Yield Curve Smoothing and Residual Variance of Fixed Income Positions

Raphaël DOUADY

Centre d'Economie de la Sorbonne, Paris 1 University, CNRS

Riskdata, 6 rue de l'Amiral de Coligny 75001 Paris

January 12, 2011

ABSTRACT   We model the yield curve in any given country as an object lying in an infinite-dimensional Hilbert space, the evolution of which is driven by what is known as a *cylindrical Brownian motion.* We assume that volatilities and correlations do not depend on rates (which hence are Gaussian). We prove that a principal component analysis (PCA) can be made. These components are called *eigenmodes* or *principal deformations* of the yield curve in this space. We then proceed to provide the best approximation of the curve evolution by a Gaussian Heath-Jarrow-Morton model that has a given finite number of factors. Finally, we describe a method, based on finite elements, to compute the eigenmodes using historical interest rate data series and show how it can be used to compute approximate hedges which optimise a criterion depending on transaction costs and residual variance.

0.1   Introduction

Infinity is a word that economists usually do not like. Nothing, in economy, can be considered either as infinitely large, or as infinitely small. The size of worldwide markets is finite, as well as the total number of various stocks and bonds. Conversely, transactions cannot be infinitely close in time (a minimum time period is required between two transactions on the same asset) and price variations cannot be less that a "tick", neither can they be infinitely large. Nevertheless, two seminal articles introduced, perhaps unwillingly, infinity into the finance literature[1]. The first one is

---

[1] One should add that the theoretical justification of arbitrage theory is itself an – actually questionable – "infinity" argument: if some true arbitrage opportunity existed, it could be implemented with an infinite nominal amount, hence reducing it to zero.

Robert C. Merton's article on continuous time finance [34, 1973] (see also [35] and ref. cit.). Indeed, considering the possibility of continuous time trading replaces the setting of a finite set of random innovations by the infinite-dimensional Wiener space of Brownian motions. Although we just said that such trading strategies are physically impossible to execute, we consider this theory as financially extremely significant. The reason is that when one wants to study the results of a given trading strategy over a time period that is huge compared to the minimal trading interval, then any discrete time approach based on the maximal trading frequency would be, for an equivalent numerical precision, much more complex to implement than the corresponding continuous time limit. Consequently, though practical hedging should be performed with a view of optimizing a near future situation, according to all particularities of the market at the present time, pricing, which is based on an average of the resulting wealth of forecasted hedging strategy, is better handled in a continuous time framework[2].

The second article is Heath-Jarrow-Morton interest rate model [26]. This model, which we call H.J.M. in the sequel, "summarizes" information about the interest rate market (Libor, Libor futures, swaps, fixed income assets, etc.) into a "yield curve" or, more precisely, a curve of "forward spot rates". The knowledge of this curve is equivalent, through a simple integration with respect to maturity, to the price of zero-coupon bonds of any maturity. Again, here the word "any" means a continuum of maturities. The set of possible curves is an infinite dimensional functional space and the market cannot be described by a finite set of "market variables", although only finitely many assets are traded. One could argue that the model can be reduced to a finite-dimensional subspace and that the knowledge of a finite number of variables is enough to describe the whole market. In fact, this argument does not hold. Even if, as explicated in Heath-Jarrow-Morton article, the infinitesimal evolution of the curve is given by a finite number of "factors", the support of the distribution of possible curves after an arbitrarily small, but finite, amount of time is in general equal to the whole functional space[3]. Economically speaking, one must understand that both the finite number of factors and the continuous curve are introduced for the sake of simplicity, but none of them corresponds to reality: the total number of asset prices is finite, and the number of sources of noise, though also finite, is much larger than what is currently implemented in most trading floors.

A few articles describe capital markets as a random field. The first one that came to our knowledge is Kennedy [30]. This model is a Gaussian random field which could be considered as a generalization of the Gaussian H.J.M. model with volatility factors which do not depend on the level of the

---

[2] "Hedge in discrete time finance, price in continuous time finance" (N. Taleb, 1996).

[3] In [37], Musiela gives an example of a one-factor HJM model with this property.

rates. In the example he provides, the forward spot rate $f(t, T)$ is a "Brownian sheet", which is in contradiction with statistical evidence. Indeed, the process $f(t, T)$ has a very asymmetrical behavior with respect to its two variables. For fixed T, it behaves as a stochastic process with respect to the current date t, but at a given date t the curve $T \longmapsto f(t, T)$ is generally smooth. In the random field framework, this can be achieved by requiring the correlation function of the field to be smooth along the diagonal in the transverse direction. See Bricio-Hernandez [9] for a theoretical study of this topic and Turner [44] where a statistical study of the correlation function, showing this smoothness along the diagonal, is performed.

In this article, we develop another approach towards an infinite dimensional model, based on so-called "cylindrical Brownian motions". These processes are a generalization multidimensional Brownian motions to infinite dimensional Hilbert spaces. They were introduced by Gaveau in 1953 (see [24]). We refer to Yor [46] and to Da Prato-Zapczyk [15, p. 96] for a complete presentation of this theory. Our model can be seen as a limit case of H.J.M. model in infinite dimensions. As in Brace-Gatarek-Musiela model (B.G.M.), we consider the term structure of interest rates as an object in a certain functional space. We then proceed to study the motion of the vector representing this object. For simplicity reasons, we chose to work in a framework where rates are Gaussian. However this theory can be easily generalized to B.G.M. log-normal setting, or to any specified diffusion process for the term structure in which rate volatilities depend on the global term structure. Obviously certain technical assumptions apply.

Under very natural hypothesis[4], we show that this type of motion can always be decomposed into an (infinite) sum of one-dimensional Brownian motions, which we call *eigenmodes* or *principal deformations.* This turns out to be a principal component analysis (P.C.A.) of the motion. Listing all the works on the yield curve P.C.A. would be impossible. Let us mention the initial study (as it came to our knowledge) of Litterman and Scheinkman [32], the theoretical article of the Banque de France [23], and the statistical analysis cited in this article. It is shown in sect. 0.8 that the n-factor H.J.M. model that best reproduces an infinite dimensional diffusion of the yield curve, in the sense of minimizing the variance of the error, is provided by the truncated P.C.A.

R. C. Merton said in his preface to *Continuous-time Finance: "The continuous time model is a watershed between the static and dynamic models of finance".* Similarly, we could say that this "functional analysis" of term structure models — a less polemical term than "infinite dimensions", though representing the same thing — is a watershed between one-dimensional and multi-dimensional arbitrage pricing.

---

[4]We assume that the price of zero-coupons always depends continuously on the maturity, and that their variance is finite at all time.

v

From the point of view of risk management, this approach allows a sub-stantial reduction of the computational burden relative to the usual "buck-eting" method, while not losing any precision on market data fitting and risk evaluation. For this purpose, performing a P.C.A. of a statistically estimated variance-covariance matrix of the yield curve movements is of little help because of the high instability of this matrix. We recommend to choose a fixed series of basic deformations of the yield curve, which could be inspired by Fourier analysis or wavelets. A thorough historical analysis has to be performed to find the minimum number of terms one needs in order to reproduce, up to a tightly controlled error, all possible variations, even in case of crisis. In [16], we provide an example of such basic deforma-tions for which seven terms are sufficient to reproduce the variations of all exchange quoted Euro-dollar futures over a 10-years period with an error that never exceeds two basis points. The same number of terms applies to cash and swap rate variations from one month until thirty years.[5] In comparison, Basel committee recommends to use 13 buckets in yield curve deformations.

Option pricing theory faces an unexpected difficulty in the infinite di-mensional setting. Even if the whole volatility structure of the yield curve — that is, rate volatilities and correlations — is deterministic and known, some options may not support perfect replication, although they may have an "arbitrage price". In fact one will seek a sequence of "almost replicating" strategies, with a wealth variance tending to 0 and converging sequence of initial price. This leads us to introduce the notion of "quasi-arbitrage", that is, a sequence of trading strategies with returns bounded from below and wealth variance tending to 0. Only in the absence of quasi-arbitrage (A.Q.A. assumption) will an equivalent risk-neutral probability exist. Then an option price is the risk-neutral expectation of its discounted pay-off. This theoretical impossibility to perfectly replicate options does not create more difficulty in practical dynamic hedging than the inability to implement of a purely continuous time dynamic hedging. In a sense, it induces even less risk for, as mentioned above, the "spatial uncertainty", which measures how rate interpolations can be inaccurate, is much less unpredictable than the time uncertainty, which measures rate variations between two dynamic hedging transactions. Actually, traders often use linear interpolations to evaluate, when necessary, the rate to apply on a period which corresponds to no standard products. This practice justifies our approach based on the yield curve regularity.

This article is organized as follows. After preliminaries and notations

---

[5] However, it does not correctly shows bond yield variations because, for economical reasons that are not the topic of this article, each bond price is subject to its own individual source of noise and the "smooth curve principle" only applies up to a limit of 20-30 bp's error size that connot be captured by smooth functions of the maturity. Note that usual one-year buckets face the same inaccuracy.

(sect 0.3), we first expose (sect.0.5) the infinite dimensional diffusion of the yield curve and study the existence of an absolutely continuous risk-neutral probability, introducing the notion of quasi-arbitrage. In sect. 0.7, we show that such a model can be seen as a limit case of H.J.M. finite dimensional model (in fact an extended version of the strict H.J.M. framework). In particular, we show in sect. 0.8 the possibility of performing an infinite dimensional P.C.A. In sect. 0.10, we provide basic option pricing formulas. The last part (sect. 0.11 and 0.12) is devoted to numerical methods for the three following purposes: option pricing and hedging, calibration, P.C.A. computation. Hedges optimizing a cost vs. residual variance criterion are provided when transaction costs apply.

In this study, we assume that rates only follow (infinite dimension) diffusion processes. In particular, we exclude jumps and other processes not driven by Brownian motions.

0.2   History, Tribute and Recent Bibliography

A first version of this article was written in 1994, while the author was working Société Générale in the fixed income and foreign exchange derivatives trading room. This version has been finalized in 2001, but was never published before. This work raised the interest of Marek Musiela, who at that time was teaching at the University of New South Wales in Sydney, and was at the origin of the long and fruitful scientific relationship between the author and Marek Musiela. Modeling yield curves with infinitely many risk factors raised a lot of skepticism from the mathematical finance community at this time, and Marek Musiela was one of the first person to perceive this deep fact that infinite dimensions do not come as an *increase in complexity* of the models but, on the contrary, as the necessary *path to the most parsimonious* models of a complex reality.

Since then, many authors wrote on infinite dimensions modeling of the term structure, and also of other aspects of financial markets, such as volatility surfaces. Let us cite here only major references, which the reader is invited to consult, as well as other articles cited in those references. One of the most complete study on this topic is Damir Filipovic's thesis [21]. A good statistical study of the regularity properties of the yield curve in the US has been performed by Bouchaud & al. [5]. Another approach to infinite dimensions modeling is through stochastic partial differential equations (SPDE), see Cont's article [12] on this matter. In [37], Musiela and Sondermann pointed out that even a one-factor model can lead to a yield curve lying in an infinite dimensional space.

As general references to interest rate modeling , we recommend books by Musiela and Rutkowski [38], Brigo and Mercurio [10] and Rebonato [41], who also wrote in 2003 a thorough survey of interest rate modeling , as it

appeared at that time [42].

0.3   Notations and Definitions

The framework described here is the classical framework of Heath-Jarrow-Morton [26], Brace-Gatarek-Musiela [6], El Karoui-Myneni-Viswanathan [20] and Jamshidian [29]. It is extended in the sense that each individual forward spot rate or discount factor is driven by its own Brownian motion (possibly correlated to others).

0.3.1   Term Structure of Interest Rates

We denote by $z(t, T)$ the discount factor applying on the period $[t, T]$, that is, the price at time t of an asset delivering one unit of numeraire at time T and let :

$$\ell(t, T) = -\log z(t, T) \qquad\qquad y(t, T) = \frac{1}{T - t}\, \ell(t, T) \qquad (1)$$

so that $y(t, T)$ is the continuously-compounded *zero-coupon rate,* or *yield,* over the time period $[t, T]$. One has $z(t, t) = 1$, $\ell(t, t) = 0$, $z(t, T) > 0$. We shall not make any assumption about positive interest rates, for our model will be of Gaussian type, which does not prevent from negative rates, although with a very low probability. Such an assumption may be achieved by changing the volatility structure (see for instance [4], [6] for a log-normal structure, and [13], [18] for a χ^2-type distribution).

The *spot rate* is the value of the zero-coupon rate when $T = t$:

$$r(t) = y(t, t)$$

In the H.J.M. framework, the *forward spot rate* is considered :

$$f(t, T) = \frac{\partial \ell(t, T)}{\partial T}$$

It is linked to the zero-coupon price and yield by the formulas :

$$z(t, T) = \exp\left(-\int_t^T f(t, s)\, ds \right) \qquad (2)$$

$$y(t, T) = \frac{1}{T - t} \int_t^T f(t, s)\, ds \qquad (3)$$

In this article, we only consider continuously compounded rates, namely, those defined via the *logarithm* of zero-coupon prices. The usual rates, with finite compounding periods, are computed from those by simple formulas.

The functions z, ℓ, y and f will be considered as various languages to describe the same term structure. They will always be linked by the formulas above.

We denote by $\beta(t, T)$ the savings account at date T initiated at time t :

$$\beta(t, T) = \exp\left(\int_t^T r(s)\, ds\right)$$

The value of $\beta(t, T)$ is only known at date T.

0.3.2 Risk-neutral Probability

In the sequel, for any random process X_t, we denote by $\mathbf{E}_{\mathcal{P}}[X_T \,|\, t]$ the conditional expectation of X_T knowing the past until date t under the probability measure $\mathcal{P}$. We shall simply write $\mathbf{E}[X_T]$ if there is no ambiguity about t and $\mathcal{P}$.

An origin of time $t_0 = 0$ is fixed once for all, as well as a maximum maturity date of assets $T_{\max}$. For any Ito process X_t, we set:

$$X_t = X_0 + \int_0^t dX_u$$

In particular, we shall write indifferently:

$$dX_t = \mu(t)\, dt + \sigma(t)\, dW_t \iff X_t = X_0 + \int_0^t \mu(u)\, du + \int_0^t \sigma(u)\, dW_u$$

where W_t is Brownian motion under $\mathcal{P}$ or another probability and $\mu(t)$, $\sigma(t)$ are predictable processes.

We let now $\mathcal{P}$ denote the real, or historical, probability. In the absence of arbitrage opportunities, for any maturity T, there exists a "risk-neutral" probability $\mathcal{Q}_T$ equivalent to $\mathcal{P}$, such that the discount factor $z(t, T)$ is the expectation at date t of $\beta(t, T)^{-1}$:

$$z(t, T) = \mathbf{E}_{\mathcal{Q}_T}\left[\beta(t, T)^{-1} \,|\, t\right]$$

Under this probability, the price of any asset X_t *depending only on discount factors $z(t', T)$, $t < t' < T$*, is such that $\beta(t_0, t)^{-1} X_t$ is a $\mathcal{Q}_T$-martingale for any initial date t_0. We call such assets T-assets. If the market of T-assets is complete, then $\mathcal{Q}_T$ is unique in the sense that two such probabilities would coincide on the space of T-assets. In this case, $\mathcal{Q}_T$ is characterized by the fact that $\beta(t_0, t)^{-1} z(t, T)$ is a martingale. The *risk-neutral* probability $\mathcal{Q}_T$ should not be confused with the *forward-neutral* one, which we shall denote $\mathcal{Q}'_T$ and characterized by the fact that $z(t, T)^{-1} X_t$ is a martingale for any T-asset X_t.

The Radon-Nicodym density of $\mathcal{Q}_T$ is given with respect to $\mathcal{P}$ by the mean of the "market price of risk", which a priori depends on the asset

$z(t,T)$, thus on T. In fact, for any given *finite* set of maturities $\mathbf{T}=(T_1,\ldots,T_n)$ there exists a probability $\mathcal{Q}_{\mathbf{T}}$ under which all the t_0-actualized discount factors $\beta(t_0,t)^{-1}z(t,T_i)$ are martingales. The existence, uniqueness and absolute continuity of a risk-neutral probability for an infinite set of maturities, for instance a whole interval, is in general not ensured. This will be the topic of sect. 0.3.3. When it exist, we call $\mathcal{Q}$ a (the) probability which is risk-neutral with respect to every maturity $T \in [0, T_{\max}]$.

0.3.3 Diffusion of discount factors and forward rates

Discount factors and yields

For every fixed T, the discount factor process (defined for $t \leq T$) :

$$t \longmapsto z(t,T)$$

follows an Ito process :

$$dz(t,T)|_{T \text{ fixed}} = z(t,T)\left(\mu(t,T)\,dt - \sigma(t,T)\,d\tilde{W}_t^T\right) \qquad (4)$$

where $\tilde{W}_t^T$ is a standard Wiener process in t under the probability $\mathcal{P}$, depending on the parameter T, $\mu(t,T)$ and $\sigma(t,T)$ are predictable processes for the drift and the volatility respectively (the "minus" sign is artificial and has been put for technical reasons). The identity $z(t,t)=1$ implies :

$$\sigma(t,t) = 0$$

and, because:

$$z(t,t+dt) = 1 - r(t)\,dt + O(dt^2)$$

one has:

$$\mu(t,t) = r(t)$$

Taking the logarithm of (4), we get from Ito formula:

$$d\ell(t,T)|_{T \text{ fixed}} = \left(-\mu(t,T) + \frac{1}{2}\sigma(t,T)^2\right)dt + \sigma(t,T)\,d\tilde{W}^T \qquad (5)$$

Assumption We make the following non-degeneracy assumption, which is a strong version of the completeness of the market of T-assets:

$$t < T \Longrightarrow \sigma(t,T) > 0$$

Let the process W^T be defined by:

$$W_t^T = \tilde{W}_t^T + \int_0^t \lambda(u,T)\,du \qquad \lambda(u,T) = \frac{r(u) - \mu(u,T)}{\sigma(u,T)}$$

This is a standard Brownian motion under probability $\mathcal{Q}_T$ and under $\mathcal{Q}$ if it exists (see [14] or [41]). One has:

$$dz(t,T)|_{T \text{ fixed}} = z(t,T)\left(r(t)\,dt - \sigma(t,T)\,dW_t^T\right) \tag{6}$$

and:

$$d\ell(t,T)|_{T \text{ fixed}} = \left(-r(t) + \frac{1}{2}\sigma(t,T)^2\right)dt + \sigma(t,T)\,dW^T \tag{7}$$

Which can be written, in terms of the quadratic variation process of $\ell(t,T)$:

$$d\ell(t,T)|_{T \text{ fixed}} = -r(t)\,dt + \frac{1}{2}\,d\langle\ell(t,T)\rangle + \sigma(t,T)\,dW^T \tag{8}$$

For the zero-coupon rates $y(t,T)$, we get :

$$dy(t,T) = \left(\frac{y(t,T) - r(t)}{T-t} + \frac{T-t}{2}\,\sigma_y(t,T)^2\right)dt + \sigma_y(t,T)\,dW^T \tag{9}$$

where:

$$\sigma_y(t,T) = \frac{1}{T-t}\,\sigma(t,T)$$

Forward rates

Assume now that, following Heath-Jarrow-Morton [26], forward spot rates themselves follow an Ito process :

$$df(t,T)|_{T \text{ fixed}} = \mu_f(t,T)\,dt + \sigma_f(t,T)\,d\tilde{W}_f^T \tag{10}$$

where $\tilde{W}_f^T$ are Brownian motions under $\mathcal{P}$ and $\mu_f(.,T)$, $\sigma_f(.,T)$ are predictable processes depending on the maturity T such that, for any $t_0 \leq t \leq T \leq T_{\max}$:

$$\mathbf{E}\left(\int_t^T |\mu_f(t,s)|\,ds\right) < \infty \qquad \mathbf{E}\left(\left(\int_t^T |\sigma_f(t,s)|\,ds\right)^2\right) < \infty \tag{11}$$

We also assume that the family $(\tilde{W}_f^T)_T$ has independent increments, that is, for any T, T' the increment $d\tilde{W}_f^{T'}(t)$ is independent of $\tilde{W}_f^T(t)$. The *instantaneous correlation function* $\rho(t,T,T')$ is defined by:

$$\rho(t,T,T') = \text{Corr}_{\mathcal{P}}\left(d\tilde{W}_f^T(t), d\tilde{W}_f^{T'}(t)\right)$$

or, in terms of cross-variation process:

$$d\left\langle \tilde{W}_f^T(t), \tilde{W}_f^{T'}(t)\right\rangle = \rho(t,T,T')\,dt$$

Obviously, $\rho(t,T,T) = 1$, $|\rho(t,T,T')| \leq 1$ and $\rho(t,T,T') = \rho(t,T',T)$ for any (t,T,T'). Moreover, for any sequence of maturities $(T_1, \ldots, T_n)$, the matrix (ρ_{ij}) where $\rho_{ij} = \rho(t,T_i,T_j)$ is symmetric and positive.

Assumption We shall assume that, for any $(t; T_1, \ldots, T_n)$ such that $t < T_i \neq T_j$ for any $i \neq j$, the matrix (ρ_{ij}) is positive definite, and:

$$\sigma_f(t, T_i) > 0$$

that is, no finite combination of forward spot rates has, at no time, zero volatility.

In order to define "risk-neutral" Brownian motions dW_f^T, we shall make heuristic calculations that will be made rigorous below. We wish to write:

$$dW_f^T = \lambda_f(t, T) \, dt + d\tilde{W}_f^T$$

Identifying the $\mathcal{P}$-martingale component of $\ell(t, T + \delta T) - \ell(t, T)$ with that of $f(t, T)\delta T$, we get:

$$\sigma(t, T + \delta T) \, d\tilde{W}^{T+\delta T} - \sigma(t, T) \, d\tilde{W}^T = \sigma_f(t, T) \, \delta T \, d\tilde{W}_f^T + O(\delta T^2)$$

Assuming that the risk-neutral probability $\mathcal{Q}$ exists, one must have, by taking $\mathcal{Q}$-expectation:

$$\mu(t, T) - \mu(t, T + \delta T) = \sigma_f(t, T) \, \delta T \, \lambda_f(t, T) + O(\delta T^2)$$

which leads, when $\delta T \to 0$, to:

$$\lambda_f = -\frac{1}{\sigma_f} \frac{\partial \mu}{\partial T}$$

On the other hand, by taking the $\mathcal{P}$-expectation in (5) and (10), we get:

$$\mu_f(t, T) \, \delta T = \mu(t, T) - \mu(t, T + \delta T) + \frac{1}{2}\sigma(t, T + \delta T)^2 - \frac{1}{2}\sigma(t, T)^2 + O(\delta T^2)$$

therefore, letting $\delta T \to 0$, one gets:

$$\frac{\partial \mu}{\partial T} = -\mu_f + \sigma \frac{\partial \sigma}{\partial T}$$

and, finally:

$$df(t, T)|_{T \text{ fixed}} = \left(\mu_f + \frac{\partial \mu}{\partial T}\right) dt + \sigma_f \, dW_f^T = \sigma \frac{\partial \sigma}{\partial T} \, dt + \sigma_f \, dW_f^T \quad (12)$$

The link between volatilities σ and σ_f is as follows. From:

$$\ell(t, T) = \int_t^T f(t, s) \, ds$$

we deduce:

$$d \langle \ell(t, T) \rangle = \iint_{(u,v) \in [t, T]^2} d \langle f(t, u), f(t, v) \rangle \, du \, dv$$

that is:

$$\sigma(t,T)^2 = \iint_{(u,v)\in[t,T]^2} \sigma_f(t,u)\sigma_f(t,v)\rho(t,u,v)\,dudv$$

Again, this formula, which generalizes H.J.M., will be rigorously proved later on, after having set the formalism of function valued random processes. It provides another expression of the risk-neutral drift of forward spot rates:

$$\sigma(t,T)\frac{\partial\sigma}{\partial T}(t,T) = \sigma_f(t,T)\int_t^T \sigma_f(t,u)\rho(t,u,T)\,du \qquad (13)$$

from which we deduce the following:

Proposition 1 *One has:*

$$\left|\frac{\partial\sigma}{\partial T}(t,T)\right| \leq (T-t)\,\sigma_f(t,T)$$

with equality in the case of a local one-factor Ho and Lee model, that is, $\sigma_f(t,u)$ doesn't depend on u and $\rho(t,u,v) = 1$ for any (u,v).

Proof Consider the scalar product on functions $\phi : [t,T] \to \mathbb{R}$:

$$\phi.\psi = \iint_{(u,v)\in[t,T]^2} \phi(u)\psi(v)\rho(t,u,v)\,dudv$$

Schwarz inequality between functions σ_f and 1 implies:

$$\left(\int_t^T \sigma_f(t,u)\rho(t,u,T)\,du\right)^2 \leq \sigma(t,T)^2 \iint_{(u,v)\in[t,T]^2} \rho(t,u,v)\,dudv$$

$$\leq (T-t)^2\,\sigma(t,T)^2$$

Equality occurs only if $\sigma_f(t,.)$ is constant and $\rho(t,.,.) \equiv 1$. $\square$

It is worthy of note that equations (12) and (13), which generalize results obtained by Heath-Jarrow-Morton [26] and by Brace-Gatarek-Musiela [6], only assume that $\mathcal{Q}$ is a risk-neutral probability, but do not require that the whole yield curve evolution is driven by a finite number of Brownian motions.

0.3.4   Function Valued Random Processes

The yield curve at time t is defined as the function :

$$\mathbf{y}_t : x \longmapsto y(t,t+x)$$

It is defined on an interval $I = [0, M]$ which we shall suppose fixed, for instance $M = 30$ years. The variable t therefore spans the interval $[0, T_{\max} - M]$. The yield curve $\mathbf{y}_t$ belongs to some functional space H^y over I. We shall define H^y more precisely later on. For the moment, we only assume it is a Banach space, equipped with a norm $\|.\|_{H^y}$ and contained in the space of continuous functions. This defines a random process $\mathbf{y}_t$ with values in H^y.

Similarly, we define functions $\mathbf{z}_t \in H^z$, $\mathbf{l}_t \in H^\ell$, $\mathbf{f}_t \in H^f$. These functions are linked to $\mathbf{y}_t$ and between themselves by equations (1) and (2). The spaces H^z, H^ℓ and H^f should also be linked in a similar way. For example, if $H^f = C^0(I)$ then H^ℓ is the space of C^1 functions vanishing at 0, H^y is the space of continuous functions on I, of class C^1 on $(0, M]$ and whose derivative (with respect to x) is $O(\frac{1}{x})$ at 0, and $H^z = C^1(I)$, or the affine subspace of functions taking value 1 at 0. Note that the correspondence between $\mathbf{y}_t$, $\mathbf{l}_t$ and $\mathbf{f}_t$ are linear, unlike that with $\mathbf{z}_t$.

In order to define function valued processes $\mathbf{y}_t$, $\mathbf{z}_t$, $\mathbf{l}_t$ and $\mathbf{f}_t$, we shall use the formalism of so-called *cylindrical Brownian motions* which appears to be best suited for our purposes. A static portfolio made only of linear assets — bonds, swaps, F.R.A., but not options — can be seen as a finite combination of Dirac masses on H^z, corresponding to payment dates and amounts.

In fact, in [47], Yor proved that, in order to define a cylindrical Brownian motion in the infinite dimensional space H^y (this will be our theoretical setting), one needs to choose a Hilbert space, for instance $L^2(I, \mu)$, where μ is a measure on I, or a Sobolev space with respect to a measure on I.

Remark 2 *The choice of the space H^y or, equivalently, of its norm, that is, of the Sobolev exponent and of the measure μ is one of the most important issues. Indeed, this norm measures the risk and should be in accordance with the most probable moves of the yield curve. Generally speaking, we shall see that the most appropriate choice for μ is linked to the distribution in maturities of the significant quoted rates, while the Sobolev exponent, which stands for the curve smoothness, results from market practice and can be deduced in a rather reliable way from the statistics.*

Although we have not yet defined processes in H^y, we see that its drift will depend on the diffusion with fixed $x = T - t$, that is, with slipping maturity T. One has :

$$dy(t, t+x)|_{x \text{ fixed}} = dy(t, t+x)|_{t+x \text{ fixed}} + \frac{\partial y}{\partial T}(t, t+x)\, dt$$

$$= \left(\frac{1}{x}\left(f(t, t+x) - r(t)\right) + \frac{x}{2}\,\sigma_y(t, t+x)^2 \right) dt + \sigma_y(t, t+x)\, dW^T$$

The same holds for z, ℓ and f (provided the function f is differentiable) :

$$df|_{x \text{ fixed}} = df|_{T \text{ fixed}} + \frac{\partial f}{\partial T}\, dt$$

$$d\ell|_{x \text{ fixed}} = \left(f(t, t+x) - r(t) + \frac{1}{2}\sigma(t, t+x)^2 \right) dt + \sigma(t, t+x)\, dW^T$$

$$dz|_{x \text{ fixed}} = z(t, t+x) \left((r(t) - f(t, t+x))\, dt - \sigma(t, t+x)\, dW^T \right)$$

0.4   Market data on the term structure

Before setting an abstract framework for the term structure of interest rates, we shall first make a short presentation of the market data that provides it. This will either justify certain modeling and assumptions, or show their limits. We mean by market data the prices of assets which are interest rate dependent, or quoted rates which are linked to these prices by a standard formula. Although they are well known, we shall list these assets in order to examine the features of each one from the point of view of their incidence on the yield curve smoothness. For a complete study of market data on the term structure and on their interrelations, we refer to Anderson-Breedon-Deacon-Derry-Murphy [1]. The reader who is familiar with the fixed income market is advised to jump directly to the conclusions of this section.

0.4.1   Bonds

A bond delivers a "coupon" C at dates $T_1, \cdots, T_n$ (where $T_k = T_0 + k\,\delta T$, $\delta T = 3$, 6 or 12 months) and the principal N at T_n. The *coupon rate* R is defined by the formula :

$$C = R\,N\,\delta T$$

Hence, its price at time $t < T_1$ is, or should be :

$$P_{\text{bond}}(t) = N \left(z(t, T_n) + R\,\delta T \sum_{k=1}^{n} z(t, T_k) \right)$$

where $z(t, T)$ is the discount factor between dates t and T.

Remark 3 *In practice, some bonds have prices trading above or below the theoretically determined price, owing to taxation, institutional factors or simply liquidity reasons.*

The market also quotes very liquid bond futures, which cannot be theoretically perfectly linked to discount factors for two reasons. First, there is a system of margin calls, which make them (slightly) sensitive to the covariance between the bond price and the short rate. Second, they quote the value of the cheapest bond in a given pool, hence they involve an option associated to the possible change in the "cheapest-to-deliver" (most of the time, this option is far from being at the money).

0.4.2 Swaps

As it is well known, a swap is an exchange of a fixed interest rate loan with a variable rate one, both of the same principal and the same maturity. The variable leg can be replicated by a rolling loan of the principal over the whole period. The following formula gives the price that should be paid at the beginning by the side paying the fixed rate in order to enter an asset swap[6] with fixed rate R and settlement dates $(T_1, \ldots, T_n)$, $T_k = T_0 + k\,\delta T$:

$$P_{\text{swap}}(t) = N\left(z(t, T_0) - z(t, T_n) - R\,\delta T \sum_{k=1}^{n} z(t, T_k)\right)$$

The *swap rate* is the value of R that cancels the price (for other fixed rates, this is an asset swap) :

$$R(t, n\delta T) = \frac{z(t, T_0) - z(t, T_n)}{\delta T \sum_{k=1}^{n} z(t, T_k)}$$

Bond prices and swap rates provide an information on the value of a given discount factor with respect to an average of others with a shorter maturity. One usually uses a "boot strapping" method to compute discount factors, indeed, errors are at each step multiplied by the coupon rate and do not accumulate.

0.4.3 Cash and future short rates

A cash rate is the rate of a loan without coupon (a "zero-coupon"). All cash rates from the "overnight" to one year[7] are permanently quoted on market screens.

Future contracts are *forward rate agreements* (FRA) on a 3 months loan on prescribed periods[8], with a system of *margin calls.* Their quotation has

[6] As usual, settled in advance, paid in arrears; there are also swaps paid in advance.
[7] Precisely O/N, next day, 1 and 2 weeks, 1, 2, 3, 6, 9 and 12 month (all intermediaries are immediately given by market makers on request).
[8] So-called IMM dates, around mid-March, June, September and December.

been recently extended to a four-years forward period in Europe (France, GB, Germany) and in Japan, and 10 years in the USA. The price of futures is slightly different from that of the corresponding FRA (no more than a few basis points) because of the convexity induced by the margin calls.

An FRA rate on the period $[T, T']$ evaluated at date $t < T$ is given by :

$$\text{FRA}(t, T, T') = \frac{1}{T' - T} \left((T' - t) \, y(t, T') - (T - t) \, y(t, T) \right)$$

If T and T' are close, the FRA gives an estimate of the forward spot rate, that is, of the *derivative*[9] of the yield curve with respect to maturity. Practically speaking, if T is worth several years, then $T' = T + 3\,months$ can be considered as close.

0.4.4   STRIP, or the decomposition of bonds

In the USA, T-bonds can be "stripped", that is, fragmented into the principal and the coupons, which can be negotiated separately. The same holds (since a more recent period) in other countries for government bonds. The coupons are called *Strips.* Theoretically, this gives the level of every zero-coupon rate. However, they are much less liquid than the rest of the market and, in fact, market makers use that other information to price them. Furthermore, their price is a little affected by the fact that, in order to reconstruct a full T-bond, one is not allowed to replace the principal by an accumulation of coupons at the same date. Because of the "reconstruction" opportunity that it provides, the principal is slightly more expensive than its "theoretical" price. The stripping generates large amounts of data, but it must be used with care.

The behavior of coupon prices is more surprising. For short maturities, they are below the theoretical price, but, at some point (between 10 and 20 years) they pass to the other side. To explain this fact, let $\tilde{z}_P(t, T_0, T_n, \delta T, R)$ be the price at date t of the principal of a bond that has coupons of size R at dates T_0, $T_0 + \delta T, \ldots, T_n = T_0 + n\delta T$, and let $\tilde{z}_C(t, T_k)$ be the price of a strip of maturity T_k. Reconstructing the bond provides a relation between these prices and the theoretical prices :

$$\tilde{z}_N(t, T_0, T_n, \delta T, R) - z(t, T_n) = R \, \delta T \sum_{k}^{n} \left(z(t, T_k) - \tilde{z}_C(t, T_k) \right)$$

This difference should always be positive, but not too big. If one sees the right hand side as a discrete approximation of an integral, we see that the "liquidity spread" on the principal price is equal to the algebraic area[10] between the theoretical zero-coupon price curve and the strip curve (counted

---

[9] In the sense of the mathematical derivative of a function.

[10] That is, counted with signs.

negatively if the strip curve goes above). When one tries to fit a constant liquidity spread for the principal, then the accumulation effect of the strips must be compensated on the long term part. Of course, the behavior of rates is symmetrical.

Remark 4 *An important observation is that market makers on futures and on strip markets have a tendency to* smooth *the curve with respect to T, as if some kind of elasticity tried to erase possible angles.*

0.4.5 Conclusion

After these observations, we look at the yield curve as an object lying in some functional space H that has either infinite dimensions or at least a large one. The discount factors and the yields are *implicit variables,* in the sense that explicit data are not reliable (see comment on strips). We must therefore take into account other linear and nonlinear functions of the rates. Remark 4 tends to indicate that the space H should consist of differentiable functions.

We tend to see the term structure of interest rates as a "smooth skeleton" given by averaging the available information, with some noise due partly to rounding to the nearest basis point (or to the bid/ask spread), partly to the particularities of each market. What will be described in a theoretical framework is the evolution of the "smooth skeleton", for the "noise" can be considered as bounded and does not represent a risk that should be hedged according to the usual Black-Scholes theory based on diffusion processes. In practice, cash and swap rates are extremely close to a smooth curve (about 1-2 bp), while each government bond has its own spread (on the positive side) over the cash-swap curve, and this series of spreads cannot be modeled as a curve.

As we mentioned in the introduction, we shall see how assuming first that H is infinite-dimensional allows us to find very good approximation subspaces of rather low dimension, through standard finite element techniques.

0.5   Brownian motions in a Hilbert space

Defining a Brownian motion in an infinite dimensional space is not a simple task. Indeed, such a space is not locally compact and there is no Lebesgue measure[11] on it, thus no Gaussian density (although Gaussian probability measures exist). To overcome this difficulty, we shall refer to the formalism developed in 1973 by Gaveau [24] and well explained in Da Prato-Zapczyk [15, p. 96] and in Yor [46].

---

[11] That is, a uniformly distributed measure, invariant by translation.

Yor gives three different definitions of a Brownian motion in a Hilbert space H and shows that these three frameworks are equivalent. In our situation, the most natural one is the so-called *cylindrical Brownian motion* : for any $h \in H$, a real valued centered Brownian motion $B_t(h)$ of volatility $||h||^2_H$ is given[12].

Intuitively, if $B_t \in H$, then:

$$B_t(h) = h.B_t$$

In fact, it is not difficult to see that, if H is infinite dimensional, B_t cannot belong to H for every t (see sect. 0.7) and that:

$$\mathbf{E}\left[||B_t||^2_H\right] = \infty$$

The process B_t will have to lie in a super-space[13] of H in which H will be dense.

In our situation, bond prices, swap rates, etc., and, generally speaking, portfolios containing the assets described above, are of the type $\mu(\mathbf{z}_t)$, where μ is a linear form on H^z . This corresponds to a stochastic integral with respect to a Brownian motion $B_t(\mu)$.

Although they are Hilbert spaces, we shall from now on make a distinction between the space $H = H^z$ of curves[14] and its dual, the space H^* of portfolios, to which the linear forms μ belong.

Of course, in order to define a cylindrical Brownian motion, we need to know $h.B_t$ for *every* $h \in H$ (or equivalently every $\mu \in H^*$). Here, only *some* μ are given, but we shall assume that they span a dense subspace, so that an entire cylindrical Brownian motion can be uniquely defined by extension.

We shall also assume that the price process of any static portfolio is an Ito process, in particular, it has a finite variance.

Remark 5 *The linear forms of the above type (sect. 0.3.4) contain Dirac masses. Hence they do not belong to the space L^2 but to the Sobolev space of distributions H^{-1} (or even to $H^{-\frac{1}{2}-\varepsilon}$, $\varepsilon > 0$)[15].*

We shall now give a more formal definition to these two assumptions. We fix $s > \frac{1}{2}$ and we assume that $H = H^s$ and hence, that $H^* = H^{-s}$. The choice of the regularity parameter s will be discussed in sect. 0.9.

---

[12]i.e., its variance at date t is $t\,||h||^2_H$.

[13]i.e. a space containing H .

[14]We drop the superscript z to ease notations. If a process is defined in H^z, we get the corresponding processes in H^y and H^f by applying formulas (2) and (3).

[15]When s is a positive integer, H^s is the space of functions whose s-th derivative belongs to L^2 and H^{-s} is the dual of H^s for the L^2 dot product. For non integer s, this space is defined by the mean of Fourier transform. If $s > \frac{1}{2}$ then functions lying in H^s are continuous, and Dirac masses belong to H^{-s}.

0.6 Assumptions

0.6.1 Almost Complete Market

ACM We assume that the set of traded assets is dense in H^* for the weak topology and that if the sequence μ_n of traded assets weakly tends to $\mu \in H^*$, then the price processes of assets μ_n converge in L^2.

For instance, if $H = H^s$, $s > \frac{1}{2}$, then the space of finite combinations of Dirac masses is dense in H^*.

0.6.2 Finite Variance

FV : For every $\mu \in H^*$, we assume that $\mu(\mathbf{z}_t)$ follows an Ito process driven by a Brownian motion $\tilde{B}_t(\mu)$:

$$d\mu(\mathbf{z}_t) = \mu\left(\mathbf{a}(t, \mathbf{z})\right) dt + dB_t(\mu) \tag{14}$$

$$dB_t(\mu) = \sigma_\mu(t, \mathbf{z}) \, d\tilde{B}_t(\mu) \qquad \mathbf{a}(t, \mathbf{y}) : x \longmapsto a(t, t + x, \mathbf{z})$$

In particular $\frac{1}{dt} \operatorname{Var} \mu(d\mathbf{z}_t) < \infty$ for any $\mu \in H^*$.

The relevance of these two hypotheses has been discussed at the end of the introduction.

0.6.3 Gaussian Rates

We shall also add the (less natural but convenient) assumption that volatility does not depend on the level of rates.

Gauss : $\operatorname{Var} dB_t(\mu)$ does not depends on the yield curve $\mathbf{y}$.

This implies that for any $\mu \in H^*$, the distribution of $\mu(\mathbf{y}_t)$ is Gaussian[16].

0.7 Principal Component Analysis

0.7.1 The Volatility Operator

On the space H^* of portfolios, there is a natural time dependent bilinear form $\tilde{Q}_t$ induced by the cross-variation process of the stochastic part of two portfolios :

$$\tilde{Q}_t(\mu, \nu) = \frac{\langle dB_t(\mu), dB_t(\nu) \rangle}{dt}$$

---

[16] $\mu(\mathbf{y}_t)$ linearly depends on rates. If their distribution is Gaussian, then so is that of $\mu(\mathbf{y}_t)$ and it has a (very low) probability of becoming negative. But a bond the price of which is given by $\mu(\mathbf{z}_t)$, where μ is a positive measure, will always have a positive price.

According to the CV hypothesis, $\tilde{Q}$ only depends on μ and ν. We shall denote by Q_t the quadratic form associated with $\tilde{Q}_t$:

$$Q_t(\mu) = \frac{\langle dB_t(\mu) \rangle}{dt}$$

Obviously, the quadratic form Q is positive. If its rank is finite, then we find the usual Gaussian H.J.M. model with a finite number of factors. On the contrary, we shall assume that it is non degenerate (any portfolio moves, even slightly, none is rigorously hedged). This assumption allows us to consider Q as a new *norm*[17] on H^*. When completing the space H^* with respect to this norm (we do not change notations), we get a cylindrical Brownian motion $B_t(\mu)$.

In such a situation, Yor [46, prop. I.4.2] shows that this motion can be realized : one can find a super-space[18] V of the dual H of H^* and a process B_t with values in V, almost surely continuous for the norm of V, such that :

$$\mathbf{E}[\,||B_t||_V^2\,] < \infty$$

and

$$\mu.dB_t = \mu(d\mathbf{y}_t - \mathbf{b}(t, \mathbf{y}_t)dt)$$

for any $t > 0$ and any $\mu \in V^* \subset H^*$.

It should be reminded that V is the space where lie the *yield curves* $\mathbf{y}_t$ and H^* that of *linear forms* (or portfolios) μ, and that:

$$V^* \subset H^* \qquad\qquad H \subset V$$

We shall see in sect. 0.9 the variety of possible spaces V.

It is now possible, in a rigorous mathematical language, to say that the random process $\mathbf{y}_t$ is determined by the following stochastic differential equation, driven by the cylindrical Brownian motion B_t :

$$d\mathbf{y}_t = \mathbf{b}(t, \mathbf{y}_t)\, dt + dB_t$$

Remark 6 *There are different Hilbert spaces, with different dot products. If $\mu \in V^*$ is a linear form, thus a measure[19] on the time interval $[0, T_{\max}]$*

---

[17] If Q is degenerate, then one may slightly modify it to get a new norm on H^* :

$$||\mu||^2 = \varepsilon||\mu_1||_{L^2}^2 + Q(\mu_2) \qquad \mu = \mu_1 + \mu_2 \qquad \mu_1 \in \ker Q \qquad h_2 \in \ker Q^\perp \qquad \varepsilon > 0$$

[18] Note that if V is a super space of H, then its norm is dominated by that of H :

$$||h||_V \leq \text{cst } ||h||_H$$

for any vector $h \in H$ should be able to be measured in V (the inclusion $H \hookrightarrow V$ is continuous).

[19] Riesz representation theorem.

that has a density χ_μ (possibly generalized, that is, with Dirac masses), and if $B_t \in V$, then one has :

$$(\mu.B_t)_{V^* \times V} = \int_0^M \chi_\mu(x) \, B_t(x) \, dx$$

However, H^* (hence also H) is a space which is specifically adapted to the quadratic form Q, for it has been completed with respect to it.

0.7.2 Principal Component Analysis

We shall now look at Q_t as a quadratic form on V^*. There exists a positive symmetric operator A_t on V^* such that for any $\mu \in V^*$:

$$Q_t(\mu) = \mu . A_t \mu \qquad \text{(dot product in V^*)}$$

Proposition 7 *The operator A_t has a finite trace. In particular, it is compact.*

Proof. Let $(\mu_n)_{n \in \mathbb{N}}$ be an orthonormal basis of V^*. Then:

$$Q_t(\mu_n) = \frac{1}{dt} \mathbf{E}[\, (\mu_n . dB_t)^2 \,]$$

hence:

$$\begin{aligned}
\operatorname{Tr} Q_t &= \sum_{n=0}^{\infty} Q_t(\mu_n) \\[2mm]
&= \frac{1}{dt} \mathbf{E}[\sum_{n=0}^{\infty} (\mu_n . dB_t)^2] \\[2mm]
&= \frac{1}{dt} \mathbf{E}[\, \|dB_t\|_V^2 \,] \quad < \infty \quad \blacksquare
\end{aligned}$$

Corollary 8 *The operator A_t is diagonalizable in an orthonormal basis, because its spectrum is discrete.*

One can thus find an orthonormal basis $(\bar{\mu}_n)_{n \in \mathbb{N}}$ of V^* and a sequence of eigenvalues $(\lambda_n)_{n \in \mathbb{N}}$ such that, if $\mu \in V^*$ and $\mu = \sum a_n \mu_n$, then:

$$Q_t(\mu) = \sum_{n=0}^{\infty} \lambda_n \, a_n^2$$

Because A_t is positive with a finite trace, one has:

$$\forall n \in \mathbb{N}, \quad \lambda_n > 0 \qquad \text{and} \qquad \sum_{n=0}^{\infty} \lambda_n < \infty$$

WE assume that the eigenvalues λ_n are decreasingly ordered : $\lambda_{n+1} \leq \lambda_n$ (possibly repeated if they are multiple).

We go back to the space V of yield curves and denote by $(\psi_n)_{n \in \mathbb{N}}$ the dual basis[20] of $(\mu_n)_{n \in \mathbb{N}}$. By definition :

$$dB_t(x) = \sum_{n=0}^{\infty} dv_n(t)\, \psi_n(x)$$

where

$$dv_n(t) = \mu_n \,.\, dB_t$$

Consequently :

$$d\,\langle B_t(x)\,,\, B_t(x') \rangle = dt \sum_{n=0}^{\infty} \lambda_n\, \psi_n(x)\, \psi_n(x')$$

Definition 9 *The functions*

$$\varphi_n = \sqrt{\lambda_n}\, \psi_n$$

will be called eigenmodes, *or* principal deformations *of the yield curve* $\mathbf{y}_t$.

0.7.3 Infinite Dimensional H.J.M. Representation

The covariance of two zero-coupon rates is given by the formula :

$$\langle d\mathbf{y}_t(x)\,,\, d\,\mathbf{y}_t(x') \rangle = dt \sum_{n=0}^{\infty} \varphi_n(x)\, \varphi_n(x')$$

When setting $x = x'$, we get:

$$\sigma(t, t+x)^2 = \sum_{n=0}^{\infty} \varphi_n(x)^2 \tag{15}$$

Let:

$$w_n(t) = \frac{1}{\sqrt{\lambda_n}}\, v_n(t)$$

The w_n are independent standard Brownian motions (i.e. with volatility 1) and :

$$d\mathbf{y}_t(x) = \frac{1}{x}\left(f(t, t+x) - r(t) + \frac{1}{2}\sum_{n=0}^{\infty} \varphi_n(x)^2 \right) + \sum_{n=0}^{\infty} \varphi_n(x)\, dw_n(t) \tag{16}$$

Under this form, we clearly see the P.C.A. of the yield curve process.

---

[20] $\mu_n(\psi_p) = 1$ if $n = p,$ 0 otherwise.

Multiplying this equation by x then deriving it with respect to x yields the Brace-Musiela equation on forward spot rates, generalized to an infinite summation. Let:

$$\mathbf{f}_t(x) = f(t, t+x) = \mathbf{y}_t(x) + x\,\frac{d\mathbf{y}_t}{dx}(x) \qquad\qquad \mathbf{f}'_t(x) = \frac{d\mathbf{f}_t}{dx}(x) = \frac{\partial f}{\partial T}(t, t+x)$$

$$\Phi_n(x) = x\,\varphi_n(x) \qquad\qquad \varphi_n^f(x) = \frac{d\Phi_n}{dx}(x) = \varphi_n(x) + x\,\frac{d\varphi_n}{dx}(x)$$

Then we have (see Brace-Musiela [7]):

$$d\mathbf{f}_t(x) = \left(\mathbf{f}'_t(x) + \sum_{n=0}^{\infty} \Phi_n(x)\,\varphi_n^f(x) \right) dt + \sum_{n=0}^{\infty} \varphi_n^f(x)\,dw_n(t) \qquad (17)$$

From these equations, we deduce the initial Brownian motions W^T and W^{fT} :

$$dW^T = \frac{1}{\sigma(x)} \sum_{n=0}^{\infty} \varphi_n(x)\,dw_n(t) \qquad\qquad dW^{fT} = \frac{1}{\sigma^f(x)} \sum_{n=0}^{\infty} \varphi_n^f(x)\,dw_n(t)$$

with

$$\sigma^f(x)^2 = \sum_{n=0}^{\infty} \varphi_n^f(x)^2 \qquad\qquad \sigma(x)\,\sigma^f(x)\,\mathrm{Corr}\left[dW^T, dW^{fT}\right] = \sum_{n=0}^{\infty} \varphi_n(x)\,\varphi_n^f(x)$$

Remark 10 *Notice the double orthogonality:*

1. *The eigenmodes φ_n are orthogonal in V ,*

2. *The Brownian motions w_n are independent.*

Remark 11 *The P.C.A. (the eigenmodes, etc.) depends on the space V , that is in fact on its norm, which, according to Da Prato-Zapczyk [15], can be any norm such that Q has a finite trace in V (see sect. 0.9).*

0.8 Optimal representation with an N-factor model

We fix an integer N and we define $\mathbf{y}_t^N \in V$ by $\mathbf{y}_0^N = \mathbf{y}_0$ and by the stochastic differential equation:

$$d\mathbf{y}_t^N(x) = \frac{1}{x}\left(\mathbf{f}_t^N(x) - \mathbf{y}_t^N(0) + \frac{1}{2}\sum_{n=0}^{N-1} \varphi_n(x)^2 \right) + \sum_{n=0}^{N-1} \varphi_n(x)\,dw_n(t)$$

Theorem 12 *The curve* $\mathbf{y}_t^N(x)$ *is the best approximation of* $\mathbf{y}_t(x)$ *in the sense of the norm of V the evolution of which is described by N Brownian motions. More precisely, if $\mathbf{u}_t$ is the solution of an SDE with values in V, driven by N real valued Brownian motions, then*[21]:

$$\mathbf{E}[\,||d\mathbf{y}_t - d\mathbf{u}_t||_V^2] \geq \mathbf{E}[\,||d\mathbf{y}_t - d\mathbf{y}_t^N||_V^2] = dt \sum_{n=N}^{\infty} \lambda_n$$

and

$$\max_{||\mu||_{V^*}=1} \operatorname{Var}\left[\mu.(d\mathbf{y}_t - d\mathbf{u}_t)\right] \geq \max_{||\mu||_{V^*}=1} \operatorname{Var}\left[\mu.(d\mathbf{y}_t - d\mathbf{y}_t^N)\right] = \lambda_N \, dt$$

Proof. The first thing to do is to characterize the linear combinations of N Brownian motions. A one-dimensional Brownian motion in V is a process:

$$(\omega, t) \in \Omega \times \mathbb{R}_+ \longmapsto \theta_t(\omega) = X_t(\omega)\,\xi$$

where Ω is the space of randomness, $X_t(\omega)$ is a real valued Brownian motion and $\xi \in V$.

We fix a date t and consider a time interval $\delta t > 0$ that will later tend to 0. One can see that $\delta\theta = \theta_{t+\delta t} - \theta_t$ as an element of a tensor product:

$$\delta\theta = \delta X \otimes \xi \in L^2(\Omega) \otimes V$$

Let

$$\Theta = L^2(\Omega)\hat{\otimes}V$$

be the *Hilbert-Schmidt completion* of this tensor product. If $(e_n)_{n\in\mathbb{N}}$ is an orthonormal basis of $L^2(\Omega)$ and if $(f_n)_{n\in\mathbb{N}}$ is one of V, then $L^2(\Omega) \otimes V$ is endowed with the norm for which $(e_p \otimes f_q)_{(p,q)\in\mathbb{N}^2}$ is an orthonormal basis (this norm does not depend on the chosen bases), and completed with respect to this norm[22]. Let $\zeta \in \Theta$ be a *random curve,* then:

$$||\zeta||_\Theta^2 = \mathbf{E}[\,||\zeta||_V^2]$$

In particular, $\delta B \in \Theta$ and one has:

$$||\delta B||_\Theta^2 = \sum_{n=0}^{\infty} \operatorname{Var}\left[\delta w_n\right] ||\varphi_n||_V^2 = \delta t \sum_{n=0}^{\infty} \lambda_n$$

---

[21] For a random process $X(t)$ in $\mathbb{R}$, with a Meyer decomposition $X = \langle X \rangle + \tilde{X}$ into a process with finite variation and a martingale, we set:

$$\mathbf{E}[dX] = d\langle X \rangle \qquad \operatorname{Var} dX = d\left\langle \tilde{X}^2 \right\rangle \qquad dX^{\text{stoch.}} = d\tilde{X}$$

[22] For instance, when $V = L^2(I)$, then the elements of the tensor product $L^2(\Omega) \otimes V$ are functions defined on $\Omega \times I$ and the Hilbert-Schmidt completion is nothing else but $L^2(\Omega \times I)$.

Denote by $\Gamma_N \subset \Theta$ the cone of N-tensors, that is the sum of N simple tensors:

$$\Gamma_N = \left\{ \sum_{i=1}^{N} X_i \otimes \xi_i \mid X_i \in L^2(\Omega) \,,\ \xi_i \in V \,,\ i = 1, \cdots, N \right\}$$

In order to show the first inequality, it is enough to prove that, for any sufficiently small $\delta t > 0$, the distance between δB and the set Γ_N is reached at $\delta B^N = B^N(t + \delta t) - B^N(t)$, where B^N is the Brownian motion defined by:

$$B^N(t) = \sum_{n=0}^{N-1} \varphi_n(x)\, w_n(t)$$

Indeed, by definition:

$$\frac{1}{dt} \mathbf{E}[\,||d\mathbf{y}_t - d\mathbf{u}_t||_V^2] = \lim_{\delta t \to 0} \frac{1}{\delta t} \mathbf{E}[\,||\delta \mathbf{y}_t^{\text{stoch.}} - \delta \mathbf{u}_t^{\text{stoch.}}||_V^2]$$

$$\frac{1}{dt} \mathbf{E}[\,||d\mathbf{y}_t - d\mathbf{y}_t^N||_V^2] = \lim_{\delta t \to 0} \frac{1}{\delta t} \mathbf{E}[\,||\delta \mathbf{y}_t^{\text{stoch.}} - \delta \mathbf{y}_t^{N\,\text{stoch.}}||_V^2]$$

To show that δB^N is the closest point of Γ_N to δB, we identify elements of Θ with linear operators from $L^2(\Omega)$ to V by setting:

$$(X \otimes \xi)\,.\,Y = \operatorname{Cov}[X, Y]\,\xi$$

The Hilbert-Schmidt norm is then given by:

$$||u||_\Theta^2 = \operatorname{Tr}\left({}^t u\, u\right)$$

($^t u$ is the transposed of u), and Γ_N is made of operators whose rank is less than or equal to N.

Lemma 13 *Assume that there exists $\bar{u} \in \Gamma_N$ such that:*

$$||\delta B - \bar{u}||_\Theta = \operatorname{dist}(\delta B\,,\,\Gamma_N)$$

Then the image $\operatorname{Im}\bar{u}$ is stable under $\delta B\,{}^t \delta B$.

Corollary 14 *$\operatorname{Im}\bar{u}$ is spanned by eigenvectors of $\delta B\,{}^t \delta B$ (that is, the φ_n or linear combinations between φ_n's corresponding to the same eigenvalue if it is multiple).*

Proof of lemma. We know that $\delta B - \bar{u}$ is orthogonal in $\bar{u}$ to Γ_N (with respect to the dot product in Θ). For any endomorphism α of $L^2(\Omega)$, the rank of $\bar{u} + \bar{u}\alpha$ remains bounded by N, thus:

$$\operatorname{Tr}\left({}^t\alpha\,{}^t\bar{u}\,(\delta B - \bar{u})\right) = 0$$

and

$$^t\bar{u}\,(\delta B - \bar{u}) = 0$$

Symmetrically, using endomorphisms of V, we see that:

$$\bar{u}\,(^t\delta B - {}^t\bar{u}) = 0$$

Combining these two identities, we get:

$$\delta B\,{}^t\delta B\,\bar{u} = \bar{u}\,\delta B\,{}^t\delta B = \bar{u}\,{}^t\bar{u}\,\bar{u}$$

The lemma follows. ∎

   End of proof. Going back to the tensor product, if $\bar{u}$ exists, it can be written as:

$$\bar{u} = \sum_{i=1}^{N} X_i \otimes \varphi_{n_i}$$

(if some eigenvalues are multiple, one might have to change the corresponding φ_n into another orthonormal basis of the eigenspace; this does not affects the decomposition of δB). Let:

$$J = \{n_1, \cdots, n_N\}$$

One has

$$\delta B - \bar{u} = \sum_{i=1}^{N}(w_{n_i} - X_i) \otimes \varphi_{n_i} + \sum_{n \notin J} w_n \otimes \varphi_n$$

and

$$||\delta B - \bar{u}||_\Theta^2 \geq \sum_{n \notin J} \lambda_n \geq \sum_{n=N}^{\infty} \lambda_n = ||\delta B - \delta B^N||_\Theta^2$$

This would end the proof if we knew that $\bar{u}$ exists. It is the case if Θ is finite dimensional. Let q be an integer which, later, will tend to infinity. We set:

$$E_q = \text{Vect}(w_0, \cdots, w_q) \qquad\qquad V_q = \text{Vect}(\varphi_0, \cdots, \varphi_q)$$

and let

$$\pi_q : \quad L^2(\Omega) \quad \longrightarrow \quad E_q$$

$$\pi_q' : \qquad V \qquad \longrightarrow \quad V_q$$

be the orthogonal projections. it is easy to check that:

$$\pi_q \otimes \pi_q' : \qquad \Theta \qquad \longrightarrow \quad \Theta_q = E_q \otimes V_q$$

$$X \otimes \xi \quad \longmapsto \quad \pi_q(X) \otimes \pi_q'(\xi)$$

is the orthogonal projection of Θ onto Θ_q and that:

$$\Gamma_{N,q} = \pi_q \otimes \pi_q'\left(\Gamma_N\right) = \left\{\sum_{i=1}^{N} X_i \otimes \xi_i \mid X_i \in E_q \, , \ \xi_i \in V_q \, , \ i = 1, \cdots, N\right\}$$

Therefore, if $\bar{u}_q \in \Gamma_{N,q}$ realizes the distance from $\pi_q \otimes \pi_q'\left(\delta B\right)$ to $\Gamma_{N,q}$ then:

$$\mathrm{dist}(\delta B \, , \, \Gamma_N)^2 \geq ||\pi_q \otimes \pi_q'\left(\delta B\right) - \bar{u}_q||_{\Theta_q}^2 \geq \sum_{n=N}^{q} \lambda_n$$

(the first inequality comes from the fact that an orthogonal projection does not increases distances). This lower bound is valid for every q, hence:

$$\mathrm{dist}(\delta B \, , \, \Gamma_N)^2 \geq \sum_{n=N}^{\infty} \lambda_n = ||\delta B - \delta B^N||_{\Theta}^2$$

Second inequality. We notice that the quantity to minimize is the usual operator norm of the transpose of $\delta B - \bar{u}$, as an operator from V to $L^2(\Omega)$, that is:

$$\max_{||\zeta||_V = 1} ||\left({}^t\delta B - {}^t\bar{u}\right)(\zeta)||_{L^2(\Omega)}$$

As the rank of ${}^t\bar{u}$ is at most N, its kernel has a co-dimension greater than or equal to N and:

$$\ker {}^t\bar{u} \cap \mathrm{Vect}(\varphi_0, \cdots, \varphi_N) \neq \{0\}$$

Let ζ be an element of this intersection such that $||\zeta||_V = 1$. One has:

$$||\left({}^t\delta B - {}^t\bar{u}\right).\zeta||_{L^2(\Omega)} = ||{}^t\delta B.\zeta||_{L^2(\Omega)} \geq \lambda_N$$

When $\bar{u} = \delta B^N$ the equality is implied the orthogonality of the φ_n. ∎

0.9   Possible choice in the Hilbert space V

In the statement of the previous theorem, we mentioned: *in the sense of the norm of V*. Indeed, Yor's construction leaves out some latitude on the choice of this space. According to Da Prato-Zapczyk [15], V^* can be any subspace of H^* provided with a Hilbert norm with respect to which Q has a finite trace. Then V is the super-space of H which is the dual of V^*. In particular, if V fits, any super-space of V fits. Nevertheless, a smallest acceptable space does not exist.

Remark 15 *In all these kinds of considerations, defining such or such space where such object lies is somewhat abstract, for, in the reality, these objects are finite dimensional and lie in any reasonable space. The very "physical" meaning of this type of statement relies in the evaluation of*

the corresponding norms. It has a meaning to say that such norm has a reasonable value or is extremely large, that such measurement of an error is bounded, while we have no clue of another evaluation. We understand the previous analysis in this context.

The linear forms μ which intervene when differentiating the prices of bonds and swaps are all of the kind *"integral over an interval + Dirac mass"*[23], thus their principal singularity is a Dirac mass. In other term, the price of a zero-coupon is well defined. We deduce that V is always contained in $H^{\frac{1}{2}}$ (since $V^* \supset H^{-\frac{1}{2}}$, see sect. 0.5, rem. 5, note 15). In countries in which a large number of futures is traded, like the United States (over a range of 10 years), we know better: a future contract evaluates in fact a forward short rate (3 months), that is an approximation of the derivative of the zero-coupon rate. As these also follow an Ito process with finite variance, we conclude that $V \subset H^{\frac{3}{2}}$. Yet one may choose any wider space.

The reality of markets is, on the contrary, oriented towards more smoothness: an tight analysis on US future curve shows that the $H^{\frac{3}{2}}$-norm of this curve, that is the $H^{\frac{5}{2}}$-norm of the yield curve, is almost always bounded (with however a slight difficulty due to the tic discretization and to a regular shift on the value of December contracts). A similar observation can be made on the French data series on OAT and BTAN bonds.

Remark 16 *The choice of the V-norm should be made carefully, in particular according to the profile of one's portfolio, indeed, as we already said, the P.C.A. is optimal with respect to this norm, and eigenmodes depend on its choice.*

In practice, we shall minimize a *least square criterion,* possibly weighted, on the prices of assets we are dealing. this criterion provides a quadratic form on the space of yield curve, which is the most natural choice as a norm for V. For instance, assume that we are dealing a series of bonds $B_1, \ldots, B_n$ the price of which is, at the first order, approximated by the measures $\mu_1, \ldots, \mu_n$ and that the least square criterion weights the bond B_i with a coefficient α_i (to take into account an unequal distribution of the portfolio). The norm on V can be set to:

$$||\mathbf{y}||_V^2 = \sum_{i=1}^{n} \alpha_i \, \mu_i(\mathbf{y})^2$$

In fact, this can be only a semi-norm (it may vanish for $\mathbf{y} \neq 0$). If the number of bonds is sufficient, and if their duration is well distributed, such a drawback will be avoided. Otherwise, one has to combine this sum with

---

[23] Here, we see that we are again more concerned with the general profile of μ, rather than with details like knowing wether the distribution of coupons is continuous or discrete. This really makes a small difference in their value.

an L^2-like norm, directly on zero-coupon rates:

$$||\mathbf{y}||_V^2 = \int_0^M \mathbf{y}(x)^2 \, m(x) \, dx$$

the weight $m(x) > 0$ being again adapted to the portfolio profile.

When dealing with futures, calendar spreads, etc., one should rather choose a Sobolev norm, that is a norm (still of Hilbert type) involving the derivative $\mathbf{y}'(x) = \dfrac{d\mathbf{y}}{dx}(x)$.

0.10 Option pricing

Jamshidian's [29] and Brace-Musiela's [7] formulae can easily be generalized to an infinite number of factors. The results match those of Kennedy [30]. We give in this section the expectation and the variance of any zero-coupon, as well as the covariance of any pair of such. In our model, rates are Gaussian and the zero-coupons have a log-normal distribution. Therefore, these data are sufficient to evaluate the price of any plain vanilla option (put or call) on any portfolio which is a linear combination of zero-coupons. This includes caplets, floorlets, options on bonds, swaps, and even on Forward Rate Agreements (options on "yield curve spreads").

Our model being a limit of Gaussian H.J.M. models with a finite number of factors, expectations, variances and covariances provided the N-factor H.J.M. model (see [7]) tend, when N tends to infinity, to a limit which corresponds to the model driven by the cylindrical Brownian motion B_t.

The option prices computed this way are of course arbitrage prices (provided the model fits the reality), but there is a little difficulty. Assume that, in the reality, interest rates satisfy the diffusion equation (16). If we try to hedge a cap against N modes of deformation using an approximation of the reality by an N-factor model, we get a price C_N and, as the hedge is not perfect, also a variance v_N. When N tends to infinity, v_N tends to 0 and the price C_N has a limit C, which is the price we propose. However, although there is a theoretically infinite number of hedging instruments P_i, $i = 1, 2, \ldots$ the N-factor model will use only N of them to cancel N hedge ratios Δ_i^N, $i = 1, \ldots, N$. When N tends to infinity, the hedge ratios tend to a well defined limit Δ_i, $i = 1, \ldots$ but it may happen that :

$$\sum_{i=1}^{\infty} |\Delta_i| \, P_i = +\infty$$

while, because of high correlations (due to the fact that the variance of Q is finite), the management cost (theoretical, that is transaction cost free) of this "infinite" portfolio remains finite. In practice, "infinite" means a

prohibitive high value. Besides, the presence of transaction costs makes a rigorous replication strategy impossible (but this remark is valid even for an option on a single asset).

Equations (7), then (15) and (16) provide the diffusion of logarithms of "forward zero-coupon" $z^F(t,T,T') = z(t,T')/z(t,T)$ when T and T' are fixed :

$$d \log z^F(t,T,T') = \frac{1}{2} \sum_{n=0}^{\infty} \left((T-t)^2 \varphi_n (T-t)^2 - (T'-t)^2 \varphi_n (T'-t)^2 \right) dt$$

$$+ \sum_{n=0}^{\infty} \left((T-t)\, \varphi_n(T-t) - (T'-t)\, \varphi_n(T'-t) \right) dw_n$$

The first series converges absolutely, while the second one converges absolutely as a function of t with values in $L^2(\Omega)$. If we set $\delta T = T' - t$ and $\Phi_n(x) = x\, \varphi_n(x)$, we get:

$$\mathbf{E}\left[z(T, T+\delta T)\,|\,t \right] = \frac{z(t, T+\delta T)}{z(t,T)} \exp\left(\sum_{n=0}^{\infty} \int_0^{T-t} \left(\Phi_n(s)\, \Phi_n(s+\delta T) - \Phi_n(s+\delta T)^2 \right) ds \right)$$

$$(18)$$

$$\text{Var}\left[\log z(T, T+\delta T)\,|\,t \right] = \sum_{n=0}^{\infty} \int_0^{T-t} \left(\Phi_n(s) - \Phi_n(s+\delta T) \right)^2 ds \qquad (19)$$

$$\text{Cov}\left[\log z(T, T+\delta T_1), \log z(T, T+\delta T_2)\,|\,t \right] =$$

$$\sum_{n=0}^{\infty} \int_0^{T-t} \left(\Phi_n(s) - \Phi_n(s+\delta T_1) \right) \left(\Phi_n(s) - \Phi_n(s+\delta T_2) \right) ds \qquad (20)$$

By truncating these summations at rank $N-1$, we obtain the expectations, variances and covariances of zero-coupons as if the yield curve were the approximate one $\mathbf{y}_t^N$.

0.11 Computation of eigenmodes

0.11.1 Reconstruction and smoothing of the yield curve

In order to perform the P.C.A. of the yield curve out of historical data series, we first need to restrict ourselves to the finite dimension through a finite element method. We shall thus approximate the yield curve by a

function depending on a finite number n of parameters: polynomial, spline, piecewise linear, etc. The main point is that, because of the usually high heteroskedasticity, one needs a fixed type of approximation, and not an approximation that depends itself on historical data. This second kind of simplification will namely be provided by the P.C.A. we are going to undertake.

Let $(\mathcal{E}^n)_{n \in \mathbb{N}}$ be a *Galerkin decomposition* of V. Each $\mathcal{E}^n$ is an n-dimensional subspace of V contained in the next one $\mathcal{E}^{n+1}$ and the union of all the $\mathcal{E}^n$ is dense in V. By the Schmidt orthonormalization procedure[24], one can find an orthonormal basis $(L_n)_{n \in \mathbb{N}}$ of V adapted to this decomposition: for any n, $(L_1, \ldots, L_n)$ is a basis of $\mathcal{E}^n$. The subspaces $\mathcal{E}^n$ are endowed with the same norm as V.

From now on, the dimension n is fixed. In practice, if we consider the bid/offer spread as a limit for precision, then most of the time, one can find an acceptable 6 to 8 dimensional Galerkin subspace. Each yield curve $\mathbf{y}$ will then be approximated by its orthogonal projection $\mathbf{y}^n$ onto the subspace $\mathcal{E}^n$. As the basis $(L_1, \ldots, L_n)$ is orthonormal, the approximate (or smoothen, see remark 17 below) curve is given by the simple formula:

$$\mathbf{y}^n = \sum_{i=1}^{n} (\mathbf{y}.L_i)_V \, L_i$$

This allows to identify the movement of the yield curve $\mathbf{y}_t$ with an n-dimensional random process:

$$\vec{a}(t) = ((\mathbf{y}_t.L_1)_V, \ldots, (\mathbf{y}_t.L_n)_V) \in \mathbb{R}^n$$

An important issue is that the norm and dot product of the space V should be easily computable out of explicit available data.

Remark 17 *If elements of $\mathcal{E}^n$ are smooth yield curves for any n, then the approximation of a curve $\mathbf{y}$ by an element $\mathbf{y}^n \in \mathcal{E}^n$ is by construction a smoothing of the yield curve.*

0.11.2 Eigenmode computation from the historical series

After this first dimension reduction we made, we shall only be able to compute the first n eigenmodes. Moreover, the last ones will lack precision. Nevertheless, this does not rise a big problem, because most of the time, the quasi-totality of the variance of the motion (more than 99%) is borne by the first three modes. Anyway, we shall have to leave out some variance

---

[24] If $(J_1, \ldots)$ is a basis adapted to the Galerkin decomposition, that is, $(J_1, \ldots, J_n)$ spans $\mathcal{E}^n$, but not necessarily orthonormal, the orthonormal $(L_1, \ldots)$ basis is built by first normalising J_1, then moving J_2 parallel to J_1 to make it orthogonal, and normalising, and so forth.

because of the impossibility to perform a strict time continuous hedge. Our aim will therefore to compute only the first three or four eigenmodes.

The previous section shows how to identify the yield curve with a vector evolving in $\mathbb{R}^n$:

$$\vec{a}(t) = (a_1(t), \cdots, a_n(t)) \in \mathbb{R}^n$$

Remark 18 *The identification:*

$$\mathbf{y}^n = \sum_{i=1}^n a_i L_i \in \mathcal{E}^n \longmapsto \vec{a} = (a_1, \cdots, a_n) \in \mathbb{R}^n$$

is an isometry when $\mathcal{E}^n$ is provided with the norm $\| \; \|_V$ and $\mathbb{R}^n$ with its usual Euclidean norm, for base functions L_i are orthonormal with respect to $\| \; \|_V$. Therefore, eigenmodes in $\mathcal{E}^n$ and in $\mathbb{R}^n$ are identical.

We now fix $\delta t > 0$. When t varies, the vectors:

$$\vec{a}'(t) = \frac{1}{\sqrt{\delta t}} \left(\vec{a}(t + \delta t) - \vec{a}(t) \right)$$

form a cloud of points in $\mathbb{R}^n$ the principal axes of which are the "historical" eigenmodes. Indeed, let π_n be the orthogonal projection of V onto $\mathcal{E}^n$. The quadratic form Q^n defined on the dual $\mathcal{E}^{n*}$ of $\mathcal{E}^n$ by:

$$Q^n(\mu) = \frac{1}{\delta t} \; \mathrm{Var} \left[\mu(\mathbf{y}^n_{t+\delta t}) \,|\, t \right]$$

is obtained from Q by the transpose of the projection π_n :

$$Q^n = Q \circ {}^t\pi_n \qquad {}^t\pi_n : \mathcal{E}^{n*} \longrightarrow V^*$$

It is a well known result that, in this situation, the eigenspaces of Q^n (that is its P.C.A.) tend, in the weak sense (that is index wise[25]), to those of Q . In practice, taking $n = 7$ or 8 gives a very good approximation of the first four modes.

Definition 19 *The matrix S of Q^n in the basis $(L_1, \cdots, L_n)$ is called the* covariance matrix *of the process $\vec{a}(t)$ (or $\mathbf{y}_t$). It is defined by:*

$$S_{ij} = \frac{1}{\delta t} \; \mathrm{Cov} \left[\mathbf{y}^n_{t+\delta t}.L_i \, , \, \mathbf{y}^n_{t+\delta t}.L_j \,|\, t \right]$$

---

[25] Let $(\lambda_1, \ldots)$ be the eigenvalues of Q and $(\lambda_1^n, \ldots, \lambda_n^n)$ be those of Q^n . For fixed k , then λ_k^n tends to λ_k as n tends to ∞ and, if λ_k is not multiple, then the corresponding eigenvector φ_k^n tends to φ_k . When there is some multiplicity, then the whole eigenspace corresponding to λ_k^n tends to that corresponding to λ_k (they have the same dimension if n is sufficiently large).

Because we took an orthonormal basis, this matrix represents the quadratic form Q^n and its diagonalization provides the eigenmodes: if $\vec{u} = (u_1, \ldots, u_n)$ is an eigenvector of S associated with the eigenvalue λ then $\sum u_i L_i$ is the eigenmode (in the $\mathcal{E}^n$ approximation) associated with the same eigenvalue λ.

Remark 20 *This is a purely historical evaluation of the covariance matrix and of the eigenmodes. If one is concerned with Vega hedging, he should rather try to perform an implicit evaluation of the factors out of the market prices of options, or mix the two methods.*

0.12   Reduction of the dimension

The previous analysis provides two opportunities to reduce the dimension of the overall space of yield curves. The first one relies in the projection onto the Galerkin subspace $\mathcal{E}^n$. It corresponds, as we said in remark 17, to smoothing the yield curve. This reduction should not depend on the movement of the yield curve. Indeed, the hedges we are going to compute do not take into account the errors made at this step, hence only a serious statistical analysis can insure that these errors are bounded in any market state, even catastrophic.

The second reduction is performed after the principal component analysis of the move of the approximated curve $\mathbf{y}_t^n$. Once principal deformations have been determined, we just keep the first d of them, $d = 2$ or 3. This way, we get an H.J.M. model with d factors. The space to which the curve belongs is still $\mathcal{E}^n$ but, infinitesimally, there are only d types of possible deformations. Nevertheless, as it is very unlikely that these deformations form the basis of a Markovian model[26], the movement, even reduced to d sources of randomness, may explore the whole space $\mathcal{E}^n$. Consequently, its complexity (for instance to price a swaption) is the dimension n and not d (keeping only d factors still simplifies computations, but not as much as one could have hoped).

0.12.1   The drift term and the real option pricing

When projecting the whole yield curve movement on one of the Galerkin subspaces $\mathcal{E}^n$, another problem arises. The real statistical drift of the diffusion could belong to this subspace, or be also projected onto it. But in order to fulfill the AAO hypothesis, the risk-neutral drift is imposed and there is no reason why it should lie in $\mathcal{E}^n$. A first solution would be to

---

[26] For this, they should show an exponential or polynomial shape with respect to the maturity x (see [19]).

choose special subspaces, such that if a yield curve $\mathbf{y}^n \in \mathcal{E}^n$ then the corresponding drift $\mathbf{b}(t, \mathbf{y}^n)$ (see eqn. (14), sect. 0.5) also belongs to $\mathcal{E}^n$. This solution is that adopted by N. El Karoui in number of papers (see [11] and [19] to [20]), and by other authors, beginning with Vasiček [45]. Although we respect the relevance of this approach, which is well adapted to implicit evaluation of the volatility structure, our own experience shows that one would rather take spaces $\mathcal{E}^n$ that really fit well the data, whatever they look like (we give in sect. 0.12.3 examples of Galerkin spaces that appeared to be efficient).

We now have two solutions. Either we consider that the space $\mathcal{E}^n$ is here only to size the deformation factors, but we keep the entire risk-neutral drift, and the formulae (18) to (20) compute options on zero-coupons and on portfolios of such. This is the approach of Brace-Musiela [7] and Jamshidian [29]. For instance, an option on a 10 years quarterly swap leads to a 40-dimensional integral, hard to compute. One solution would be Monte-Carlo technique, or deterministic low discrepancy sequences (Sobol, etc.). Another is again to diagonalize the 40-dimensional covariance matrix of the zero-coupons involved in the swap, and to compute the integral only on the three dimensional subspace spanned by the eigenvectors corresponding to the three biggest eigenvalues, by a Gauss-Legendre interpolation.

Remark 21 *The faster an option evaluation, the better it allows to compute implicit deformation factors out of option market prices (see remark 20).*

The second solution is to give up the strict AAO assumption. Indeed, the theoretical arbitrages that one could achieve in such a setting are impossible to realize in practice because of transaction costs. In other words, even if the vector $\mathbf{b}(t, \mathbf{y}^n)$ does not belong to $\mathcal{E}^n$, it is so close to its projection $\mathbf{b}^n(t, \mathbf{y}^n) = \pi_n(\mathbf{b}(t, \mathbf{y}^n))$ that the difference cannot be made into a real free lunch. Therefore, it is possible to let the curve evolve with the "almost risk-neutral" drift $\mathbf{b}^n(t, \mathbf{y}^n)$ and compute as well vanilla options, as exotic ones (barriers, etc.) by Monte-Carlo techniques or PDE discretization inside the space $\mathcal{E}^n$.

0.12.2 Practical Option Hedging

Practical option hedging is always an optimization between costs and residual risks. In practice, dynamic hedging will never be able to totally offset risks, therefore some level of residual risk must be accepted. Moreover, in an optimal hedging strategy, the various sources of risk are comparable. There are mostly four sources of such residual risks:

- Discrete time dynamic hedging

- Uncertain volatility and correlations

- Shocks and non-diffusion processes

- Hedging only finitely many risk factors

Comparing the size of residual risk stemming from these sources leads to the optimal choice of number of risk factors to choose. In practice, one first assess the minimum amount of residual risk one cannot avoid by any dynamic hedging, then the level of acceptable residual risk with respect to the corresponding transaction costs. Finally, equally splitting this acceptable level of risk across the four sources above, implement the appropriate dynamic hedging strategy in order to achieve the targeted level of risk.

0.12.3 Difficulties

Galerkin space

The first difficulty is to find good Galerkin subspaces $\mathcal{E}^n$ in order to optimize the *computation/accuracy* ratio of the model. Let us mention the following series, with their advantages and drawbacks:

- *Polynomials of degree $n-1$*. Arbitrage free, but not performing: it cannot at the same time the variety of short term rates and the barely changing behavior of long term rates.

- *Decreasing exponential $e^{-\alpha_k x}$, $\alpha_k \geq 0$*. So-called generalized Vasiček (see [19]). Arbitrage free and better than the previous one. Good for implicit evaluation of factors, because of possibility of rather fast evaluation of swaptions.

- *Cubic splines* (piecewise third degree polynomials we C^2 fit at junction). Good for fitting the prices of assets with rather a small number of parameters, but not arbitrage free. Dimension n equals $3 + number$ *of splines*. Most common: three splines (see Turner [44]).

- *Polynomials of degree $n-1$ with a change of variable on the maturity*. Also non arbitrage free. One of the most efficient is the *Log* change for it fits the fast changing shape of the short term part and the very regular one of the long term part. This idea was first suggested by P. Gaye, and appears to have several theoretical justifications, one of them being that the yield curve and the curve of forward spot rates belong to the same space (the forward spot rate is the derivative of the zero-coupon rate with respect to the *Log* of the maturity). It is also better than cubic splines because it does not particularize any maturity and thus gives a "nicer" smoothing.

Instability of eigenmodes

In this first formulation, our model assumes that volatilities and correlations are constant. This fact is obviously denied by most statistical analysis. Studies we made on the French curve show that the plane spanned by the first two modes almost follows a Brownian path in the manifold of 2-dimensional planes of $\mathcal{E}^n$ for $n = 6$ (polynomials in Log). Short term options, or even European ones (options on futures, swaps, bonds, etc.) can afford a homoskedastic model, provided we don't mix the maturity of the various options, whereas caps and floors really need a heteroskedastic model (see Brace-Gatarek-Musiela [6]).

Similarly, it is of importance to detect "changes of regime", that is situations where we need to take into account a larger number of factors of a bigger Galerkin subspace to keep the "noise" under control.

Statistical evaluation of drift and volatility

Evaluating the diffusion coefficients (drifts, volatilities and correlations) of a multidimensional process, especially when data are not always of good quality, can easily become a challenge, see Genon-Catalot & Jacod [25]. One again needs to optimize the value of n in order to guaranty a correct estimation of the coefficients, as well as a good control on the noise.

ARCH and GARCH models should also be considered.

Note that the "tick" discretization introduces its own noise for large n, and without it, the move of prices inside the bid/offer spread is rather erratic.

Mixing historical and implicit data

Historical data are useful to price illiquid vanilla options (options on bonds, etc.), for we seldom have a Vega hedging and we need to forecast the behavior of volatility. For any other option (liquid, exotic when liquid vanilla exist, etc.) one need an implicit evaluation of the factors out of the market data on the prices of liquid options. The problem is that the number of parameters to estimate can be larger than the number of reliable data. This means that we necessarily need to mix implicit data with historical ones, through the optimization of some penalty function that weights both.

Incoherence between historical and implicit data, or even among implicit data, can sometimes give rise to (quasi) arbitrage opportunities, provided transaction costs allow to enter the setting up of such position.

0.13 References

[1] N. Anderson, F. Breedon, M. Deacon, A. Derry, G. Murphy *Estimating and Interpreting the Yield Curve* J. Wiley & sons, New York, 1996.

[2] L. Bachelier *Théorie de la spéculation* Ann. sci. de l'E.N.S. Paris nᵒ17, 1900, pp. 21-86.

[3] T. Björk, Yu. Kabanov, W. Runggaldier *Bond Market Structure in the Presence of Marked Point Processes* preprint Univ. of Padua, Italy, 1996.

[4] F. Black, E. Derman, W. Toy *A one-factor model of interest rates and its application to Treasury Bond options* Financial Analist Journal, Jan-Feb 1990, 33-39.

[5] J.-P. Bouchaud, N. Sagna, R. Cont, N. El Karoui, M. Potters *Phenomenology pf the Interest Rate Curve* Preprint SSRN, 1997

[6] A. Brace, D. Gatarek, M. Musiela *The market model of interest rate dynamics* Preprint Univ. of New South Wales, Sydney, 1995.

[7] A. Brace, M. Musiela *A multifactor Gauss-Markov implementation of Heath, Jarrow and Morton* Math. Finance, Vol. 4, nᵒ3, July 1994, pp. 259-283.

[8] F. Black, M. Scholes *The pricing of Options and Corporate Liabilities* J. of Polit. Economy, 81, 1973, pp. 635-654.

[9] D. Bricio-Hernandez *Lectures on Probability and Second Order Random Fields* World Scientific, Series on Adv. in Math. for Appl. Sci., Vol. 30, Singapore, 1995.

[10] D. Brigo and F. Mercurio *Interest Rate Models: Theory and Practice*, 2ⁿᵈ ed., Springer Finance, Heidelberg, 2006

[11] T. Cherif, N. El Karoui, R. Myneni, R. Viswanathan *Arbitrage pricing and hedging of quanto options and interest rate claims with quadratic Gaussian state variables* Preprint, Lab. of proba., Univ. of Paris VI, 1995.

[12] R. Cont *Modeling Term Structure Dynamics: An Infinite Dimensional Approach* Preprint Ecole Polytechnique, Paris, 2002

[13] J.C. Cox, J.E. Ingersoll, S.A. Ross *A theory of the term structure of interest rates* Econometrica, nᵒ53 (2), 1985.

[14] R.A. Dana, M. Jeanblanc-Piqué *Marchés financiers en temps continu* Economica, coll. Recherche en Gestion, Paris 1994. In English *Financial Markets in Continuous Time* Springer Finance, 2007

[15] G. Da Prato, J. Zapczyk *Stochastic equations in infinite dimensions* Cambridge University Press, 1992.

[16] R. Douady *A Fourier-log Analysis of US Euro-dollar Futures* Working paper, 1997.

[17] D. Duffie, R. Kan *A yield-factor model of interest rates* Preprint Stanford Univ., Sept. 1993.

[18] N. El Karoui, P. Durand *Quadratic Gaussian Model of Interest Rates and Quanto Options* Preprint, Labo. de proba. de Paris VI, 1997.

[19] N. El Karoui, V. Lacoste *Multifactor model of the term structure of interest rates* Preprint, Labo. de proba. de Paris VI, 1992.

[20] N. El Karoui, R. Myneni, R. Visnawathan *Arbitrage pricing and hedging of interest rate claim with state variables* Preprint, Labo. de proba. de Paris VI et Univ. de Stanford, février 1992.

[21] D; Filipovic *Consistency Problems for Heath-Jarrow-Morton Interest Rate Models* Springer Lecture Notes in Mathematics, vol 1760, 2004

[22] A. Frachot *Cours sur les taux d'intérêt* preprint ENSAE, 1993.

[23] A. Frachot, D. Janci, V. Lacoste *Factor analysis of the term structure: a probabilistic approach* Preprint, Labo. de la Banque de France, 1992.

[24] B. Gaveau *Intégrale stochastique radonifiante* C.R. Acad. des Sciences (Paris), t. 276, mai 1973.

[25] V. Genon-Catalot, J. Jacod *On the estimation of the diffusion coefficients for multidimensional diffusion processes* Annales de l'I.H.P., 1993, n°29, pp. 119-152.

[26] D. Heath, R. Jarrow, A. Morton *Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation* Econometrica, n°60, 1992, pp. 77-106.

[27] T.S. Ho, S.B. Lee *Term structure movements and pricing interest rate contingent claims* J. of Finance, 41, 1986, pp. 1011-1029.

[28] J. Hull, A. White *Pricing interest rate derivatives* The Review of Financial Studies, n°3, 1990, pp. 573-592.

[29] F. Jamshidian *Option and future evaluation with deterministic volatilities* Math. Finance, Vol. 3, n°2, April 1993, pp. 149-159.

[30] D.P. Kennedy *The term structure of interest rates as a Gaussian random field* Math. Finance, Vol. 4, n°3, July 1994, pp. 247-258.

[31] D. Lamberton, B. Lapeyre *Introduction au calcul stochastique appliqué à la finance* Ellipse, coll. Mathématiques et applications, n°9, Paris 1991.

[32] R. Litterman, J. Scheinkman *Common factors affecting bond returns* Technical Report 62, Goldman-Sachs Financial Strategies Group, septembre 1988.

[33] F. Longstaff, E Schwartz *Interest rate volatility and the term structure: a two factor general equilibrium model* Journal of Finance, n^o47, 1992, pp. 1259-1282.

[34] R. Merton *Theory of Rational Option Pricing* Bell Journal of Economics and Management Science, n^o4, 1973, pp. 141-183.

[35] R. Merton *Continuous Time Finance* Basil Blackwell, USA, 1991.

[36] K. Miltersen, K. Sandmann, D. Sondermann *Closed form solutions for the term structure derivatives with log-normal interest rates* Preprint Univ. Bonn, 1994.

[37] M. Musiela, D. Sondermann *Different Dynamical Specifications of the Term Structure of Interest Rates and their Implications* Preprint Univ. Bonn, 1994.

[38] M. Musiela, M. Rutkowski *Martingale Methods in Financial Modeling* 2nd ed., Springer, 2005

[39] M. Musiela, M. Rutkowski *Continuous-time Term Structure Models* Preprint Univ. New South Wales, 1996

[40] W. Press, S. Teukolsky, W. Vettering, B. Flannery *Numerical recipes in C* 2nd ed., Cambridge University Press, 1992.

[41] R. Rebonato *Interest-Rate Option Models* 2nd ed., J. Wiley & Sons, Chichester UK, 1998.

[42] R. Rebonato *Interest-Rate Term-Structure Pricing Models: a Review* Preprint, 2003

[43] S. Schaefer, E.S. Schwartz, *A two factor model of the term structure: An approximate analytical solution* J. of Finan. and Quant. Anal., 19, 1984, pp. 413-424.

[44] S.R.E. Turner *Modeling interest rates and assessing risk of derivative securities* Preprint, Univ. of Cambridge, Statistical Lab., 1993.

[45] O.A. Vasiček *An equilibrium characterisation of the term structure* J. of Finan. Economics, 5, 1977, pp. 177-188.

[46] M. Yor *Existence et unicité de diffusions à valeur dans un espace de Hilbert* Annales I.H.P., vol. X, n^o1, 1974, pp. 55-58.

[47] M. Yor *Sur les intégrales stochastiques à valeur dans un espace de Banach* Annales I.H.P., vol. X, n^o1, 1974, pp. 31-36.