (15.4) are all simple to understand and to apply, they are ultimately flawed as they violate the basic no-arbitrage conditions

$$P(0,T_{m-1},T) = \mathrm{E}^{T_{m-1}}\left(P(t,T_{m-1},T)\right), \quad (15.5)$$

$$P(0,T_{m-1},T_m) = \mathrm{E}^{T_{m-1}}\left(P(t,T_{m-1},T_m)\right). \quad (15.6)$$

Any interpolation scheme, once we apply the expected value operator, imposes a certain relationship on discount bonds at time 0, a relationship that, in general, will not be satisfied by actual market prices. Taking as an example (15.2), applying the expectation operator $\mathrm{E}^{T_{m-1}}$ and using (15.5), we obtain

$$P(0,T_{m-1},T) = \frac{T_m - T}{T_m - T_{m-1}} + \frac{T - T_{m-1}}{T_m - T_{m-1}}P(0,T_{m-1},T_m).$$

a relationship between time 0 discount bond prices that is unlikely to be satisfied *a priori*.

15.1.2 Back Stub, Arbitrage-Free Interpolation

To ensure that observable relationships between time 0 discount bond prices are respected, consider choosing an arbitrary constant $\alpha(T)$ and setting

$$P(t,T_{m-1},T) = P(0,T_{m-1},T) + \alpha(T)\left(P(t,T_{m-1},T_m) - P(0,T_{m-1},T_m)\right). \quad (15.7)$$

Clearly, the additive scheme (15.7) will preserve (15.5) as long as (15.6) is satisfied — and (15.6) is essentially a no-arbitrage condition for discount bonds maturing on the tenor structure and is guaranteed by the LM model construction itself.

We can regard $\alpha(\cdot)$ as a function of maturity time[1]; for consistency $\alpha(T_m)$ must equal 1 and $\alpha(T_{m-1}) = 0$, but beyond this there are few restrictions on $\alpha(T)$. Yet it is not advisable to specify $\alpha(T)$ arbitrarily, as this may affect model dynamics in unintended ways. To devise a reasonable approach to the definition of $\alpha(T)$, we note from (15.7) that

$$dP(t,T_{m-1},T) = O(dt) + \alpha(T)\,dP(t,T_{m-1},T_m).$$

[1] And, implicitly, calendar time, which we ignore here as we work with a fixed t.

On the other hand, in an HJM model we have

$$dP\left(t,T_{m-1},T\right)/P\left(t,T_{m-1},T\right) = O\left(dt\right) + \left(\sigma_P(t,T_{m-1}) - \sigma_P\left(t,T\right)\right)^\top dW(t),$$

and

$$dP\left(t,T_{m-1},T_m\right)/P\left(t,T_{m-1},T_m\right) = O\left(dt\right) + \left(\sigma_P(t,T_{m-1}) - \sigma_P\left(t,T_m\right)\right)^\top dW(t),$$

from which we conclude that $\alpha(T)$ is linked to the ratio of forward discount bond volatilities. Exploiting this link, we may define $\alpha(T)$ from, for instance. the equation

$$P\left(t,T_{m-1},T\right)\left\|\sigma_P(t,T_{m-1}) - \sigma_P\left(t,T\right)\right\| = \alpha(T)P\left(t,T_{m-1},T_m\right)\left\|\sigma_P(t,T_{m-1}) - \sigma_P\left(t,T_m\right)\right\|.$$

Then[2]

$$\begin{aligned}\alpha(T) &= \frac{P\left(t,T_{m-1},T\right)}{P\left(t,T_{m-1},T_m\right)}\frac{\left\|\sigma_P(t,T_{m-1}) - \sigma_P\left(t,T\right)\right\|}{\left\|\sigma_P(t,T_{m-1}) - \sigma_P\left(t,T_m\right)\right\|}\\ &\approx \frac{P\left(0,T_{m-1},T\right)}{P\left(0,T_{m-1},T_m\right)}\frac{\left\|\sigma_P(t,T_{m-1}) - \sigma_P\left(t,T\right)\right\|}{\left\|\sigma_P(t,T_{m-1}) - \sigma_P\left(t,T_m\right)\right\|}.\end{aligned} \tag{15.8}$$

As we have seem in Chapter 14, the LM model does not uniquely define all the bond volatilities in (15.8) and we would need to interpolate available volatilities to compute (15.8). Note that (15.8) turns the problem of interpolating bond prices into a problem of interpolating bond *volatilities* instead. This point of view is advantageous as there are few, if any, arbitrage restrictions on interpolating the volatilities of bonds, as opposed to the bonds themselves. One can, for example, choose a linear interpolation to obtain $\sigma_P(t,T)$ from $\sigma_P(t,T_{m-1})$ and $\sigma_P(t,T_m)$ or, for a more sophisticated scheme, draw inspiration from the shape of forward volatilities in a mean-reverting one-factor Gaussian model. To explore the latter idea, recall that in the one-dimensional Gaussian model with constant mean reversion (Section 10.1.2),

$$\sigma_P(t,T) = \sigma(t)\frac{1-e^{-\varkappa(T-t)}}{\varkappa}, \tag{15.9}$$

where $\varkappa$ is the mean reversion and $\sigma(t)$ is the short rate volatility. Plugging into (15.8) and rearranging we get

[2]As discount bond volatility vectors $\sigma_P(t,T)$ are generally non-deterministic, a suitable approximation is required to make sense of this formula. This is most easily done by freezing any state variables appearing in bond volatilities, such as Libor rates, at their time 0 values.

$$\alpha(T) \approx \frac{P\left(0, T_{m-1}, T\right)}{P\left(0, T_{m-1}, T_m\right)} \frac{1 - e^{-\varkappa(T - T_{m-1})}}{1 - e^{-\varkappa(T_m - T_{m-1})}}.$$

The mean reversion parameter $\varkappa$ could be either set as part of the user input, or obtained by best-fitting the Gaussian parametric form (15.9) to the volatility structure of the LM model.

15.1.3 Back Stub, Interpolation Inspired by the Gaussian Model

Above, we used a volatility parameterization inspired by the Gaussian model to construct an interpolation scheme for bond prices. An alternative, more direct, approach to extracting information from a Gaussian model in the interpolation exercise is to let the bond reconstitution formulas from the Gaussian model form the basis for interpolation. To demonstrate, recall that in the one-factor Gaussian model with constant mean reversion $\varkappa$ and short rate volatility $\sigma(t)$ (see Section 10.1.2.2),

$$\begin{aligned} P\left(t, T_{m-1}, T\right) = P\left(0, T_{m-1}, T\right) \exp\left(-G\left(T_{m-1}, T\right) e^{-\varkappa(T_{m-1} - t)} x(t)\right) \\ \times \exp\left(-\frac{1}{2}\left(G\left(t, T\right)^2 - G\left(t, T_{m-1}\right)^2\right) y(t)\right), \end{aligned} \tag{15.10}$$

and

$$\begin{aligned} P(t, T_{m-1}, T_m) = P(0, T_{m-1}, T_m) \exp\left(-G(T_{m-1}, T_m) e^{-\varkappa(T_{m-1} - t)} x(t)\right) \\ \times \exp\left(-\frac{1}{2}\left(G(t, T_m)^2 - G(t, T_{m-1})^2\right) y(t)\right), \end{aligned} \tag{15.11}$$

where

$$G\left(t, T\right) = \frac{1 - e^{-\varkappa(T - t)}}{\varkappa}, \tag{15.12}$$

$$y(t) = e^{-2\varkappa t} \int_0^t e^{2\varkappa s} \sigma(s)^2 \, ds, \tag{15.13}$$

and $x(t)$ is the short rate state. Eliminating $x(t)$ in (15.10) and (15.11) defines a relationship between bond prices,

$$\begin{aligned} &\frac{1}{G(T_{m-1}, T)} \left(\ln \frac{P(t, T_{m-1}, T)}{P(0, T_{m-1}, T)} + \left(G(t, T)^2 - G(t, T_{m-1})^2\right) \frac{y(t)}{2}\right) \\ &= \frac{1}{G(T_{m-1}, T_m)} \left(\ln \frac{P(t, T_{m-1}, T_m)}{P(0, T_{m-1}, T_m)} + \left(G(t, T_m)^2 - G(t, T_{m-1})^2\right) \frac{y(t)}{2}\right) \end{aligned}$$

or, after a few additional manipulations,

$$P(t,T_{m-1},T) = P(0,T_{m-1},T)\left(\frac{P(t,T_{m-1},T_m)}{P(0,T_{m-1},T_m)}\right)^{\frac{G(T_{m-1},T)}{G(T_{m-1},T_m)}}$$
$$\times \exp\left(\frac{1}{2}\frac{G(T_{m-1},T)}{G(T_{m-1},T_m)}\left(G(t,T_m)^2 - G(t,T_{m-1})^2\right)y(t)\right)$$
$$\times \exp\left(-\frac{1}{2}\left(G(t,T)^2 - G(t,T_{m-1})^2\right)y(t)\right). \tag{15.14}$$

In (15.14), the volatility $\sigma(t)$ and the mean reversion $\varkappa$ can be obtained, for example, by fitting the Gaussian volatility structure to the volatilities of Libor rates generated by the LM model itself. High level of precision is not required here; we can take, say, $\varkappa = 0$ and $\sigma(t) = \|\sigma_m(t)\|$ where $\sigma_m(t)$ is the vector of volatilities for the m-th Libor rate $L_m(t)$ (the comment of footnote 2 applies here as well, to Libor volatilities).

The scheme (15.14) would be arbitrage-free in the context of a Gaussian model, i.e. if the expected value in (15.5) were computed in the Gaussian model. In the LM model the equality (15.5) would not hold exactly, but would be a good approximation as long as the choice of the volatility/mean reversion in the scheme is reasonably consistent with the actual LM model volatility structure. While we have no strong opinions on the matter, on the whole (15.14) tends to be our preferred choice for the back-stub interpolation as it is nearly arbitrage-free and is perhaps somewhat easier to implement and maintain than other schemes from Section 15.1.2.

15.1.4 Front Stub, Zero Volatility

Having considered various options for the back stub, let us now focus on the front stub. The simplest way to "complete" the LM model volatility specification is to specify that the front stub bond has no volatility, i.e.

$$\sigma_P\left(t,T_{q(t)}\right) \equiv 0 \tag{15.15}$$

in the notation of Section 14.2.3. This automatically specifies the front stub interpolation scheme, as

$$P\left(t,T_{q(t)}\right) = P\left(T_{q(t)-1},t,T_{q(t)}\right),$$

where the right-hand side can be computed from the previous results on *back* stub interpolation.

The choice (15.15) was first proposed in Brace et al. [1997] and implies that the continuously rolling money market account $\beta(t)$ coincides with the discrete numeraire $B(t)$, whereby the risk-neutral measure is identical to the spot Libor measure. The bond volatilities for "core" bonds — that is, the bonds paying on the dates in the tenor structure — are explicitly given by

$$\sigma_P(t,T_n) = \sum_{j=q(t)}^{n-1} \frac{\tau_j \sigma_j(t)}{1+\tau_j L_j(t)}.$$

While certainly technically convenient, (15.15) leads to unrealistic dynamics of the yield curve. To give an example, assume that the LM model is specified with a 6 month tenor; i.e. τ_n's are all approximately 0.5 and $T_1 = 0.5$. For a 3 month option on a 3 month rate, i.e. a caplet with the payoff

$$(L(0.25, 0.25, 0.5) - K)^+, \tag{15.16}$$

(15.15) implies that the rate $L(t, 0.25, 0.5)$ has no volatility for $t \in [0, 0.25]$, i.e. the option with the payoff (15.16) will be priced at its intrinsic value. Clearly, this is unrealistic and will not be consistent with an actual market price. This example is not as contrived as one might think, as many complex derivatives have at least some exposure to short-expiry, short-tenor volatility. The specification (15.15) forces such volatility to zero, and cannot be recommended.

The analysis of the shortcomings of (15.15) puts into focus the requirements that one would wish to impose on all "good" front-stub interpolation schemes. In particular, we will be looking for schemes that recover both market-implied forward rates, as well as volatilities for rates with non-standard expiries and tenors. With "non-standard", we here refer to rates that are not aligned with the tenor structure. For example, for a 6 month tenor LM model, we can examine the dynamics and values of 3 month tenor forward rates maturing in 1 month, 2 months, ..., or, in 3 months time, we can look at rates with 3 month, 6 month, 9 month, ..., tenors. Note that such an analysis will also help uncover problems with the *back*-stub interpolation, should there be any.

15.1.5 Front Stub, Exogenous Volatility

At a fundamental level, the volatilities $\sigma_P(t, T_{q(t)})$ for all t constitute extra information that is required to specify the model behavior between tenor dates. This information would define the dynamics of the stub bonds $P(t, T_{q(t)})$ (for all t) which, in principle, is all that is required to define the values of *all* bonds, as for any $0 \le s < t$,

$$P(s,t) = P(s, T_{q(s)}) P(s, T_{q(s)}, T_{q(t)}) \mathrm{E}_s^{T_{q(t)}} \left(\frac{1}{P(t, T_{q(t)})} \right). \tag{15.17}$$

As this formula specifies the values of all discount bonds, in principle it makes back-stub interpolation methods from earlier in the section redundant. In practice, however, calculation of the expected value in (15.17) is non-trivial, especially for $s \ll t$, so the application of (15.17) is best left for the case of $t - s$ being small, when accurate approximations are easier to obtain.

The information on the stub bond volatilities $\sigma_P(t, T_{q(t)})$ should be supplied *in addition* to the basic Libor rate volatility structure of the LM model. To make such exogenous volatility specification easier on the model user, we recommend that some additional structure is added through

simplifying assumptions. For instance, an obvious simplification is to assume that the stub volatilities are both time-homogeneous and identical for all periods in the tenor structure; in other words, $\sigma_P(t,u) = \sigma_{\text{stub}}(u-t)$ for some chosen $\sigma_{\text{stub}}(\tau)$, $t \le u < T_{q(t)}$. This reduces the problem to specifying a stub volatility function of only one argument; this function could be calibrated to short-dated options on short-tenor rates.

In the scheme above, we notice that σ_{stub} is vector-valued. To reduce the specification burden further, we can inherit the correlation structure from that of the core Libor forward rates, and set

$$\sigma_{\text{stub}}(u-t) = \frac{\|\sigma_{\text{stub}}(u-t)\|}{\|\sigma_1(t)\|}\sigma_1(t),$$

where $\sigma_1(t)$ is the vector of volatilities for the front Libor rate $L_1(t)$. With this scheme, we only need to specify a *scalar* function $\|\sigma_{\text{stub}}(\tau)\|$.

The dimensionality of the additional model inputs could be further reduced by using a particular parametric form. Drawing inspiration — yet again — from the one-factor Gaussian model, we could for instance specify

$$\|\sigma_{\text{stub}}(\tau)\| = \overline{\sigma}_{\text{stub}} \frac{1 - e^{-\varkappa_{\text{stub}}\tau}}{\varkappa_{\text{stub}}} \tag{15.18}$$

for given constants $\overline{\sigma}_{\text{stub}}$ and $\varkappa_{\text{stub}}$. With this, we have reduced the specification of the front-stub volatility function to just two constants.

Once the stub volatility function is specified, it needs to be incorporated into a Monte Carlo simulation. Suppose we need to perform a time step from t to t' where t is one of the tenor dates, $t = T_{m-1}$ and $t' < T_m$. Then, in addition to the standard MC step (see (14.64))

$$\widehat{L}_n(t') = \widehat{L}_n(t) + \sigma_n(t)^\top \mu_n(t)\,(t'-t) + \widehat{M}_n(t,t'),\ n = m,\ldots,N-1,$$

we also need an update equation for $P(t',T_m)$. Over the period $s \in [T_{m-1},T_m]$, the spot numeraire coincides, up to a constant, with the bond price $P(t,T_m)$ (see (14.8)). Hence, in spot measure, the forward bond price $P(s,t',T_m)$, $t \le s \le t'$, satisfies

$$\begin{aligned} dP(s,t',T_m)/P(s,t',T_m) &= \|\sigma_{\text{stub}}(s,t',T_m)\|^2\,ds \\ &\quad + \sigma_{\text{stub}}(s,t',T_m)^\top\,dW^B(s), \end{aligned}$$

where we have defined

$$\sigma_{\text{stub}}(s,t',T_m) \triangleq \sigma_{\text{stub}}(T_m - s) - \sigma_{\text{stub}}(t'-s).$$

A simple log-Euler scheme for the bond is given by

$$\begin{aligned} \widehat{P}(t',T_m) &= \widehat{P}(t,t',T_m)\exp\left(\sigma_{\text{stub}}(t,t',T_m)^\top\sqrt{t'-t}Z\right) \\ &\quad \times \exp\left(\frac{1}{2}\|\sigma_{\text{stub}}(t,t',T_m)\|^2(t'-t)\right), \end{aligned} \tag{15.19}$$

where Z is a draw from the standard one-dimensional Gaussian distribution. This scheme, together with the update equations for $\{\widehat{L}_n(t)\}_{n=m}^{N-1}$, defines the Monte Carlo step. Extensions to more sophisticated discretization schemes for $\widehat{P}(t', T_m)$ are straightforward, but rarely needed.

Since we assumed that $t = T_{m-1}$, the term $\widehat{P}(t, t', T_m)$ in (15.19) is available at time t from Libor rates using *back*-stub interpolation methods discussed earlier in the section. If we, in addition, need to calculate $\widehat{P}(t'', T_m)$ for some t'', $t' < t'' < T_m$, we can either apply (15.19) to step from t to t'' directly, or to suitably modify (15.19) to step from t' to t''. In the latter case we will, however, need to be able to simulate other short-dated discount bond values, i.e. $P(t', u)$ for $u < T_m$ (something that may be required for other purposes as well); this can be incorporated into the time-stepping algorithm in the same way as for $P(t', T_m)$.

More pragmatically, instead of simulating a whole collection of short-tenor bonds $P(\cdot, u_1), \ldots, P(\cdot, u_k)$, $t' < u_1 < \ldots < u_k \leq T_m$, that may be required at time t', we may choose to propagate a single discount bond with the shortest time to maturity, i.e. $P(\cdot, u_1)$, with other bond prices subsequently obtained by one of the *back*-stub interpolation schemes discussed earlier. Alternatively we could propagate the stub bond $P(\cdot, T_m)$ and use (15.17) to approximate all $P(t', u_1), \ldots, P(t', u_{k-1})$ (with, perhaps, some approximation to calculate the expected values):

$$P(t', u_l) = P(t', T_m)\mathrm{E}_{t'}^{T_m}\left(\frac{1}{P(u_l, T_m)}\right).$$

Proceeding by either of these methods, the yield curve at time t', as well as the spot numeraire, are then fully defined by the augmented state vector $\{P(t', T_{q(t')}), L_{q(t')}(t'), \ldots, L_{N-1}(t')\}$. Moreover, by construction, the Monte Carlo scheme recovers the expected values of short-tenor rates and their volatilities, as specified by the stub volatility function (within Monte Carlo error and up to discretization bias, of course). The disadvantage of the scheme is the need to carry an extra state variable for each time step.

Remark 15.1.1. Parameterizations such as (15.18) introduce volatility functions (of stub bonds) that are not constant between tenor dates. A similar idea could be applied to core Libor rate volatilities as well, as an extension of the simple piecewise constant interpolation we discussed in Section 14.5.3. This could be required to be able to match shorter-dated options on Libor rates, e.g. a 3 month option on a 6 month Libor rate in an LM model with 6 months between dates in the tenor structure. For example, drawing inspiration from a Gaussian model we could amend (14.42) by a final-period Libor volatility specification of the form

$$\|\lambda_k(t)\| = e^{\varkappa_{\text{short}}(t-T_{k-1})}\|\lambda_k(T_{k-1})\| = e^{\varkappa_{\text{short}}(t-T_{k-1})}\|\lambda_{k,k}\|, \ t \in [T_{k-1}, T_k),$$

where $\varkappa_{\text{short}}$ is user-specified or calibrated to short-dated options.

15.1.6 Front Stub, Simple Interpolation

As an alternative to the scheme in the previous section, one could contemplate avoiding simulation of the stub bond altogether by interpolating it from the original state vector of Libor rates. To demonstrate, let us assume the same setup as in the previous section. Then, at time t', simulated forward bond prices $P(t', T_n, T_{n+1})$, $n \geq m = q(t')$, are available; the goal is to construct $P(t', T_m)$ from these. A simple interpolation scheme could, for example, specify that the simulated instantaneous forward rates $f(t', u)$ are constant over $u \in [t', T_{m+1}]$. This would link $P(t', T_m)$ to $P(t', T_m, T_{m+1})$ in the following way:

$$P(t', T_m) = P(t', T_m, T_{m+1})^{\frac{T_m - t'}{T_{m+1} - T_m}} . \tag{15.20}$$

This scheme, while simple, does not satisfy our notion of a "good" one. In particular, it does not recover the time 0 forward discount bond value, as generally

$$P(0, t', T_m) \neq \mathrm{E}^{t'} \left(P(t', T_m, T_{m+1})^{\frac{T_m - t'}{T_{m+1} - T_m}} \right) .$$

Nor does the scheme recover the market volatility of $P(t', T_m)$, as now

$$\ln P(t', T_m) = \frac{T_m - t'}{T_{m+1} - T_m} \ln P(t', T_m, T_{m+1}) ,$$

and the volatilities of $\ln P(t', T_m)$ and $\ln P(t', T_m, T_{m+1})$ are then linked in a specific way that does not necessarily hold in the market.

Both of the shortcomings above can be addressed in a more sophisticated interpolation scheme, to be discussed in the next section. Before proceeding, however, we note that a similar, albeit somewhat more general, scheme has been proposed by Schlögl [2002]:

$$\frac{1}{P(t', T_m)} = 1 + (T_m - t') \left(\xi(t') L_{m-1}(T_{m-1}) + (1 - \xi(t')) L_m(t') \right) ,$$

where an essentially arbitrary deterministic function $\xi(t)$ is chosen so that $1 = \xi(T_{m-1}) \geq \xi(t') \geq \xi(T_m) = 0$. While the volatility of the stub bond $P(t', T_m)$ could be manipulated by the choice of the function $\xi(t)$, the forward discount bond value is still not recovered by the model.

15.1.7 Front Stub, Interpolation Inspired by the Gaussian Model

The idea of employing a bond reconstitution formula from a Gaussian model, already used for the back stub, can be applied to the front stub as well. Here we assume that for short tenors, the LM model can be locally approximated by a one-factor Gaussian model. Recall that in the latter,

$$P(t', T_m) = P(0, t', T_m) \exp\left(-G(t', T_m)\, x(t') - \frac{1}{2} G(t', T_m)^2\, y(t')\right)$$

and

$$\begin{aligned} P(t', T_m, T_{m+1}) &= P(0, T_m, T_{m+1}) \\ &\quad \times \exp\left(-\left(G(t', T_{m+1}) - G(t', T_m)\right) x(t')\right) \\ &\quad \times \exp\left(-\frac{1}{2}\left(G(t', T_{m+1})^2 - G(t', T_m)^2\right) y(t')\right). \end{aligned}$$

Solving the second equation for $x(t')$ and substituting into the first one, we obtain

$$\begin{aligned} \ln \frac{P(t', T_m)}{P(0, t', T_m)} + \frac{1}{2} G(t', T_m)^2\, y(t') &= \frac{G(t', T_m)}{G(t', T_{m+1}) - G(t', T_m)} \\ &\quad \times \left(\ln \frac{P(t', T_m, T_{m+1})}{P(0, T_m, T_{m+1})} + \frac{1}{2}\left(G(t', T_{m+1})^2 - G(t', T_m)^2\right) y(t')\right), \end{aligned}$$

so that

$$\begin{aligned} P(t', T_m) &= P(0, t', T_m) \left(\frac{P(t', T_m, T_{m+1})}{P(0, T_m, T_{m+1})}\right)^{\frac{G(t', T_m)}{G(t', T_{m+1}) - G(t', T_m)}} \\ &\quad \times \exp\left(\frac{1}{2} G(t', T_m)\, G(t', T_{m+1})\, y(t')\right). \end{aligned} \tag{15.21}$$

The interpolation scheme depends on two parameters, the mean reversion $\varkappa$ in (15.12) and the short rate volatility $\sigma(t)$ in (15.13) of the Gaussian model. The latter could be approximated by the volatilities of the front Libor rate, ensuring that the forward price of the front-stub discount bond is approximately recovered by the model. The mean reversion plays an important role here as it defines the relative magnitude of the volatility of (log) $P(t', T_m)$ in relationship to (log) $P(t', T_m, T_{m+1})$. As such, $\varkappa$ can be used to set the stub bond volatility to, or near, its market-implied value.

Empirical evidence shows a close fit between time 0 market-observed short-tenor rates and those computed from the model in the manner described above. Also, with the right choice of mean reversion, the market-implied front-stub volatility is recovered as well. We recommend this scheme for most applications.

15.2 Advanced Approximations for Swaption Pricing via Markovian Projection

Having wrapped up the topic of interpolation, let us now go back to the problem of approximate pricing of vanilla options — such as swaptions

— in LM models. While the pricing approximation for swaptions derived in Section 14.4.2 has proved to be remarkably successful for calibration purposes, it does have limitations, especially for longer-dated options on longer-dated swaps and for swaptions that are not at-the-money. There are several improvements that could be made to the basic approximation. For instance, in (14.33) the skew of a swap rate is taken to be the same as the (common) skew of all Libor rates, yet numerical simulation shows that this is not the case. For instance, in a log-normal LM model, a long-expiry, long-tenor swap rate would have a "super log-normal" skew, i.e. implied log-normal swaption volatilities would trend up with the strike. Hence, a more accurate estimate of the swap rate skew from Libor rate skews would be useful. Moreover, the accuracy of the swap rate volatility calculations can be improved by a more careful analysis than that of Section 14.4.2.

Before proceeding, let us first enlarge the model setup from Section 14.2.5 somewhat. Specifically, we wish to address the fact that the specification of the model in Section 14.2.4 (or Section 14.2.5) uses the same time-homogeneous local volatility function for all Libor rates. Such a setup implies that swaptions of all tenors and expiries have essentially the same volatility skew, a model feature that is inconsistent with current market reality. For the Libor market model to be able to match the swaption volatilities for all expiries, tenors *and strikes*, it is necessary to assume that different Libor rates have different local volatility functions, and that those functions explicitly depend on time. In the stochastic volatility model, it may also be necessary to assume that the volatility of variance is time-dependent, so that the curvatures of smiles of different swaptions are allowed to differ. More advanced methods are then required to derive approximations to swap rate volatilities in such a, more generic, specification. In total, we therefore consider the following generalization[3] of the specification (14.15)–(14.16):

$$dz(t) = \theta\left(z_0 - z(t)\right)dt + \eta(t)\psi\left(z(t)\right)\,dZ(t), \tag{15.22}$$

$$dL_n(t) = \sqrt{z(t)}\varphi_n\left(t, L_n(t)\right)\lambda_n(t)^{\top}dW^{T_{n+1}}(t), \quad n = 1, \ldots, N-1. \tag{15.23}$$

In particular, the volatility of variance parameter $\eta(t)$ depends on t, and the DVFs $\varphi_n(t,x)$ are now specific to each Libor rate (i.e. depend on n) and may also depend on time t. Without loss of generality, we assume

$$\varphi_n\left(t, L_n(0)\right) = 1.$$

As in other applications of DVF modeling, we assume that the functions $\varphi_n(t,x)$ are well-approximated by their first-order expansions,

[3] As before, we here assume zero correlation between $Z(t)$ and $W(t)$, but relax this assumption in Section 15.6.

$$\varphi_n(t,x) \approx 1 + b_n(t)\,(x - L_n(0))\,, \tag{15.24}$$
$$b_n(t) \triangleq \frac{\partial}{\partial x}\varphi_n(t, L_n(0))\,.$$

In practical applications, this usually means that $\varphi_n(t,x)$'s are either linear or power functions, see Table 14.1.

As we did in Section 14.6.2.1, let us denote by $\mathbf{L}(t)$ the vector of all Libor rates, i.e. $\mathbf{L}(t) = (L_1(t), \ldots, L_{N-1}(t))^\top$, with the convention that $L_i(t) \equiv L_i(T_i)$ for $i < q(t)$. Continuing with the notations of Section 14.4.2, let $S(t) = S_{j,k-j}(t)$ be a particular swap rate. The dynamics of $S(t)$ in the model (15.22)–(15.23) are easy to write down,

$$dS(t) = \sqrt{z(t)}\lambda_S(t, \mathbf{L}(t))^\top\, dW^A(t), \tag{15.25}$$

where

$$\lambda_S(t, \mathbf{L}(t)) = \sum_{n=j}^{k-1} \frac{\partial S(t)}{\partial L_n(t)} \varphi_n(t, L_n(t))\,\lambda_n(t), \tag{15.26}$$

and $W^A(t)$ is a Brownian motion in the annuity measure Q^A for $S(t)$. Moreover, the dynamics can be written in a one-dimensional form,

$$dS(t) = \sqrt{z(t)}\,\|\lambda_S(t, \mathbf{L}(t))\|\, dY^A(t), \tag{15.27}$$

where $Y^A(t)$ is a one-dimensional Brownian motion in Q^A.

The following result serves as a starting point for various useful approximations.

Proposition 15.2.1. *For the purposes of European swaption valuation, the dynamics of the swap rate $S(t)$ in Q^A are approximately given by the following displaced log-normal stochastic volatility SDE*

$$dS(t) \approx \sqrt{z(t)}p_S(t)\,(1 + b_S(t)\,(S(t) - S(0)))\, dY^A(t), \tag{15.28}$$

with

$$p_S(t) = \left\|\lambda_S\left(t, \mathrm{E}^A(\mathbf{L}(t))\right)\right\|, \tag{15.29}$$

$$b_S(t) = \frac{1}{p_S(t)} \sum_{n=j}^{k-1} \frac{\partial \left\|\lambda_S\left(t, \mathrm{E}^A(\mathbf{L}(t))\right)\right\|}{\partial L_n(t)} \frac{\int_0^t \left(\lambda_n(u)^\top \lambda_S(u, \mathbf{L}(0))\right) du}{\int_0^t \|\lambda_S(u, \mathbf{L}(0))\|^2\, du}. \tag{15.30}$$

Proof. The proof relies on standard results on Markovian projection, see Appendix A. From (A.18), the European options on $S(t)$ in the model (15.27) have the same values as in the Markovian model

$$dS(t) = \sqrt{z(t)}\left(\mathrm{E}^A\left(\|\lambda_S(t, \mathbf{L}(t))\|^2 \middle| S(t)\right)\right)^{1/2} dY^A(t).$$

First, we approximate

$$\left(\mathrm{E}^A\left(\|\lambda_S(t,\mathbf{L}(t))\|^2\middle| S(t)\right)\right)^{1/2} \approx \mathrm{E}^A\left(\|\lambda_S(t,\mathbf{L}(t))\| \,|\, S(t)\right),$$

and then linearize $\|\lambda_S(t,\mathbf{L}(t))\|$ around $\mathrm{E}^A(\mathbf{L}(t))$,

$$\begin{aligned}\|\lambda_S\left(t,\mathbf{L}(t)\right)\| \approx &\left\|\lambda_S\left(t,\mathrm{E}^A\left(\mathbf{L}(t)\right)\right)\right\| \\ &+ \left(\nabla\left\|\lambda_S\left(t,\mathrm{E}^A\left(\mathbf{L}(t)\right)\right)\right\|\right)\left(\mathbf{L}(t)-\mathrm{E}^A\left(\mathbf{L}(t)\right)\right),\end{aligned}$$

where $\nabla = (\frac{\partial}{\partial L_1(t)},\ldots,\frac{\partial}{\partial L_{N-1}(t)})$ is the (row-vector) gradient.

Let us introduce the Gaussian approximation

$$d\widehat{L}_n(t) = \lambda_n(t)^\top dW^A(t), \quad d\widehat{S}(t) = \lambda_S\left(t,\mathbf{L}(0)\right)^\top dW^A(t), \tag{15.31}$$

so that we can approximate

$$\begin{aligned}\mathrm{E}^A\left(\mathbf{L}(t)-\mathrm{E}^A\left(\mathbf{L}(t)\right)\middle| S(t)=s\right) &\approx \mathrm{E}^A\left(\widehat{\mathbf{L}}(t)-\mathbf{L}(0)\middle|\widehat{S}(t)=s\right) \\ &= \frac{\mathrm{E}^A\left(\left(\widehat{\mathbf{L}}(t)-\mathbf{L}(0)\right)\left(\widehat{S}(t)-S(0)\right)\right)}{\mathrm{E}^A\left(\left(\widehat{S}(t)-S(0)\right)\left(\widehat{S}(t)-S(0)\right)\right)}\left(s-S(0)\right),\end{aligned}$$

where we have defined $\widehat{\mathbf{L}}(t) = (\widehat{L}_1,\ldots,\widehat{L}_{N-1})^\top$. The result follows. □

The price of a European swaption is given by the value of the option on $S(T_j)$ (times the annuity). The model (15.22) and (15.28) is a stochastic volatility model with time-dependent parameters. Using the methods of Chapter 9, effective, time-constant parameters can be derived, to facilitate fast pricing of European swaptions, as well as calibration of the model parameters to their market values. We do not repeat the relevant formulas here, but simply note that the total volatility, skew and volatility of variance of any swap rate are available as functions of the model parameters.

15.2.1 Advanced Formula for Swap Rate Volatility

In this section, we use the results of Proposition 15.2.1 to derive useful formulas for the swaption volatility. Recall (15.28), the approximate SDE for the swap rate used for European option pricing. The function $p_S(t)$ in (15.29) is the (time-dependent) swap rate volatility,

$$\begin{aligned}p_S(t) &= \left\|\lambda_S\left(t,\mathrm{E}^A\left(\mathbf{L}(t)\right)\right)\right\| \\ &= \left\|\sum_{n=j}^{k-1}\left.\frac{\partial S(t)}{\partial L_n(t)}\right|_{\mathbf{L}(t)=\mathrm{E}^A(\mathbf{L}(t))}\varphi_n\left(t,\mathrm{E}^A\left(L_n(t)\right)\right)\lambda_n(t)\right\|,\end{aligned}$$

where we have used (15.26). The equivalent quantity in the standard approximation of Section 14.4.2 is given by (compare to (14.33))

$$p_{S,\text{standard}}(t) = \left\| \sum_{n=j}^{k-1} \frac{\partial S(t)}{\partial L_n(t)}\bigg|_{\mathbf{L}(t)=\mathbf{L}(0)} \varphi_n\left(t, L_n(0)\right) \lambda_n(t) \right\|.$$

Hence, the improvements over the standard approximation come from evaluating the actual volatility function of the swap rate ($\lambda_S(t, \mathbf{L}(t))$) at $\mathbf{L}(t) = \mathrm{E}^A(\mathbf{L}(t))$ rather than at $\mathbf{L}(t) = \mathbf{L}(0)$; this is similar to the improvements obtained for the quasi-Gaussian model in Section 13.1.4.2. Clearly $\mathrm{E}^A(\mathbf{L}(t)) \neq \mathbf{L}(0)$; the difference can be approximated with the help of the following proposition.

Proposition 15.2.2. *For $j \leq n \leq k-1$, the expected value of the n-th Libor rate in the annuity measure is approximately given by*

$$\mathrm{E}^A\left(L_n(t)\right) = L_n(0)\left(1 + c_n(t)\right),$$

where

$$c_n(t) = \frac{1}{L_n(0)Q_n(0)} \sum_{i=j}^{k-1} \frac{\partial Q_n(0)}{\partial L_i(0)} \int_0^t \left(\lambda_i^\top(s)\lambda_n(s)\right) ds, \; Q_n(t) = \frac{A(t)}{P\left(t, T_{n+1}\right)},$$

with $A(t)$ defined in (14.29).

Proof. We have,

$$\mathrm{E}^A\left(L_n(t)\right) = Q_n^{-1}(0)\mathrm{E}^{T_{n+1}}\left(Q_n(t)L_n(t)\right).$$

Both $Q_n(t)$ and $L_n(t)$ are martingales in $\mathrm{Q}^{T_{n+1}}$. Applying Gaussian approximations,

$$d\widehat{Q}_n(t) \approx \lambda_{Q_n}(t)^\top dW^{T_{n+1}}(t), \quad d\widehat{L}_n(t) \approx \lambda_n^\top(t)\, dW^{T_{n+1}}(t),$$

where

$$\lambda_{Q_n}(t) = \sum_{i=j}^{k-1} \frac{\partial Q_n(0)}{\partial L_i(0)} \lambda_i(t),$$

we obtain

$$\mathrm{E}^{T_{n+1}}\left(Q_n(t)L_n(t)\right) \approx Q_n(0)L_n(0) + \int_0^t \left(\lambda_{Q_n}(s)^\top \lambda_n(s)\right) ds,$$

and the statement follows. □

The idea of employing $\mathrm{E}^A(\mathbf{L}(t))$ instead of $\mathbf{L}(0)$ in the standard swap rate volatility approximation can be used independently of any considerations involving time-dependent skews. In particular, the more accurate approximation can be applied directly in the statement of Proposition 14.4.3 when defining the value of $\lambda_S(t)$ in (14.33). The differences between the two formulas are small, but become noticeable for swaptions of longer-dated expiries and tenors.

15.2.2 Advanced Formula for Swap Rate Skew

In the approximate SDE (15.28) for the swap rate used for European option pricing, the parameter $b_S(t)$ controls the skew of the volatility smile. Define

$$v_n(t) = \frac{\partial S(t)}{\partial L_n(t)}, \quad n = j, \ldots, k-1,$$

(compare to (14.31)), and

$$v_{n,n'}(t) = \frac{\partial^2 S(t)}{\partial L_n(t) \partial L_{n'}(t)}, \quad n, n' = j, \ldots, k-1.$$

Proposition 15.2.3. *The time-dependent swaption skew $b_S(t)$ is approximately given by*

$$b_S(t) = \sum_{i,n=j}^{k-1} \frac{r_{S,n}(t)}{r_{S,i}(t)} v_{i,n}(0) \xi_i(t) + \sum_{i=j}^{k-1} b_i(t) v_i(0) \xi_i(t),$$

where

$$r_{S,i}(t) = \lambda_i(t)^\top \lambda_S\left(t, \mathbf{L}(0)\right), \quad r_S(t) = \left\|\lambda_S\left(t, \mathbf{L}(0)\right)\right\|^2,$$
$$\xi_i(t) = \frac{r_{S,i}(t) \int_0^t r_{S,i}(u)\, du}{r_S(t) \int_0^t r_S(u)\, du}.$$

Proof. Recall from Proposition 15.2.1,

$$\begin{aligned} b_S(t) &= \frac{1}{p_S(t)} \sum_{i=j}^{k-1} \frac{\partial \left\|\lambda_S\left(t, \mathbf{L}(t)\right)\right\|}{\partial L_i(t)}\bigg|_{\mathbf{L}(t)=\mathrm{E}^{\Lambda}(\mathbf{L}(t))} \frac{\int_0^t r_{S,i}(u)\, du}{\int_0^t r_S(u)\, du} \\ &= \sum_{i=j}^{k-1} \frac{\partial \ln \left\|\lambda_S\left(t, \mathbf{L}(t)\right)\right\|}{\partial L_i(t)}\bigg|_{\mathbf{L}(t)=\mathrm{E}^{\Lambda}(\mathbf{L}(t))} \frac{\int_0^t r_{S,i}(u)\, du}{\int_0^t r_S(u)\, du}. \end{aligned}$$

We have

$$\begin{aligned} \frac{\partial \ln \left\|\lambda_S\left(t, \mathbf{L}(t)\right)\right\|}{\partial L_i(t)} &= \frac{1}{\left\|\lambda_S\left(t, \mathbf{L}(t)\right)\right\|^2} \\ \times \sum_{n,n'=j}^{k-1} &\left(\lambda_n(t)^\top \lambda_{n'}(t)\right) \left(v_n(t) \varphi_n\left(t, L_n(t)\right)\right) \frac{\partial}{\partial L_i(t)} \left(v_{n'}(t) \varphi_{n'}\left(t, L_{n'}(t)\right)\right). \end{aligned}$$

While using $\mathrm{E}^A(\mathbf{L}(t))$ instead of $\mathbf{L}(0)$ in the calculations of the swaption volatility (see the previous section) results in noticeable improvements in the approximation quality, this turns out to not be the case for approximations to the skew. Furthermore, as $\varphi_n(t, L_n(0)) = 1$ for any n, the formulas resulting

from evaluating the expression for $b_S(t)$ above at $L_n(0)$ are compact and convenient, which in practice will justify any slight deterioration in precision. With this in mind, we note that

$$\frac{\partial}{\partial L_i(t)}\left(v_{n'}(t)\varphi_{n'}\left(t, L_{n'}(t)\right)\right)\Big|_{\mathbf{L}(t)=\mathbf{L}(0)} = v_{i,n'}(0) + 1_{\{i=n'\}}v_{n'}(0)b_{n'}(t)$$

(recall (15.24) for the definition of b_n's). Hence,

$$\begin{aligned}\frac{\partial \ln \|\lambda_S\left(t, \mathbf{L}(t)\right)\|}{\partial L_i(t)}\Big|_{\mathbf{L}(t)=\mathbf{L}(0)} &= \frac{1}{r_S(t)}\sum_{n,n'=j}^{k-1}\left(\lambda_n(t)^\top\lambda_{n'}(t)\right)v_n(0)v_{i,n'}(0)\\ &\quad + \frac{1}{r_S(t)}b_i(t)v_i(0)\sum_{n=j}^{k-1}\left(\lambda_n(t)^\top\lambda_i(t)\right)v_n(0).\end{aligned}$$

We recognize

$$\sum_{n=j}^{k-1}\left(\lambda_n(t)^\top\lambda_{n'}(t)\right)v_n(0) = \lambda_{n'}(t)^\top\lambda_S\left(t, \mathbf{L}(0)\right) = r_{S,n'}(t),$$

so that

$$\begin{aligned}&\frac{\partial \ln \|\lambda_S\left(t, \mathbf{L}(t)\right)\|}{\partial L_i(t)}\Big|_{\mathbf{L}(t)=\mathbf{L}(0)}\\ &\qquad = \left(\sum_{n'=j}^{k-1} v_{i,n'}(0)\frac{r_{S,n'}(t)}{r_S(t)}\right) + b_i(t)v_i(0)\frac{r_{S,i}(t)}{r_S(t)}.\end{aligned}$$

The result follows. □

Remark 15.2.4. The swaption skew consists of two parts, one that involves second-order derivatives of the swap rate with respect to Libor rates and captures overall convexity of a swap rate with respect to Libor rates, and the other being a weighted average of Libor skews.

Remark 15.2.5. Even with Libor rates sharing a common skew, i.e. in the "classic" LM model of Sections 14.2.4, 14.2.5, the swaption skew is not exactly equal to the (shared) Libor skews. If $b_i(t) = b$ for all i and t, then

$$b_S(t) = \sum_{i,n=j}^{k-1}\frac{r_{S,n}(t)}{r_{S,i}(t)}v_{i,n}(0)\xi_i(t) + b\sum_{i=j}^{k-1}v_i(0)\xi_i(t).$$

Even if the convexity term is ignored, we have

$$b_S(t) = b\sum_{i=j}^{k-1}v_i(0)\xi_i(t) \neq b.$$

15.2.3 Skew and Smile Calibration in LM Models

With pricing issues out of the way, let us now turn our attention to calibration, and see what modifications to the calibration algorithm of Section 14.5 are required by the more general model specification (15.22)–(15.23). We assume a typical set of swaption volatilities is given, and we have available a collection of market-implied volatility smiles across strikes for swaptions of different expiries and tenors. The model (15.22)–(15.23) has enough parameter flexibility to match

- At-the-money volatilities of all European swaptions, using the volatility structure $\|\lambda_n(t)\|$.
- Slopes of volatility smiles (skews) of all European swaptions, using the skew structure $b_n(t)$ (see (15.24)).
- Curvatures of smiles for swaptions of different expiries, for a given tenor, using term structure of volatilities of variance $\eta(t)$.

Of course, all these parameters are in addition to the correlation structure of Libor rates, as in the standard LM model case. Note that, for the last point, technically there is no flexibility in the model to change the volatility smile curvature for swaptions of the same expiry but different tenors. This is not really a serious limitation as the curvature of the smile tends to be constant across tenors for a given expiry.

Assume, as in Section 14.5, that we have chosen calibration targets that include N_S swaptions, $V_{\text{swaption},1}, V_{\text{swaption},2}, \ldots, V_{\text{swaption},N_S}$, and let us ignore caps; the considerations below extend trivially to cover them as well. Unlike previously, however, let us assume that each target includes not one swaption, but a collection of them of different strikes. Hence, we redefine the calibration as the goal to match volatility smiles of N_S swaptions.

Having chosen a (vanilla) stochastic volatility model for European swaptions such as (8.3)–(8.4), the target volatility smiles can be summarized by a collection of market-implied SV parameters, namely volatilities $\widehat{\lambda}_{S_i}$, skews $\widehat{b}_{S_i}$ and volatilities of variance $\widehat{\eta}_{S_i}$ for $i = 1, \ldots, N_S$ (a common mean reversion of volatility parameter is assumed). Recall that in Section 14.5.2 we denoted by G a grid of the Libor volatilities $\|\lambda_n(t)\|$. To be able to generalize, redefine $G_\lambda = G$ and, in the same spirit, define G_b to be the grid of Libor skews $b_n(t)$, and G_η (a vector) to be the discretized term structure of volatilities of variance $\eta(t)$. The formulas from the previous section allow us to compute term volatilities, skews and volatilities of variance from the model. We denote them by

$$\overline{\lambda}_{S_i}(G_\lambda, G_b, G_\eta), \quad \overline{b}_{S_i}(G_\lambda, G_b, G_\eta), \quad \overline{\eta}_{S_i}(G_\lambda, G_b, G_\eta).$$

One can incorporate the skew and smile calibration in the algorithm of Section 14.5 by adding extra terms to the calibration norm, replacing (14.54) with

$$\begin{aligned}\mathcal{I}\left(G_\lambda, G_b, G_\eta\right) = & \frac{w_{S,\lambda}}{N_S}\sum_{i=1}^{N_S}\left(\overline{\lambda}_{S_i}\left(G_\lambda, G_b, G_\eta\right) - \widehat{\lambda}_{S_i}\right)^2 \\ & + \frac{w_{S,b}}{N_S}\sum_{i=1}^{N_S}\left(\overline{b}_{S_i}\left(G_\lambda, G_b, G_\eta\right) - \widehat{b}_{S_i}\right)^2 \\ & + \frac{w_{S,\eta}}{N_S}\sum_{i=1}^{N_S}\left(\overline{\eta}_{S_i}\left(G_\lambda, G_b, G_\eta\right) - \widehat{\eta}_{S_i}\right)^2 \\ & + \ldots,\end{aligned} \tag{15.32}$$

with dots denoting various regularization terms for G_λ, G_b, and G_η. This, however, is not necessarily the best approach, as it increases the number of degrees of freedom in the non-linear optimization problem quite substantially, thus potentially significantly reducing the speed at which it could be solved. It is much better to take advantage of the structure of the problem and solve for volatilities, skews and volatilities of variance *separately* rather than all at the same time.

To motivate the method, we note that the impact of changes in the Libor skews G_b on term swaption volatilities, or their volatilities of variance, is rather small. Likewise, changes in Libor volatilities G_λ have only a small impact on term swaption skews, and so on. This near-orthogonality allows us to solve for various parameters sequentially. To facilitate this approach, we define three norms

$$\mathcal{I}_\lambda\left(G_\lambda, G_b, G_\eta\right) = \frac{w_{S,\lambda}}{N_S}\sum_{i=1}^{N_S}\left(\overline{\lambda}_{S_i}\left(G_\lambda, G_b, G_\eta\right) - \widehat{\lambda}_{S_i}\right)^2 + \ldots,$$

$$\mathcal{I}_b\left(G_\lambda, G_b, G_\eta\right) = \frac{w_{S,b}}{N_S}\sum_{i=1}^{N_S}\left(\overline{b}_{S_i}\left(G_\lambda, G_b, G_\eta\right) - \widehat{b}_{S_i}\right)^2 + \ldots,$$

$$\mathcal{I}_\eta\left(G_\lambda, G_b, G_\eta\right) = \frac{w_{S,\eta}}{N_S}\sum_{i=1}^{N_S}\left(\overline{\eta}_{S_i}\left(G_\lambda, G_b, G_\eta\right) - \widehat{\eta}_{S_i}\right)^2 + \ldots,$$

and modify the algorithm from Section 14.5.7 as such.

1. First, make a guess for G_b and G_η, denoted by G_b^0 and G_η^0. The guesses could be quite approximate. For example the grid G_b^0 could be set to the same value, an average of swaption term skews $N_S^{-1}\sum_{i=1}^{N_S}\widehat{b}_{S_i}$, and the same for G_η^0.
2. Perform steps 1–5 from Section 14.5.7 with $\mathcal{I} = \mathcal{I}_\lambda(G_\lambda, G_b^0, G_\eta^0)$ until G_λ^1, the solution, is found. Note that we keep the skew and volatility of variance grids constant throughout
3. Minimize $\mathcal{I}_b(G_\lambda^1, G_b, G_\eta^0)$ by iterating over G_b; denote the solution by G_b^1.
4. Minimize $\mathcal{I}_\eta(G_\lambda^1, G_b^1, G_\eta)$ by iterating over G_η; denote the solution by G_η^1.

Typically, the triple of parameters $(G^1_\lambda, G^1_b, G^1_\eta)$ provides a good overall solution. If a better fit is desired, the steps could be iterated, starting with (G^1_b, G^1_η) on Step 1. If the number of iterations is more than 1, it could be beneficial to stop after Step 2 (of the second or subsequent iterations) in order to have the best possible fit to the volatilities, usually the most important target.

15.3 Near-Markov LM Models

The LM model, even if driven by a single Brownian motion, is not Markovian in a low number of state variables. Rather, the state vector comprises all Libor rates and whatever state variables are needed for the stochastic volatility, if present. This is readily seen from the expression for the drift of each Libor rate (see e.g. (14.7)) which involves multiple other rates.

This state of affairs, however, has not stopped some researchers from attempting to *approximate* an LM model with a low-dimensional Markovian one. These attempts mostly involve i) restricting the volatility structure to a "separable" form, similar to that used to specify Markovian models in Chapters 12 and 13; and ii) removing path dependence in the Libor drifts by either "freezing" them at time 0 values or by employing various tricks not unlike those discussed previously in the context of drift estimation for large time steps in Monte Carlo.

The practical usefulness of such approximations for pricing and risk managing of derivatives is limited. The imposed restrictions on the volatility structure remove many of the main LM advantages in terms of calibration flexibility, and the necessity of approximating the Libor rate drifts makes the model arbitrageable and problematic to use for anything other that short-dated derivatives. Fundamentally, there is also something misplaced about trying to force LM models into a low-dimensional Markovian setting: if such a setting is desired, there are really no appealing reasons to use an LM model in the first place. Instead, one should from the outset pick one of the many perfectly good low-dimensional Markovian models we have covered earlier in the book. The models in Chapter 13 are particularly appealing, we think.

That said, a low-dimensional Markov approximation to an LM model can, however, still find use as a type of a model-based control variate. We discuss this application in more detail in Chapter 25; this chapter also covers the mechanics of the various steps involved in creating near-Markov LM models.

15.4 Swap Market Models

In a nutshell, the LM modeling principle revolves around using models for simple forward rates (Libor) that become tractable in properly selected

martingale measures. Longer-dated swap rates can be constructed iteratively by, in effect, adding up the individual forward rates that constitute the LM model primitives. As we have seen in Section 14.4.2, the swap rates constructed in this manner rarely have tractable dynamics, and swaption pricing formulas will nearly always involve approximations and/or numerical methods.

An alternative modeling approach due to Jamshidian [1997] turns the LM philosophy on its head by using forward swap rates as the fundamental model primitives, and constructing individual Libor rates as differences of such swap rates. This is similar to the idea behind swap Markov-functional model construction that we briefly discussed in Appendix 11.A.3 of Chapter 11.

By specifying tractable dynamics for forward swap rates, swaption pricing formulas now often can be done exactly, whereas cap pricing must be done by approximation. While this so-called *swap market model* framework has some proponents (e.g. Galluccio and Hunter [2003], Galluccio et al. [2005]), it is fair to say that the approach has attracted much less interest from practitioners and academics than has the LM model. Our treatment of swap market models shall therefore be brief.

Given a tenor structure $0 = T_0 < T_1 < \ldots < T_N$, our market primitives are now the swap rates

$$S_j(t) = S_{j,N-j}(t), \quad 1 \leq j < N,$$

where the notation is similar to that used in Section 14.4.2:

$$S_j(t) = \frac{P(t,T_j) - P(t,T_N)}{A_j(t)}, \quad A_j(t) = A_{j,N-j}(t) = \sum_{n=j}^{N-1} P(t,T_{n+1})\tau_n.$$

We emphasize that all forward swaps here have identical terminal maturity, namely T_N. We also emphasize that the information content of the primitives of a swap market model is identical to that of an LM model, as the forward Libor rates curve can be uniquely constructed from knowledge of $S_i(t)$, for all $i \in [q(t), N-1]$. To show this, write for any j

$$\begin{aligned} L_{j-1}(t) &= \frac{S_{j-1}(t)A_{j-1}(t) - S_j(t)A_j(t)}{P(t,T_j)\tau_{j-1}} \\ &= S_{j-1}(t)\frac{A_{j-1}(t)}{P(t,T_j)\tau_{j-1}} - S_j(t)\frac{A_{j-1}(t) - P(t,T_j)\tau_{j-1}}{P(t,T_j)\tau_{j-1}} \\ &= S_j(t) + (S_{j-1}(t) - S_j(t))\,\frac{A_{j-1}(t)}{P(t,T_j)\tau_{j-1}}. \end{aligned}$$

As

$$\frac{P(t,T_{n+1})}{P(t,T_j)} = \frac{P(t,T_{j+1})}{P(t,T_j)}\frac{P(t,T_{j+2})}{P(t,T_{j+1})}\cdots\frac{P(t,T_{n+1})}{P(t,T_n)} = \prod_{k=j}^{n}(1 + L_k(t)\tau_k)^{-1},$$

it follows from the definition of $A_{j-1}(t)$ that

$$L_{j-1}(t) = S_j(t) + (S_{j-1}(t) - S_j(t)) \left(1 + \sum_{n=j}^{N-1} \prod_{k=j}^{n} (1 + L_k(t)\tau_k)^{-1} \frac{\tau_n}{\tau_{j-1}} \right). \tag{15.33}$$

Starting from $L_{N-1}(t) = S_{N-1}(t)$, equation (15.33) gives us an iterative formula to construct the Libor forward curve at time t, as claimed.

In the swap market model framework, the modeler specifies dynamics on the swap forward rates in their respective annuity measures. Let $W^{A_j}(t)$ be an m-dimensional Brownian motion in the annuity measure induced by $A_j(t)$. As shown earlier, $S_j(t)$ is a martingale in this measure, and we can write

$$dS_j(t) = \sigma_{S_j}(t)^\top \, dW^{A_j}(t), \quad j = q(t), \ldots, N-1, \tag{15.34}$$

for some adapted volatility function $\sigma_{S_j}(t)$ specific to the j-th swap rate. As for an LM model, we could now proceed to specialize the model by using DVF or SV type specifications for $\sigma_{S_j}(t)$, generally keeping an eye out for models that allow construction of easy-to-compute expressions for payer swaption prices

$$V_{\text{swaption},j}(t) = A_j(t) \mathrm{E}_t^{A_j} \left((S_j(T_j) - c)^+ \right).$$

As we mentioned earlier, no exact pricing formulas for caplets will normally exist (as should be obvious from the complicated form of (15.33)), a situation that also holds for any swaption that does not involve a swap maturing at time T_N. For these instruments, approximative formulas must be devised if a quick calibration of the model is desired. See Galluccio et al. [2005] for some details on this.

In the context of the LM models, Section 14.2.2 derived relations between the different forward martingale measures, allowing us to describe the dynamics of all forward rates in a single measure, as required in Monte Carlo simulations, say. For the swap market models, starting with the specification (15.34) we can derive similar relations between the different annuity measures, ultimately allowing for simulation of all swap rates $S_1, S_2, \ldots, S_{N-1}$ in a common measure. For details, the reader is referred to Jamshidian [1997] and Section 14.4 of Musiela and Rutkowski [1997].

15.5 Evolving Separate Discount and Forward Rate Curves

A single yield curve for discounting and calculating Libor rates is not always compatible with no-arbitrage constraints of cross-currency markets, nor is it

particularly realistic in stressed marked conditions, such as those experienced during the subprime crisis of 2007–2009. As explained in Section 6.5.2.2, separating the discounting curve from the forward, or index, curve will ensure that linear instruments (i.e. swaps and bonds) are correctly priced at time 0. In this section we consider how to incorporate the idea of curve separation into a *dynamic* model of interest rates. While we use an LM setting for some of the material in this section, the basic ideas are generic and can be applied to virtually all models in this book.

15.5.1 Basic Ideas

Suppose two yield curves are given at time 0, the *discount* curve $P(0,T)$, $T \geq 0$, and the *index* curve $\widetilde{P}(0,T)$, $T \geq 0$ (note the change of notation from $P^{(L)}$ in Section 6.5.2.2 to $\widetilde{P}$ for convenience). The index curve corresponds to a particular Libor tenor τ, and is defined through the requirement that forward Libor rates of tenor τ must equal the conditional expected values of the future spot Libor rates,

$$\widetilde{L}\left(t,T,T+\tau\right) = \mathrm{E}_t^{T+\tau}\left(L\left(T,T,T+\tau\right)\right), \tag{15.35}$$

where $\mathrm{E}_t^{T+\tau}$ denotes expectation in the $(T+\tau)$-forward measure and, by definition,

$$\widetilde{L}(t,T,T+\tau) = \frac{1}{\tau}\left(\frac{1}{\widetilde{P}(t,T,T+\tau)} - 1\right). \tag{15.36}$$

For emphasis, we have added a tilde to the regular Libor rate notation to highlight the link to the index curve. Note that we take (15.35) to hold for arbitrary values of $T \geq t$.

It is important to point out that using different Libor tenors τ in the equation above would lead to *different* models of two-curve evolution, and our first choice shall focus on what tenor we actually want to use. In the next section we will be looking at extending the *HJM* instantaneous forward rate formalism, so our choice will be influenced by that fact. Later, in Section 15.5.3, we will consider Libor market models that, not surprisingly, lead to a somewhat different choice.

As we initially focus on adding a second curve evolution in the HJM setting, it is convenient to consider a specific choice of $\tau = 0$ that corresponds to instantaneous forward rates. In particular, we note that for $\tau \to 0$, $\widetilde{L}(t,T,T+\tau)$ converges to $\widetilde{f}(t,T)$, so according to (15.35), for the next section we shall require that

$$\widetilde{f}(t,T) = \mathrm{E}_t^T\left(\widetilde{f}(T,T)\right). \tag{15.37}$$

As discussed in Chapter 6 (see (6.44)–(6.45)), it is convenient to represent the index curve through an additive spread ε in continuously compounded forward rates. Specifically, at time 0 we write

$$\widetilde{P}(0,T) = P(0,T)\, e^{\int_0^T \varepsilon(0,u)\, du}, \quad T \geq 0. \tag{15.38}$$

We already hinted in Section 6.5.2.4 that a standard way of including the spread in a dynamic model is to assume that it evolves deterministically. We will ultimately make the same recommendation here, but it is still instructive to develop a generic dynamic model of two-curve evolution first. In developing such a model, we have several possible choices for the model primitives. For example, we could impose dynamics on the index and discount curves in separation. Alternatively, we could impose dynamics on the spread and just one of the two curves, deriving the remaining curve from these two primitives. It is often desirable to have direct control of the spread process — e.g. to ensure that its dynamics and range are in line with historical observations — so we adopt the latter approach here. As for the choice of which yield curve to use as a direct model ingredient, this is largely a matter of taste and convenience. We shall demonstrate both approaches below, using the discount curve as the primary curve in the HJM model setting of Section 15.5.2; and the index curve as the primary curve in the LM model setting of Section 15.5.3.

15.5.2 HJM Extension

The framework for discount bonds is unchanged by the presence of an index curve, and we can therefore start with the standard HJM dynamics for the discount factors (see Section 4.4.1) in the risk-neutral measure Q,

$$dP(t,T)/P(t,T) = r(t)\, dt - \sigma_P(t,T)^\top\, dW(t),$$

where $W(t)$ is a d-dimensional Brownian motion and $\sigma_P(t,T)$ is a d-dimensional discount bond volatility function. Importantly, the T-forward measure Q^T is determined solely by the dynamics of the discount curve because the numeraire, the zero-coupon discount bond $P(t,T)$, is just a particular point on the time t discount curve. In particular, the T-forward measure is defined by the familiar requirement that

$$dW^T(t) = dW(t) + \sigma_P\,(t,T)\; dt$$

be a driftless Brownian motion.

As for the instantaneous forward rates $f(t,T)$, the standard HJM result (see Lemma 4.4.1) shows that their dynamics are given by

$$df(t,T) = \sigma_f(t,T)^\top \sigma_P(t,T) dt + \sigma_f(t,T)^\top\, dW(t). \tag{15.39}$$

Let $\varepsilon(t,u)$ be the *forward rate spread* defined by extending (15.38) to arbitrary t,

$$\widetilde{P}(t,T) = P(t,T) e^{\int_t^T \varepsilon(t,u)\, du}, \quad T \geq t. \tag{15.40}$$

Treating ε as a forward rate spread is justified by considering $\widetilde{f}(t,T)$, the instantaneous forward rates calculated from the index curve $\widetilde{f}(t,T) = -\partial \ln \widetilde{P}(t,T)/\partial T$, and observing that

$$\widetilde{f}(t,T) = f(t,T) - \varepsilon(t,T), \quad 0 \le t \le T. \tag{15.41}$$

Let us endow $\varepsilon(t,T)$ with the following dynamics[4],

$$d\varepsilon(t,T) = \alpha_\varepsilon(t,T)\,dt + \sigma_\varepsilon(t,T)^\top\,dW(t), \tag{15.42}$$

for some, yet unspecified, adapted processes $\alpha_\varepsilon(t,T)$ and $\sigma_\varepsilon(t,T)$, the latter d-dimensional. Not surprisingly, the drift term in (15.42) cannot be set arbitrarily.

Proposition 15.5.1. *To satisfy the martingale restrictions of (15.37), the drift of $\varepsilon(t,T)$ in (15.42) must obey*

$$\alpha_\varepsilon(t,T) = \sigma_\varepsilon(t,T)^\top \sigma_P(t,T). \tag{15.43}$$

Proof. According to (15.37), $\widetilde{f}(t,T)$ must be a martingale in the T-forward measure. Its dynamics in that measure follow from (15.39)–(15.42),

$$\begin{aligned} d\widetilde{f}(t,T) &= \left(\sigma_f(t,T)^\top \sigma_P(t,T) - \alpha_\varepsilon(t,T)\right) dt \\ &\quad + \left(\sigma_f(t,T) - \sigma_\varepsilon(t,T)\right)^\top dW^T(t) - \left(\sigma_f(t,T) - \sigma_\varepsilon(t,T)\right)^\top \sigma_P(t,T)\,dt, \end{aligned}$$

and the result in the proposition follows by setting the dt term to zero. □

The SDEs for various quantities related to the index curve can now be derived from the discount and spread parameters.

Proposition 15.5.2. *Define*

$$\widetilde{\alpha}_f(t,T) \triangleq \widetilde{\sigma}_f(t,T)^\top \sigma_P(t,T), \quad \widetilde{\sigma}_f(t,T) \triangleq \sigma_f(t,T) - \sigma_\varepsilon(t,T).$$

The dynamics of $\widetilde{f}(t,T)$ in the risk-neutral measure are then

$$d\widetilde{f}(t,T) = \widetilde{\alpha}_f(t,T)\,dt + \widetilde{\sigma}_f(t,T)^\top\,dW(t).$$

Further, set $\widetilde{r}(t) = \widetilde{f}(t,t)$ and

[4] Arguably, diffusive dynamics do not reflect the observed movements of the spread particularly well, as the spread between discount and index curves tends to remain stable for extended period of times, followed by sometimes violent dislocations. On the other hand, our needs in capturing the distribution of the spread within a model are usually quite modest, rarely exceeding the requirement to capture an overall level of its dispersion. As long as the volatility of the spread is reasonable, this requirement will be satisfied by a diffusion model.

$$\widetilde{\alpha}_P(t,T) \triangleq \int_t^T \widetilde{\sigma}_f(t,u)^\top \left(\widetilde{\sigma}_P(t,u) - \sigma_P(t,u)\right) du,$$
$$\widetilde{\sigma}_P(t,T) \triangleq \int_t^T \widetilde{\sigma}_f(t,u)\, du.$$

Then

$$d\widetilde{P}(t,T)/\widetilde{P}(t,T) = \widetilde{r}(t)\,dt + \widetilde{\alpha}_P(t,T)\,dt - \widetilde{\sigma}_P(t,T)^\top dW(t).$$

Proof. The result for $d\widetilde{f}(t,T)$ follows directly from (15.39) and (15.41)–(15.43). From the equation

$$\widetilde{P}(t,T) = e^{-\int_t^T \widetilde{f}(t,u)du},$$

it follows that $Y(t,T) = \ln(\widetilde{P}(t,T))$ satisfies

$$\begin{aligned} dY(t,T) &= \widetilde{f}(t,t)\,dt - \int_t^T \widetilde{\alpha}_f(t,u)\,du\,dt - \widetilde{\sigma}_P(t,T)^\top dW(t) \\ &= \widetilde{r}(t)\,dt - \widetilde{\sigma}_P(t,T)^\top dW(t) - \int_t^T \widetilde{\sigma}_f(t,u)^\top \sigma_P(t,u)\,du\,dt. \end{aligned}$$

An application of Ito's lemma (to e^Y) then shows that

$$d\widetilde{P}(t,T)/\widetilde{P}(t,T) = \widetilde{r}(t)\,dt + \widetilde{\alpha}_P(t,T)\,dt - \widetilde{\sigma}_P(t,T)^\top dW(t),$$

where

$$\widetilde{\alpha}_P(t,T) = -\int_t^T \widetilde{\sigma}_f(t,u)^\top \sigma_P(t,u)\,du + \frac{1}{2}\widetilde{\sigma}_P(t,T)^\top \widetilde{\sigma}_P(t,T).$$

Integration by parts shows that

$$\widetilde{\sigma}_P(t,T)^\top \widetilde{\sigma}_P(t,T) = 2\int_t^T \widetilde{\sigma}_f(t,u)^\top \widetilde{\sigma}_P(t,u)\,du,$$

and the result follows. □

Remark 15.5.3. Note the presence of an "extra" drift term $\widetilde{\alpha}_P(t,T)$ in the dynamics for the pseudo-bonds $\widetilde{P}(t,T)$; sometimes, in an analogy with foreign exchange markets, this term is called a *quanto correction.*

For future use, let us define

$$Z(t,T) \triangleq e^{-\int_t^T \varepsilon(t,u)\,du}, \tag{15.44}$$

so that

$$\widetilde{P}(t,T) = P(t,T)/Z(t,T), \quad 0 \le t \le T. \tag{15.45}$$

The drift and diffusion terms in the process for $Z(t)$,

$$dZ(t,T)/Z(t,T) = (r(t) - \widetilde{r}(t))\, dt + \alpha_Z(t,T)\, dt - \sigma_Z(t,T)^\top\, dW(t), \quad (15.46)$$

are given by

$$\alpha_Z(t,T) = -\widetilde{\alpha}_P(t,T) - \sigma_Z(t,T)^\top \widetilde{\sigma}_P(t,T), \quad \sigma_Z(t,T) = \int_t^T \sigma_\varepsilon(t,u)\, du. \quad (15.47)$$

Options linked directly to the spread between the index and discount curves are rarely, if ever, traded, so a high level of sophistication in basis spread dynamics is seldom required. A simple one-factor Gaussian model for the spread, say, is more than sufficient for most applications. Nevertheless, richer dynamics are possible. Recall that if the forward rate volatility function is of separable form,

$$\sigma_f(t,T) = g(t)h(T),$$

where $g(\cdot)$ is a $d \times d$ matrix-valued process and $h(\cdot)$ is a d-dimensional deterministic vector-valued function (see e.g. (4.44), (12.2) or (13.70)), then the discount curve admits a d-dimensional Markovian representation, see, for example, Proposition 13.3.1. If, in addition, the forward rate spread volatility function $\sigma_\varepsilon(t,T)$ is also of separable form with the *same* $g(t)$ but a different $h_\varepsilon(T)$,

$$\sigma_\varepsilon(t,T) = g(t)h_\varepsilon(T),$$

then the forward rate spread curve $\varepsilon(t,T)$ and, by extension, the index curve also admit Markovian representations with the same Markovian state variables. This fact opens up a possibility of building efficient Markovian models of joint evolution of the discount and the index curve with non-deterministic spread. We leave this line of inquiry for the reader to explore, and instead move on to the changes required to extend the LM model framework to support two-curve dynamics.

15.5.3 Applications to LM Models

We now return to the LM framework, and continue with the notations of Section 14.2. In particular, we assume that a tenor structure (14.1) has been specified. Clearly, in the LM setup, it is natural to assume that (15.36) is satisfied for the Libor tenor used in the definition of the LM model. This would correspond to extending the LM model to drive both the discounting curve and the *index curve that corresponds to the Libor tenor used in LM model definition.* We remind the reader that if multiple index curve dynamics are required, each will have to be driven by its own LM model (of a given Libor tenor).

As discussed earlier, we here choose as our primary model variables (in addition to the spread) the set of "regular" Libor rates $\widetilde{L}_n(t) \triangleq \widetilde{L}(t, T_n, T_{n+1})$,

$n = 0, \ldots, N-1$, as defined off the index curve in (15.36). Apart from giving ourselves a chance to present a slightly different twist on the material of the previous section, the choice of the index Libors as building blocks allows us to use the volatilities of $\{\widetilde{L}_n(t)\}_{n=0}^{N-1}$ that are more directly related to the market-observable volatility information, i.e. the volatilities of swap rates of various maturities and tenors. Conveniently, for each n, $\widetilde{L}_n(t)$ is a martingale in T_{n+1}-forward measure by construction, see (15.35). Therefore, in close analogy to (14.5), we can define the dynamics by

$$d\widetilde{L}_n(t) = \widetilde{\sigma}_n(t)^\top \, dW^{n+1}(t), \tag{15.48}$$

where $W^{n+1}(t)$ is an m-dimensional Brownian motion in $\mathrm{Q}^{T_{n+1}}$ and $\widetilde{\sigma}_n(t)$ is an m-dimensional adapted process.

The discrete money market account $B(t)$, see (14.8), induces a useful common measure for all Libor rates in the single-curve case. The extension to the two-curve framework is straightforward: we apply (14.8) literally, i.e. use simply compounded forward rates computed off the discount curve in the definition of $B(t)$. This corresponds exactly to the actual trading strategy of re-investing into deposits of a given tenor, with the value of the strategy solely determined by the discount curve:

$$B(t) = P\left(t, T_{q(t)}\right) \prod_{n=0}^{q(t)-1} \frac{1}{P\left(T_n, T_{n+1}\right)}. \tag{15.49}$$

We define the spot measure Q^B accordingly. Interestingly, a measure change from any T-forward measure to the spot measure is determined by the dynamics of discount factors $P(t,T)$ only, and we see that if $\widetilde{L}_n$'s satisfy (15.48), then in the spot measure Q^B, the process for $\widetilde{L}_n$ is given by

$$d\widetilde{L}_n(t) = \widetilde{\sigma}_n(t)^\top \left(\sum_{j=q(t)}^{n} \frac{\tau_j \sigma_j(t)}{1 + \tau_j L_j(t)} dt + dW^B(t) \right), \tag{15.50}$$

where $W^B(t)$ is an m-dimensional Brownian motion in measure Q^B. Here $L_j(t)$ are simply compounded forward rates calculated off the discount curve, and $\sigma_j(t)$ are their volatilities. It would, however, be more convenient to have the drift in terms of the Libor rate model primitives $\widetilde{L}_j(t)$ and the process for the spread.

Before deriving the result for the drifts, let us decide what variables to use to represent the spread evolution. Using the instantaneous forward spread curve $\varepsilon(t,T)$ is not convenient in the LM framework, and we settle on *forward bond ratios* $Z^n(t)$, $n = 0, \ldots, N-1$, as state variables, defined by

$$Z^n(t) \triangleq Z(t, T_{n+1}) / Z(t, T_n), \tag{15.51}$$

with the $Z(t,T)$'s given in (15.44). This choice of state variables keeps the LM framework in line with (but not exactly equivalent to!) the HJM extension

from Section 15.5.2, although other choices are possible; we comment on that at the end of the section.

Let us assume that forward bond ratios follow the dynamics

$$dZ^n(t)/Z^n(t) = O(dt) - \sigma_{Z^n}(t)^\top dW^B(t) \tag{15.52}$$

in the spot measure, with $\sigma_{Z^n}(t)$'s being their volatilities. Note that $\sigma_{Z^n}(t) = \sigma_Z(t, T_{n+1}) - \sigma_Z(t, T_n)$, where $\sigma_Z(t, T)$ are defined in (15.46). The drifts of forward bond ratios, not surprisingly, are not free parameters. We have the following result on the dynamics of Libor rates and forward bond ratios.

Proposition 15.5.4. *If the Libor rates $\widetilde{L}_n(t)$ satisfy (15.48), and forward bond ratios $Z^n(t)$ follow the dynamics of (15.52), then in the spot measure Q^B, the process for the state variables of the model is given by*

$$d\widetilde{L}_n(t) = \widetilde{\sigma}_n(t)^\top \left(\widetilde{\mu}_n(t)\, dt + dW^B(t)\right), \tag{15.53}$$

$$\begin{aligned} dZ^n(t)/Z^n(t) &= \sigma_{Z^n}(t)^\top \\ &\quad \times \left(\left(\frac{\tau_n \widetilde{\sigma}_n(t)}{1 + \tau_n \widetilde{L}_n(t)} + \sigma_{Z^n}(t) - \widetilde{\mu}_n(t)\right) dt - dW^B(t)\right), \end{aligned} \tag{15.54}$$

where

$$\widetilde{\mu}_n(t) = \sum_{j=q(t)}^{n} \left(\frac{\tau_j \widetilde{\sigma}_j(t)}{1 + \tau_j \widetilde{L}_j(t)} + \sigma_{Z^j}(t)\right)$$

and $W^B(t)$ is an m-dimensional Brownian motion in measure Q^B.

Proof. Clearly

$$\frac{1}{1 + \tau_j L_j(t)} = \frac{1}{1 + \tau_j \widetilde{L}_j(t)} Z^j(t). \tag{15.55}$$

Differentiating (15.55) and matching the dW terms, we obtain that

$$\frac{\tau_j \sigma_j(t)}{1 + \tau_j L_j(t)} = \frac{\tau_j \widetilde{\sigma}_j(t)}{1 + \tau_j \widetilde{L}_j(t)} + \sigma_{Z^j}(t),$$

and (15.53) follows.

To derive the drift in (15.54), we note that $(1 + \tau_n L_n(t)) Z^n(t) = 1 + \tau_n \widetilde{L}_n(t)$ is a Q^{n+1}-martingale. Hence, in T^{n+1}-forward measure we have the following dynamics,

$$dZ^n(t)/Z^n(t) = \sigma_{Z^n}(t)^\top \left(\frac{\tau_n \sigma_n(t)}{1 + \tau_n L_n(t)}\, dt - dW^{n+1}(t)\right).$$

Switching to the spot measure, we obtain (15.54). □

The simulation scheme for the two-curve LM model is similar to the one-curve case, with obvious modifications. The initial values of the model primitives, the Libor rates $\widetilde{L}_n(0)$ and the forward bond ratios $Z^n(t)$, $n = 0, \ldots, N-1$, are derived from the initial values of the discount and index curves. Having specified the Libor rate volatilities $\widetilde{\sigma}_n(t)$ and forward bond spread volatilities $\sigma_{Z^n}(t)$, $n = 1, \ldots, N-1$, we simulate the Libor rates and the bond spreads using SDEs from Proposition 15.5.4. At each time t, we may construct an index curve $\widetilde{P}(t,T)$, $T \geq t$, from the simulated Libor rates $\widetilde{L}_n(t)$, $n = 0, \ldots, N-1$, as in the one-curve case in Chapter 14. From the simulated index curve and the simulated bond spread curve $Z(t,T)$, $T \geq t$, the discount curve $P(t,T)$, $T \geq t$, is calculated via (15.40). The discount curve is used to update the numeraire by (15.49) and, together with the index curve, to calculate market rates needed for evaluating derivative payoffs. In particular, forward swap rates are calculated by projecting Libor rates off the index curve and discounting them off the discount curve (compare to (6.39)), so a rate fixing at T_j covering $k-j$ periods is calculated by

$$S_{j,k-j}(t) = \frac{\sum_{i=j}^{k-1} \tau_i P(t, T_{i+1}) \widetilde{L}_i(t)}{\sum_{i=j}^{k-1} \tau_i P(t, T_{i+1})}. \tag{15.56}$$

In the two-curve setup, the derivation of swaption pricing approximations for model calibration can proceed as for the single-curve case. For this purpose, first rewrite (15.56) in terms of model primitives, to obtain

$$S_{j,k-j}(t) = \frac{\sum_{i=j}^{k-1} \tau_i Z(t, T_{i+1}) \widetilde{P}(t, T_{i+1}) \widetilde{L}_i(t)}{\sum_{i=j}^{k-1} \tau_i Z(t, T_{i+1}) \widetilde{P}(t, T_{i+1})}.$$

Proceeding as in the standard single-curve case and ignoring contributions from the $Z(t, T_{i+1})$ terms in the swap rate dynamics — a respectable approximation even with stochastic spreads — we obtain the following dynamics in the appropriate annuity measure

$$dS_{j,k-j}(t) = \sum_{n=1}^{N-1} \left. \frac{\partial S_{j,k-j}(t)}{\partial \widetilde{L}_n(t)} \right|_{t=0} \widetilde{\sigma}_n(t)^\top dW^{A_{j,k-j}}(t). \tag{15.57}$$

This formula can be used as the starting point for the usual European swaption approximations. Compared to earlier results in Section 14.4.2, the presence of the basis spread leads to slightly altered expressions for the weights $\partial S(t)/\partial \widetilde{L}_n(t)|_{t=0}$; the exact form of these weights is left for the reader to derive.

To conclude our discussion of dynamic two-curve modeling in the LM framework, we note that our decision to use forward bond spreads $Z^j(t)$ as variables driving the spread between the index and discount curves is not the only choice. For example, we could have used the spreads between the simply compounded forward rates computed off the index and discount curves,

$\widetilde{L_j}(t) - L_j(t)$, $j = 0, \ldots, N-1$, instead. This would lead to a somewhat different, but equally tractable, extension. As this area of interest rate modeling is still rapidly evolving, consensus on what are the right variables to use has not emerged yet.

15.5.4 Deterministic Spread

It is not hard to see that the values of derivative securities that do not directly reference the spread between the index and the discount curve — which is the majority of them — respond approximately linearly to this spread. For one, the spread is usually confined to a rather tight range, ensuring that a linear approximation may be adequate. Moreover, changes in the spread will normally affect the discount curve substantially more than the index curve, as the latter is predominantly calibrated to linear market instruments such as FRAs and swaps. Values of most securities, even exotic ones, tend to be approximately linear to changes in discounting.

A fairly self-evident rule states that if the dependence of a value of a security on a given market factor is approximately linear, there is little reason to model that factor with a stochastic process for the purposes of valuation and hedging of that security. As a consequence, it is not uncommon to assume deterministic evolution for the spread between the discount and index curves. In the framework we developed, this may be achieved by setting the volatility of the spread $\sigma_\varepsilon(t,T)$ to zero. The two-curve LM model simplifies accordingly. Obviously, one does not need to simulate $\varepsilon(t,T)$ anymore, as

$$\varepsilon(t,T) = \varepsilon(0,T), \quad Z(t,T) = Z(0,T)/Z(0,t)$$

for all t, T such that $0 \leq t \leq T$. Furthermore, the drift term in (15.53) reduces to the standard single-curve expression, and discount factors $P(t,T)$ are obtained from pseudo-discount ones by

$$P(t,T) = \frac{P(0,t,T)}{\widetilde{P}(0,t,T)} \widetilde{P}(t,T), \quad T \geq t.$$

Derivation of swaption pricing expressions proceeds as in Section 15.5.3 where the randomness in the spreads was already ignored in (15.57).

15.6 SV Models with Non-Zero Correlation

Let us return to the SV-type LM model we considered in Section 14.2.5. In the spot measure, we recall that forward rate dynamics are postulated to be of the form

$$dL_n(t) = \sqrt{z(t)}\varphi\left(L_n(t)\right)\lambda_n(t)^\top\left(\sqrt{z(t)}\mu_n(t)\,dt + dW^B(t)\right),$$

$$\mu_n(t) = \sum_{j=q(t)}^{n} \frac{\tau_j\varphi\left(L_j(t)\right)\lambda_j(t)}{1+\tau_j L_j(t)},$$

with

$$dz(t) = \theta\left(z_0 - z(t)\right)dt + \eta\psi\left(z(t)\right)dZ(t).$$

See Section 14.2.5 for further details on the notation.

In previous work, we assumed that the scalar Brownian motion $Z(t)$ was independent of all components of the m-dimensional Brownian motion $W(t)$. This assumption allowed us to switch from the spot measure into more convenient forward measures, without altering the form of the process for $z(t)$. We shall now briefly consider relaxing this assumption. In particular, let us assume a non-zero deterministic correlation vector $\rho(t)$ between $Z(t)$ and $W^B(t)$. In this case, Lemma 14.2.6 tells us the dynamics of $z(t)$ in the forward measure $\mathrm{Q}^{T_{n+1}}$. As we defined $\langle dW^B(t), dZ(t)\rangle = \rho(t)\,dt$, then

$$dz(t) = \theta\left(z_0 - z(t) - \frac{\eta}{\theta}\psi\left(z(t)\right)\sqrt{z(t)}\rho(t)^\top\mu_n(t)\right)dt + \eta\psi\left(z(t)\right)dZ^{n+1}(t),$$

where $Z^{n+1}(t)$ in a $\mathrm{Q}^{T_{n+1}}$-Brownian motion. In certain important cases, this expression can be simplified and approximated somewhat, as demonstrated in the following corollary.

Corollary 15.6.1. *When $\psi(z(t)) = \sqrt{z(t)}$, we have*

$$dz(t) = \theta\left(z_0 - \alpha(t)z(t)\right)dt + \eta\sqrt{z(t)}\,dZ^{n+1}(t), \tag{15.58}$$

where

$$\alpha(t) = 1 + \frac{\eta}{\theta}\rho(t)^\top\mu_n(t) \approx 1 + \frac{\eta}{\theta}\rho(t)^\top\sum_{j=q(t)}^{n}\frac{\tau_j\varphi\left(L_j(0)\right)\lambda_j(t)}{1+\tau_j L_j(0)}. \tag{15.59}$$

Notice that in Corollary 15.6.1 we have used an approximation under which the multiplier $\alpha(t)$ in the drift term of (15.58) is approximately deterministic. The approximation is based on the same "freeze along forwards" idea as was used in Section 14.4.2 and elsewhere, and makes the SDE for $z(t)$ affine. As such, from the results in Chapter 8 we would expect that a Fourier-based approach would allow for (approximate) caplet pricing if the dynamics for $L_n(t)$ are themselves affine, e.g. if $\varphi(x) = ax + b$. We verify this below for the log-normal case ($\varphi(x) = x$); for displaced log-normal dynamics the result follows along similar lines.

Proposition 15.6.2. *Define $X_n(t) = \ln L_n(t)$ where*

$$dL_n(t)/L_n(t) = \sqrt{z(t)}\lambda_n(t)^\top dW^{n+1}(t),$$

$$dz(t) = \theta\left(z_0 - \alpha(t)z(t)\right)dt + \eta\sqrt{z(t)}dZ^{n+1}(t),$$

for a deterministic function $\alpha(t)$ *and* $\langle dW^{n+1}(t), dZ^{n+1}(t)\rangle = \rho(t)\,dt$. *The moment-generating function*

$$\Psi_{X_n}(u) = \mathrm{E}^{T_{n+1}}\left(e^{uX_n(T_n)}\right)$$

can be written as

$$\Psi_{X_n}(u) = e^{A(0,T)+X_n(0)B(0,T)+z(0)C(0,T)},$$

where A, B, *and* C *solve a system of Riccati ODEs*

$$\begin{aligned}
0 &= A'(t,T) + \theta z_0 C(t,T),\\
0 &= B'(t,T),\\
0 &= C'(t,T) - \frac{1}{2}\|\lambda_n(t)\|^2 B(t,T) + \frac{1}{2}\|\lambda_n(t)\|^2 B(t,T)^2\\
&\quad - \theta\alpha(t)C(t,T) + \frac{1}{2}\eta^2 C(t,T)^2 + \eta\lambda_n(t)^\top\rho(t)B(t,T)C(t,T),
\end{aligned}$$

with the terminal conditions

$$A(T,T) = 0, \quad B(T,T) = u, \quad C(T,T) = 0.$$

Proof. Positing the form

$$\mathrm{E}^{T_n+1}\left(e^{uX_n(T_n)}\middle| X_n(t) = x, z(t) = v\right) = e^{A(t,T)+xB(t,T)+vC(t,T)}$$

and substituting it into the Feynman-Kac PDE that corresponds to the dynamics of $(X_n(t), z_n(t))$, we obtain the following equation

$$\begin{aligned}
0 &= A'(t,T) + xB'(t,T) + vC'(t,T)\\
&\quad - \frac{1}{2}v\|\lambda_n(t)\|^2 B(t,T) + \frac{1}{2}v\|\lambda_n(t)\|^2 B(t,T)^2\\
&\quad + \theta(z_0 - \alpha(t)v)C(t,T) + \frac{1}{2}\eta^2 vC(t,T)^2\\
&\quad + v\eta\lambda_n(t)^\top\rho(t)B(t,T)C(t,T).
\end{aligned}$$

Combining the terms in x and v together, the result follows. □

The result of the proposition above can be combined with Theorem 8.4.1 to allow for analytical pricing of caplets by Fourier methods. Pricing of swaptions follows a similar line of attack. To start, we write down the drift of $z(t)$ under the corresponding annuity measure; while somewhat more complicated in appearance, it will still be linear in z after application of the "freezing" technique. Then, the moment-generating function for the logarithm of the swap rate is available, allowing for analytic pricing of swaptions via Fourier methods. We trust the reader to fill in missing details.

15.7 Multi-Stochastic Volatility Extensions

15.7.1 Introduction

In the specification (15.22)–(15.23) of the LM model, a single stochastic volatility process $\sqrt{z(t)}$ is used to scale the diffusion coefficients of all forward rates. As such, the volatility structure of the model is only allowed (nearly) parallel moves up and down. While sufficiently rich to introduce the volatility smile for all European swaptions, one has to wonder about the limitations of this one-factor specification.

The value of many exotic interest rate derivatives, sometimes called "first generation" exotics, is primarily linked to the overall *level* of the interest rate curve. This class includes, for example, Bermudan swaptions, callable inverse floaters, or callable range accruals on a Libor or CMS rate. For such instruments, having a single stochastic volatility factor applied to all rates is typically adequate.

On the other hand, interest rate exotics linked to the spread of two CMS rates, such as CMS spread callable swaps (see Section 5.13.3) or CMS spread TARNs (see Section 5.15.2) derive their value from the distribution of the *slope* of the interest rate curve. Just like a single-factor model is unsuitable for pricing such exotics — being unable to represent the changes in the slope of the curve — a common stochastic volatility factor applied to all rates does not always allow for sufficient control over the distribution of the slope of the interest rate curve. In particular, such a specification does not typically allow for much control over the finer features of the *volatility smile of the spread*, e.g. its slope or curvature.

We will have quite a bit more to say about modeling the smile of a CMS spread later on in Chapter 17, but let us take these observations as a rationale (or excuse) for sketching an extension of the LM model that allows for some de-correlation in stochastic volatility factors applied to different rates. While many roads could be taken, the route we shall suggest here

- Incorporates the features necessary for realistic CMS spread modeling.
- Remains relatively parsimonious.
- Can still be calibrated using analytical approximations to prices of caps and swaptions.

We must point out that the whole area of multi-dimensional stochastic volatility interest rate modeling is quite new, so we keep the discussion suitably brief, with details to be filled by future research.

15.7.2 Setup

One can take a view, admittedly not inconsistent with the philosophy of LM modeling, that each Libor rate should be driven by its own stochastic variance process. However, it should be obvious that such specification

would be quite unwieldy, and would likely not lend itself easily to closed-form approximations for swaption prices. With the point of our proposed extension focused primarily on controlling the smiles of CMS spreads, we instead consider a parsimonious extension that involves only *two* stochastic variance processes.

We define

$$dz^i(t) = \theta^i \left(z_0^i - z^i(t)\right) dt + \eta^i \sqrt{z^i(t)}\, dZ^i(t), \quad i = 1, 2. \tag{15.60}$$

Note the time-independence of parameters; while a more general specification is certainly possible, we keep our focus on more important details. Moreover, we require

$$\left\langle dZ^1(t), dZ^2(t) \right\rangle = 0.$$

For the applications we have in mind, we need to allow for non-zero correlation between Brownian motions driving the Libor rates and those driving the stochastic variances. Hence, (15.60) is understood to hold under Q^B, the spot Libor measure only. Under the same measure, we assume that the Libor rates follow

$$\begin{aligned} dL_n(t)/\varphi\left(L_n(t)\right) &= \sqrt{z^1(t)}\lambda_n^1(t)^\top \left(dW^1(t) + \mu_n^1(t)\, dt\right) \\ &\quad + \sqrt{z^2(t)}\lambda_n^2(t)^\top \left(dW^2(t) + \mu_n^2(t)\, dt\right), \quad n = 1, \ldots, N-1. \end{aligned} \tag{15.61}$$

Here, $W^1(t)$ and $W^2(t)$ are two independent copies of a d-factor Brownian motion. Moreover, we assume that

$$\left\langle dW^i(t), dZ^i(t) \right\rangle = \chi^i, \quad i = 1, 2, \tag{15.62}$$

but also

$$\left\langle dW^1(t), dZ^2(t) \right\rangle = \left\langle dW^2(t), dZ^1(t) \right\rangle = 0. \tag{15.63}$$

Under the T_{n+1}-forward measure, the Libor rate $L_n(t)$ is a martingale,

$$\begin{aligned} dL_n(t)/\varphi\left(L_n(t)\right) &= \sqrt{z^1(t)}\lambda_n^1(t)^\top\, dW^{1,T_{n+1}}(t) \\ &\quad + \sqrt{z^2(t)}\lambda_n^2(t)\, dW^{2,T_{n+1}}(t), \quad n = 1, \ldots, N-1. \end{aligned} \tag{15.64}$$

15.7.3 Pricing Caplets and Swaptions

Following the same techniques used in Section 15.6 above, it is straightforward to show that in the T_{n+1}-forward measure, the dynamics of the stochastic variance processes are

$$dz^i(t) = \theta^i \left(z_0^i - z^i(t)\right) dt - \eta^i \nu^{i,n+1}\left(t, \mathbf{L}(t)\right) z^i(t)\, dt + \eta^i \sqrt{z^i(t)}\, dZ^{i,T_{n+1}}(t),$$

$i = 1, 2$, where

$$\nu^{i,n+1}(t,\mathbf{L}(t)) = \left(\chi^i\right)^\top \sum_{j=q(t)}^{n} \frac{\tau_j \varphi\left(L_j(t)\right) \lambda_j^i(t)}{1+\tau_j L_j(t)}.$$

Freezing the drifts, in the same manner as in Corollary 15.6.1, we get

$$dz^i(t) = \theta^i \left(z_0^i - z^i(t)\right) dt - \eta^i \nu^{i,n+1}(t,\mathbf{L}(0))\, z^i(t)\, dt + \eta^i \sqrt{z^i(t)}\, dZ^{i,T_{n+1}}(t), \tag{15.65}$$

and we obtain that (15.64), (15.65) constitute an affine specification (thanks to (15.63)). Hence, the moment-generating function could be represented in a quasi-closed form of an exponential of coefficients computable from Riccati equations, and prices of caplets could be obtained by Fourier inversion of the moment-generating function. We omit straightforward details.

As far as swaptions are concerned, as is the case in many other LM model specifications, we could derive the dynamics of a swap rate of essentially the same form as that of the Libor rates. Then, the same arguments as above can be applied to compute European swaption prices.

15.7.4 Spread Options

The intention of introducing two stochastic variance processes above is to be able to control the volatility smile[5] of the spread option, while keeping the smiles of individual swap rates fixed and in calibration with observed swaption values. As shown later in Chapter 17, such control can be achieved if we have a mechanism to affect i) the correlation of the stochastic variance processes affecting the two rates in the spread; and ii) the "cross" correlations between a forward rate and the variance process of the other forward rate in the spread. The specification outlined above allows for this, as we shall now demonstrate.

Consider two forward Libor rates, $L_n(t)$ and $L_m(t)$, $n \neq m$. For simplicity we assume that in (15.61) $\varphi(x) \equiv 1$ and $\lambda_k^i(t) \equiv \lambda_k^i$, $i = 1,2$, $k = n,m$. Focusing on the diffusion terms only, we have (drifts and probability measure are irrelevant),

$$dL_k(t) = O(dt) + \sqrt{z^1(t)}\left(\lambda_k^1\right)^\top dW^1(t) + \sqrt{z^2(t)}\left(\lambda_k^2\right)^\top dW^2(t),\ k = n,m, \tag{15.66}$$

$$dz^i(t) = O(dt) + \eta^i \sqrt{z^i(t)}\, dZ^i(t), \quad i = 1,2. \tag{15.67}$$

In the following easily proven result, we rewrite the dynamics in the form convenient for correlation analysis.

Proposition 15.7.1. *Assume that L_n and L_m satisfy (15.66)–(15.67) above, with the correlation structure as in (15.62) and (15.63). The joint dynamics*

[5] One way to define such a smile is in terms of the implied Bachelier volatility of the spread itself; see Sections 14.4.3 and 17.4.1 for more details.

of the two Libor rates and their stochastic variance processes can then be written as

$$dL_k(t) = O(dt) + \sqrt{u^k(t)}dU^k(t),\ du^k(t) = O(dt) + \sqrt{\eta^k(t)}dX^k(t),\ k = n, m,$$

where

$$u^k(t) = z^1(t)\left\|\lambda_k^1\right\|^2 + z^2(t)\left\|\lambda_k^2\right\|^2,$$
$$\eta^k(t) = z^1(t)\left\|\lambda_k^1\right\|^4 \left(\eta^1\right)^2 + z^2(t)\left\|\lambda_k^2\right\|^4 \left(\eta^2\right)^2,$$

and

$$dU^k(t) = \frac{1}{\sqrt{u^k(t)}}\left(\sqrt{z^1(t)}\left(\lambda_k^1\right)^\top dW^1(t) + \sqrt{z^2(t)}\left(\lambda_k^2\right)^\top dW^2(t)\right),$$
$$dX^k(t) = \frac{1}{\sqrt{\eta^k(t)}}\left(\sqrt{z^1(t)}\left\|\lambda_k^1\right\|^2 \eta^1\, dZ^1(t) + \sqrt{z^2(t)}\left\|\lambda_k^2\right\|^2 \eta^2\, dZ^2(t)\right).$$

It follows from Proposition 15.7.1 that our model setup allows for essentially independent control of the correlations between the variance processes u^n and u^m, as well as the correlations between L_m and u^n and L_n and u^m as should be clear from the expression for one of the correlations (others are similar):

$$\mathrm{Corr}(dL^n(t), du^m(t)) = \frac{1}{\sqrt{u^n(t)\eta^m(t)}} \times \left(z^1(t)\|\lambda_m^1\|^2\eta^1(\lambda_n^1)^\top\chi^1 + z^2(t)\|\lambda_m^2\|^2\eta^2(\lambda_n^2)^\top\chi^2\right).$$

The results extend to swap rate spreads in a predictable, and largely mechanical, fashion, allowing us to set up an LM calibration that targets quantities linked to the shape of smile of various CMS spread options of interest. As the reader might expect, the formulas become rather cumbersome and we do not list them here.

At this point we have gathered enough results for our later discussion on spread option pricing, so we conclude the analysis here. We return to spread options in stochastic volatility models in Chapter 17.

15.7.5 Another Use of Multi-Dimensional Stochastic Volatility

In models for foreign exchange or for equity prices, multi-dimensional stochastic volatility is typically used not as a way to refine spread option pricing, but rather as i) a mechanism to induce multiple[6] time-scales in the mean-reverting behavior of volatility; or ii) to control the evolution of the volatility smile through time. A discussion of multiple time-scales in empirical data

[6]Basically this means that κ_1 is either much larger or much smaller than κ_2.

can be found in Perello et al. [2004], and Kainth and Saravanamuttu [2007] (among others) discuss applications specific to foreign exchange, with an emphasis on smile dynamics. We also note that Andersen and Brotherton-Ratcliffe [2005] introduce a tractable alternative to (15.61) that has perfect correlation between the variances of all Libor forwards, but still introduces multiple time-scales in the variance dynamics; this setup (the details of which we omit) is mainly useful for applications where the correlation between variances of different forwards is thought to be high.