Chapter 12

Markovian Term-Structure Models

In these notes we return to interest rates carrying what the volatility chapters taught us. The Heath-Jarrow-Morton framework of chapter 8 is completely general and, for that reason, computationally useless: its state is the whole curve. We derive the one restriction on the volatility that collapses it to a finite state, obtain the quasi-Gaussian models, and take a mathematical excursion into why models of this shape can be solved at all.

12.1 Why HJM Is Not Enough

Chapter 8 derived the Heath-Jarrow-Morton drift condition: under the risk neutral measure, an arbitrage-free evolution of the instantaneous forward curve must take the form

$$df(t,T) = \sigma(t,T)\left(\int_t^T \sigma(t,s)\,ds\right)dt + \sigma(t,T)\,dW_t. \tag{12.1}$$

This is a complete answer to the question chapter 8 asked. Choose any volatility structure whatsoever and the drift is determined; there is no further freedom and no further constraint. As a piece of theory it could hardly be better.

As a thing to compute with it is close to unusable. The state of the model at time t is the entire function $T \mapsto f(t,T)$. Nothing smaller will do: to take the next step, (12.1) needs $\sigma(t,\cdot)$ integrated over the whole remaining curve, and in general σ may depend on the curve's current shape. So the process is Markov in an infinite dimensional state and in nothing less.

For pricing a European payoff this is survivable, though not because simulation somehow escapes the dimension. It does not. Discretise the curve onto a few hundred maturities and every one of them is a state variable that has to be stored and stepped, because (12.1) needs $\sigma(t,\cdot)$ integrated over the whole remaining curve before it can take a step, and if σ depends on the curve's shape then the previous step's answer is what supplies it. A simulated HJM path is a simulated *surface*.

What makes that bearable is the shape of the cost rather than its size. Carrying three hundred tenors costs three hundred times a one-factor model per step, which is linear in the dimension; and chapter 19's measurement is that a Monte Carlo error is set by the payoff's variance and the path

count and does not grow with the number of state variables at all. So a large state multiplies the cost of each path and does not multiply the number of paths needed. A European payoff also never asks what anything is worth at an intermediate date: one sweep from today to expiry, average at the end, and the state's dimension is bookkeeping.

Remark (When even the stepping is unnecessary)**.** If $\sigma(t,T)$ is deterministic, (12.1) can be integrated in closed form:

$$f(T,T') = f(0,T') + \int_0^T \sigma(u,T') \left(\int_u^{T'} \sigma(u,s)\, ds \right) du + \int_0^T \sigma(u,T')\, dW_u,$$

in which the drift term is a number and the stochastic integral is Gaussian with a deterministic variance. The whole curve at T is then a jointly Gaussian vector that can be drawn in one shot, with no time stepping at all. It is the dependence of σ on the curve that forces the step-by-step simulation, and with it the need to carry the surface.

Backward induction is where the dimension stops being bookkeeping. A Bermudan swaption must be valued by comparing, at each exercise date, the value of exercising against the value of continuing, and that comparison has to be made state by state — so it needs the value *at* a state rather than an average over states. And regression needs a small set of variables to regress against, which is exactly what an infinite dimensional state does not offer, and nothing in the model tells us which handful of curve summaries would do.

So the question of this chapter is: what must we give up for the curve to be a function of a handful of numbers?

12.2 What a Finite State Requires

Two separate things have to be true before the curve collapses onto a handful of numbers. The first is that the model be Markov in finitely many variables. The second is that its volatility depend on how long a rate has left to run rather than on the date it matures. The exponential comes from the two together.

Take $\sigma(t,T)$ deterministic for this section. The state-dependent case is what the model is eventually for, and it is returned to at the end.

12.2.1 Markov Is Exactly Finite Rank

Integrating (12.1), the curve at time t is

$$f(t,T) \;=\; \underbrace{f(0,T) + \int_0^t \alpha(u,T)\, du}_{\text{deterministic, so free}} \;+\; \underbrace{\int_0^t \sigma(u,T)\, dW_u}_{\text{the part needing a state}} \tag{12.2}$$

for every maturity $T \geq t$. With σ deterministic the drift is deterministic too, so the whole question is the second term: a continuum of random variables, one per maturity, each an integral over the whole past.

Lemma 12.1 (Finite state, finite rank, and a sum of products are the same thing)**.** *Let σ be deterministic and square integrable. The following are equivalent.*

(i) For every t, the curve $f(t,\cdot)$ is recoverable from $f(0,\cdot)$ and n random variables.

(ii) The functions $u \mapsto \sigma(u,T)$, as T varies, span a space of dimension n.

(iii) σ is a sum of n products and no fewer,

$$\sigma(u,T) \;=\; \sum_{i=1}^{n} \phi_i(u)\,\psi_i(T). \tag{12.3}$$

In that case the state variables are $X_t^i = \int_0^t \phi_i(u)\,dW_u$, and

$$f(t,T) \;=\; f(0,T) + \int_0^t \alpha(u,T)\,du + \sum_{i=1}^{n} \psi_i(T)\,X_t^i. \tag{12.4}$$

Proof. (i) $\Leftrightarrow$ (ii) is the Itô isometry. The map $h \mapsto \int_0^t h(u)\,dW_u$ is linear, and

$$\mathbb{E}\left[\left(\int_0^t h(u)\,dW_u\right)^2\right] = \int_0^t h^2(u)\,du,$$

so it is an isometry and in particular injective. A family of such integrals therefore spans a space of exactly the dimension spanned by the family of integrands. Recovering (12.2) for every T from n numbers is thus precisely the statement that the integrands span n dimensions.

(ii) $\Leftrightarrow$ (iii) is linear algebra with no probability in it. If the sections span n dimensions, choose a basis $\phi_1, \dots, \phi_n$ of that span. Each section is in the span, so

$$\sigma(\,\cdot\,,T) = \sum_{i=1}^{n} \psi_i(T)\,\phi_i,$$

where the coefficients depend on which section we took, that is on T alone; and that is (12.3). Conversely (12.3) exhibits every section as a combination of the same n functions. Substituting into (12.2) and pulling $\psi_i(T)$ out of the integral, since it does not involve u, gives (12.4). □

Nothing was assumed about the shape of σ to get (12.3). The product structure is not an ansatz; it is what having finite rank *means*, and lemma 12.1 says that is exactly what a finite state is.

12.2.2 Finite Rank Alone Forces Nothing

It is tempting to stop here and conclude that a Markov model needs an exponential decay. It does not.

Example 12.1 (A one-state model with an arbitrary maturity shape)**.** Let $\sigma(t,T) = \psi(T)$ for any square-integrable ψ whatsoever — the volatility of the forward rate maturing at T is $\psi(T)$, and does not change as t advances. This is rank one, with $\phi_1 \equiv 1$, so lemma 12.1 gives a single state,

$$X_t = \int_0^t dW_u = W_t, \qquad f(t,T) = f(0,T) + \int_0^t \alpha(u,T)\,du + \psi(T)\,W_t.$$

The drift condition determines α from ψ as always, so the model is arbitrage free; it is Markov in one variable; and ψ is unrestricted. Take $\psi(T) = 1/(1+T)$ if a concrete one is wanted. Measuring the rank of the kernel confirms it is one, for that ψ and for anything else put in its place.[1]

So a Markov model can carry any maturity shape at all. What is wrong with example 12.1 is not its mathematics but its economics: the volatility attached to the rate maturing in 2055 is the same number in 2054 as it is today. A thirty year forward rate and a rate one day from fixing are given whatever volatility their maturity *dates* happen to have been assigned, with no relation between them. Real forward rates do not behave that way — what governs a rate's volatility is how long it has left to run.

Structure. This is chapter 10's complaint arriving in the rates market. There the trouble was a local volatility surface whose skew decayed along the calendar, so a forward starting option saw a flatter smile than a spot one; here it is a volatility whose shape is pinned to dates rather than to tenors, so a forward curve does not look like a spot curve. Both are the same failure: a parameter carrying calendar time where the market carries only elapsed time.

12.2.3 The Second Requirement

So we impose it, as a modelling demand and not as a theorem.

Definition 12.2 (Time-homogeneous volatility)**.** The forward rate volatility is time homogeneous if it depends on the maturity only through the time left to it,

$$\sigma(t,T) \;=\; \sigma_r(t,\omega)\, g(T-t), \tag{12.5}$$

for a *maturity profile* g and a level σ_r carrying whatever randomness there is.

Now put the two requirements together. Taking σ_r deterministic and non-vanishing, the sections of (12.5) are $u \mapsto \sigma_r(u)g(T-u)$ on $[0,t]$; the factor σ_r is common to all of them and so cannot change the dimension of their span, which leaves the sections of $g(T-u)$. Writing

$$T - u = \underbrace{(T-t)}_{a} + \underbrace{(t-u)}_{b} \tag{12.6}$$

turns a section into $b \mapsto g(a+b)$, indexed by $a \geq 0$. This is not a trick to make the algebra work. It is what time homogeneity does to the kernel: the sections of a kernel that depends only on the difference of its arguments *are* the shifts of one function, and the substitution merely names them.

So lemma 12.1 becomes, under definition 12.2: the model is Markov in n states if and only if the shifts of g span a space of dimension n. Write g_a for the shifted profile,

$$g_a(b) \;=\; g(a+b),$$

and V for the linear span of the g_a over $a \geq 0$. Expanding g_a in a basis $g_1, \ldots, g_n$ of V gives the form (12.3) again, now with the shift structure visible:

$$g(a+b) \;=\; \sum_{i=1}^{n} c_i(a)\, g_i(b), \qquad X_t^i = \int_0^t \sigma_r(u)\, g_i(t-u)\, dW_u. \tag{12.7}$$

Taking $a = 0$ shows $g \in V$: the profile is one of the shapes its own shifts span.

[1] the test that pins this.

12.2.4 Rank One Is the Exponential

Proposition 12.3. *Under definition 12.2, a model Markov in one state per factor has $g(x) = e^{-\kappa x}$, and no other profile will do.*

Proof. $\dim V = 1$ means V is spanned by g itself, so (12.7) reads $g(a+b) = c(a)\,g(b)$. Setting $b = 0$ gives $c(a) = g(a)/g(0)$, and normalising $g(0) = 1$ — the constant can be absorbed into σ_r — leaves

$$g(a+b) = g(a)\,g(b). \tag{12.8}$$

This is Cauchy's exponential equation, whose measurable solutions are $g(x) = e^{-\kappa x}$ and nothing else.[2] □

That is the model this chapter uses.

Definition 12.4 (Separable volatility)**.** The forward rate volatility is separable if it factorises as

$$\sigma(t,T) = \sigma_r(t,\omega)\, e^{-\int_t^T \kappa(u)\,du}, \tag{12.9}$$

with the first factor allowed to depend on time and on the state of the world, and the second a purely deterministic decay in the maturity.

Take κ constant for the rest of this chapter, so the decay is $e^{-\kappa(T-t)}$. Proposition 12.3 is what earns it: given time homogeneity, the exponential is not one convenient choice among many but the only one a single state permits.

Remark (Whether a varying mean reversion is allowed)**.** (12.9) writes $e^{-\int_t^T \kappa(u)du}$ and the proposition below concludes $e^{-\kappa x}$. Let's see how these two sit together.

Write $G(s) = e^{-\int_0^s \kappa(u)du}$, so that the decay in (12.9) is $G(T)/G(t)$ and

$$\sigma(t,T) = \frac{\sigma_r(t,\omega)}{G(t)} \cdot G(T).$$

That is a function of t times a function of T, whatever κ does — an outer product, of rank one as a kernel in (t,T).[3] So a time-dependent mean reversion is permitted and costs nothing: the state is still (x,y) and everything derived below goes through with $e^{-\kappa(T-t)}$ read as $G(T)/G(t)$.

What it costs is time homogeneity, which is the proposition's hypothesis rather than its conclusion. The maturity profile at date t is $x \mapsto G(t+x)/G(t)$, and that depends on t as well as on x: the shape a rate's volatility follows as it ages is a different shape depending on when one starts watching. Read as a fixed profile in time to maturity it is not an exponential at all, and its shifts span a space of dimension well above one.[4]

Which makes this example 12.1 again, in the notation the chapter actually uses. One state, because the volatility is indexed by the maturity date; no fixed profile, because it is not indexed by time remaining. The economic objection carries across unchanged, and so does the practical reason desks do it anyway: a varying κ is one of the cheapest ways to fit a term structure of volatility exactly, and what is bought with the fit is paid for in forward volatility, which is then whatever the calibration happened to leave behind.

[2] Measurability is not a formality one can drop: without it the equation has pathological solutions built from a Hamel basis. It is not a real restriction on a volatility.

[3] `computed`.

[4] `computed`.

12.2.5 Rank n Is a Differential Equation

Theorem 12.5 (Quasi-exponentials, and nothing else)**.** *Suppose g is smooth and $\dim V = n$ is finite. Then g satisfies a linear differential equation of order n with constant coefficients,*

$$g^{(n)} + \alpha_{n-1} g^{(n-1)} + \cdots + \alpha_1 g' + \alpha_0 g = 0, \tag{12.10}$$

and consequently

$$g(x) = \sum_j p_j(x)\, e^{\lambda_j x}, \tag{12.11}$$

a finite sum of polynomials times exponentials, with the λ_j the roots of the characteristic polynomial of (12.10) and $\deg p_j$ one less than the multiplicity of λ_j. Complex roots appear in conjugate pairs and contribute $e^{\mu x}\cos(\omega x)$ and $e^{\mu x}\sin(\omega x)$. Conversely every such g has $\dim V$ equal to the order of the equation it satisfies.

Proof. V is closed under shifting: a typical element is $h(x) = \sum_k \lambda_k\, g(x + a_k)$, and $h(x + s) = \sum_k \lambda_k\, g(x + a_k + s)$ is again a combination of shifts of g, so again in V.

That makes V closed under differentiation. For $h \in V$,

$$h'(b) = \lim_{s \to 0} \frac{h(b+s) - h(b)}{s},$$

and every difference quotient on the right lies in V, because both $x \mapsto h(x + s)$ and h do and V is a linear space. A finite dimensional subspace is closed, so the limit lies in V too.

So $D = d/dx$ is a linear operator on an n dimensional space. By Cayley-Hamilton it satisfies its own characteristic polynomial p, of degree n, on all of V; and $g \in V$. That is (12.10), with $p(\lambda) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_0$, and (12.11) is its classical solution. For the converse, the shifts of a solution of (12.10) are again solutions, so V sits inside a solution space of dimension n.[5] □

Read (12.10) back as a statement about the model: the *order of the differential equation the maturity profile obeys is the number of state variables.* Nothing else about g matters. Proposition 12.3 is the case $n = 1$, where (12.10) is $g' + \kappa g = 0$.

12.2.6 What Rank Two and Three Look Like

The cheapest way to see the general case is to expand $g(a + b)$ by hand. For $g(x) = x^k e^{-\kappa x}$ the binomial theorem gives

$$\underbrace{(a+b)^k e^{-\kappa(a+b)}}_{g(a+b)} \;=\; \sum_{j=0}^{k} \underbrace{\binom{k}{j} a^{k-j} e^{-\kappa a}}_{c_j(a)}\; \underbrace{b^j e^{-\kappa b}}_{g_j(b)}, \tag{12.12}$$

which is (12.7) exactly, with basis $g_j(b) = b^j e^{-\kappa b}$ for $j = 0, \ldots, k$. So $x^k e^{-\kappa x}$ needs $k + 1$ states, and (12.10) agrees: $(D + \kappa)^{k+1} g = 0$, a root of multiplicity $k + 1$.

[5] Smoothness is assumed rather than derived. It can be got from $\dim V$ being finite with more work, by solving (12.7) for the c_i at n well-chosen values of b and bootstrapping; but a volatility profile that is not smooth is not a modelling choice anyone makes.

The case $k = 1$ is the one that matters in practice, since a volatility peaking at some tenor rather than decaying from the front is closer to what is quoted. Written out, (12.12) is

$$g(a+b) \;=\; \underbrace{a e^{-\kappa a}}_{c_0(a)} \cdot \underbrace{e^{-\kappa b}}_{g_0(b)} \;+\; \underbrace{e^{-\kappa a}}_{c_1(a)} \cdot \underbrace{b e^{-\kappa b}}_{g_1(b)},$$

so the hump is Markov in the two states of (12.7) built from $g_0(b) = e^{-\kappa b}$ and $g_1(b) = be^{-\kappa b}$. The first is Hull-White's own state; the hump keeps it and adds one. This is an identity per increment rather than in a limit, so it holds pathwise and exactly.[6]

Everything else follows straight from the characteristic polynomial:

maturity profile $g(x)$	**its equation**	**states**
$e^{-\kappa x}$	$(D+\kappa)\, g = 0$	1
$x\, e^{-\kappa x}$	$(D+\kappa)^2 g = 0$	2
$x^2 e^{-\kappa x}$	$(D+\kappa)^3 g = 0$	3
$e^{-\kappa_1 x} + e^{-\kappa_2 x}$	$(D+\kappa_1)(D+\kappa_2)\, g = 0$	2
$e^{-\kappa x}\cos(\omega x)$	$\big(D^2 + 2\kappa D + \kappa^2 + \omega^2\big) g = 0$	2
$1/(1+x)$	none	no finite number
$e^{-\kappa\sqrt{x}}$	none	no finite number

Computing the rank of $(a, b) \mapsto g(a+b)$ numerically returns exactly this column,[7].

The last two rows deserve a word, since a theorem excluding nothing would hardly need proving. Neither function is pathological: both are smooth, positive and decaying, and either would be a reasonable thing to fit to a volatility term structure.

Example 12.2 (A profile with no finite realisation)**.** Take $g(x) = 1/(1+x)$ and suppose some finite combination of its shifts vanishes,

$$\sum_{j=1}^{m} \lambda_j \, \frac{1}{1 + a_j + b} = 0 \quad \text{for all } b \geq 0,$$

with the a_j distinct. The left side is a rational function of b with simple poles at $b = -(1+a_j)$, all distinct. Multiplying by $(1 + a_1 + b)$ and letting $b \to -(1+a_1)$ kills every term but the first and leaves $\lambda_1 = 0$; repeating kills them all. So no finite set of shifts is dependent, $\dim V$ is infinite, and there is no Markov realisation in any number of states.

Set that beside example 12.1, which used the same $1/(1+T)$ and got a model with *one* state. Nothing about the function changed. What changed is whether it was read as a shape in calendar time or a shape in time to maturity, and that alone moves the model from one state to infinitely many. It is as sharp a statement as this section has of where the exponential actually comes from: not from Markovianity, which tolerates any shape, but from insisting that the shape ride along with the maturity.

So being Markov in finitely many variables does not force separability. It forces *finite rank*, by lemma 12.1; time homogeneity then turns finite rank into a differential equation, by theorem 12.5; and separability is the case $n = 1$ of the two together. Multi-factor Cheyette models are the case $n > 1$, at n times the state and with nothing else changed.

[6] the pathwise test.

[7] quasigaussian::realisation_dimension.

Structure. This is chapter 14's subject arriving early. There the question is which models are solvable, and the answer is that a family of functions preserved by the dynamics turns a partial differential equation into a few ordinary ones. Here the dynamics is an infinite dimensional one on the curve, the family is V, and preservation is closure under shifting; the payoff is the same, an infinite dimensional object collapsing onto finitely many coordinates. Tractability keeps turning out to be an invariance, and the invariant object keeps turning out to be a small linear space — here small enough that its dimension is literally the number of state variables.

Remark (What survives when the level is stochastic)**.** Everything above assumed σ deterministic, and (12.9) explicitly permits a σ_r depending on the state — which is what makes the model useful, since it is where the smile comes from.

Sufficiency is untouched. Nothing in theorem 12.6 uses determinism: the integrals defining x, y and A are pathwise, and hold for any adapted σ_r. So a rank-n profile with an exponential shape and an arbitrary state-dependent level still gives a finite Markov state, which is the direction the chapter actually needs. The construction is safe.

Necessity is what the Itô isometry was buying, and that is where the argument breaks — though not because the isometry fails. It does not: $\mathbb{E}\big[(\int h\,dW)^2\big] = \mathbb{E}\big[\int h^2 du\big]$ holds for every adapted square integrable integrand, random ones included. What stops working is the use the proof makes of it.

For deterministic σ the family $\{u \mapsto \sigma(u,T)\}_T$ is a family of *fixed* vectors in $L^2[0,t]$, and the isometry embeds that space linearly and injectively into $L^2(\Omega)$. So the span of the random variables has exactly the dimension of the span of the functions, and both are spans over $\mathbb{R}$ — constant coefficients. That is the notion (12.4) needs, because there the curve is rebuilt by *deterministic* loadings $\psi_i(T)$ on n random variables. Linear algebra is the right tool because, with σ deterministic, the curve is Gaussian and everything in sight is linear.

Let σ_r depend on the state and neither remains true. The integrands are themselves random, so $\{\sigma(\cdot,T)\}_T$ is no longer a deterministic family in a fixed space; and a curve recoverable from n state variables is recoverable by a *measurable* function of them, which for a non-Gaussian curve has no reason to be linear. The span that matters is then over the state-measurable coefficients rather than over $\mathbb{R}$ — a module rather than a finite dimensional vector space — and a rank computed over $\mathbb{R}$ is not counting it. The isometry survives intact; what it no longer does is convert "n state variables" into "n linearly independent integrands". Lemma 12.1 loses its proof there.

The replacement is Lie-algebraic. Reparametrising the curve by time to maturity, $r_t(x) = f(t, t+x)$ — the Musiela form — turns (12.1) into a stochastic partial differential equation whose drift contains the transport term $\partial/\partial x$, because holding a tenor fixed means sliding along the curve as time passes. The model has a finite dimensional realisation exactly when the Lie algebra generated by its drift and volatility vector fields is finite dimensional.

Applied to a *deterministic* volatility, that criterion reproduces theorem 12.5. Bracketing the volatility field against the transport term differentiates the maturity profile; bracketing again differentiates it again. Finite dimensionality of the algebra therefore says that the derivatives of g span a finite dimensional space — which is (12.10), and so still the quasi-exponentials. The elementary proof and the geometric one arrive at the same object, V closed under d/dx, by different roads.

Once σ_r depends on the curve, those brackets acquire further terms from the volatility field's own dependence on the state, and the criterion stops reducing to a condition on g alone.

- What the chapter uses is the *if* direction, and it holds regardless. An exponential profile with

any adapted level whatever gives the two-state model of theorem 12.6.

- What is given up is the *only if*. For deterministic σ, lemma 12.1 is an equivalence, so we know that nothing outside the quasi-exponentials could have worked. With a stochastic level we no longer have that proof, and whether some cleverly state-dependent level could rescue a profile that is not quasi-exponential is a question nothing here settles. It has to be settled model by model, from the Lie algebra.

The deterministic characterisation is due to Bjork and Christensen, the state-dependent analysis to Bjork and Svensson, and the general geometry of invariant manifolds to Filipovic and Teichmann.

With that, we can find the state.

Theorem 12.6 (Cheyette, quasi-Gaussian state)**.** *Under* (12.9)*, define*

$$x_t = r_t - f(0,t), \qquad y_t = \int_0^t \sigma_r^2(u)\, e^{-2\kappa(t-u)}\, du.$$

Then (x, y) *is Markov, with*

$$dx_t = \big(y_t - \kappa x_t\big)\, dt + \sigma_r(t)\, dW_t, \qquad dy_t = \big(\sigma_r^2(t) - 2\kappa y_t\big)\, dt, \tag{12.13}$$

both starting from zero, and the entire curve is recovered from them by

$$P(t,T) = \frac{P(0,T)}{P(0,t)} \exp\Big(- G(t,T)\, x_t - \tfrac{1}{2} G^2(t,T)\, y_t \Big), \qquad G(t,T) = \frac{1 - e^{-\kappa(T-t)}}{\kappa}. \tag{12.14}$$

Proof. Integrate (12.1) from 0 to t with the separable volatility. The stochastic part is

$$\int_0^t \sigma_r(u) e^{-\kappa(T-u)}\, dW_u = e^{-\kappa(T-t)} \int_0^t \sigma_r(u) e^{-\kappa(t-u)}\, dW_u,$$

which is the factorisation the previous remark promised: the maturity has come outside, leaving a single history-dependent quantity. Call that quantity X_t.

For the drift, the inner integral in (12.1) is

$$\int_u^T \sigma_r(u) e^{-\kappa(s-u)}\, ds = \sigma_r(u)\, \frac{1 - e^{-\kappa(T-u)}}{\kappa} = \sigma_r(u)\, G(u,T),$$

so the accumulated drift is $\int_0^t \sigma_r^2(u) e^{-\kappa(T-u)} G(u,T)\, du$. Since $\kappa G(u,T) = 1 - e^{-\kappa(T-u)}$, the integrand is

$$\tfrac{1}{\kappa}\, \sigma_r^2(u) \Big[e^{-\kappa(T-u)} - e^{-2\kappa(T-u)} \Big],$$

and splitting both exponentials at t as before gives

$$\int_0^t \alpha(u,T)\, du \;=\; \frac{1}{\kappa} \Big[e^{-\kappa(T-t)} A_t \;-\; e^{-2\kappa(T-t)} y_t \Big], \qquad A_t = \int_0^t \sigma_r^2(u)\, e^{-\kappa(t-u)}\, du. \tag{12.15}$$

Note that *two* accumulators have appeared, A at rate κ and y at rate 2κ, so the state looks three dimensional at this point.

It is not. Setting $T = t$ in (12.15) and adding the stochastic part,

$$r_t = f(t,t) = f(0,t) + \frac{A_t - y_t}{\kappa} + X_t, \qquad \text{so} \qquad x_t = X_t + \frac{A_t - y_t}{\kappa}.$$

That is a relation among the three, and using it to eliminate A_t from (12.15),

$$\begin{aligned} f(t,T) - f(0,T) &= e^{-\kappa(T-t)} X_t + \frac{1}{\kappa}\Big[e^{-\kappa(T-t)} A_t - e^{-2\kappa(T-t)} y_t\Big] \\ &= e^{-\kappa(T-t)} \underbrace{\Big[X_t + \frac{A_t - y_t}{\kappa}\Big]}_{x_t} + \frac{y_t}{\kappa}\, e^{-\kappa(T-t)}\Big[1 - e^{-\kappa(T-t)}\Big], \end{aligned}$$

in which A_t has cancelled. The bracket on the right is $\kappa\, G(t,T)$, so

$$f(t,T) = f(0,T) + e^{-\kappa(T-t)}\, x_t + e^{-\kappa(T-t)} G(t,T)\, y_t. \tag{12.16}$$

Two states, not three, because the short rate itself supplies the third relation.

Integrating (12.16) in T and exponentiating, using $P(t,T) = \exp(-\int_t^T f(t,s)ds)$ from chapter 7, gives (12.14): the $\int_t^T e^{-\kappa(s-t)}ds$ produces $G(t,T)$, and

$$\int_t^T e^{-\kappa(s-t)} G(t,s)\, ds = \frac{1}{\kappa}\int_0^{T-t} \big(e^{-\kappa v} - e^{-2\kappa v}\big) dv = \tfrac{1}{2} G^2(t,T)$$

produces the quadratic term, after cancelling against the initial curve.

Finally the dynamics. Differentiating the three pieces separately,

$$dX_t = -\kappa X_t\, dt + \sigma_r\, dW_t, \qquad dA_t = \big(\sigma_r^2 - \kappa A_t\big) dt, \qquad dy_t = \big(\sigma_r^2 - 2\kappa y_t\big) dt,$$

the last of which is already the second equation of (12.13). For the first, $d\big[(A_t - y_t)/\kappa\big] = (2y_t - A_t)\, dt$, so

$$dx_t = \big(-\kappa X_t + 2y_t - A_t\big) dt + \sigma_r\, dW_t = \Big(y_t - \kappa\big[X_t + (A_t - y_t)/\kappa\big]\Big) dt + \sigma_r\, dW_t,$$

and the bracket is x_t. Note that A has cancelled again: it is needed to write the curve down and not to step it forward. □

Three things about this deserve emphasis, because they are what make the model usable rather than merely finite.

First, y has no dW in it. It is not a risk factor and cannot be shocked; it is a running accumulator of variance, and its only job is to make the pair Markov. It is exactly the memory that (12.1)'s drift needs and that x alone cannot supply.

Second, (12.14) reconstructs *every* discount factor as an explicit function of two numbers. So every forward rate, every swap rate and every annuity is a closed-form function of (x, y). That is what makes backward induction possible: a two-dimensional grid, or a regression on (x, y) in a Monte Carlo, and the exercise decision can be made state by state.

Third, today's curve enters only through the ratio $P(0,T)/P(0,t)$, which is read straight off chapter 7's bootstrap. The initial curve is matched exactly and automatically, with nothing to calibrate.

Example 12.3 (Climbing down the ladder to Hull-White)**.** The quickest way to believe Theorem 12.6 is to take the one case where we already know the answer.

Set $\sigma_r(t) = \sigma$, a constant. The equation for y in (12.13) has no randomness in it at all, so it is an ordinary linear differential equation, $\dot{y} = \sigma^2 - 2\kappa y$ with $y_0 = 0$. Solving,

$$y_t = \frac{\sigma^2}{2\kappa}\Big(1 - e^{-2\kappa t}\Big),$$

which is a known function of time and not a state variable at all. The model has collapsed to one dimension.

What is left is

$$dx_t = \big(y_t - \kappa x_t\big)dt + \sigma\, dW_t,$$

a Gaussian Ornstein-Uhlenbeck process with a time-dependent drift, and since $r_t = f(0,t) + x_t$, the short rate is

$$dr_t = \big(\theta(t) - \kappa r_t\big)dt + \sigma\, dW_t, \qquad \theta(t) = \frac{\partial f(0,t)}{\partial t} + \kappa f(0,t) + y_t.$$

This is exactly the Hull-White short rate equation from chapter 8, with θ assembled from the initial curve and the model's own accumulated variance — which is what chapter 8 found there too, by a completely different route.

And (12.14) becomes

$$P(t,T) = \frac{P(0,T)}{P(0,t)} \exp\Big(- G(t,T)\big(r_t - f(0,t)\big) - \tfrac{1}{2}G^2(t,T)\, y_t\Big),$$

an exponential of something affine in r_t — the affine term structure chapter 8 observed and did not explain. The next section explains it.

Remark (Where Hull-White sits)**.** Set σ_r deterministic. Then y_t is a deterministic function of time and drops out as a state variable, leaving a one-factor Gaussian model with an affine bond formula — which is the Hull-White model of chapter 8, met there through the short rate and arrived at here from the other direction. The ladder is:

HJM (general, infinite state) $\longrightarrow$ impose exponential separability $\longrightarrow$ quasi-Gaussian (Markov in (x, y)) $\longrightarrow$ make σ_r deterministic $\longrightarrow$ Hull-White.

Each arrow is a restriction bought for a computational gain, and knowing which one you are standing on is most of knowing what your model can and cannot do.

12.3 Excursion: Affine Models and Why They Solve

Formula (12.14) has a particular shape — an exponential of something linear in the state, plus a correction — and it is not a coincidence. It is an instance of a general structure that accounts for essentially every interest rate model with a closed form.

Definition 12.7 (Affine term structure model)**.** A model with state $X_t \in \mathbb{R}^n$ has an affine term structure if

$$P(t,T) = \exp\big(A(t,T) + B(t,T)^\top X_t\big)$$

for deterministic functions A and B.

The question is which state processes produce this, and the answer is clean.

Theorem 12.8 (Duffie and Kan)**.** *Suppose the short rate is affine in the state, $r_t = \delta_0 + \delta^\top X_t$, and the state is a diffusion whose drift and* covariance *are both affine in the state:*

$$dX_t = \big(a + bX_t\big)dt + \Sigma(X_t)\, dW_t, \qquad \big[\Sigma(x)\Sigma(x)^\top\big]_{ij} = c_{ij} + d_{ij}^\top x.$$

Then the model has an affine term structure, and A and B solve a system of ordinary differential equations — Riccati equations — in the maturity.

Proof. The bond price satisfies the pricing equation of chapter 5 with this model's generator — discounted, it is a martingale — which written out for a multi-factor state is

$$\frac{\partial P}{\partial t} + (a + bx)^\top \nabla P + \frac{1}{2}\operatorname{tr}\big(\Sigma\Sigma^\top \nabla^2 P\big) - r\,P = 0, \qquad P(T,T) = 1.$$

Substitute the guess $P = \exp(A + B^\top x)$. Every derivative brings down a factor of the exponential, which cancels throughout, leaving

$$\dot{A} + \dot{B}^\top x + (a + bx)^\top B + \frac{1}{2}B^\top\big(c + d^\top x\big)B - \delta_0 - \delta^\top x = 0,$$

where the dots are derivatives in t. Now the key step: this must hold for *every* value of x, and every term is either constant in x or linear in it. A polynomial vanishing identically has vanishing coefficients, so the constant and linear parts must separately be zero:

$$\begin{aligned} \dot{A} + a^\top B + \tfrac{1}{2}B^\top c\, B - \delta_0 &= 0, \\ \dot{B} + b^\top B + \tfrac{1}{2}B^\top d\, B - \delta &= 0, \end{aligned}$$

with $A(T,T) = 0$ and $B(T,T) = 0$. These are ordinary differential equations in one variable, quadratic in the unknown — Riccati equations. □

Remark (What "solvable" actually means here)**.** The pricing problem for a general model is a partial differential equation in n space dimensions plus time, and solving it costs work that grows exponentially in n. What the affine structure achieves is to reduce that PDE to a system of $n + 1$ *ordinary* differential equations, which cost essentially nothing and are often available in closed form.

The mechanism is the separation in the proof: because the coefficients are affine and the guess is exponential-affine, the state variable x appears only linearly and can be matched off, leaving equations for functions of maturity alone. The state has been eliminated from the problem entirely.

The condition is genuinely restrictive. The drift being affine is mild. The requirement that the *covariance* be affine in the state is what excludes almost everything: a volatility proportional to x makes the covariance proportional to x^2, and the match fails. This is why the tractable models are the Gaussian ones (Σ constant) and the square-root ones ($\Sigma \propto \sqrt{x}$, so the covariance is proportional to x) and essentially nothing else. Heston's variance process in chapter 10 is a square-root process for exactly this reason, and its closed-form characteristic function is Theorem 12.8 applied to the log-price rather than to a bond.

Structure (Why affine gives Riccati, in one line)**.** The system above is usually presented as what falls out of substituting an ansatz. See why the ansatz works, because the reason generalises and the algebra does not.

Apply the generator of chapter 3 to an exponential, $f(x) = e^{\psi x}$, for a diffusion whose drift is $b + \beta x$ and whose variance is $a + \alpha x$:

$$\mathcal{L}e^{\psi x} = \Big[\underbrace{b\psi + \tfrac{1}{2}a\psi^2}_{\text{constant}} + x\big(\underbrace{\beta\psi + \tfrac{1}{2}\alpha\psi^2}_{\text{coefficient of } x}\big)\Big]e^{\psi x}.$$

The generator has mapped an exponential-affine function to an *affine multiple* of itself. It does not say the family is mapped into itself: $(c_0 + c_1 x)\, e^{\phi+\psi x}$ is not of the form $e^{\phi'+\psi' x}$, so $\{e^{\phi+\psi x}\}$ is not preserved by $\mathcal{L}$. What has happened is better suited to the purpose.

Regard the family as a two-dimensional surface in function space, with coordinates (ϕ, ψ). Differentiating along each coordinate gives the tangent directions,

$$\frac{\partial}{\partial\phi}e^{\phi+\psi x} = e^{\phi+\psi x}, \qquad \frac{\partial}{\partial\psi}e^{\phi+\psi x} = x\, e^{\phi+\psi x},$$

so the tangent space at any point of the surface is precisely the set of affine multiples of that point. The display above therefore says that $\mathcal{L}$ carries each point of the surface into *its own tangent space*: the vector field is tangent to the surface.

That is exactly the condition for the surface to be invariant under the flow $\partial_t u = \mathcal{L}u$, which is the object we actually care about — invariance under the semigroup, not under the generator. A flow tangent to a two-dimensional surface is a pair of ordinary differential equations in its coordinates, whatever the underlying problem was. And because both sides of $\partial_t u = \mathcal{L}u$ now lie in the same tangent space, they can be compared in its basis $\{1, x\} \cdot e^{\phi+\psi x}$, which is all the next line does. Chapter 14 takes this distinction as its starting point and makes it general.

Reading the bracket gives them directly. The constant term is ϕ' and the coefficient of x is ψ', so

$$\phi' = b\psi + \tfrac{1}{2}a\psi^2, \qquad \psi' = \beta\psi + \tfrac{1}{2}\alpha\psi^2,$$

and the second is quadratic in ψ — a Riccati equation — for the single reason that the *variance* was allowed to depend on the state. Vasicek has $\alpha = 0$, the quadratic term disappears, and the equation is linear; the square-root model has $\alpha \neq 0$ and it is not.

So "affine" is not a description of the coefficients so much as a statement about the generator, and the Riccati equation is not a coincidence but the shadow of a two-dimensional invariant family. Chapter 14 asks which other families work.

Example 12.4 (Solving the Riccati equations for Vasicek)**.** The abstract statement is more convincing once the equations have actually been solved once, and in one dimension they can be.

Take $n = 1$, $r_t = x_t$, and $dx_t = \kappa(\bar{x} - x_t)dt + \sigma\, dW_t$: the Vasicek model. In the notation of Theorem 12.8, $a = \kappa\bar{x}$, $b = -\kappa$, the covariance is the constant σ^2 so $c = \sigma^2$ and $d = 0$, and $\delta_0 = 0$, $\delta = 1$. Writing $\tau = T - t$ so the equations run forward in τ from zero, they become

$$\begin{aligned} \frac{dB}{d\tau} &= -\kappa B - 1, & B(0) &= 0, \\ \frac{dA}{d\tau} &= \kappa\bar{x}B + \tfrac{1}{2}\sigma^2 B^2, & A(0) &= 0. \end{aligned}$$

Because $d = 0$ the first equation is *linear* — the quadratic term of the Riccati equation has vanished — and it integrates immediately:

$$B(\tau) = -\frac{1 - e^{-\kappa\tau}}{\kappa},$$

which is $-G$ from (12.14), arrived at yet again. The second equation is then a plain integral of known functions, giving A in closed form.

So a bond price in the Vasicek model costs two evaluations of an exponential. This is what Theorem 12.8 is worth: the same problem posed as a partial differential equation would have to be solved numerically on a grid, for every maturity, every time the parameters moved.

Notice also exactly where the linearity came from. It came from $d = 0$, that is, from the covariance not depending on the state — the Gaussian case. In the square-root case $d \neq 0$, the quadratic term survives, and B solves a genuine Riccati equation. It still has a closed form, which is why the square-root process is the other tractable one, but the algebra is a page rather than a line.

Remark (Quadratic Gaussian models)**.** There is one more family that solves, and it looks at first like a counterexample. Take X to be an ordinary Gaussian process — constant Σ — but make the short rate *quadratic* in it:

$$r_t = X_t^\top Q X_t + \delta^\top X_t + \delta_0.$$

Guessing $P = \exp(A + B^\top x + x^\top C x)$ and repeating the argument of Theorem 12.8, the exponential again cancels and one matches the constant, linear *and quadratic* coefficients in x, obtaining ordinary differential equations for A, B and a matrix Riccati equation for C. The model solves.

It is not really an exception. Adjoin the products $X_i X_j$ to the state; then r is affine in the enlarged state, and the enlarged state's drift and covariance are affine in it too, because X is Gaussian. Quadratic Gaussian models are affine models wearing a smaller state.

What one buys is a smile. With Q non-zero, the rate is a quadratic function of a Gaussian variable, so it is not Gaussian, and it is bounded below if Q is positive definite — which gives a distribution with genuine skew and curvature while keeping closed-form bonds.

12.4 Where the Smile Lives

Return now to the quasi-Gaussian state equations (12.13). Nothing so far has said what σ_r is, and this is where the volatility chapters come back.

The model is called quasi-Gaussian because it is Gaussian whenever σ_r is deterministic, and departs from Gaussian only through the state dependence of σ_r.

σ_r	name	what it gives
$\sigma(t)$	Hull-White	Gaussian, closed forms, no smile
$\sigma(t)\big[a(t) + b(t)x\big]$	linear quasi-Gaussian	a skew; b is the skew dial
$\sigma(t)\big[a(t) + b(t)x + c(t)x^2\big]$	quadratic quasi-Gaussian	skew and curvature: a genuine smile
$\sqrt{z_t}\,\big[a + bx + cx^2\big]$	stochastic-vol quasi-Gaussian	vol-of-vol, forward smile, real vega

The state x is a one-dimensional diffusion with volatility $\sigma_r(x)$, so it is a local volatility model — of the state rather than of the underlying, but the same object chapter 9 studied. Two results carry straight over.

Gyongi's theorem says the distribution of x_T depends on the volatility only through its value at each level, so the shape of σ_r as a function of x is precisely what shapes the distribution, and therefore the smile. Nothing else about the model can affect it.

And chapter 10's midpoint rule says how: the implied volatility at a strike is σ_r averaged over the journey from the forward to that strike. Averaging preserves shape and halves slopes. So a σ_r that is

- constant averages to a constant, and the smile is flat;
- linear in x averages to something linear, and the smile is a straight tilt at half the slope;
- quadratic in x averages to something quadratic, and the smile acquires curvature.

The quadratic term is the first shape that survives averaging as a *bend*, which is why it is the first one that produces a smile rather than a skew.

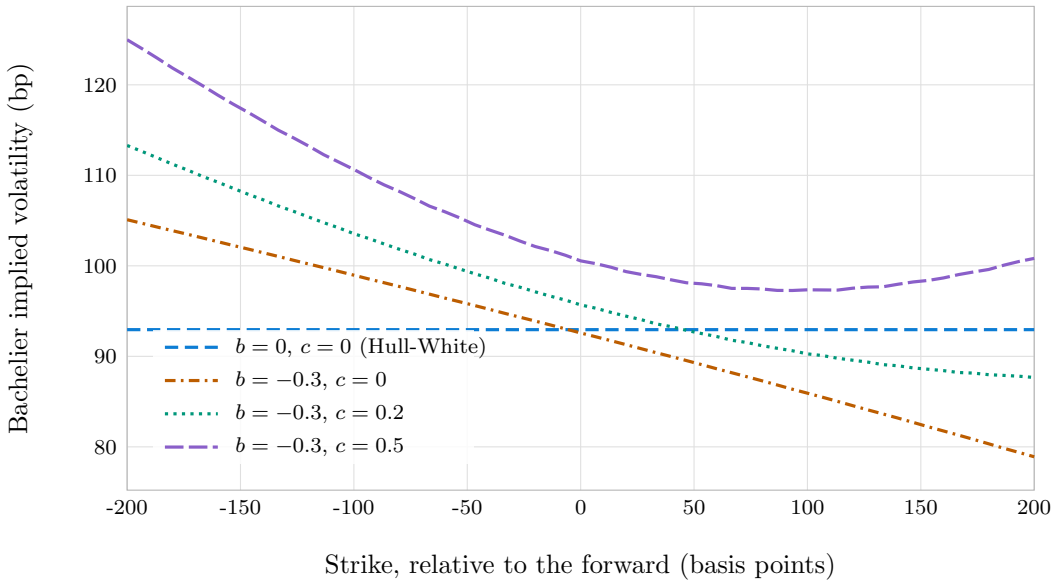

Figure 12.1: The smile a quadratic σ_r produces. Each curve solves the model rather than approximating it: the Fokker-Planck equation for the state is integrated forward and the options priced against the resulting density, so the curvature shown is the curvature the model has. With $b = c = 0$ the model is Hull-White and the Bachelier smile is exactly flat. The linear term tilts it without bending it. The quadratic term bends it, and more so as c grows.
Drawn by `QuasiGaussian::sigma_r`, `QuasiGaussian::density` and `QuasiGaussian::smile`.

Two details stand out.

The first is that the bent smile is not centred at the money. Its lowest point does not sit at a strike equal to the forward, and the averaging rule says exactly where it sits. Averaging $\sigma_r(u) = \sigma(1 + bu + cu^2)$ over the journey from the forward out to u,

$$\frac{1}{u}\int_0^u \sigma\big(1 + bz + cz^2\big)\,dz \;=\; \sigma\left(1 + \frac{b}{2}\,u + \frac{c}{3}\,u^2\right),$$

so the tilt is halved and the bend is divided by three. The smile is therefore least at

$$u_{\min} = -\frac{3b}{4c}, \tag{12.17}$$

against $-b/(2c)$ for σ_r itself: the vertex moves *outward* by half again, because averaging damps the quadratic term harder than the linear one and so leaves the tilt relatively stronger.[8]

The plot spans ± 200 basis points, which at these parameters is ± 0.89 of a $\sigma\sqrt{T}$ move. Both bent curves carry $b = -0.3$, which matters because (12.17) reads the vertex off the ratio and naming only c would name the wrong half of it. The $b = -0.3$, $c = 0.5$ curve bottoms out at $u_{\min} = 0.45$, a hundred basis points above the forward, and that is on the picture. The $b = -0.3$, $c = 0.2$ curve bottoms out at $u_{\min} = 1.125$, some two hundred and fifty basis points above — off the right edge, so what is drawn is only the descent towards it.

Remark (Which mean, given that chapter 10 says harmonic)**.** The average taken above is the arithmetic one, and the short-expiry theorem of chapter 10 gives the implied volatility as the *harmonic* mean of the local volatility along the path. The two are not the same rule, so here is why the simpler one is being used and what it costs.

Write $\sigma_r = \sigma(1 + \varepsilon)$ with $\varepsilon(u) = bu + cu^2$. Expanding the reciprocal,

$$\left\langle \tfrac{1}{1+\varepsilon} \right\rangle^{-1} \;=\; 1 + \langle\varepsilon\rangle - \mathrm{Var}(\varepsilon) + O(\varepsilon^3),$$

so the harmonic mean is the arithmetic mean less the variance of the perturbation: the two agree to first order in (b, c) and part company only at second. And the vertex is a first-order statement — it is the stationary point of $\langle\varepsilon\rangle$ — so (12.17) is what either rule gives at leading order.

What the difference is worth can be measured rather than bounded, since the density is solved exactly. Across the parameter sets used here the arithmetic rule places the vertex within 4.4% of the solved minimum and the harmonic rule within 3.4%.[9] The harmonic rule is better everywhere and by very little, and both leave a residual larger than the gap between them — because the theorem is a short-maturity limit and these are five year options. The choice of mean is not what the approximation is losing.

The consequence for a desk is that b and c do not act independently. The tilt and the bend share one vertex, and (12.17) places it by the *ratio* b/c rather than by either coefficient alone, so fitting a skew and a curvature is one fit rather than two.

The second is the units. The dials multiply $u = x/\text{scale}$ rather than x itself, and the scale is the typical move $\sigma\sqrt{T}$. That number is not chosen for convenience; it is the spread of the state.

Calculation 12.9 (Where $\sigma\sqrt{T}$ comes from)**.** Hold σ_r at σ for a moment, so x is the Hull-White state $dx = (y - \kappa x)dt + \sigma\, dW$. It is Gaussian, and integrating its variance gives

$$\mathrm{sd}(x_T) \;=\; \sigma\sqrt{\frac{1 - e^{-2\kappa T}}{2\kappa}} \;\longrightarrow\; \sigma\sqrt{T} \qquad \text{as } \kappa T \to 0. \tag{12.18}$$

So $\sigma\sqrt{T}$ is how far the state typically travels by T, in the limit of no mean reversion, and u counts standard deviations of the state. At $\sigma = 100$ basis points and $\kappa = 0.03$ the approximation is

[8] the test that pins this.

[9] measured.

worth 99% of the exact spread at one year and 93% at five, falling to 68% at thirty — where mean reversion has begun to bite and the true spread is heading for its ceiling of $\sigma/\sqrt{2\kappa}$, about 410 basis points, while $\sigma\sqrt{T}$ goes on growing.

With that, $u = 1$ means a one standard deviation move of the state, and b and c become dimensionless: b is how much the volatility tilts over one typical move, c how much it bends over one. Those are questions with comparable answers at every expiry, which is exactly what coefficients in fixed rate units would not give. A c multiplying x^2 in basis points describes a gentle bend where the state reaches a hundred of them and a violent one where it reaches three hundred, so the same number would mean something different at one year and at ten — a statement about the parametrisation rather than about the model. Normalising by the distance the state actually covers removes that.

It is a convention for reading the dials rather than a claim that one pair of dials fits every expiry. The model carries a single scale, and (12.18) is the number to set it to for the expiry being fitted.

Remark (Then why not simply use a local volatility model?)**.** Dupire's construction fits every quoted option exactly and needs no dials at all, so the resemblance above invites an obvious question. The answer is that the two are not competing for the same job.

They model different objects. A local volatility model gives the marginal distribution of one underlying. This gives the entire curve — every forward rate, every bond, every swap rate, at every date — out of the same two numbers. One can certainly fit a Dupire surface to the ten year swap rate, but it says nothing about the two year, nothing about the same ten year rate seen from a future date, and nothing that guarantees the two fits are consistent with any single arbitrage-free model. Here they are one model by construction.

The volatility is indexed differently. Dupire's $\sigma_{\text{loc}}(t, K)$ is a function of calendar time and strike. Here $\sigma_r(x)$ is a function of the *state*, with no t in it and no strike. That is the same distinction chapter 10 turned on: a shape written in the calendar decays along the calendar, and a shape written in the state does not. What forward smile term structure this model has therefore comes from its own state settling down — y starts at zero and climbs towards $\sigma^2/2\kappa$ over a mean reversion time — and not from a dated coefficient. It is the same caveat chapter 10 attached to starting Heston away from its long-run variance.

The state is small, and that is the entire point of the chapter. Two numbers fit on a grid, so backward induction works and a Bermudan can be priced. Dupire's surface is not the obstacle there; the obstacle is that a local volatility model of a swap rate is not a model of the curve the callable depends on.

And the fit is worse. This is the cost. Three parameters fit the *shape* of a smile; they do not reprice a strip of quotes to the penny, and chapter 10's local-stochastic volatility construction exists precisely because desks want both. What is bought for that is stability: a shape with three dials cannot develop the local ripples that differentiating a noisy surface twice in strike produces, which is the instability chapter 9 warns about.

The cap on σ_r does not bind anywhere in this figure — at these coefficients the volatility reaches about 167 basis points at the edge of the plot against a ceiling of 500 — which is the point to restate the warning below rather than assume the model is safe because one picture looked reasonable. Push c to three or four and the cap is all that stands between the model and a volatility that grows without limit.

The last row of the table adds an independent variance factor, usually a square-root process $dz = \theta(1 - z)dt + \eta\sqrt{z}\, dZ$, chosen to be square-root for the reason the previous section gave. The parallel with chapter 11 is exact: the polynomial in x is the leverage function, carrying the fit, and z is the stochastic part, carrying the dynamics. The same division of labour, applied to a curve.

Remark (The quadratic term will try to explode)**.** The standard existence theorem asks for a diffusion coefficient of *at most* linear growth: locally Lipschitz coefficients obeying $|\sigma(x)| \le C(1 + |x|)$ give a unique strong solution for all time. A quadratic coefficient grows faster than that, so the theorem does not apply — the growth bound is the hypothesis being broken, not the conclusion being met.

Losing a sufficient condition is not by itself losing the solution. It does. For a driftless diffusion $dX = \sigma(X)dW$, Feller's test says the process reaches infinity in finite time with positive probability exactly when

$$\int^{\infty} \frac{x}{\sigma^2(x)}\, dx \;<\; \infty,$$

and the two cases sit on either side of that. With $\sigma(x) = x$ the integrand is $1/x$ and the integral diverges, which is why geometric Brownian motion never explodes. With $\sigma(x) = x^2$ it is x^{-3} and the integral converges: the process does reach infinity in finite time, and the model's moments go with it.

A quadratic σ_r can also go negative, which is meaningless.

Production implementations therefore cap and floor σ_r, or replace the raw polynomial by a smooth interpolation between an absolute and a proportional regime. This is not a numerical detail bolted on afterwards; a quadratic-volatility model without it is not well posed.

12.5 More Than One Factor

Everything above used a single Brownian motion, and that carries a consequence which is easy to miss and expensive to ignore.

Look again at (12.16). With one factor, the whole curve is driven by the single state x_t, so every point of it moves in the same direction on any given day, by an amount $e^{-\kappa(T-t)}$ that depends only on maturity. The two year rate and the thirty year rate are *perfectly correlated*. The curve can shift and it can steepen as it shifts, but it cannot do the two independently, and it can never twist.

That is empirically false. A principal component analysis of a year of Treasury curve changes finds three shapes, of which a one-factor model can represent only the first.

Calculation 12.10 (How the components were found, and what a loading is)**.** The figure is a principal component analysis. The same decomposition returns in chapter 13 as the low-rank form imposed on a correlation matrix.

The data. One row per business day, one column per quoted tenor, and the entries are daily *changes* rather than levels. So the observations are days and the features are tenors — thirteen of them — and a single observation is one day's move of the whole curve, a vector $M \in \mathbb{R}^{13}$.

The decomposition. Subtract each tenor's mean change and form the covariance $C = \mathbb{E}[MM^{\top}]$, thirteen by thirteen and symmetric positive semi-definite. Any such matrix diagonalises with orthonormal eigenvectors,

$$C \;=\; \sum_{j} \lambda_j\, v_j v_j^{\top}, \qquad v_j^{\top} v_k = \delta_{jk}, \tag{12.19}$$

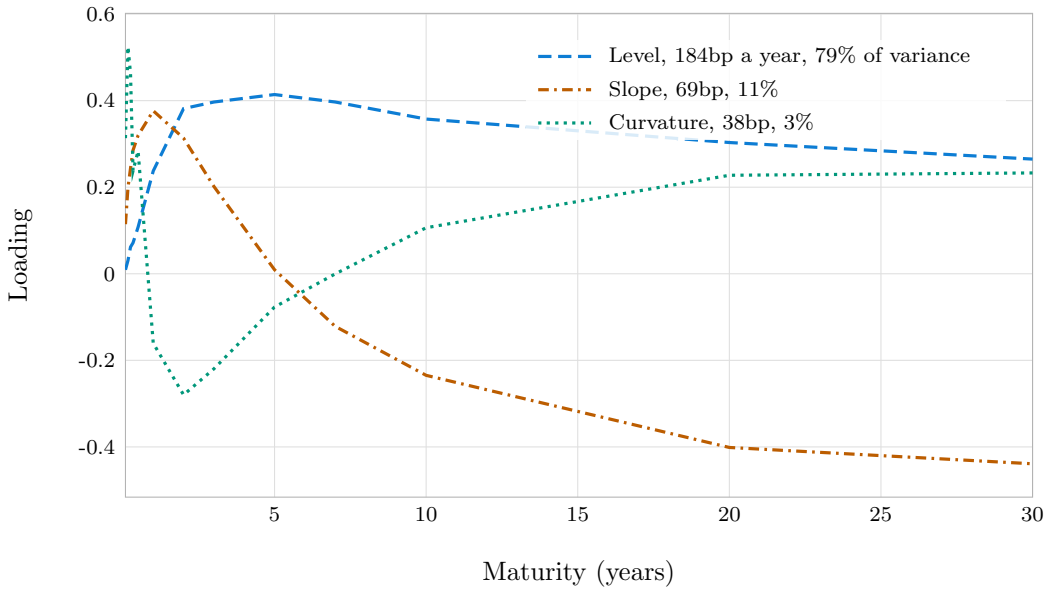

Figure 12.2: The three leading components of daily changes in the US Treasury par curve, over the year to the date shown. The first moves every tenor the same way and is the only shape (12.16) can produce — the one-factor loading $e^{-\kappa(T-t)}$ is a positive decay across maturities, so it lives entirely in this component. The second crosses zero once: the short end and the long end move in opposite directions, which is a steepening, and a one-factor model asserts it does not happen. The third crosses twice, the wings against the belly. Read the legend for what each costs to omit: the second and third carry a seventh of the variance of the first between them, which sounds ignorable and is not, because a hedged book has removed its exposure to the first.

Drawn by `curve_factors`. US Department of the Treasury, daily par yield curve rates, 2026-01-02 to 2026-08-07 (par yield, semiannual coupon, actual/actual). Retrieved from https://home.treasury.gov/interest-rates-data-csv-archive.

and the computation is nothing more than reading off the leading three: take the top eigenvector by power iteration, subtract $\lambda_1 v_1 v_1^\top$, and repeat.[10]

What that says about the curve. Equation (12.19) is exactly the statement that the day's move can be written

$$M_i \;=\; \sum_j \sqrt{\lambda_j}\, v_{j,i}\, W_j, \tag{12.20}$$

with W_j uncorrelated and of unit variance — multiply (12.20) out and its covariance is (12.19). So the analysis has rewritten thirteen correlated tenor moves as thirteen *independent* drivers, and the jth eigenvector is the vector of loadings of the tenors on the jth driver. That is the sense in which a curve has factors at all, and it is what the plotted shapes are.

Why three. Truncating (12.19) after d terms is the best rank-d approximation of C there is, in the sense that no other matrix of rank d is closer in Frobenius norm. So keeping three factors is not a convenient simplification but the optimal one at that rank, and the share the three account for is exactly what the approximation retains. Chapter 13 imposes the same truncation on a correlation matrix for a different reason — there it is the market's failure to identify all the entries rather than a wish to see the shapes — and inherits the same consequence, which the next paragraphs draw out: d factors admit exactly d shapes of curve move, and a payoff sensitive to a $(d+1)$th is priced as though it did not exist.

The omission is expensive for any product whose payoff depends on the shape of the curve rather than its level. The clearest case is a constant maturity swap spread option, which pays on the difference between a long and a short swap rate.

Perfect correlation does *not* make the spread deterministic. The two rates load on the state by different amounts — a short swap by nearly all of it, a long one by less, since the loading $e^{-\kappa(T-t)}$ has further to decay — so the difference still moves. What perfect correlation does is put a floor under it, and put the model on the floor.

Calculation 12.11 (How much a factor costs)**.** For two rates with volatilities σ_1 and σ_2 and correlation ρ, the spread has volatility

$$\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\,\sigma_1\sigma_2} \;\longrightarrow\; \big|\sigma_1 - \sigma_2\big| \qquad \text{as } \rho \to 1, \tag{12.21}$$

and the right-hand side is the *minimum* of the left over all ρ. A one-factor model sets $\rho = 1$, since every rate is an increasing function of the same x, so it always delivers exactly that minimum.

Taking the loading of an n-year swap rate as the average of $e^{-\kappa u}$ over its life, $\beta(n) = (1 - e^{-\kappa n})/(\kappa n)$, at $\kappa = 3\%$:[11]

spread	$\rho = 1$ **(one factor)**	$\rho = 0.9$	**ratio**
2y against 10y	11%	44%	4.0
2y against 30y	32%	49%	1.5

as a percentage of the two year rate's own volatility. So the one-factor model gives the two-year-against-ten-year spread about a quarter of the volatility a plausible correlation would, and prices the option accordingly. Not zero — a quarter.

[10] `risk::curve_factors`.

[11] `quasigaussian::spread_volatility`.

Note where the damage is worst. It is not the widest spread but the narrowest, because there the two loadings are close, so $|\sigma_1 - \sigma_2|$ is nearly nothing while a correlation below one leaves two almost uncancelled volatilities behind. The 2s10s trade is both the standard one and the one a single factor handles worst.

And there is no dial. Once σ, b and c are fitted to swaptions, (12.21) is determined: the model has no parameter that moves the correlation, so it cannot be marked to a spread option's price even in principle. That is chapter 20's question answered in the negative — the payoff is a bet on decorrelation, and the model has not assumed a view on decorrelation so much as assumed it away.

The construction generalises without difficulty. With n Brownian motions, take $x_t \in \mathbb{R}^n$ and let y_t become an $n \times n$ matrix of accumulated covariances:

$$dx_t = \big(y_t \mathbf{1} - \kappa x_t\big)dt + \sigma_r^\top dW_t, \qquad \frac{dy_t}{dt} = \sigma_r^\top \sigma_r - \kappa y_t - y_t \kappa,$$

and the bond reconstruction keeps its shape with the scalar products becoming quadratic forms:

$$P(t,T) = \frac{P(0,T)}{P(0,t)} \exp\Big(- G(t,T)^\top x_t - \tfrac{1}{2} G(t,T)^\top y_t \, G(t,T)\Big).$$

The cost is dimensionality. The state is x together with the distinct entries of the symmetric matrix y, so $n + n(n+1)/2$ variables: two factors means five states, three means nine. Five is comfortable in a Monte Carlo with a regression on the states, and marginal on a grid; nine is Monte Carlo only. Since y carries no randomness of its own, its entries are often frozen at their expected paths in practice, which brings the effective dimension back down to n at some cost in accuracy.

What two factors buy is exactly the decorrelation the one-factor model cannot express, and with it the ability to price curve-shape products and to get a Bermudan's exercise boundary right — since the decision to exercise depends on the shape of the curve at that moment and not only on its level.

12.6 Calibration, and the Parameter Nobody Can See

Finally, how such a model is fitted — because the pattern of chapters 10 and 11 appears here a third time, and by now it should be recognisable.

- **The initial curve** is matched exactly and for free, by (12.14).

- $\sigma(t)$, the level, is bootstrapped to a strip of at-the-money European swaptions chosen to match the exotic's own exercise schedule.

- b **and** c, the skew and smile parameters, are fitted to the smile around those same swaptions.

- **The vol-of-vol** is fitted to how the smile's curvature varies with expiry, and is usually remarked occasionally rather than daily.

- κ, the mean reversion, is not fitted to European swaptions at all.

That last point deserves a closer look. European swaptions are very nearly insensitive to κ: it can be moved a long way and the fit compensated by adjusting $\sigma(t)$, leaving vanilla prices essentially unchanged. But κ controls how strongly rates of different maturities move together, and therefore how much a Bermudan is worth over the European swaptions it could become.

So κ is the single largest model risk in a callable book, and it is invisible in the instruments used to calibrate. It is set instead from the term structure of rate correlations, historically or from a principal component analysis of the curve, or from constant-maturity-swap spread option prices, or by matching a Bermudan-to-European ratio the desk believes in.

If that sounds familiar, it should. It is β from chapter 10 and the mixing weight from chapter 11, in different clothing: a parameter that the calibration set cannot see and the product depends on. The recurring lesson of these last four chapters is that a model is not determined by fitting it, and the part that fitting leaves undetermined is usually the part that decides the answer.

Remark (The alternative is to stop insisting on a state)**.** Everything in this chapter was bought with one decision: keep the Markov state small enough to put on a grid. That decision is what constrains the volatility structure, what makes the vanilla fit approximate, and what leaves κ carrying the correlation that the calibration cannot see.

Chapter 13 takes the opposite decision. It models the quoted rates directly, so the calibration instruments price exactly and the correlation between rates is specified rather than inferred from a mean reversion — and it gives up the state entirely, so the exercise decision this chapter handles by backward induction becomes a regression on the simulated paths. The two chapters are the same trade seen from either end, and neither dominates: this one prices the exercise exactly and the vanillas approximately, that one the reverse.

References

- Bjork, T., & Svensson, L. (2001). On the existence of finite-dimensional realizations for nonlinear forward rate models. *Mathematical Finance*, 11(2), 205–243.

- Bjork, T., & Christensen, B. J. (1999). Interest rate dynamics and consistent forward rate curves. *Mathematical Finance*, 9(4), 323–348.

- Filipovic, D., & Teichmann, J. (2004). On the geometry of the term structure of interest rates. *Proceedings of the Royal Society A*, 460(2041), 129–167.

- Cheyette, O. (2001). Markov representation of the Heath-Jarrow-Morton model. Available at SSRN 6073.

- Ritchken, P., & Sankarasubramanian, L. (1995). Volatility structures of forward rates and the dynamics of the term structure. *Mathematical Finance*, 5(1), 55–72.

- Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6(4), 379–406.

- Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6), 1343–1376.

- Andersen, L. B. G., & Piterbarg, V. V. (2010). *Interest Rate Modeling*, Volumes 1–3. Atlantic Financial Press.