10.1098/rspa.2000.0722

Interest rates and information geometry

By Dorje C. Brody[1] and Lane P. Hughston[2]

[1] *Blackett Laboratory, Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK and Centre for Mathematical Science, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK* (dorje@ic.ac.uk)

[2] *Department of Mathematics, King's College London, Strand, London WC2R 2LS, UK* (lane.hughston@kcl.ac.uk)

Received 5 July 2000; accepted 21 November 2000

The space of probability distributions on a given sample space possesses natural geometric properties. For example, in the case of a smooth parametric family of probability distributions on the real line, the parameter space has a Riemannian structure induced by the embedding of the family into the Hilbert space of square-integrable functions, and is characterized by the Fisher–Rao metric. In the non-parametric case the relevant geometry is determined by the spherical distance function of Bhattacharyya. In the context of term-structure modelling, we show that the derivative of the discount function with respect to the time left until maturity gives rise to a probability density. This follows as a consequence of the positivity of interest rates. Therefore, by mapping the density functions associated with a given family of term structures to Hilbert space, the resulting metrical geometry can be used to analyse the relationship of yield curves to one another. We show that the general arbitrage-free yield-curve dynamics can be represented as a process taking values in the convex space of smooth density functions on the positive real line. It follows that the theory of interest rate dynamics can be represented by a class of processes in Hilbert space. We also derive the dynamics for the central moments associated with the distribution determined by the yield curve.

Keywords: interest rate models; Heath–Jarrow–Morton framework; arbitrage-free term-structure movements; differential geometry and statistics

1. Introduction

The theory of interest rates has gone through two major developments in recent decades. Following initial investigations by Merton (1973) and others, the first decisive advance culminated in the work of Vasicek (1977), who was able to give a fairly general characterization of the arbitrage-free dynamics of a family of discount bonds, indexed by their maturity. The well-known model that bears his name appears as an exact solution obtained with specializing assumptions. In the wake of Vasicek's work were a number of other specific interest rate models, of varying degrees of usefulness and tractability, including, for example, the CIR model (Cox *et al*. 1985) and its generalizations. The next significant line of development, following the general martingale characterization of arbitrage-free asset pricing by Harrison & Kreps

(1979) and Harrison & Pliska (1981), was instigated with the recognition by Ho & Lee (1986) that the initial term structure might be specified essentially arbitrarily, a feature that has important practical implications. This insight was incorporated into the HJM framework (Heath *et al*. 1992), which constituted a major advance in the subject, providing a general model-independent basis for the analysis of interest rate dynamics and the pricing of interest rate derivatives.

Since then there have been numerous further developments. These include, for example, the infinite-dimensional or 'string-type' models of Kennedy (1994), Santa-Clara & Sornette (1997), Filipović (2000) and others; the positive interest rate models of Flesaker & Hughston (1996); the potential approach of Rogers (1997); the so-called market models (Sandmann & Sondermann 1996; Brace *et al*. 1996, 1997; Jamshidian 1997); and the geometric analysis of the space of yield curves undertaken by Björk & Svensson (2001).

Nevertheless, no criterion has emerged, based on the extensive econometric evidence available, that allows for the identification of a clearly preferred class of models in a rational way. On these grounds it makes sense to try to cast the general interest rate framework into a new form, with the idea that certain models might thus become recognizable as more natural on mathematical and economic grounds.

With this end in mind, the purpose of the present article is to propose a novel application of information geometry to interest rate theory. The main results are:

(i) the construction of a geometric measure for how 'different' two term structures are from one another;

(ii) a characterization of the evolutionary trajectory of the term structure as a measure-valued process;

(iii) the derivation of dynamics for the moments of the term structure; and

(iv) a reformulation of arbitrage-free interest rate dynamics in terms of a class of processes on Hilbert space.

The paper is organized as follows. In § 2 we review the basic idea of information geometry and its role in estimation theory. The geometry of the normal distribution is considered in detail as an illustration. In § 3 a remarkable characterization of the discount function in terms of an abstract probability density function is introduced in proposition 3.1. This allows us to apply information geometric techniques to determine the deviation between different term structures within a given model. In this connection, in § 4 we consider a class of flat rate models as examples.

The material of the first four sections of the paper is essentially static, i.e. set in the present, whereas in § 5 we investigate the dynamics of the density function that generates the term structure. This is carried out in such a way that the resulting dynamics is manifestly arbitrage-free. Our key result here is formula (5.15), in which we establish that the dynamics of the term structure can be characterized as a measure-valued process. This idea is developed further in proposition 5.1.

In § 6 we introduce an analogue of the classical principal components analysis for yield curves, and in propositions 6.1 and 6.2 we derive formulae for the evolution of the first two moments of the term-structure density process. Then, making use of the information geometry introduced earlier, in § 7 we map the dynamics developed in § 5 to Hilbert space. Our main result here is proposition 7.1, which shows how this can be achieved.

2. Information geometry

Because some of the mathematical techniques we employ here may not be familiar to those working in finance, it will be appropriate to begin with a few background remarks. It has long been known (see, for example, Amari 1985; Kass 1989; Murray & Rice 1993) that a useful approach to statistical inference is to regard a parametric model as a differentiable manifold equipped with a metric. The recognition that a parametric family of probability distributions has a natural geometry associated with it arose in the work of Mahalanobis (1936), Bhattacharyya (1943) and Rao (1945).

Suppose that X is a continuous random variable taking values on the real line $\mathbb{R}^1$, and that $\rho(x)$ is a density function for X. Because $\rho(x)$ is non-negative and has integral unity, it follows that the square-root likelihood function

$$\xi(x) = \sqrt{\rho(x)} \tag{2.1}$$

exists for all x, and satisfies the normalization condition

$$\int_{-\infty}^{\infty} (\xi(x))^2 \, \mathrm{d}x = 1. \tag{2.2}$$

We see that $\xi(x)$ can be regarded as a unit vector in the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^1)$. Now let $\rho_1(x)$, $\rho_2(x)$ denote a pair of density functions on $\mathbb{R}^1$, and $\xi_1(x)$, $\xi_2(x)$ the corresponding Hilbert space elements. Then the inner product

$$\cos\phi = \int_{-\infty}^{\infty} \xi_1(x)\xi_2(x) \, \mathrm{d}x \tag{2.3}$$

defines an angle ϕ that can be interpreted as the *distance* between the two probability distributions. More precisely, if we write $\mathcal{S}$ for the unit sphere in $\mathcal{H}$, then ϕ is the spherical distance between the points on $\mathcal{S}$ determined by the vectors $\xi_1(x)$ and $\xi_2(x)$. The maximum possible distance, corresponding to non-overlapping densities, is given by $\phi = \pi/2$. This follows from the fact that $\xi_1(x)$ and $\xi_2(x)$ are non-negative functions, and thus define points on the positive orthant of $\mathcal{S}$. We remark that an alternative way of expressing (2.3) is

$$\cos\phi = 1 - \frac{1}{2}\int_{-\infty}^{\infty} (\xi_1(x) - \xi_2(x))^2 \, \mathrm{d}x, \tag{2.4}$$

which makes it apparent that the angle ϕ measures the extent to which the two distributions are distinct.

The spherical distance in the work of Bhattacharyya introduced above is applicable in a non-parametric context. In the case of a parametric family of distributions we can develop matters further. Let us write $\rho(x, \theta)$ for the parametrized density function. Here θ stands for a set of parameters θ^i $(i = 1, \dots, r)$. By varying θ we obtain an r-dimensional submanifold $\mathcal{M}$ in $\mathcal{S}$ determined by the unit vectors $\xi(x, \theta) \in \mathcal{H}$. The parameters θ^i are local coordinates for $\mathcal{M}$.

The key point that we require in the following (cf. Dawid 1977) is that the spherical geometry of $\mathcal{S}$ induces a Riemannian geometry on $\mathcal{M}$, for which the metric tensor $g_{ij}(\theta)$ is given, in local coordinates, by

$$g_{ij}(\theta) = \int_{-\infty}^{\infty} \frac{\partial\xi(x,\theta)}{\partial\theta^i}\frac{\partial\xi(x,\theta)}{\partial\theta^j} \, \mathrm{d}x. \tag{2.5}$$

By use of definition (2.1), we see that an alternative expression for $g_{ij}(\theta)$ is

$$g_{ij}(\theta) = \frac{1}{4}\int_{-\infty}^{\infty} \rho(x,\theta)\frac{\partial \ln \rho(x,\theta)}{\partial \theta^i}\frac{\partial \ln \rho(x,\theta)}{\partial \theta^j}\,\mathrm{d}x, \tag{2.6}$$

which shows (cf. Brody & Hughston 1998) that the metric g_{ij} is, apart from the factor of $\frac{1}{4}$, the Fisher information matrix, i.e. the covariance matrix of the parametric gradient of the log-likelihood function (Fisher 1921). We refer to $g_{ij}(\theta)$ as the Fisher–Rao metric on the statistical model $\mathcal{M}$.

The significance of the Fisher–Rao metric in estimation theory is well known. Suppose that $\tau(\theta)$ is some given function of the parameters, and that the random variable T represented by the function $T(x)$ on $\mathbb{R}^1$ is an unbiased estimator for $\tau(\theta)$ in the sense that

$$\int_{-\infty}^{\infty} \rho(x,\theta)T(x)\,\mathrm{d}x = \tau(\theta). \tag{2.7}$$

The variance of the estimator T is defined, as usual, by

$$\mathrm{var}[T] = \int_{-\infty}^{\infty} \rho(x,\theta)(T(x)-\tau(\theta))^2\,\mathrm{d}x. \tag{2.8}$$

Then a set of fundamental bounds on $\mathrm{var}[T]$, independent of the choice of the estimator $T(x)$, can be obtained by applying the operator $\sum_i \alpha^i\partial_i$ to (2.7), letting α^i be arbitrary. By use of (2.1) and the Schwartz inequality for $L^2(\mathbb{R}^1)$, we obtain

$$g_{ij}\,\mathrm{var}[T] \geqslant \frac{1}{4}\frac{\partial \tau}{\partial \theta^i}\frac{\partial \tau}{\partial \theta^j}. \tag{2.9}$$

This matrix inequality is interpreted as saying that if we subtract the right-hand side from the left, the result is non-negative definite. It follows that if the random variables Θ^i $(i = 1,\dots,r)$ are unbiased estimators for the parameters θ^i, satisfying

$$\int_{-\infty}^{\infty} \rho(x,\theta)\Theta^i(x)\,\mathrm{d}x = \theta^i, \tag{2.10}$$

then the covariance matrix of the estimators is bounded by the inverse Fisher information matrix:

$$\mathrm{cov}[\Theta^i,\Theta^j] \geqslant \tfrac{1}{4}g^{ij}. \tag{2.11}$$

The Riemannian metric (2.5) introduced above can be used to define a distance measure between two distributions belonging to a given parametric family. This measure is invariant in the sense that it is unaffected by a reparametrization of the distributions. The distance is calculated by integrating the infinitesimal line element $\mathrm{d}s$ along the geodesic connecting the two points in the statistical manifold $\mathcal{M}$, where

$$\mathrm{d}s^2 = \sum_{i,j} g_{ij}\,\mathrm{d}\theta^i\,\mathrm{d}\theta^j. \tag{2.12}$$

The geodesics with respect to a given metric g_{ij} are the solutions of the differential equation

$$\frac{\mathrm{d}^2\theta^i}{\mathrm{d}u^2} + \Gamma^i_{jk}\frac{\mathrm{d}\theta^j}{\mathrm{d}u}\frac{\mathrm{d}\theta^k}{\mathrm{d}u} = 0 \tag{2.13}$$

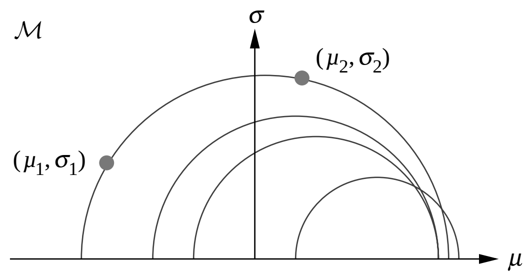

Figure 1. *Geodesic curves for normal distributions.* The statistical manifold $\mathcal{M}$ in this case is the upper half-plane parametrized by μ and σ. We have $-\infty < \mu < \infty$ and $0 < \sigma < \infty$. The shortest path joining the two normal distributions $\mathcal{N}(\mu_1, \sigma_1)$ and $\mathcal{N}(\mu_2, \sigma_2)$ is given by the unique semicircular arc through the given two points and centred on the boundary line $\sigma = 0$.

for the curve $\theta^i(u)$ in $\mathcal{M}$, subject to the given boundary conditions at the two end points. Here, we have written

$$\Gamma^i_{jk} = \tfrac{1}{2} g^{il}(\partial_j g_{kl} + \partial_k g_{jl} - \partial_l g_{jk}), \tag{2.14}$$

where $\partial_i = \partial/\partial\theta^i$, and the inverse metric g^{ij}, also appearing in (2.11), satisfies $g^{ij} g_{jk} = \delta^i_k$, where δ^i_k is the Kronecker delta. Note that in equations (2.13) and (2.14) above, and henceforth in this article, we employ the Einstein summation convention on repeated indices.

Let us consider, as an explicit example, the manifold $\mathcal{M}$ corresponding to the normal distributions $\mathcal{N}(\mu, \sigma)$ on $\mathbb{R}^1$, with mean μ and standard deviation σ. For the parametrized density function we have

$$\rho(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{2.15}$$

A straightforward computation, making use of (2.6), gives

$$\mathrm{d}s^2 = \frac{1}{\sigma^2}(\mathrm{d}\mu^2 + 2\,\mathrm{d}\sigma^2) \tag{2.16}$$

for the line element, which is defined on the upper half-plane $-\infty < \mu < \infty$, $0 < \sigma < \infty$. The resulting Riemannian geometry is that of hyperbolic space, which is a homogeneous manifold with constant negative curvature. The geometry of this space has been studied extensively, and has many intriguing properties. For the distance function in the case of a pair of normal distributions $\mathcal{N}(\mu_1, \sigma_1)$, $\mathcal{N}(\mu_2, \sigma_2)$ we obtain

$$D(\rho_1, \rho_2) = \frac{1}{\sqrt{2}} \log \frac{1 + \delta_{1,2}}{1 - \delta_{1,2}}, \tag{2.17}$$

where the function $\delta_{1,2}$, defined by

$$\delta_{1,2} = \sqrt{\frac{(\mu_2 - \mu_1)^2 + 2(\sigma_2 - \sigma_1)^2}{(\mu_2 - \mu_1)^2 + 2(\sigma_2 + \sigma_1)^2}}, \tag{2.18}$$

lies between 0 and 1. The geodesics, in particular, are given in general by semicircular arcs centred on the boundary line $\sigma = 0$ (this line itself is not part of the manifold $\mathcal{M}$), as illustrated in figure 1. An exceptional situation arises when $\mu_1 = \mu_2$, for which the geodesic is a straight line given by constant μ, and we have

$$D(\rho_1, \rho_2) = \frac{1}{\sqrt{2}}\left|\log\frac{\sigma_1}{\sigma_2}\right|. \tag{2.19}$$

We refer the reader to Burbea (1986), where metric and distance computations have been carried out explicitly for other families of distributions.

3. Discount bond densities

Our goal now is to make use of the analysis presented in the previous section to construct a natural metric on the space of yield curves. In doing so we shall take advantage of a remarkable 'probabilistic' characterization of discount bonds, which we proceed to describe here.

Let $t = 0$ denote the present, and P_{0T} a smooth family of discount bonds, where T is the maturity date ($0 \leqslant T < \infty$), and $P_{00} = 1$. For positive interest we require

$$0 < P_{0T} \leqslant 1, \qquad \frac{\partial}{\partial T}P_{0T} < 0, \tag{3.1}$$

and we assume that $P_{0T} \to 0$ as T goes to infinity. A term structure that satisfies these conditions will be said to be 'admissible'. These conditions can, in fact, be relaxed slightly: P_{0T} need not be strictly smooth, nor strictly decreasing; but for most of the present discussion we shall stick with the assumptions indicated.

The interesting point that arises here, of which we shall make extensive use in the discussions that follow, is that the discount function P_{0T} can be viewed as a complementary probability distribution. In other words, we think of the maturity date as an abstract random variable X, and for its distribution we write

$$\boldsymbol{P}[X < T] = 1 - P_{0T}. \tag{3.2}$$

It should be clear that this can be done if and only if the positive interest rate conditions given in (3.1) hold. As a consequence we are able to embody the positive interest property in a fundamental way in the structure of the theory. Indeed, this basic economic property is essential if we wish to treat the yield curve consistently and naturally as a kind of mathematical object in its own right. Now let us introduce the function $\rho(T)$ defined by

$$\rho(T) = -\frac{\partial}{\partial T}P_{0T}. \tag{3.3}$$

Clearly, we have $\rho(T) > 0$ and

$$\int_0^\infty \rho(T)\,\mathrm{d}T = 1, \tag{3.4}$$

from which we infer that $\rho(T)$ can be consistently viewed as a probability density function. It follows from the defining equation (3.3) that the term-structure density $\rho(T)$ is the product of the instantaneous forward rate and the discount function itself. Now clearly if $\rho_1(T)$ and $\rho_2(T)$ are admissible term-structure densities, and if A and B are non-negative constants satisfying $A + B = 1$, then $A\rho_1(T) + B\rho_2(T)$ is also an admissible term-structure density. Putting these ingredients together, we see that the term structure of interest rates can be given the following general characterization.

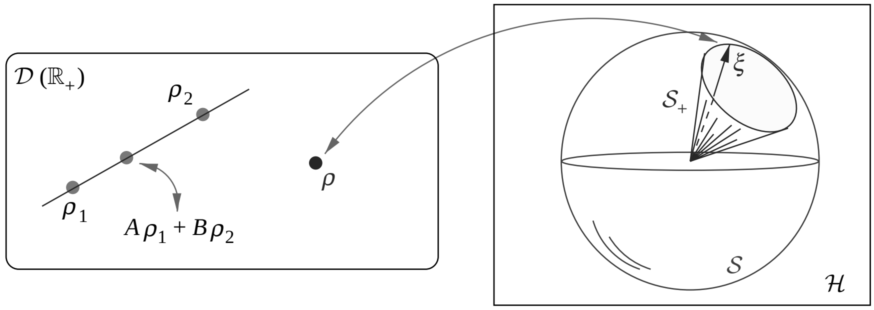

Figure 2. *The system of admissible term structures.* A smooth positive interest term structure can be regarded as a point in $\mathcal{D}(\mathbb{R}^1_+)$, the convex space consisting of smooth density functions on $\mathbb{R}^1_+$. Associated with each point $\rho \in \mathcal{D}(\mathbb{R}^1_+)$ there is a ray ξ in the positive orthant $\mathcal{S}_+$ of the unit sphere $\mathcal{S}$ in the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^1_+)$. A dynamical trajectory on $\mathcal{D}(\mathbb{R}^1_+)$ can then be mapped to a corresponding trajectory in $\mathcal{S}_+$.

Proposition 3.1. *The system of admissible term structures is isomorphic to the convex space $\mathcal{D}(\mathbb{R}^1_+)$ of smooth density functions on the positive real line.*

At first glance it may seem odd to think of the discount function in this manner. However, it gives us the advantage of being able to apply the tools of information geometry in an unexpected way, as we indicate in what follows.

In particular, there is a map from the space $\mathcal{D}(\mathbb{R}^1_+)$ of such term-structure densities to the positive orthant $\mathcal{S}_+$ of the unit sphere $\mathcal{S}$ in the Hilbert space $\mathcal{H}$, as indicated in figure 2. Therefore, given two yield curves we can calculate the distance between them. This can be carried out either in a non-parametric sense, by use of the Bhattacharyya spherical distance, or in a parametric sense, by use of the Fisher–Rao distance. In the former case, first we calculate the corresponding term-structure densities $\rho_1(T)$ and $\rho_2(T)$. These are then mapped to $\mathcal{S}_+$ by taking the square roots, and their distance $\phi(\rho_1, \rho_2)$ is given by

$$\phi(\rho_1, \rho_2) = \cos^{-1} \int_0^\infty \sqrt{\rho_1(T)\rho_2(T)}\, \mathrm{d}T. \tag{3.5}$$

In the parametric case we regard the given parametric family of yield curves as defining a statistical model $\mathcal{M} \subset \mathcal{S}_+$, and the distance between the two yield curves within the given family is then defined by the Fisher–Rao metric.

4. Flat term structures

To provide some illustrations of the principles set forth in the previous section we consider here properties of yield curves for which the term structure is *flat*. Such yield curves, which are of various types, are on the whole too simple for use in practical modelling. Nevertheless, they are of interest as examples, because many of the relevant computations can be carried out explicitly.

In this connection we begin by introducing a representation of the discount function as a Laplace transform

$$P_{0T} = \int_0^\infty \mathrm{e}^{-rT} \psi(r)\, \mathrm{d}r \tag{4.1}$$

for some function $\psi(r)$. Thus we think of the discount function P_{0T} as being given by a weighted superposition of elementary discount functions, each of the form e^{-rT} for some value of r. Taking the limit $T \to 0$, we find that $\psi(r)$ must satisfy $\int_0^\infty \psi(r)\,dr = 1$. In general the inverse Laplace transform $\psi(r)$ need not be positive. However, if we restrict our consideration to non-negative functions, then $\psi(r)$ can be interpreted as a density function, and by various choices of $\psi(r)$ we are led to some interesting candidates for term structures.

First we consider the case where $\psi(r)$ is a Dirac δ-function concentrated at a point, that is, $\psi(r) = \delta(r-R)$. A direct substitution gives $P_{0T} = \exp(-RT)$, corresponding to a 'flat' term structure with a continuously compounded rate R for each value of the maturity date T. If the density function $\psi(r)$ is given by an exponential distribution $\psi(r) = \tau\exp(-\tau r)$, with parameter τ, then one sees that τ must have dimensions of time, and a short calculation gives $P_{0T} = \tau/(\tau + T)$, which also corresponds to a flat term structure, in this case with a simple percentage yield of τ^{-1} for all maturities. We see that the characteristic time-scale τ allows us to define an interest rate $R = \tau^{-1}$, which turns out to be the characteristic interest rate of the resulting structure, and we can write $P_{0T} = 1/(1 + RT)$ for the discount function.

We note that flatness is not a completely unambiguous notion, because having a uniform continuously compounded yield for all maturities is not the same thing as having a uniform simple yield for all maturities. Both define plausible, albeit distinct, systems of discount bonds. This example illustrates how, by superposing term structures of the elementary form $\exp(-RT)$ for various maturities, we can obtain other reasonable-looking and well-behaved term structures. We mention one more example, which contains the previous two examples as special cases. Consider the standard gamma distribution, with parameters κ and λ, defined for non-negative values of r by the density function

$$\psi(r) = \frac{1}{\Gamma(\kappa)}\lambda^{\kappa} r^{\kappa-1}\exp(-\lambda r). \tag{4.2}$$

In this case, we can verify that the resulting system of discount bonds is given by

$$P_{0T} = \left(\frac{\lambda}{\lambda + T}\right)^{\kappa}, \tag{4.3}$$

which assumes a more recognizable form if we set $\lambda = \kappa\tau$, where τ again defines a characteristic time-scale, and κ is a dimensionless number. Then we have

$$P_{0T} = \left(1 + \frac{RT}{\kappa}\right)^{-\kappa}, \tag{4.4}$$

where $R = \tau^{-1}$. The system of discount bonds arising here can also be interpreted as a flat term structure, in this case with a constant annualized rate of interest R assuming compounding at the frequency κ over the life of each bond (κ need not be an integer). It is not difficult to check that for $\kappa = 1$ this reduces to the case of a flat rate on the basis of a simple yield, whereas in the limit $\kappa \to \infty$ we recover the case of a flat rate on the basis of continuous compounding.

Now we shall apply the ideas of statistical geometry to make comparisons between various term structures of the form (4.4). For density function $\rho(T) = -\partial_T P_{0T}$ we obtain

$$\rho(T,R) = R\left(1+\frac{RT}{\kappa}\right)^{-(\kappa+1)}. \tag{4.5}$$

Here we find it convenient to label the density function by the flat rate R. Note that in the limit $\kappa \to \infty$ we have $\rho(T,R) \to R\mathrm{e}^{-RT}$. First consider the non-parametric separation between different term structures in this model as determined by the spherical distance of Bhattacharyya given in formula (3.5). Let us write $\rho_i(T) = \rho(T, R_i)$ for $i = 1, 2$. A direct integration leads to the expression

$$\phi(\rho_1,\rho_2) = \cos^{-1}\left(\frac{\sqrt{R_1R_2}}{R_1-R_2}\log\frac{R_1}{R_2}\right) \tag{4.6}$$

for the distance when $\kappa = 1$, whereas in the limit $\kappa \to \infty$ (continuous compounding) we have

$$\phi(\rho_1,\rho_2) = \cos^{-1}\left(\frac{2\sqrt{R_1R_2}}{R_1+R_2}\right). \tag{4.7}$$

It is interesting to observe that the bracketed term in (4.7) is given by the ratio of the geometric and arithmetic means of the two rates.

Alternatively, we can view (4.5) as a family of distributions parametrized by the flat rate R. Then it is natural to consider the Fisher–Rao distance between the two term structures characterized by R_1 and R_2. A straightforward calculation leads to the simple distance formula

$$D(R_1,R_2) = \sqrt{\frac{\kappa}{\kappa+2}}\log\frac{R_2}{R_1}, \tag{4.8}$$

where we have assumed $R_2 \geqslant R_1$.

5. Interest rate dynamics

The formalism we have developed so far is essentially a static one, set in the present. Now we turn to the problem of developing a dynamical theory of interest rates. The idea is that, at each instant of time, the yield curve is characterized by a term-structure density according to the scheme described in the previous sections. Then, as time passes, the density function evolves randomly. As a consequence we obtain a measure-valued process. In particular, we obtain a process on $\mathcal{D}(\mathbb{R}^1_+)$. Our goal in this section is to determine a set of conditions on this process necessary and sufficient to ensure that the resulting interest rate dynamics will be arbitrage-free.

We shall assume that the reader is familiar with the general theory of interest rate dynamics as laid out, for example, in Carverhill (1994), Rogers (1995), Hughston (1996*a*), Baxter (1997), Musiela & Rutkowski (1997), Rebonato (1998), Nielsen (1999), Brody (2000), James & Webber (2000) or Hunt & Kennedy (2000). For the general discount bond dynamics, let us write

$$\mathrm{d}P_{tT} = \mu_{tT}\,\mathrm{d}t + \Sigma_{tT}\,\mathrm{d}W_t, \tag{5.1}$$

where μ_{tT} and Σ_{tT} are the *absolute drift* and *absolute volatility processes*, respectively, for a bond with maturity T. Here, W_t is a vector Brownian motion, and Σ_{tT}

is a vector process, and there is an inner product implied between Σ_{tT} and $\mathrm{d}W_t$. We need not specify the dimensionality of the Brownian motion, which might be infinite, and indeed in some respects the infinite-dimensional setting is the most natural one. In fact, it suffices for our purposes merely to assume that P_{tT} is a one-parameter family of continuous semimartingales on the given probability space, with respect to the given filtration. However, for simplicity of exposition we shall have in mind the case where the relevant stochastic basis is generated by a multidimensional Brownian motion. Here, as in Flesaker & Hughston (1997*a*, *b*), we regard the discount bond dynamics as the natural starting position, rather than, say, the instantaneous forward rate dynamics (Heath *et al*. 1992), which we need not consider here directly. We shall assume nevertheless that the processes μ_{tT} and Σ_{tT} are both smooth in the variable T, and that sufficiently strong technical conditions are in place to ensure that the instantaneous forward rate processes are semimartingales, as in the HJM framework.

In order to extend the analysis of the previous section it is convenient to introduce what is sometimes referred to as the 'Musiela parametrization' (Musiela 1993), given by

$$B_{tx} = P_{t,t+x}, \tag{5.2}$$

where $T = t + x$ represents the maturity date of the bond, and hence x is the time left until maturity. Thus B_{tx} is the value at time t of a discount bond that has x years left to mature. This choice of parametrization has already been shown to be useful in the geometric analysis of interest rates (Björk & Christensen 1999; Björk & Gombani 1999; Björk & Svensson 2001; Björk 2001) and is also a key idea in the treatment of infinite-dimensional models (Filipović 2000). We note that $B_{t0} = 1$ for all t, and that $B_{tx} \to 0$ as $x \to \infty$. It follows that

$$\rho_t(x) = -\frac{\partial}{\partial x} B_{tx} \tag{5.3}$$

is a measure-valued process in the sense that, for each value of t, the random function $\rho_t(x)$ satisfies $\rho_t(x) > 0$ and the normalization condition

$$\int_0^\infty \rho_t(x)\,\mathrm{d}x = 1. \tag{5.4}$$

Here we have chosen the notation $\rho_t(x)$, which makes the x dependence more prominent, to emphasize the fact that, for each value of t, and conditional on information given up to time t, $\rho_t(x)$ is a density function, though we might have written ρ_{tx} instead. As a consequence, $\rho_t(x)$ describes a process on $\mathcal{D}(\mathbb{R}^1_+)$. By consideration of (5.1) and (5.2) we deduce for the dynamics of B_{tx} that

$$\mathrm{d}B_{tx} = (\mathrm{d}P_{tT})|_{T=t+x} + \frac{\partial}{\partial x} B_{tx}\,\mathrm{d}t, \tag{5.5}$$

and thus, by use of (5.1), that

$$\mathrm{d}B_{tx} = (\mu_{t,t+x} + \partial_x B_{tx})\,\mathrm{d}t + \Sigma_{t,t+x}\,\mathrm{d}W_t, \tag{5.6}$$

where $\partial_x = \partial/\partial x$. Differentiating this expression with respect to x and introducing the measure-valued process $\rho_t(x)$ according to formula (5.3) we therefore obtain

$$\mathrm{d}\rho_t(x) = (-\partial_x \mu_{t,t+x} + \partial_x \rho_t(x))\,\mathrm{d}t - \partial_x \Sigma_{t,t+x}\,\mathrm{d}W_t. \tag{5.7}$$

A further simplification is then achieved by introducing the notation

$$\beta_{tx} = -\partial_x \mu_{t,t+x} \tag{5.8}$$

and

$$\omega_{tx} = -\partial_x \Sigma_{t,t+x}, \tag{5.9}$$

which gives us

$$\mathrm{d}\rho_t(x) = (\beta_{tx} + \partial_x \rho_t(x))\,\mathrm{d}t + \omega_{tx}\,\mathrm{d}W_t. \tag{5.10}$$

In the foregoing discussion we have not yet imposed the arbitrage-free condition. This is given by the drift constraint

$$\mu_{tT} = r_t P_{tT} + \Sigma_{tT}\lambda_t, \tag{5.11}$$

where λ_t is the process for the market price of risk. We note that λ_t, like Σ_{tT}, is a vector process. However, λ_t does not depend on the maturity T. The absence of arbitrage ensures the existence of λ_t. For our purposes we do not need to insist that the bond market is complete: all we require is the existence of a pricing kernel, or equivalently the existence of a self-financing 'natural numeraire' portfolio with value process N_t, such that P_{tT}/N_t is a martingale for each value of T (cf. Long 1990; Constantinides 1992; Flesaker & Hughston 1997*c*). The numeraire process satisfies

$$\frac{\mathrm{d}N_t}{N_t} = (r_t + \lambda_t^2)\,\mathrm{d}t + \lambda_t\,\mathrm{d}W_t, \tag{5.12}$$

and the corresponding pricing kernel is given by $1/N_t$. As a consequence of the constraint (5.11) we then have

$$\mu_{t,t+x} = r_t B_{tx} + \Sigma_{t,t+x}\lambda_t, \tag{5.13}$$

and therefore, by differentiation of this expression with respect to x, we obtain

$$\beta_{tx} = r_t \rho_t(x) + \omega_{tx}\lambda_t. \tag{5.14}$$

Inserting (5.14) in (5.10) we are thus able to express the dynamics of the density function $\rho_t(x)$ in the form

$$\mathrm{d}\rho_t(x) = (r_t \rho_t(x) + \partial_x \rho_t(x))\,\mathrm{d}t + \omega_{tx}(\mathrm{d}W_t + \lambda_t\,\mathrm{d}t). \tag{5.15}$$

Before proceeding further, let us verify, as a consistency check, that the dynamics given by (5.15) preserves the normalization condition on $\rho_t(x)$, given by (5.4). Integrating the right-hand side of (5.15) with respect to x and equating the drift and volatility terms separately to zero leads to the relations

$$r_t + \int_0^\infty \partial_x \rho_t(x)\,\mathrm{d}x = 0 \tag{5.16}$$

and

$$\int_0^\infty \omega_{tx}\,\mathrm{d}x = 0, \tag{5.17}$$

which must hold for all t. Condition (5.16) is satisfied because $\rho_t(x) \to 0$ as $x \to \infty$ and

$$\rho_t(0) = r_t. \tag{5.18}$$

Condition (5.17) is satisfied because, by definition, we have $\omega_{tx} = -\partial_x \Sigma_{t,t+x}$, and the absolute volatility $\Sigma_{t,t+x}$ vanishes both as $x \to 0$ (a maturing bond has a definite value and thus has no absolute volatility), and as $x \to \infty$ (a bond with infinite maturity has no value, and hence no absolute volatility).

Summing up matters so far, we see that in (5.15) we are able to cut the standard HJM arbitrage-free interest rate dynamics in the form of a measure-valued process $\rho_t(x)$ subject to the constraints (5.16) and (5.17). At first glance the role of the short rate r_t in (5.15) seems anomalous, because it might appear that it has to be specified separately. However, by virtue of (5.18) we can incorporate r_t directly into the dynamics of $\rho_t(x)$.

In fact, there is another way of expressing (5.15) which is very suggestive, and ties in naturally with the Hilbert space approach to dynamics introduced in § 7. First we note that (5.16) can be rewritten in the form

$$r_t = -\int_0^\infty \rho_t(x)\partial_x \ln \rho_t(x)\,\mathrm{d}x. \tag{5.19}$$

In other words, r_t is minus the expectation of the gradient of the log-likelihood function. Here the expectation is taken with respect to $\rho_t(x)$ itself. Writing E_ρ for this abstract expectation, we have

$$\mathrm{d}\rho_t(x) = \rho_t(x)(\partial_x \ln \rho_t(x) - E_\rho[\partial_x \ln \rho_t(x)])\,\mathrm{d}t + \omega_{tx}\,\mathrm{d}W_t^*, \tag{5.20}$$

where $\mathrm{d}W_t^* = \mathrm{d}W_t + \lambda_t\,\mathrm{d}t$. We note that W_t^* has the interpretation of being a Brownian motion with respect to the risk-neutral measure associated with the given pricing kernel. In the risk-neutral measure, for which the term involving λ_t effectively disappears, the remaining drift for $\rho_t(x)$ is determined by the deviation of $\partial_x \ln \rho_t(x)$ from its abstract mean.

Let us now examine the volatility term ω_{tx}, appearing in (5.20), more closely, with a view to gaining a better understanding of the significance of the volatility constraint (5.17). Because $\rho_t(x)$ must remain positive for all values of x, the coefficient of $\mathrm{d}W_t^*$ in (5.20) must be of the form

$$\omega_{tx} = \rho_t(x)\sigma_{tx} \tag{5.21}$$

for some bounded process σ_{tx}, to ensure that ω_{tx} dies off appropriately for values of x such that $\rho_t(x)$ approaches zero. As a consequence, we can write (5.15) in the quasi-lognormal form

$$\frac{\mathrm{d}\rho_t(x)}{\rho_t(x)} = (r_t + \partial_x \ln \rho_t(x))\,\mathrm{d}t + \sigma_{tx}\,\mathrm{d}W_t^*, \tag{5.22}$$

and for the constraint (5.17) we have

$$E_\rho[\sigma_{tx}] = 0, \tag{5.23}$$

which can be satisfied by writing

$$\sigma_{tx} = \nu_{tx} - E_\rho[\nu_{tx}], \tag{5.24}$$

where ν_{tx} is an exogenously specifiable unconstrained process. Here, for any process A_{tx} we define

$$E_\rho[A_{tx}] = \int_0^\infty \rho_t(x) A_{tx}\,\mathrm{d}x. \tag{5.25}$$

The results established above can then be summarized as follows.

Proposition 5.1. *The general admissible term-structure evolution based on the information set generated by a multidimensional Brownian motion W_t is given by a measure-valued process $\rho_t(x)$ in $\mathcal{D}(\mathbb{R}^1_+)$ satisfying*

$$\frac{\mathrm{d}\rho_t(x)}{\rho_t(x)} = (\partial_x \ln \rho_t(x) - E_\rho[\partial_x \ln \rho_t(x)])\,\mathrm{d}t + (\nu_{tx} - E_\rho[\nu_{tx}])(\mathrm{d}W_t + \lambda_t\,\mathrm{d}t), \tag{5.26}$$

where the processes λ_t and ν_{tx} are specified exogenously, along with the initial term-structure density $\rho_0(x)$.

An advantage of the particular expression (5.26) given for the dynamics above is that the preservation of the normalization condition on $\rho_t(x)$ is evident by inspection, because this is equivalent to the relation

$$E_\rho\left[\frac{\mathrm{d}\rho_t(x)}{\rho_t(x)}\right] = 0. \tag{5.27}$$

An alternative expression for (5.26), which brings out more explicitly the nonlinearities in the dynamics, is given by

$$\mathrm{d}\rho_t(x) = (\partial_x\rho_t(x) + \rho_t(0)\rho_t(x))\,\mathrm{d}t + \rho_t(x)\left(\nu_{tx} - \int_0^\infty \rho_t(y)\nu_{ty}\,\mathrm{d}y\right)\mathrm{d}W_t^*, \tag{5.28}$$

where $\mathrm{d}W_t^* = \mathrm{d}W_t + \lambda_t\,\mathrm{d}t$, as defined earlier.

6. Moment analysis

The characterization of the yield curve as an abstract probability density enables us to develop an analogue of the classical 'principal component' analysis often used in the study of yield-curve dynamics. To this end we let $\rho_t(x) = -\partial_x P_{t,t+x}$ be the density process associated with an admissible family of discount bond prices, and define the moment processes

$$\bar{x}_t = \int_0^\infty x\rho_t(x)\,\mathrm{d}x \tag{6.1}$$

and

$$\bar{x}_t^{\prime(n)} = \int_0^\infty x^n \rho_t(x)\,\mathrm{d}x \tag{6.2}$$

for $n \geqslant 2$, along with the central moment processes

$$\bar{x}_t^{(n)} = \int_0^\infty (x - \bar{x}_t)^n \rho_t(x)\,\mathrm{d}x. \tag{6.3}$$

It is important to note that in some cases the relevant moments may not exist. For example, in the case of a continuously compounded flat yield curve given at $t = 0$ by the density function $\rho_0(x) = R\mathrm{e}^{-Rx}$, we have

$$\bar{x}_0 = R^{-1}, \qquad \bar{x}_0^{(2)} = R^{-2}, \qquad \bar{x}_0^{(3)} = 3R^{-3}, \qquad \bar{x}_0^{(4)} = 9R^{-4}$$

for the first four central moments. On the other hand, in the example of the simple flat term structure for which $\rho_0(x) = R/(1+Rx)^2$ we find that none of the moments exist, on account of the fatness of the tail of the distribution. In fact, for the flat rate term structures with compounding frequency κ the moments exist only up to order $\kappa - 1$.

The first four moments, if they exist, are the mean, variance, skewness and kurtosis of the distribution of the abstract random variable X characterizing the yield curve. At $t = 0$ the mean $\bar{x}_0$ determines a characteristic time-scale associated with the given term structure, and its inverse $1/\bar{x}_0$ can be thought of as an associated characteristic yield. The difference $\bar{x}_0^{(2)} - (\bar{x}_0)^2$ then measures the departure of the given term structure from flatness on a continuously compounded basis. This is on account of the fact that in the case of an exponential distribution the variance is given by the square of the mean.

It is legitimate to conjecture that for some purposes the specification of, for example, the first three or four moments will be sufficient to provide a satisfactory representation of the term structure. One way of implementing this idea is to introduce the entropy S_ρ of the given distribution, defined by

$$S_\rho = -\int_0^\infty \rho(x) \ln \rho(x) \, \mathrm{d}x. \tag{6.4}$$

Because $\rho(x)$ has dimensions of inverse time, S_ρ is defined only up to an overall additive constant. Therefore, the difference of the entropies associated with two yield curves has an invariant significance.

For yield-curve calibration we propose that $\rho(x)$ should be chosen such that S_ρ is maximized subject to the constraints of the data available, along with any other appropriate priors we wish to impose. For example, if we are given as data only the mean $\bar{x}_0$, then the maximum entropy term structure is $\rho_0(x) = R\mathrm{e}^{-Rx}$, where $R = 1/\bar{x}_0$.

In this connection, it is also of interest to study the dynamics of the moments in the case of a general admissible arbitrage-free term structure. We examine here, in particular, the mean and the variance processes. For this purpose we introduce a simplified notation $v_t = \bar{x}_t^{(2)}$ for the variance process, i.e.

$$v_t = \int_0^\infty x^2 \rho_t(x) \, \mathrm{d}x - (\bar{x}_t)^2, \tag{6.5}$$

where the mean process $\bar{x}_t$ is given as in (6.1). We assume that both $\rho_t(x)$ and the discount bond volatility $\Sigma_{t,t+x}$ fall off to zero sufficiently rapidly to ensure that

$$\lim_{x\to\infty} x^n \rho_t(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} x^n \Sigma_{t,t+x} = 0 \quad \text{for } n = 1, 2,$$

and that the integrals

$$\int_0^\infty x^n \rho_t(x) \, \mathrm{d}x \quad \text{and} \quad \int_0^\infty x^{n-1} \Sigma_{t,t+x} \, \mathrm{d}x$$

exist for $n = 1, 2$. A straightforward calculation then leads us to the following proposition.

Proposition 6.1. *The first moment $\bar{x}_t$ of an admissible, arbitrage-free term structure satisfies the dynamical law*

$$\mathrm{d}\bar{x}_t = (r_t\bar{x}_t - 1)\,\mathrm{d}t + \bar{\Sigma}_t\,\mathrm{d}W_t^*, \tag{6.6}$$

where $\bar{\Sigma}_t = \int_0^\infty \Sigma_{t,t+x}\,\mathrm{d}x$.

Proof. Starting with (5.22) and (6.1) we have

$$\begin{aligned}\mathrm{d}\bar{x}_t &= \int_0^\infty x\,\mathrm{d}\rho_t(x)\,\mathrm{d}x \\ &= \left(r_t\bar{x}_t + \int_0^\infty x\partial_x\rho_t(x)\,\mathrm{d}x\right)\mathrm{d}t - \left(\int_0^\infty x\partial_x\Sigma_{t,t+x}\,\mathrm{d}x\right)\mathrm{d}W_t^* \end{aligned} \tag{6.7}$$

by use of (5.9). Then, integrating by parts and using the assumed asymptotic behaviours for $\rho_t(x)$ and $\Sigma_{t,t+x}$, we obtain the desired result. ∎

The first moment $\bar{x}_t$ has the economic interpretation of being the value, at time t, of a perpetual annuity that pays one unit of currency per annum on a continuous basis. The assumed asymptotic conditions ensure this price is finite. We note that there is a critical level $\bar{x}_t^*$ for the first moment given by

$$\bar{x}_t^* = \frac{1}{r_t}(1 - \lambda_t\bar{\Sigma}_t). \tag{6.8}$$

When $\bar{x}_t > \bar{x}_t^*$ the drift of $\bar{x}_t$ is positive, and the drift increases further as $\bar{x}$ increases. On the other hand, when $\bar{x}_t < \bar{x}_t^*$, the drift of $\bar{x}_t$ is negative, and the drift decreases further as $\bar{x}_t$ decreases.

Proposition 6.2. *The second central moment v_t of an admissible, arbitrage-free term structure satisfies the dynamical law*

$$\mathrm{d}v_t = (r_t(v_t - \bar{x}_t^2) - \bar{\Sigma}_t^2)\,\mathrm{d}t + 2(\bar{\Sigma}_t^{(1)} - \bar{x}_t\bar{\Sigma}_t)\mathrm{d}W_t^*, \tag{6.9}$$

where

$$\bar{\Sigma}_t^{(1)} = \int_0^\infty x\Sigma_{t,t+x}\,\mathrm{d}x. \tag{6.10}$$

Proof. Starting with formula (6.5) for v_t we have

$$\mathrm{d}v_t = \int_0^\infty x^2\,\mathrm{d}\rho_t(x)\,\mathrm{d}x - \mathrm{d}(\bar{x}_t^2). \tag{6.11}$$

For the first term we obtain

$$\begin{aligned}\int_0^\infty x^2\,\mathrm{d}\rho_t(x)\,\mathrm{d}x = \left(r_t\int_0^\infty x^2\rho_t(x)\,\mathrm{d}x + \int_0^\infty x^2\partial_x\rho_t(x)\,\mathrm{d}x\right)\mathrm{d}t \\ - \left(\int_0^\infty x^2\partial_x\Sigma_{t,t+x}\right)\mathrm{d}W_t^*,\end{aligned} \tag{6.12}$$

where we have used (5.9) and (5.22). As a consequence of the assumed asymptotic behaviour of $\rho_t(x)$ and $\Sigma_{t,t+x}$, this becomes

$$\int_0^\infty x^2\, \mathrm{d}\rho_t(x)\, \mathrm{d}x = (r_t \bar{x}_t'^{(2)} - 2\bar{x}_t)\, \mathrm{d}t + 2\bar{\Sigma}_t^{(1)}\, \mathrm{d}W_t^*, \tag{6.13}$$

after an integration by parts. For the second term in (6.11) we have

$$\mathrm{d}(\bar{x}_t^2) = 2\bar{x}_t\, \mathrm{d}\bar{x}_t + (\mathrm{d}\bar{x}_t)^2 \tag{6.14}$$

by Ito's lemma, and thus

$$\mathrm{d}(\bar{x}_t^2) = (2r_t\bar{x}_t^2 - 2\bar{x}_t + \Sigma_t^2)\, \mathrm{d}t + 2\bar{x}_t\bar{\Sigma}_t\, \mathrm{d}W_t^* \tag{6.15}$$

by use of proposition 6.1. Combining (6.13) and (6.15), and using the definition (6.5), we obtain (6.9). ■

In this case we recall that the difference $v_t - \bar{x}_t^2$ acts as a simple measure of the extent to which the distribution deviates from the 'flat' term structure. As a consequence we see that the effect of the dynamics here is that the second central moment of the term structure tends to increase, i.e. has a positive drift, providing $v_t - \bar{x}_t^2$ is already above the level given by

$$v_t - \bar{x}_t^2 = \frac{1}{r_t}(\bar{\Sigma}_t^2 - 2\lambda_t(\bar{\Sigma}_t^{(1)} - \bar{x}_t\bar{\Sigma}_t)). \tag{6.16}$$

7. Hilbert space dynamics for term structures

Now that we have examined some of the advantages of expressing the arbitrage-free interest rate term-structure dynamics as a randomly evolving density function, let us consider how we transform to the Hilbert space representation for density functions considered in § 2. Denote by ξ_{tx} the process for the square-root likelihood function, defined by

$$\rho_t(x) = \xi_{tx}^2. \tag{7.1}$$

It follows by Ito's lemma that

$$\mathrm{d}\rho_t(x) = 2\xi_{tx}\, \mathrm{d}\xi_{tx} + (\mathrm{d}\xi_{tx})^2, \tag{7.2}$$

and hence $(\mathrm{d}\rho_t(x))^2 = 4\xi_{tx}^2(\mathrm{d}\xi_{tx})^2$. By rearranging (7.2) we thus obtain

$$\mathrm{d}\xi_{tx} = \frac{1}{2\xi_{tx}}\, \mathrm{d}\rho_t(x) - \frac{1}{8\xi_{tx}^3}(\mathrm{d}\rho_t(x))^2 \tag{7.3}$$

for the dynamics of the process ξ_{tx}, and hence

$$\mathrm{d}\xi_{tx} = \left(\partial_x\xi_{tx} + \tfrac{1}{2}r_t\xi_{tx} - \frac{1}{8\xi_{tx}^3}\omega_{tx}^2\right)\mathrm{d}t + \frac{1}{2\xi_{tx}}\omega_{tx}\, \mathrm{d}W_t^*, \tag{7.4}$$

where $\omega_{tx}^2 = \omega_{tx}\cdot\omega_{tx}$. Now suppose we define σ_{tx} by the quotient

$$\sigma_{tx} = \omega_{tx}/\xi_{tx}^2, \tag{7.5}$$

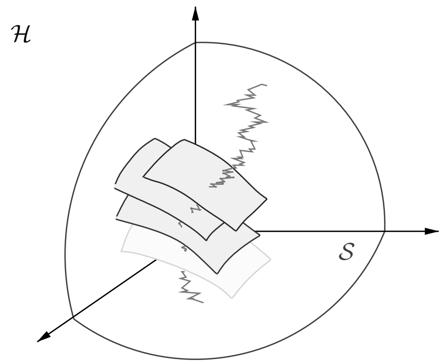

Figure 3. *Interest rate dynamics.* At each instant of time the interest rate term structure can be represented as a point on the positive orthant of the unit sphere $\mathcal{S}$ in the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^1_+)$. The associated arbitrage-free interest rate dynamics gives rise to a stochastic trajectory on this space, which is foliated by hypersurfaces corresponding to level values of the short-term interest rate.

as before, and set $\sigma^2_{tx} = \sigma_{tx} \cdot \sigma_{tx}$. Then the process for the square-root density ξ_{tx} can be written in the form

$$\mathrm{d}\xi_{tx} = (\partial_x \xi_{tx} + \tfrac{1}{2} r_t \xi_{tx} - \tfrac{1}{8} \xi_{tx} \sigma^2_{tx})\,\mathrm{d}t + \tfrac{1}{2} \xi_{tx} \sigma_{tx}\,\mathrm{d}W^*_t. \tag{7.6}$$

We recall that the volatility process σ_{tx} arising in this connection, which is given more explicitly by the ratio

$$\sigma_{tx} = \frac{\partial_x \Sigma_{t,t+x}}{\partial_x B_{tx}}, \tag{7.7}$$

can be specified exogenously, subject only to the condition that it has mean zero in the measure $\rho_t(x)$, which implies that σ_{tx} can be written in the form (5.24).

We would now like to interpret the Hilbert space dynamics in equation (7.6) more directly in a geometrical fashion. For this purpose we find it expedient to introduce an index notation, using Greek letters to signify Hilbert space operations (cf. Brody & Hughston 1998).

Thus if the function $\psi(x)$ is an element of $\mathcal{H} = L^2(\mathbb{R}^1_+)$, we denote it by ψ^α, and if $\varphi(x)$ belongs to the dual Hilbert space $\mathcal{H}^*$, we denote this by φ_α. Furthermore, their inner product is written

$$\psi^\alpha \varphi_\alpha = \int_0^\infty \psi(x)\varphi(x)\,\mathrm{d}x. \tag{7.8}$$

There is a preferred symmetric quadratic form $g_{\alpha\beta}$ on $\mathcal{H}$, given by

$$g_{\alpha\beta}\psi^\alpha\psi^\beta = \int_0^\infty (\psi(x))^2\,\mathrm{d}x, \tag{7.9}$$

which thus establishes an isomorphism between $\mathcal{H}$ and $\mathcal{H}^*$, given by $\psi^\alpha \to \psi_\alpha = g_{\alpha\beta}\psi^\beta$. Intuitively, one can think of $g_{\alpha\beta}$ as corresponding to the delta function $\delta(x, y)$, and then we have

$$g_{\alpha\beta}\psi^\alpha\varphi^\beta = \int_0^\infty \psi(x)\delta(x, y)\varphi(y)\,\mathrm{d}x\,\mathrm{d}y. \tag{7.10}$$

There are a number of Hilbert space technicalities that have to be considered for a complete exposition of the matter, but that is not our immediate concern.

If $\xi(x) > 0$ belongs to the positive orthant of $L^2(\mathbb{R}^1_+)$, then the corresponding indexed quantity ξ^α has the interpretation of a 'state vector' for the associated probability system. In that case we can think of symmetric quadratic forms as representing certain classes of random variables. More specifically, the expectation of the random variable $H_{\alpha\beta}$ in the state ξ^α is

$$E_\xi[H] = \frac{H_{\alpha\beta}\xi^\alpha\xi^\beta}{\xi_\gamma\xi^\gamma}. \tag{7.11}$$

Therefore, a state vector determines a mapping from random variables to real numbers, through (7.11). For a normalized state vector we have $\xi_\alpha\xi^\alpha = 1$, although for some purposes it is convenient to relax the normalization condition. In particular, we notice that the expectation (7.11) only depends on the direction of ξ^α.

Now suppose $\xi(x)$ is a positive function. In that case, the derivative ∂_x can be thought of as a linear operator D^α_β on $\mathcal{H}$, and we have an endomorphism given by $\xi^\alpha \to D^\alpha_\beta\xi^\beta$ provided that ξ^α lies in the domain of D^α_β. By making use of this, we can now tentatively interpret, in the language of Hilbert space geometry, the first two terms appearing in the drift in the dynamical equation (7.6). Let us begin by noting first that (5.16) can be rewritten in the form

$$\int_0^\infty \xi_{tx}\partial_x\xi_{tx}\,\mathrm{d}x = -\tfrac{1}{2}r_t. \tag{7.12}$$

This allows us to interpret the short-term interest rate process r_t in terms of the mean of the symmetric part of the operator D^α_β in the state ξ^α_t. In particular, we have

$$\frac{D_{\alpha\beta}\xi^\alpha_t\xi^\beta_t}{g_{\alpha\beta}\xi^\alpha_t\xi^\beta_t} = -\tfrac{1}{2}r_t, \tag{7.13}$$

where $D_{\alpha\beta} = g_{\alpha\gamma}D^\gamma_\beta$. Therefore, if we let $D_{(\alpha\beta)}$ denote the symmetric part of the operator D^α_β, then the abstract random variable in $\mathcal{H}$ corresponding to the short rate r_t is given by $r_{\alpha\beta} = -2D_{(\alpha\beta)}$.

Similarly, we can represent the abstract random variable x for the time left until maturity by a symmetric operator $X_{\alpha\beta}$. It is interesting to note that, with respect to the abstract probability system associated with the term-structure density, the random variables $X_{\alpha\beta}$ for the maturity date and $r_{\alpha\beta}$ for the short-term interest rate are not 'compatible'. This is an idea that has its origin in quantum theory, but can, at least in principle, arise in other contexts as well. Two random variables A and B are said to be compatible if the expression $\{\{A,C\},B\} - \{A,\{C,B\}\}$ vanishes for any random variable C, where $\{A,B\} = AB + BA$ denotes the anticommutator (Segal 1947). The lack of compatibility here indicates that the abstract probability system containing both $r_{\alpha\beta}$ and $X_{\alpha\beta}$ as random variables is not Kolmogorovian. However, the algebra of random variables generated by $X_{\alpha\beta}$ is Kolmogorovian.

Now, let $\eta(x)$ be an arbitrary element of $L^2(\mathbb{R}^1_+)$, and let η^α be the corresponding Hilbert space vector. Then evidently we have

$$\int_0^\infty \eta(x)[\partial_x\xi_{tx} + \tfrac{1}{2}r_t\xi_{tx}]\,\mathrm{d}x = \eta_\alpha\left[D^\alpha_\beta\xi^\beta_t - \left(\frac{D_{\beta\gamma}\xi^\beta_t\xi^\gamma_t}{g_{\delta\epsilon}\xi^\delta_t\xi^\epsilon_t}\right)\xi^\alpha_t\right]. \tag{7.14}$$

In other words, the first two terms of the drift in (7.6) can be replaced by the expression $\tilde{D}^{\alpha}_{\beta}\xi^{\beta}_{t}$, where

$$\tilde{D}^{\alpha}_{\beta} = D^{\alpha}_{\beta} - \left(\frac{D_{\gamma\delta}\xi^{\gamma}_{t}\xi^{\delta}_{t}}{g_{\gamma\delta}\xi^{\gamma}_{t}\xi^{\delta}_{t}}\right)\delta^{\alpha}_{\beta}, \tag{7.15}$$

where δ^{α}_{β} is the Kronecker delta. Clearly, it follows that $\tilde{D}_{\alpha\beta}\xi^{\alpha}\xi^{\beta} = 0$.

With this in mind, let us now proceed to the interpretation of the volatility process σ_{tx}. Again, σ_{tx} has the character of a linear operator acting on ξ_{tx}, subject to the constraint $E_{\rho}[\sigma_{tx}] = 0$. This can be consistently enforced if there exists a symmetric process $\nu_{t\alpha\beta}$ such that

$$\int_{0}^{\infty} \eta(x)\xi_{tx}\sigma_{tx}\,\mathrm{d}x = \eta^{\alpha}(\nu_{t\alpha\beta}\xi^{\beta}_{t} - E_{\xi}[\nu_{t}]\xi_{t\alpha}). \tag{7.16}$$

The symmetric operator-valued random process $\nu_{t\alpha\beta}$, whose existence is thus implied, is 'primitive' in the sense that it is unconstrained and can be specified exogenously. If we write

$$\sigma_{t\alpha\beta} = \nu_{t\alpha\beta} - E_{\xi}[\nu_{t}]g_{\alpha\beta}, \tag{7.17}$$

we obtain

$$\int_{0}^{\infty} \eta(x)\xi_{tx}\sigma_{tx}\,\mathrm{d}x = \eta_{\alpha}\sigma^{\alpha}_{t\beta}\xi^{\beta}_{t}, \tag{7.18}$$

and also

$$\int_{0}^{\infty} \eta(x)\xi_{tx}\sigma^{2}_{tx}\,\mathrm{d}x = \eta_{\alpha}\sigma^{\alpha}_{t\beta}\sigma^{\beta}_{t\gamma}\xi^{\gamma}_{t}. \tag{7.19}$$

Therefore, putting the various ingredients together, we obtain the following proposition.

Proposition 7.1. *The dynamics of the Hilbert space vector ξ^{α}_{t} that characterizes the term structure in an admissible, arbitrage-free interest rate framework is governed by the stochastic differential equation*

$$\mathrm{d}\xi^{\alpha}_{t} = (\tilde{D}^{\alpha}_{\beta} - \tfrac{1}{8}\sigma^{\alpha}_{t\gamma}\sigma^{\gamma}_{t\beta})\xi^{\beta}_{t}\,\mathrm{d}t + \tfrac{1}{2}\sigma^{\alpha}_{t\beta}\xi^{\beta}_{t}(\mathrm{d}W_{t} + \lambda_{t}\,\mathrm{d}t), \tag{7.20}$$

where $\tilde{D}^{\alpha}_{\beta}$ is given as in (7.15), and the adapted operator-valued process $\sigma_{t\alpha\beta}$ is expressible in the form

$$\sigma_{t\alpha\beta} = \nu_{t\alpha\beta} - \left(\frac{\nu_{\gamma\delta}\xi^{\gamma}_{t}\xi^{\delta}_{t}}{g_{\gamma\delta}\xi^{\gamma}_{t}\xi^{\delta}_{t}}\right)g_{\alpha\beta}, \tag{7.21}$$

where $\nu_{t\alpha\beta}$ is an arbitrary adapted operator-valued process.

This result shows that the evolution of the yield curve can be viewed consistently as a process on the positive orthant of the unit sphere in Hilbert space, and thus gives rise to a new way of understanding the dynamics of the term structure (see figure 3). The purpose of the quadratic term in the drift of (7.20) is to keep the process on the sphere, and in the absence of the term involving the operator D^{α}_{β} we would have a general local martingale on the sphere $\mathcal{S}$ with respect to the risk-neutral measure, where the martingale property on $\mathcal{S}$ is characterized in a standard way by use of the techniques of stochastic differential geometry (see, for example, Emery 1989; Ikeda & Watanabe 1989; Hughston 1996*b*). The term involving the operator D^{α}_{β} splits into a symmetric and an antisymmetric part. The drift generated by the antisymmetric part of D^{α}_{β} is generated by a symmetry of the sphere $\mathcal{S}$. The drift generated by the symmetric part of D^{α}_{β}, on the other hand, is a negative-gradient vector field orthogonal to surfaces in $\mathcal{S}$ generated by level values of the short rate r_t. This term therefore creates a tendency for the vector ξ^{α} to drift towards a lower interest rate, a property of the negative gradient field that is then counterbalanced by the effects of the diffusive term.

D.C.B. gratefully acknowledges financial support from The Royal Society. L.P.H. acknowledges the hospitality of the Finance Department of the Graduate School of Business of the University of Texas at Austin, where part of this work was carried out. The authors are very grateful for helpful comments by seminar participants at the Risk 2000 Conference in Paris, the Boston Derivative Securities Conference, the Finance Department of the University of Amsterdam, and the Quantitative Risk Management Conference at Carnegie Mellon University.

References

Amari, S. 1985 *Differential-geometric methods in statistics.* Springer.

Baxter, M. 1997 General interest rate models. In *Mathematics of Derivative Securities* (ed. M. A. H. Dempster & S. R. Pliska). Cambridge University Press.

Bhattacharyya, A. 1943 *Bull. Calcutta Math. Soc.* **35**, 99.

Björk, T. 2001 A geometric view of interest rate theory. In *Handbooks in mathematical finance* (ed. E. Jouini, J. Cvitanic & M. Musiela). Cambridge University Press.

Björk, T. & Christensen, B. J. 1999 *Math. Finance* **9**, 323.

Björk, T. & Gombani, A. 1999 *Finance Stochastics* **3**, 413.

Björk, T. & Svensson, L. 2001 *Math. Finance* **11**, 205.

Brace, A., Gatarek, D. & Musiela, M. 1996 The market model of interest rate dynamics. In *Vasicek and beyond: approaches to building and applying interest rate models* (ed. L. P. Hughston). London: Risk Publications.

Brace, A., Gatarek, D. & Musiela, M. 1997 *Math. Finance* **7**, 125.

Brody, D. C. 2000 *Mathematical theory of finance.* Tokyo: Nihon-Hyoronsya.

Brody, D. C. & Hughston, L. P. 1998 *Proc. R. Soc. Lond.* A **454**, 2445.

Burbea, J. 1986 *Expo. Math.* **4**, 347.

Carverhill, A. 1994 *Stoch. Rep.* **53**, 227.

Constantinides, G. M. 1992 *Rev. Finance Studies* **5**, 531.

Cox, J. C., Ingersoll, J. & Ross, S. 1985 *Econometrica* **53**, 385.

Dawid, A. P. 1977 *Ann. Statist.* **5**, 1249.

Emery, M. 1989 *Stochastic calculus on manifolds.* Springer.

Filipović, D. 2000 Consistency problems for HJM interest rate models. PhD thesis, Swiss Federal Institute of Technology, Zurich.

Fisher, R. A. 1921 *Phil. Trans. R. Soc. Lond.* A **222**, 309.

Flesaker, B. & Hughston, L. P. 1996 *Risk Mag.* **9**, 46.

Flesaker, B. & Hughston, L. P. 1997*a* Exotic interest rate options. In *Exotic options: the state of the art* (ed. L. Clewlow & C. Strickland). London: International Thomson Press.

Flesaker, B. & Hughston, L. P. 1997*b* Dynamic models of yield curve evolution. In *Mathematics of derivative securities* (ed. M. A. H. Dempster & S. R. Pliska). Cambridge University Press.

Flesaker, B. & Hughston, L. P. 1997*c* *Net Exposure* **3**, 55. (Reprinted in *The new interest rate models* (ed. L. P. Hughston) (2000). London: Risk Publications.)

Harrison, J. M. & Kreps, D. M. 1979 *J. Econom. Theory* **20**, 381.

Harrison, J. M. & Pliska, S. R. 1981 *Stoch. Proc. Appl.* **11**, 215.

Heath, D., Jarrow, R. & Morton, A. 1992 *Econometrica* **60**, 77.

Ho, T. S. Y. & Lee, S. B. 1986 *J. Finance* **41**, 1011.

Hughston, L. P. 1996*a* Introduction. In *Vasicek and beyond: approaches to building and applying interest rate models* (ed. L. P. Hughston). London: Risk Publications.

Hughston, L. P. 1996*b* *Proc. R. Soc. Lond.* A **452**, 953.

Hunt, P. J. & Kennedy, J. E. 2000 *Financial derivatives in theory and practice*. Wiley.

Ikeda, N. & Watanabe, S. 1989 *Stochastic differential equations and diffusion processes*. Amsterdam: North-Holland.

James, J. & Webber, N. 2000 *Interest rate modelling*. Wiley.

Jamshidian, F. 1997 *Finance Stochastics* **1**, 293.

Kass, R. E. 1989 *Statist. Sci.* **4**, 188.

Kennedy, D. 1994 *Math. Finance* **4**, 247.

Long, J. L. 1990 *J. Financial Econ.* **26**, 29.

Mahalanobis, P. C. 1936 *Proc. Natn. Inst. Sci. India* A **2**, 49.

Merton, R. C. 1973 *Bell J. Econ. Manag. Sci.* **4**, 141.

Murray, M. K. & Rice, J. W. 1993 *Differential geometry and statistics*. London: Chapman and Hall.

Musiela, M. 1993 Stochastic PDEs and term structure models. *Journées Internationales de Finance, IGR-AFFI*, La Baule.

Musiela, M. & Rutkowski, M. 1997 *Martingale methods in financial modelling*. Springer.

Nielsen, L. T. 1999 *Pricing and hedging of derivative securities*. Oxford University Press.

Rao, C. R. 1945 *Bull. Calcutta Math. Soc.* **37**, 81.

Rebonato, R. 1998 *Interest rate option models*, 2nd edn. Wiley.

Rogers, L. C. G. 1995 Which model of the term structure of interest rate should one use? In *Mathematical finance* (ed. M. H. A. Davis *et al.*), IMA, vol. 65. Springer.

Rogers, L. C. G. 1997 *Math. Finance* **7**, 157.

Sandmann, K. & Sondermann, D. 1996 Lognormal interest rate models: stability and methodology. Discussion Paper B-398, University of Bonn.

Santa-Clara, P. & Sornette, D. 1997 The dynamics of the forward interest rate curve with stochastic string shocks. Preprint, University of California at Los Angeles.

Segal, I. E. 1947 *Ann. Math.* **48**, 930.

Vasicek, O. 1977 *J. Finance Econ.* **5**, 177.