Università "La Sapienza" di Roma
Facoltà di Scienze Matematiche, Fisiche e Naturali
Dottorato di Ricerca in Matematica (XV ciclo)

Modelling Interest Rates for Public Debt Management

Adamo Uboldi

Contents

Introduction

This Ph.D. thesis is devoted to the application of Stochastic Differential Equations (SDE) to problems arising in Interest Rates Modelling. The work is part of a large joint project of the IAC-CNR (Institute for the Applications of Calculus of the Italian National Research Council) and the Treasury Department of the Italian Economy and Finances Ministry.
The project aims to optimize the Public Debt Management: a brief survey on this complex problem is given in chapter 3. In particular, interest rates play a key role as one of the stochastic components of the corresponding optimal control problem.

A self contained introduction to the theoretical framework of SDE is presented in chapter 1. We report just those fundamental results that are required in the following chapters, like the Feynman-Kac representation or the Kolmogorov backward and forward equations. A detailed survey on the mathematical settings of Interest Rates Theory is presented in chapter 2. We discuss the standard models for the short rate like the Cox-Ingersoll-Ross (CIR) model and the Hull-White (HW) extension of the Vasicek model. We then present the Heath-Jarrow-Morton (HJM) framework for the forward rates.
Finally, we expose pricing techniques for interest rates derivatives like *caps&floors*, that will be fully developed later.

A first aim of the present work is to study models existing in mathematical and financial literature, in order to evaluate their applicability to specific interest rates markets. In chapter 4 we analyze the features of the CIR model and of the Nelson-Siegel (NS) model. We perform a statistical analysis and a comparison between the two different approaches, in order to give some suggestions to financial institutions interested in using these models to study the dynamics of interest rates.

A second part of the thesis is oriented to propose new techniques for the application of existing models. In particular, in chapter 5, by means of a combination of the two different techniques of time series analysis and cross-section analysis, we propose a new empirical evaluation of the market price of risk, the implied variable that is crucial in the study of

treasury bonds.
First results obtained by the implementation of the method suggest that this new technique is quite robust and sensitive to new information introduced in the market.
This approach is applied to the CIR model under the so called Local Expectation Hypothesis, in order to compare results with the "standard" approaches of chapter 4.
With the same aim, in chapter 6 we propose a new model of the HJM class which is a Gaussian extension of the infinite dimensional HW model. In this chapter we face the problems related to the new model from a practical viewpoint, selecting real benchmarks for interest rates from the markets. We perform a Principal Component Analysis, in order to investigate how many factors are really relevant in the fixed income securities market. We then compare our empirical results with the existing literature, and we close this part providing some observations and indications for future directions of research.

The last part of this work is devoted to develop from a theoretical point of view the new model introduced in chapter 6. In chapter 7 we prove that this model is consistent (according to the Björk and Christensen definition) with a polynomial-exponential extension of the NS family of curves. It must be noted that consistency has been in the last years one of the most important topics of discussion in interest theory, although the lack of practical applications up to now keep the empirical value of the whole theory not fully comprehended.
With a slightly different approach, in chapter 8 we show that the model admits a finite dimensional realization, and we prove a minimality result for our choice of the family of curves. In other words, our extension of the NS family of curves is minimal: no smaller family of curves is available in order to achieve the same goal of finite dimensional realization.

The new model belongs to the class of Gaussian HJM models, meaning that there is a deterministic volatility: the functional form of the volatility allows an humped shaped volatility, thus extending the HW approach which is monotone decreasing with respect to time. This particular shape is clearly suggested by real data on the volatility of Treasury bonds.
The main drawback is represented by the increasing of computational complexity and by the difficulty of symbolic calculations in order to obtain closed formulas for derivatives prices.

In chapter 9 we face one of the most important problem in Mathematical Finance, that is the pricing of derivative securities. We apply our new model to the pricing of *caps&floors*, the most important derivative product in fixed income markets. Closed for-

mulas for the implied volatility (the common way to price *caps&floors*) are developed by resorting to the technique of the change of numeraire.
An empirical analysis is thus performed with in-sample and out of sample tests: we show that interest rates market presents a very peculiar behavior, as the implied volatility of *caps&floors* heavily depends on the evolution of the underlying benchmark rates. This empirical result confirms that the fixed income market is very different from the widely studied stock markets, and it encourages the search for completely different mathematical approaches.

Our fit of the total variance is particularly precise, and the higher complexity of calculations is justified by this important empirical evidence. It is crucial to remark that although our choice of a deterministic volatility is restrictive from a theoretical point of view, empirical results oriented to forecast interest rates evolution seems to be very good.

Chapter 1

Introduction to SDE in Mathematical Finance

1.1 Stochastic Differential Equations

Stochastic Differential Equations (SDE) are a key tool in Mathematical Finance: we thereafter introduce the subject in a simplified way. We try to keep this introduction self-contained by showing only those parts of the whole theory that we will discuss in the next chapters.
Applications to Interest Rates Modelling are simply cited, and the wide mathematical machinery will be fully developed in chapter 2.
For preliminaries of Stochastic Calculus, such as Itô Formula, Semi-Martingales, Brownian Motion and Stochastic Integral, an intuitive approach can be found in [55], while for an exhaustive treatment see [46].
A good introduction oriented to Mathematical Finance is [48].

We set all this discussion on a filtered probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$: if X is a stochastic process such that we have $X(t) \in \mathcal{F}_t$ for all $t \geq 0$, then we say that X is adapted to the filtration $\{F_t\}_{t\geq 0}$. In general we will deal with $\mathcal{F}_t^X$, the filtration generated by a stochastic process X: in particular we will heavily use F_t^W, the internal filtration of the standard Brownian motion (or Wiener process) W_t.
Consider as given the following objects (where $M(n,d)$ denote the class of $n \times d$ matrices):

- A (column-vector valued) function $\mu : R_+ \times R^n \to R^n$.

- A function $\sigma : R_+ \times R^n \to M(n,d)$.

- A real (column) vector $x_0 \in R^n$.

We now want to investigate whether there exists a stochastic process X which satisfies the stochastic differential equation (SDE)

$$(1.1) \qquad \begin{cases} dX_t &= \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \\ \\ X_0 &= x_0. \end{cases}$$

From a formal point of view, the previous form is simply a shorthand notation: to be more precise we want to find a process X satisfying the integral equation

$$(1.2) \qquad X_t = x_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \text{for all } t \geq 0.$$

We must recall that in order to guarantee the existence of the stochastic integral in (1.2), the argument must belong to $\mathcal{L}^2[0,t] = \{Y|Y \text{ is } \mathcal{F}_t^W\text{-adapted and } \int_0^t (Y^2(s))ds < +\infty\}$.
The standard method for proving the existence of a solution to the SDE above is to construct an iteration scheme of Cauchy-Picard type, as for deterministic differential equations (for the deterministic case, a classic reference is [45]).
The idea is to define a sequence of processes $X^0, X^1, X^2, \ldots$ according to the recursive definition:

$$(1.3) \qquad \begin{cases} X_t^0 &\equiv x_0, \\ \\ X_t^{n+1} &= x_0 + \int_0^t \mu(s, X_s^n)ds + \int_0^t \sigma(s, X_s^n)dW_s. \end{cases}$$

One can "naturally" expect that the sequence $\{X^n\}_{n=l}^\infty$ will converge to some limiting process X, and that this X is a solution to the SDE. This construction can in fact be carried out, but the proof requires some inequalities, taking account of the stochastic nature of the problem.
We only give the result.

Proposition 1.1 *Suppose that there exists a constant K such that the following conditions are satisfied for all x, y and t:*

$$\begin{aligned} \|\mu(t,x) - \mu(t,y)\| &\leq K\|x - y\|, \\ \\ \|\sigma(t,x) - \sigma(t,y)\| &\leq K\|x - y\|, \\ \\ \|\mu(t,x)\| - \|\sigma(t,x)\| &\leq K(1 + \|x\|). \end{aligned}$$

Then there exists a unique solution to the SDE (1.1). The solution has these properties:

- *X is $\mathcal{F}_t^W$-adapted.*

- *X has continuous trajectories.*

- *X is a Markov process.*

- *There exists a constant C such that*

$$\mathbb{E}(\|X_t\|^2) \leq Ce^{Ct}(1 + \|x_0\|^2).$$

Roughly speaking, the fact that the solution X is $\mathcal{F}_t^W$-adapted means that for each fixed t the process value X_t is a functional of the Wiener trajectory on the interval $[0, t]$, and in this way a SDE induces a transformation of the space $C[0, +\infty)$ into itself, where a Wiener trajectory $W.(w)$ is mapped to the corresponding solution trajectory $X.(w)$.
In general this transformation, which takes a Wiener trajectory into the corresponding X-trajectory, is rather complicated and it is extremely rare that one can solve an SDE in some "explicit" manner.
There are, however, a few nontrivial interesting cases where it is possible to solve in a explicit way a SDE: we now introduce two classical equations heavily used in Mathematical Finance, the Linear SDE and the Geometric Brownian Motion (GBM).

1.1.1 Linear SDE

We now study the linear SDE, which in the scalar case has the form

$$(1.4) \qquad \begin{cases} dX_t &= \alpha X_t dt + \sigma dW_t, \\ X_0 &= x_0. \end{cases}$$

This equation turns up in various physical applications, and we will also meet it below, in connection with interest rate theory: on this classical and quite simple equation is based the Vasicek Model (see [66]), the historical cornerstone of interest rate models.
In order to get some feeling for how to solve this equation we recall that the linear ODE

$$\frac{dx_t}{dt} = \alpha x_t + u_t,$$

where u is a deterministic function of time, has the solution

$$x_t = e^{\alpha t} x_0 + \int_0^t e^{\alpha(t-s)} u_s ds.$$

If we, for a moment, reason heuristically, then it is tempting to formally divide equation (1.4) by dt. This (formally) would give us

$$\frac{dX_t}{dt} = \alpha X_t + \sigma \frac{dW_t}{dt},$$

and, by analogy with the ODE above, one is led to conjecture the formal solution:

$$x_t = e^{\alpha t} x_0 + \sigma \int_0^t e^{\alpha(t-s)} \frac{dW_s}{ds} ds = e^{\alpha t} x_0 + \sigma \int_0^t e^{\alpha(t-s)} dW_s.$$

We must remark that in general tricks like this will not work, since the solution of the ODE is based on ordinary calculus, whereas we have to use Itô Calculus when dealing with SDEs. In this case, however, we have a linear structure, which means that the second order term in the Itô Formula does not come into play. Thus the solution of the linear SDE (1.4)is indeed given by the heuristically derived formula above.

1.1.2 Geometric Brownian Motion

Geometric Brownian Motion has been one of the fundamental building blocks for the modelling of asset prices: the celebrated Black-Scholes model is in fact based on a extended version of the GBM.
The equation looks as follows:

$$\begin{cases} dX_t &= \alpha X_t dt + \sigma X_t dW_t, \\ X_0 &= x_0. \end{cases} \tag{1.5}$$

Written in a slightly sloppy form we can write the equation as

$$\dot{X}_t = (\alpha + \sigma \dot{W}_t) X_t,$$

where $\dot{W}$ is the "white noise", i.e. the (formal) time derivative of the Wiener process. Thus we see that GBM can be viewed as a linear ODE, with a stochastic coefficient driven by the white noise.
Thinking to a graphical representation, one can realize that for small values of σ, the trajectory will (at least initially) stay fairly close to the expected value function $\mathbb{E}(X_t) = 1 \cdot e^{\alpha t}$, whereas for a large value of σ it will give rise to large random deviations.
Inspired by the fact that the solution to the corresponding deterministic linear equation is an exponential function of time we are led to investigate the process Z, defined by $Z_t = ln(X_t)$, where we assume that X is a solution and that X is strictly positive (see below).
The Itô Formula gives us:

$$dZ = \frac{1}{X} dt + \frac{1}{2}\Big(-\frac{1}{X^2}\Big)[dW]^2 = \frac{1}{X}(\alpha X dt + \sigma X dW) + \frac{1}{2}\Big(-\frac{1}{X^2}\Big)\sigma^2 X^2 dt,$$

and then

$$dZ = (\alpha dt + \sigma dW) - \frac{1}{2}\sigma^2 dt.$$

Thus we have the system

$$\begin{cases} dZ_t &= \left(\alpha - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t, \\ \\ Z_0 &= ln(x_0). \end{cases} \tag{1.6}$$

This equation, however, is extremely simple: since the right hand side does not contain Z it can be integrated directly to

$$Z_t = ln(x_0) + \left(\alpha - \frac{1}{2}\sigma^2\right)t + \sigma W_t,$$

which means that X is given by

$$X_t = x_0 \cdot e^{(\alpha - \frac{1}{2}\sigma^2)t + \sigma W_t}. \tag{1.7}$$

Mathematically speaking, there is a logical flaw in the reasoning above. In order for Z to be well defined we have to assume that there actually exists a solution X to equation (1.5) and we also have to assume that the solution is positive. As for the existence, this is covered by the previous proposition, but the positivity seems to present a bigger problem. We may actually avoid both these problems by regarding the calculations above as purely heuristic.

On the contrary, we define the process X by the formula (1.7). Then it is easy to show that X thus defined actually satisfies the SDE (1.5), and that $\mathbb{E}(X_t) = x_0 \cdot e^{\alpha t}$.

1.2 The infinitesimal operator

Consider, the n-dimensional SDE

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{1.8}$$

Through the Itô Formula, the process above is closely connected to a partial differential operator $\mathcal{A}$ that we are going to define. The next two paragraphs are devoted to investigating the connections between, on the one hand, the analytical properties of the operator $\mathcal{A}$, and on the other hand the probabilistic properties of the process X above.

Definition 1.2 *Given the SDE in (1.8), the partial differential operator $\mathcal{A}$, referred to as the infinitesimal operator of X, is defined, for any function $h(x)$ with $h \in C^2(R^n)$, by*

$$\mathcal{A}h(t,x) = \sum_{i=1}^{n} \mu_i(t,x)\frac{\partial h}{\partial x_i}(x) + \frac{1}{2}\sum_{i,j=1}^{n} C_{ij}(t,x)\frac{\partial^2 h}{\partial x_i \partial x_j}(x),$$

where

$$C(t,x) = \sigma(t,x)\sigma^T(t,x).$$

This operator is also known as the Dynkin operator or the Kolmogorov backward operator. We note that, in terms of the infinitesimal generator, the Itô Formula takes the form

$$df(t, X_t) = \Big(\frac{\partial f}{\partial t} + \mathcal{A}f\Big)dt + [\bigtriangledown_x f]\sigma dW_t,$$

where the gradient $\bigtriangledown_x$ is defined for $h \in C^1(R^n)$ as

$$\bigtriangledown_x h = \Big[\frac{\partial h}{\partial x_1}, \dots, \frac{\partial h}{\partial x_n}\Big].$$

1.3 Partial Differential Equations and Feynman-Kac representation

In this paragraph we will explore the intimate connection which exists between stochastic differential equations and certain parabolic partial differential equations.
Consider for example the following problem, corresponding to the so called Cauchy Problem in the deterministic case. We are given three scalar functions $\mu(t, x), \sigma(t, x)$ and $\Phi(x)$. Our task is to find a function F which satisfies the following boundary value problem on $[0, T] \times R$.

$$(1.9) \qquad \begin{cases} \frac{\partial F}{\partial t}(t,x) + \mu(t,x)\frac{\partial F}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 F}{\partial x^2}(t,x) &= 0, \\ F(T,x) &= \Phi(x). \end{cases}$$

Now, instead of attacking this problem using purely analytical tools, we will produce a so called stochastic representation formula, which gives the solution to (1.9) in terms of the solution to an SDE which is associated to (1.9) in a natural way. Thus we assume that there actually exists a solution F to (1.9). Let us now fix a point in time t and a point in space x. Having fixed these we define the stochastic process X on the time interval $[t, T]$ as the solution to the SDE

$$\begin{cases} dX_s &= \mu(s, X_s)ds + \sigma(s, X_s)dW_s, \\ X_t &= x. \end{cases}$$

and the point is that the infinitesimal generator $\mathcal{A}$ for this process is given by

$$\mathcal{A} = \mu(t,x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2}{\partial x^2},$$

which is exactly the operator appearing in the PDE above. Thus we may write the boundary value problem as

$$(1.10) \qquad \begin{cases} \frac{\partial F}{\partial t}(t,x) + \mathcal{A}F(t,x) &= 0, \\ F(T,x) &= \Phi(x). \end{cases}$$

By applying the Itô Formula to the process $F(s, X(s))$ we obtain

$$(1.11)\quad F(t, X_T) = F(t, X_t) + \int_t^T \Big(\frac{\partial F}{\partial t} + \mathcal{A}F\Big)(s, X_s)ds + \int_t^T \sigma(s, X_s)\frac{\partial F}{\partial x}(s, X_s)dW_s.$$

Since, by assumption, F actually satisfies (1.10), the time integral above will vanish. If furthermore the process $\sigma(s, X_s)\frac{\partial F}{\partial x}(s, X_s)$ is sufficiently integrable and we take expected values, the stochastic integral will also vanish. The initial value $X_t = x$ and the boundary condition $F(T, x) = \Phi(x)$ will eventually leave us with the formula

$$F(t, x) = \mathbb{E}_{t,x}(\Phi(X_T)),$$

where we have indexed the expectation operator in order to emphasize that the expected value is to be taken given the initial value $X_t = x$. Thus we have proved the following result, which is known as the Feynman-Kac stochastic representation formula.

Proposition 1.3 *Assume that F is a solution to the boundary value problem*

$$\begin{cases} \frac{\partial F}{\partial t}(t, x) + \mu(t, x)\frac{\partial F}{\partial x} + \frac{1}{2}\sigma^2(t, x)\frac{\partial^2 F}{\partial x^2}(t, x) &= 0, \\ \\ F(T, x) &= \Phi(x). \end{cases}$$

Assume furthermore that the process

$$\sigma(s, X_s)\frac{\partial F}{\partial x}(s, X_s)$$

is in $\mathcal{L}^2$, where X is defined below. Then F has the representation

$$F(t, x) = \mathbb{E}_{t,x}(\Phi(X_T)),$$

where X satisfies the SDE

$$\begin{cases} dX_s &= \mu(s, X_s)ds + \sigma(s, X_s)dW_s, \\ \\ X_t &= x. \end{cases}$$

Note that we need the integrability assumption $\sigma(s, X_s)\frac{\partial F}{\partial x}(s, X_s) \in \mathcal{L}^2$ in order to guarantee that the expected value of the stochastic integral in (1.11) equals zero.
In fact the generic situation is that a boundary value problem of the type above, a so called parabolic problem, will have infinitely many solutions. It will, however, only have one "nice" solution, the others being rather "wild", and the proposition above will only give us the "nice" solution.
We may also consider the closely related boundary value problem

$$(1.12)\quad \begin{cases} \frac{\partial F}{\partial t}(t, x) + \mu(t, x)\frac{\partial F}{\partial x} + \frac{1}{2}\sigma^2(t, x)\frac{\partial^2 F}{\partial x^2}(t, x) + rF(t, x) &= 0, \\ \\ F(T, x) &= \Phi(x). \end{cases}$$

where r is a given real number. Equations of this type appear over and over again in the study of pricing problems for financial derivatives: we will deal with this point in chapter 2, in order to obtain the tools used in chapter 9.
Inspired by the ODE technique of integrating factors we are led to multiply the entire equation (1.12) by the factor e^{rs}, and if we then consider the process $Z(s) = e^{rs}F(s, X(s))$, where X is defined as in the previous proposition, we obtain the following result.

Proposition 1.4 *Assume that F is a solution to the boundary value problem*

$$\begin{cases} \frac{\partial F}{\partial t}(t,x) + \mu(t,x)\frac{\partial F}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 F}{\partial x^2}(t,x) + rF(t,x) &= 0, \\ \\ F(T,x) &= \Phi(x). \end{cases}$$

Assume furthermore that the process

$$\sigma(s, X_s)\frac{\partial F}{\partial x}(s, X_s),$$

is in $\mathcal{L}^2$, where X is defined below.
Then F has the representation

$$F(t,x) = e^{r(T-t)}\mathbb{E}_{t,x}(\Phi(X_T)),$$

where X satisfies the SDE

$$\begin{cases} dX_s &= \mu(s, X_s)ds + \sigma(s, X_s)dW_s, \\ \\ X_t &= x. \end{cases}$$

Up to now we have only treated the scalar case, but exactly the same argument as above will give us the following general result.

Proposition 1.5 *Take as given*

- *A (column-vector valued) function $\mu : R_+ \times R^n \to R^n$.*
- *A function $C : R_+ \times R^n \to M(n,n)$, which can be written in the form*

  $$C(t,x) = \sigma(t,x)\sigma^T(t,x),$$

  for some function $\sigma : R_+ \times R^n \to M(n,d)$.
- *A real valued function $\Phi : R^n \to R$.*

- *A real number r.*

Assume that $F : R_+ \times R^n \to R$ is a solution to the boundary value problem

$$\left\{ \begin{array}{rcl} \frac{\partial F}{\partial t}(t,x) + \sum_{i=1}^n \mu_i(t,x)\frac{\partial F}{\partial x_i}(t,x) + \frac{1}{2}\sum_{i,j=1}^n C_{ij}(t,x)\frac{\partial^2 F}{\partial x_i \partial x_j}(t,x) - rF(t,x) & = & 0, \\ \\ F(T,x) & = & \Phi(x). \end{array} \right.$$

Assume furthermore that the process

$$\sum_{i=1}^n \sigma_i(s,X_s)\frac{\partial F}{\partial x_i}(s,X_s),$$

is in $\mathcal{L}^2$, where X is defined below. Then F has the representation

$$F(t,x) = e^{-r(T-t)}\mathbb{E}_{t,x}(\Phi(X_T)),$$

where X satisfies the SDE

$$\left\{ \begin{array}{rcl} dX_s & = & \mu(s,X_s)ds + \sigma(s,X_s)dW_s, \\ \\ X_t & = & x. \end{array} \right.$$

We end this paragraph with a very useful result, a technical tool heavily used in the study of SDEs.

Proposition 1.6 *Consider as given a vector process X with generator $\mathcal{A}$, and a function $F(t,x)$. Then, modulo some integrability conditions, the following hold.*

- *The process $F(t,X_t)$ is a martingale relative to the filtration $\mathcal{F}^X$ if and only if F satisfies the PDE*
$$\frac{\partial P}{\partial t} + \mathcal{A}F = 0.$$
- *For every (t,x) and $T \geq t$, we have*
$$F(t,x) = \mathbb{E}_{t,x}(F(T,X_T)).$$

1.4 The Kolmogorov backward and forward Equations

We will now use the results of the previous paragraphs in order to derive some classical results concerning the transition probabilities for the solution to an SDE, leaving undefined some technical details. A complete and detailed treatment of this subject can be found in

[46].
Suppose that X is a solution to the equation

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t,$$

with infinitesimal generator $\mathcal{A}$ given by

$$(\mathcal{A}f)(s, y) = \sum_{i=1}^{n} \mu_i(s, y)\frac{\partial f}{\partial y_i}(s, y) + \frac{1}{2}\sum_{i,j=1}^{n} C_{ij}(s, y)\frac{\partial^2 f}{\partial y_i \partial y_j}(s, y),$$

where as usual

$$C(t, x) = \sigma(t, x)\sigma^T(t, x).$$

Now consider the boundary value problem

$$\begin{cases} \left(\frac{\partial u}{\partial s} + \mathcal{A}u\right)(s, y) = 0 & (s, y) \in (0, T) \times R^n, \\ u(T, y) = I_B(y), & y \in R^n, \end{cases}$$

where I_B is the indicator function of the set B. From the previous propositions we immediately have

$$u(s, y) = \mathbb{E}_{s,y}(I_B(X_T)) = \mathbb{P}(X_T \in B | X_s = y),$$

where X is a solution of (1.8).
This argument can also be turned around, and we have thus "informally" proved the following result.

Proposition 1.7 *Let X be a solution to equation (1.8). Then the transition probabilities $\mathbb{P}(s, y, t, B) = \mathbb{P}(X_t \in B | X(s) = y)$ are given as the solution to the equation*

$$\begin{cases} \left(\frac{\partial P}{\partial s} + \mathcal{A}P\right)(s, y, t, B) = 0 & (s, y) \in (0, t) \times R^n, \\ P(t, y, t, B) = I_B(y). \end{cases} \tag{1.13}$$

Using basically the same reasoning it is easy to obtain the following corresponding result for transition densities.

Proposition 1.8 *Let X be a solution of equation (1.8). Assume that the measure $\mathbb{P}(s, y, t, dx)$ has a density $p(s, y, t, x)dx$. Then we have*

$$\begin{cases} \left(\frac{\partial p}{\partial s} + \mathcal{A}p\right)(s, y, t, B) = 0 & (s, y) \in (0, t) \times R^n, \\ p(t, y, t, x) \rightarrow \delta_x & \text{as } s \rightarrow t. \end{cases} \tag{1.14}$$

The reason that equations (1.13) and (1.14) are called backward equations is that the differential operator is working on the "backward variables" (s, y). We will now derive a corresponding "forward" equation, where the action of the differential operator is on the "forward" variables (t, x). For simplicity we consider only the scalar case.
We assume that X has a transition density. Let us then fix two points in time s and T with $s < T$. Now consider an arbitrary "test function", i.e. an infinite differentiable function $h(t, x)$ with compact support in the set $(s, T) \times R$. From the Itô Formula we have

$$h(Y, X_T) = h(s, X_s) + \int_s^T \Big(\frac{\partial h}{\partial t} + \mathcal{A}h\Big)(t, X_t)dt + \int_s^T \frac{\partial h}{\partial x}(t, X_t)dW_t.$$

Applying the expectation operator $\mathbb{E}s, y(\cdot)$, and using the fact that $h(T, x) = h(s, x) = 0$ (because of the compact support), we obtain

$$\int_{-\infty}^{+\infty} \int_s^T h(t, x)\Big(\frac{\partial}{\partial t} + \mathcal{A}\Big)h(t, x)dxdt = 0.$$

Partial integration with respect to t (for $\frac{\partial}{\partial t}$) and with respect to x (for $\mathcal{A}$) gives us

$$\int_{-\infty}^{+\infty} \int_s^T h(t, x)\Big(-\frac{\partial}{\partial t} + \mathcal{A}^*\Big)p(s, y, t, x)dxdt = 0,$$

where the adjoint operator $\mathcal{A}^*$ is defined by

$$(\mathcal{A}^* f)(t, x) = -\frac{\partial}{\partial x}(\mu(t, x)f(t, x)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2(t, x)f(t, x)).$$

Since this equation holds for all test functions we have shown the following result.

Proposition 1.9 *Assume that the solution X of equation (1.8) has a transition density $p(s, y, t, x)$. Then p will satisfy the Kolmogorov forward equation*

$$\begin{cases} \frac{\partial}{\partial t}p(s, y, t, x) &= \mathcal{A}^* p(s, y, t, x), \quad (t, x) \in (0, T) \times R, \\ p(s, y, t, x) &\to \delta_y \qquad \textit{as } t \downarrow s. \end{cases} \tag{1.15}$$

This equation is also known as the Fokker-Planck equation. The multi-dimensional version is

$$\frac{\partial}{\partial t}p(s, y, t, x) = \mathcal{A}^* p(s, y, t, x),$$

where the adjoint operator $\mathcal{A}^*$ is defined by

$$(\mathcal{A}^* f)(t, x) = -\sum_{i=1}^n \frac{\partial}{\partial x_i}(\mu_i(t, x)f(t, x)) + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}(C_{ij}(t, x)f(t, x)).$$

1.5 Applications

Stochastic Differential Equations are a powerful tool in Mathematical Finance. The equations used in literature in order to model interest rates evolutions are exposed in detail in chapter 2.
A particular equation belonging to the Feller Diffusion class is studied in detail in chapter 4 and 5, while the Gaussian class of infinite dimensional equations is developed in chapter 6 and 9.
Finally, a geometrical point of view on the state variables evolution is introduced in chapter 7 and 8.

Chapter 2

Interest Rates Modelling

Source and inspiration of this chapter is [6]. Following the classical Björk's notations, in order to keep this work self-contained we expose only those parts of the standard theory that we will use in the following chapters.
Interest Rate Modelling is a growing area in mathematical finance: for a detailed survey on IRM from a theoretical point of view see [7], and for empirical approaches see (among the others) [15, 44, 58].

2.1 Preliminaries

Self-financing portfolios

We take as given a filtered probability space $(\Omega, \mathcal{F}_t, \mathcal{P})$ carrying a finite number of stochastic processes $S_0, \ldots, S_K$. We assume that all processes are semi-martingales, and with very little loss of generality we assume that the S-processes are defined by a system of stochastic differential equations, driven by a finite number of Wiener processes.
The above setup is intended as a model of a financial market with $K+1$ traded assets, where $S_i(t)$ is interpreted as the price of one unit of asset number i at time t. The assets will often be referred as "stocks", but they could be stocks, bonds, options or any other traded financial securities.
The asset S_0 will play a particular role as "numeraire asset" and we assume henceforth that $S_0(t) > 0, \mathcal{P}$-a.s. for all $t > 0$. As a convention we also assume that $S_0 = 1$. This numeraire asset can in principle be any asset: it is chosen for simple convenience, and in most cases it is chosen as the so called "risk free" asset (for details see the next section), with $\mathcal{P}$-dynamics of the form

$$\begin{cases} dS_0(t) &= r(t)S_0(t)dt, \\ S_0(0) &= 1. \end{cases}$$

We will study portfolios (trading strategies) based on the assets above, and to this end we need some definitions.

Definition 2.1 *1. A portfolio is any locally bounded predictable vector process*

$$h(t) = [h^0(t), \dots, h^K(t)].$$

2. The value process corresponding to a portfolio h is defined by

$$V(t,h) = \sum_{i=0}^{K} h^i(t) S_i(t).$$

3. A Portfolio h is said to be self-financing if the following hold

$$dV(t,h) = \sum_{i=0}^{K} h^i(t) dS_i(t).$$

The self-financing condition is intended to formalize the intuitive idea of a trading strategy with no exogenous infusion or withdrawal of money, i.e. a strategy where the purchase of a new asset is financed solely by the sale of assets already in the portfolio.
All prices above are interpreted as being given in terms of some a priori given numeraire, and typically this numeraire is a local currency like Euro. We may of course equally express all prices in terms of some other numeraire, and we will in fact choose S_0.

Definition 2.2 *1. The discounted price process vector $Z(t) = [Z^0(t), \dots, Z^K(t)]$ is defined by*

$$Z(t) = \frac{S(t)}{S_0(t)}.$$

2. The discounted value process corresponding to a portfolio h is defined as

$$V^Z(t,h) = \frac{V(t)}{S_0(t)}.$$

As is to be expected, the property of being self-financing does not depend upon the particular choice of numeraire.

Lemma 2.3 *A portfolio is self-financing if and only if the following relation hold:*

$$dV^Z(t,h) = \sum_{i=0}^{K} h^i(t) dZ_i(t).$$

We recall that for some applications it is convenient to express a portfolio strategy in terms of the "relative weights" $[u^0, \dots, u^K]$, where $u(t)$ is defined as the proportion of the portfolio value which at time t is invested in asset i: $u^i(t) = \frac{h^i(t)S_i(t)}{V(t,h)}$.

2.1.1 Contingent claims, arbitrage and martingale measures

The basic objects of study in arbitrage theory are the financial derivatives. These are assets which in some way are defined in terms of the underlying assets $S_0, \dots, S_K$. Typical examples are options, forwards, futures, bonds, interest rate swaps, caps and floors: these derivatives will be discussed in the next sections, and in particular caps&floors will be deeply studied in chapter 9.

Definition 2.4 *A contingent T-claim is any random variable $X \in L^0(\mathcal{F}_T, \mathcal{P})$ (i.e. an arbitrary $\mathcal{F}_T$-measurable random variable). We use the notation $X \in L^0_+(\mathcal{F}_T, \mathcal{P})$to denote the set of non-negative elements of $L^0(\mathcal{F}_T, \mathcal{P})$, and $X \in L^0_{++}(\mathcal{F}_T, \mathcal{P})$ denotes the set of elements X of $L^0_+(\mathcal{F}_T, \mathcal{P})$ with $\mathcal{P}(X > 0) > 0$.*

The interpretation of this definition is that the contingent claim is a contract which specifies that the stochastic amount, X, of money is to be payed out to the holder of the contract at time T. In the sequel we will often have to impose some further integrability restrictions upon the set of contingent claims under consideration.

Definition 2.5 *An arbitrage portfolio is a self-financing portfolio h such that the corresponding value process has the properties:*

1. $V(0) = 0$,

2. $V(T) \in L^0_{++}(\mathcal{F}_T, \mathcal{P})$.

If no arbitrage portfolios exists for any $T \in \mathbb{R}_+$ we say that the model is "free of arbitrage" or "arbitrage free".

As can be seen from the definition, an arbitrage portfolio is a deterministic money-making machine, and we typically assume that there are no arbitrage possibilities in our market (more or less the definition of an "efficient" market). Our first goal is to investigate when a given model is free of arbitrage, and the main tool is the concept of a martingale measure.

Definition 2.6 *We say that a probability measure $\mathcal{Q}$ is a martingale measure if*

1. $\mathcal{Q} \sim \mathcal{P}$,

2. *The discounted price process Z is a $\mathcal{Q}$-local martingale.*

If the discounted price process Z is $\mathcal{Q}$-martingale we say that $\mathcal{Q}$ is a strong martingale measure.

We can now slightly modify our set of admissible portfolios.

Definition 2.7 *Consider a given martingale measure $\mathcal{Q}$, and a self-financing portfolio h. Then h is called $\mathcal{Q}$-admissible if $V^Z(t,h)$ is a $\mathcal{Q}$-martingale.*

Since by definition Z is a $\mathcal{Q}$-martingale, and since the V^Z-process is the stochastic integral of h with respect to Z, we see that every sufficiently integrable self financing portfolio is in fact admissible. It is of course annoying that the definition of admissibility is dependent upon the particular choice of martingale measure, but the need of the admissibility condition can be seen inside the proof of the following proposition, which is one the basic results in the theory.

Proposition 2.8 *Assume that there exist a martingale measure $\mathcal{Q}$. Then the model is free of arbitrage in the sense that there exist no $\mathcal{Q}$-admissible arbitrage portfolio.*

Proof See [7]. □

We have seen that the existence of a martingale measure implies absence of arbitrage, and a natural question is whether is a converse to this statement, i.e. if absence of arbitrage implies the existence of a martingale measure. For models in discrete time with a finite space Ω there is indeed such a result (see [38]), but for infinite models the situation is more complicated. A deep study of these problems can be found in [25], where the notion of "no arbitrage" is replaced by the notion of "no free lunch with vanishing risk".

The theory is somewhat unsatisfactory, but the general consensus seems to be that the existence of a martingale measure is informally considered to be more or less equivalent to the absence of arbitrage.

2.1.2 Hedging

We assume absence of arbitrage, or to be more precise we assume the existence of a martingale measure.We now turn to the possibility of replicating a given contingent claim in terms of a portfolio based on the underlying assets.

Definition 2.9

1. *A fixed T-claim is said to be attainable if there exist a self-financing portfolio h such that the corresponding value process have the property that*

$$V(T,h) = X, \mathcal{P} - a.s.$$

2. *The market is said to be complete if every claim is attainable.*

The portfolio h above is called a replicating or hedging portfolio for the claim and we see that, from a purely financial point of view, the replicating portfolio is indistinguishable from the claim X.

The main problem is to determine which claims can be hedged, and this is most conveniently carried out in terms of discounted prices. It can be shown that, modulo some integrability conditions, completeness is equivalent to the existence of a martingale representation theorem for the discounted price process. We quote the following result from [38].

Theorem 2.10 *If the martingale measure $\mathcal{Q}$ is unique, then the market is complete in the restricted sense that every claim X satisfying*

$$\frac{X}{S_0(T) \in L^1(\mathcal{Q}, \mathcal{F}_T)}$$

is attainable.

2.1.3 Completeness & absence of arbitrage

From the last theorem we know that the model is free of arbitrage if there exists a martingale measure, and that the model is complete if the martingale measure is unique. In this section we will give some general rules of thumb for quickly determining whether a certain model is complete and/or free of arbitrage: we want to explicitly declare that this arguments will be purely heuristic.
Let us assume that the price processes of the underlying assets are driven by R "random sources". We can not give a precise definition of what constitutes a "random source", but typical examples are driving Wiener processes, Poisson process or more general Levy processes.
When discussing these concepts, it is important to realize that

completeness and absence of arbitrage work in opposite directions.

Let for example the number of random sources R be fixed. Then every new underlying asset added to the model (without increasing R) will of course give us potential opportunity of creating an arbitrage portfolio so in order to have an arbitrage free market the number of underlying assets (apart from the numeraire asset) must be small in comparison with the number of random sources R.
On the other hand we see that every new underlying asset added to the model gives us new possibilities of replicating a given contingent claim, so completeness requires M to be large in comparison with R.
We can not formulate and prove a precise theorem, but the following rule of thumb, or "Meta-Theorem" is nevertheless extremely useful. In concrete cases it can in fact be given a precise formulation and a precise proof. We will later use the meta-theorem when dealing with problems connected with non-traded underlying assets, which often arise in interest rate theory.

Theorem 2.11 (Meta-Theorem) *Let K denote the number of underlying assets in the model excluding the numeraire asset, and let R denote the number of random sources. Generically we then have the following relations.*

1. The model is arbitrage free if and only if $K \leq R$.

2. The model is complete if an d only if $K \geq R$.

3. The model is complete and arbitrage free if and only if $K = R$.

2.1.4 Pricing

We now turn to the problem of determining a "reasonable" price process $\Pi(t, X)$ for a fixed contingent T-claim X. We assume that there exist al least a martingale measure. There are two approaches:

- The derivative should be priced in a way that is consistent with the prices of the underlying assets. More precisely we should demand that the extended market $[\Pi(X), S_0, S_1, \ldots, S_K]$ is free of arbitrage possibilities.

- If the claim is attainable, with hedging portfolio h, then the only reasonable price is given by $\Pi(t, X) = V(t, h)$.

In order to keep thing simple in the first approach, let us suppose the there exists a strong martingale measure. Letting $\mathcal{Q}$ denote such a measure and applying the definition of a strong martingale measure we obtain

$$\frac{\Pi(t, X)}{S_0(t)} = \mathbb{E}^{\mathcal{Q}}\Big(\frac{\Pi(T, X)}{S_0(T)}\Big|\mathcal{F}_t\Big) = \mathbb{E}^{\mathcal{Q}}\Big(\frac{X}{S_0(T)}\Big|\mathcal{F}_t\Big).$$

We thus have the pricing formula

$$\Pi(t, X) = S_0(t)\mathbb{E}^{\mathcal{Q}}\Big(\frac{X}{S_0(T)}\Big|\mathcal{F}_t\Big), \tag{2.1}$$

where $\mathcal{Q}$ is a martingale measure for the a priori given market $[S_0, S_1, \ldots, S_K]$. Note that different choices of $\mathcal{Q}$ will generally give rise to different price processes.
For the second approach to pricing let us assume that X can be replicated by h. since the holding of derivative contract and the holding of replicating portfolios are equivalent from a financial point of view, we have seen that the price of the derivative must be given by the formula $\Pi(t, X) = V(t, h)$. One problem is what will happen in a case when X can be replicated two different portfolios, and one would also like to know how this formula is connected to (2.1).
Defining $\Pi(t, X)$ by (2.1) we see that $\Pi(t, X)/S_0(t) = V^Z(t)$ and since, assuming enough integrability, V^Z is a $\mathcal{Q}$-martingale, we obtain that also $\Pi(t, X)/S_0(t)$ is a $\mathcal{Q}$-martingale.

Thus we again get the formula (2.1) and for an attainable claim we have in particular the formula

$$V(t,h) = S_0(t)\mathbb{E}^{\mathcal{Q}}\Big(\frac{X}{S_0(T)}\Big|\mathcal{F}_t\Big),$$

which will hold for any replicating portfolio and for any martingale measure $\mathcal{Q}$. We thus have the following result.

Proposition 2.12 *1. If we assume the existence of a strong martingale measure for the market* $[\Pi(X), S_0, S_1, \ldots, S_K]$ *the* X *must be priced according to the formula*

$$\Pi(t,X) = S_0(t)\mathbb{E}^{\mathcal{Q}}\Big(\frac{X}{S_0(T)}\Big|\mathcal{F}_t\Big),$$

where $\mathcal{Q}$ *is a strong martingale measure for* $[S_0, S_1, \ldots, S_K]$.

2. Different choices of $\mathcal{Q}$ *will generically give rise to different price processes, bur if* X *is attainable then all choices of* $\mathcal{Q}$ *will produce the same price process.*

3. for an attainable claim the price process will also be given by the formula

$$\Pi(t,X) = V(t,h),$$

where h *is the hedging portfolio. different choices of hedging portfolios (is such portfolios exist) will produce the same price process.*

Summing up, we see that in a complete market the price of any derivative will be uniquely determined by the requirement of absence of arbitrage. The price is unique precisely because the derivative is in a sense superfluous, as it can be equally well been replaced by its replicating portfolio. In particular we see that the price does not depend on any assumptions made about the risk-preferences of the agents in the market. The agents can have any attitude toward risk, as long as they prefer more (deterministic) money to less. In an incomplete market the requirement of no arbitrage is no longer sufficient to determine a unique price for the derivative. We have several martingale measures, all of which can be used to price derivatives in a way consistent with no arbitrage. The question which martingale measure one should use for pricing has a very simple answer

The martingale measure is chosen by the market.

Schematically speaking the price of a derivative is thus determined by two major factors.

1. We require that the derivative should be priced in such a way as to not introduce arbitrage possibilities into the market. This requirement is reflected by the fact that

all derivatives must be priced by formula (2.1) where the same $\mathcal{Q}$ is used for all derivatives.

2. In an incomplete market the price is also partly determined by aggregate supply and demand on the market. Supply and demand for a specific derivative are in turn determined by the aggregate risk aversion on the market, as well as by liquidity considerations and other factors. all these aspects are aggregated into the particular martingale measure used by the market.

2.2 Bonds and interest rates

2.2.1 Zero coupon bonds

In this chapter we will begin to study the particular problems which appear when we try to apply arbitrage theory to the bond market. The primary objects of investigation are zero coupon bonds, also known as pure discount bonds, of various maturities. All payments are assumed to be made in a fixed currency which, for convenience, we choose to be Euro.

Definition 2.13 *A zero coupon bond with maturity date T, also called a T-bond, is a contract which guarantees the holder 1 euro to be paid on the date T. The price at time t of a bond with maturity date T is denoted by $P(t,T)$.*

The convention that the payment at the time of maturity, known as the principal value or face value, equals one is made for computational convenience. Coupon bonds, which give the owner a payment stream during the interval $[0,T]$ are treated below. These instruments have the common property, that they provide the owner with a deterministic cash flow, and for this reason they are also known as fixed income instruments. We now make an assumption to guarantee the existence of a sufficiently rich and regular bond market.

Assumption 2.14 *We assume the following.*

- *There exists a (frictionless) market for T-bonds for every $T > 0$.*

- *The relation $P(t,t) = 1$ holds for all t.*

- *For each fixed t, the bond price $P(t,T)$ is differentiable w.r.t. time of maturity T.*

Note that the relation $P(t,t) = 1$ above is necessary in order to avoid arbitrage. The bond price $P(t,T)$ is thus a stochastic object with two variables, t and T, and, for each outcome in the underlying sample space, the dependence upon these variables is very different. For a fixed value of t, $P(t,T)$ is a function of T. This function provides the prices, at the fixed time t, for bonds of all possible maturities. The graph of this function is called "the bond price curve at t", or "the term structure at t". Typically it will be a very smooth graph, i.e. for each t, $P(t,T)$ will be differentiable w.r.t. T. The smoothness property is in fact a part of the assumptions above, but this is mainly for convenience. All models to be considered below will automatically produce smooth bond price curves.

For a fixed maturity T, $P(t,T)$ (as a function of t) will be a scalar stochastic process. This process gives the prices, at different times, of the bond with fixed maturity T, and the trajectory will typically be very irregular (like a Wiener process).

We thus see that the bond market is different from any other market, in the sense that the bond market contains an infinite number of assets (one bond type for each time of maturity). The basic goal in interest rate theory is roughly that of investigating the relations between all these different bonds. Somewhat more precisely we may pose the following general problems, to be studied below.

- What is a reasonable model for the bond market above?

- Which relations must hold between the price processes for bonds of different maturities, in order to guarantee an arbitrage free bond market?

- Is it possible to derive arbitrage free bond prices from a specification of the dynamics of the short rate of interest?

- Given a model for the bond market, how to compute prices of interest rate derivatives?

2.2.2 Interest rates

Given the bond market above, we may now define a number of interest rates, and the basic construction is as follows. Suppose that we are standing at time t, and let us fix two other points in time, S and T, with $t < S < T$. The immediate project is to write a contract at time t which allows us to make an investment of one Euro at time S, and to have a deterministic rate of return, determined at the contract time t, over the interval $[S,T]$. This can easily be achieved as follows.

1. At time t we sell one S-bond. This will give us $P(t,S)$ Euro.

2. We use this income to buy exactly $P(t,S)/P(t,T)$ T-bonds. Thus our net investment at time t equals zero.

3. At time S the S-bond matures, so we are obliged to pay out one euro.

4. At time T the T-bonds mature at one euro a piece, so we will receive the amount $P(t,S)/P(t,T)$ euro.

5. The net effect of all this is that, based on a contract at t, an investment of one euro at time S has yielded $P(t,S)/P(t,T)$ euro at time T.

6. Thus, at time t, we have made a contract guaranteeing a riskless rate of interest over the future interval $[S,T]$. Such an interest rate is called a forward rate.

We now go on to compute the relevant interest rates implied by the construction above. We will use two (out of many possible) ways of quoting forward rates, namely as continuously compounded rates or as simple rates.
The simple forward rate (or LIBOR rate) L, is the solution to the equation

$$1 + (T-S)L = P(t,S)/P(t,T),$$

whereas the continuously compounded forward rate R is the solution to the equation

$$e^{R(T-S)} = P(t,S)/P(t,T).$$

The simple rate notation is the one used in the market, whereas the continuously compounded notation is used in theoretical contexts. They are of course logically equivalent, and the formal definitions are as follows.

Definition 2.15 *1. The simple forward rate for $[S,T]$ contracted at t, henceforth referred to as the LIBOR forward rate, is defined as*

$$L(t,S,T) = -\frac{P(t,T) - P(t,S)}{(T-S)P(t,T)}.$$

2. The simple spot rate for $[S,T]$, henceforth referred to as the LIBOR spot rate, is defined as

$$L(S,T) = -\frac{P(S,T) - 1}{(T-S)P(S,T)}.$$

3. *The continuously compounded forward rate for* $[S, T]$ *contracted at* t *is defined as*

$$R(t, S, T) = -\frac{logP(t, T) - logP(t, S)}{(T - S)}.$$

4. *The continuously compounded spot rate,* $R(S, T)$*, for the period* $[S, T]$ *is defined as*

$$R(S, T) = -\frac{logP(S, T)}{(T - S)}.$$

5. *The instantaneous forward rate with maturity* T*, contracted at* t*, is defined by*

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

6. *The instantaneous short rate at time* t *is defined by*

$$r(t) = f(t, t).$$

We note that spot rates are forward rates where the time of contracting coincides with the start of the interval over which the interest rate is effective, i.e. $t = S$. The instantaneous forward rate, which will be of great importance below, is the limit of the continuously compounded forward rate when $S \to T$. It can thus be interpreted as the riskless rate of interest, contracted at t, over the infinitesimal interval $[T, T + dT]$. We now go on to define the money account process B.

Definition 2.16 *The money account process is defined by*

$$B_t = exp\Big(\int_o^t r(s)ds\Big),$$

i.e.

$$\begin{cases} dB(t) &= r(t)B(t)dt, \\ B(0) &= 1. \end{cases}$$

The interpretation of the money account is the same as before, i.e. you may think of it as describing a bank with a stochastic short rate of interest. It can also be shown that investing in the money account is equivalent to investing in a self-financing "rolling over" trading strategy, which at each time t consists entirely of "just maturing" bonds, i.e. bonds which will mature at $t + dt$. As an immediate consequence of the definitions we have the following useful formulas.

Proposition 2.17 *For* $t < s < T$ *we have*

$$P(t, T) = P(t, s)exp\Big(-\int_s^T f(t, u)du\Big),$$

and in particular

$$P(t, T) = exp\Big(-\int_t^T f(t, s)ds\Big).$$

If we wish to make a model for the bond market, it is obvious that this can be done in many different ways.

- We may specify the dynamics of the short rate (and then perhaps try to derive bond prices using arbitrage arguments).

- We may directly specify the dynamics of all possible bonds.

- We may specify the dynamics of all possible forward rates, and then use (2.17) in order to obtain bond prices.

All these approaches are of course related to each other, and we now go on to present a small "toolbox" of results to facilitate the analysis. We will consider dynamics of the following form:

- **short rate dynamics:**

$$dr(t) = a(t)dt + b(t)dW(t), \tag{2.2}$$

- **bond price dynamics:**

$$dP(t) = P(t,T)m(t,T)dt + P(t,T)v(t,T)dW(t), \tag{2.3}$$

- **forward rate dynamics:**

$$df(t) = \alpha(t)dt + \sigma(t)dW(t). \tag{2.4}$$

The Wiener process W is allowed to be vector valued, in which case the volatilities $v(t,T)$ and $\sigma(t,T)$ are row vectors. The processes $a(t)$ and $b(t)$ are scalar adapted processes, whereas $m(t,T), v(t,T), \alpha(t,T)$ and $\sigma(t,T)$ are adapted processes parameterized by time of maturity T. The interpretation of the bond price equation (2.3) and the forward rate equation (2.4) is that these are scalar stochastic differential equations (in the t-variable) for each fixed time of maturity T. Thus (2.3) and (2.4) are both infinite dimensional systems of SDEs. We will now study the formal relations which must hold between bond prices and interest rates, and to this end we need a number of technical assumptions, which we collect below in an "operational" manner.

Assumption 2.18 *For each fixed ω, t all the objects $m(t,T), v(t,T), \alpha(t,T)$ and $\sigma(t,T)$ are assumed to be continuously differentiable in the T-variable. This partial T-derivative is sometimes denoted by $m_r(t,T)$ etc.*
All processes are assumed to be regular enough to allow us to differentiate under the integral sign as well as to interchange the order of integration.

The main result is as follows: note that the results below hold, regardless of the measure under consideration, and in particular we do not assume that markets are free of arbitrage.

Proposition 2.19 *1. If $P(t,T)$ satisfies (2.3), then for the forward rate dynamics we have*

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

where α and σ are given by

$$\begin{cases} \alpha(t,T) &= v_T(t,T)v(t,T) - m_T(t,T), \\ \sigma(t,T) &= -v_T(t,T). \end{cases}$$

2. If $f(t,T)$ satisfies (2.4) then the short rate satisfies

$$dr(t) = a(t)dt + b(t)dW(t),$$

where

$$\begin{cases} a(t) &= f_T(t,t) + \alpha(t,t), \\ b(t) &= \sigma(t,t). \end{cases}$$

3. If $f(t,T)$ satisfies (2.4) then $P(t,T)$ satisfies

$$dP(t,T) = P(t,T)\Big(r(t) + A(t,T) + \frac{1}{2}\|S(t,T)\|^2\Big)dt + P(t,T)S(t,T)dW(t),$$

where $\|\cdot\|$ denotes the Euclidean norm, and

$$\begin{cases} A(t,T) &= -\int_t^T \alpha(t,s)ds, \\ S(t,T) &= -\int_t^T \sigma(t,s)ds. \end{cases}$$

Proof See [7]. □

2.2.3 Coupon bonds, swaps and yields

In most bond markets, there are only a relative small number of zero coupon bonds traded actively. The maturities for these are generally short, whereas most bonds with a longer time to maturity are coupon bearing. Despite this empirical fact we will still assume the existence of a market for all possible pure discount bonds, and we now go on to introduce and price coupon bonds in terms of zero coupon bonds.

Fixed coupon bonds

The simplest coupon bond is the fixed coupon bond. This is a bond which, at some intermediary points in time, will provide predetermined payments (coupons) to the holder of the bond. The formal description is as follows.

- Fix a number of dates, i.e. points in time, $T_0, \dots, T_n$. Here T_0 is interpreted as the emission date of the bond, whereas $T_1, \dots, T_n$ are the coupon dates.

- At time $T_i, i = 1, \dots, n$ the owner of the bond receives the deterministic coupon c_i.

- At time T_n the owner receives the face value K.

We now go on to compute the price of this bond, and it is obvious that the coupon bond can be replicated by holding a portfolio of zero coupon bonds with maturities $T_i, i = 1, \dots, n$. More precisely we will hold c_i zero coupon bonds of maturity $T_i, i = 1, \dots, n-1$ and $K + c_n$ bonds with maturity T_n, so the price $P(t)$, at a time $t < T_1$, of the coupon bond is given by

$$P(t) = KP(t, T_n) + \sum_{i=1}^{n} c_i P(t, T_i).$$

Very often the coupons are determined in terms of return, rather than in monetary (e.g. Euro) terms. The return for the i-th coupon is typically quoted as a simple rate acting on the face value K, over the period $[T_{i-l}, T_i]$. Thus, if, for example, the i-th coupon has a return equal to r_i, and the face value is K, this means that

$$c_i = r_i(T_i - T_{i-1})K.$$

For a standardized coupon bond, the time intervals will be equally spaced, i.e.

$$T_i = T_0 + i\delta,$$

and the coupon rates $r_l, \dots, r_n$ will be equal to a common coupon rate r. The price $P(t)$ of such a bond will, for $t < T_1$, be given by

$$P(t) = K\Big(P(t, T_n) + r\delta \sum_{i=1}^{n} P(t, T_i)\Big).$$

Floating rate bonds

There are various coupon bonds for which the value of the coupon is not fixed at the time the bond is issued, but rather reset for every coupon period. Most often the resetting is determined by some financial benchmark, like a market interest rate, but there are also bonds for which the coupon is benchmarked against a nonfinancial index. As an example (to be used in the context of swaps below), we will confine ourselves to discussing one of the simplest floating rate bonds, where the coupon rate r_i is set to the spot LIBOR rate $L(T_{i-l}, T_i)$. Thus

$$c_i = (T_i - Ti - 1)L(T_{i-1}, T_i)K,$$

and we note that $L(T_{i-1}, T_i)$ is determined already at time T_{i-l}, but that c_i is not delivered until at time T_i. We now go on to compute the value of this bond at some time $t < T_0$, in the case when the coupon dates are equally spaced, with $T_i - T_{i-1} = \delta$, and to this end we study the individual coupon c_i. Without loss of generality we may assume that $K = 1$, and inserting the definition of the LIBOR rate we have

$$c_i = \delta \frac{1 - P(T_{i-1}, T_i)}{\delta P(T_{i-1}, T_i)} = \frac{1}{P(T_{i-1}, T_i)} - 1.$$

The value at t of the term -1 (paid out at T_i), is of course equal to $-P(t, T_i)$, and it remains to compute the value of the term $\frac{1}{P(T_{i-1}, T_i)}$ which is paid out at T_i. This is, however, easily done in the following way.

- Buy, at time t, one T_{i-1}-bond. This will cost $P(t, T_{i-1})$.

- At time T_{i-1} you will receive the amount 1.

- Invest this unit amount in T_i-bonds. This will give you exactly $\frac{1}{P(T_{i-1}, T_i)}$ bonds.

- At T_i the bonds will mature, each at the face value 1. Thus, at time T_i, you will obtain the amount $\frac{1}{P(T_{i-1}, T_i)}$.

This argument shows that it is possible to replicate the cash flow above, using a self-financing bond strategy, to the initial cost $P(t, T_{i-1})$. Thus the value at t, of obtaining $\frac{1}{P(T_{i-1}, T_i)}$ at T_i, is given by $P(t, T_{i-1})$, and the value at t of the coupon c_i is

$$P(t, T_{i-1}) - P(t, T_i).$$

Summing up all the terms we finally obtain the following valuation formula for the floating rate bond

$$P(t) = P(t, T_n) + \sum_{i=1}^{n} [P(t, T_{i-1}) - P(t, T_i)] = P(t, T_0).$$

In particular we see that if $t = T_0$, then $P(T_0) = 1$. The reason for this formula is of course that the entire floating rate bond can be replicated through a self-financing portfolio.

Interest rate swaps

In this section we will discuss the simplest of all interest rate derivatives, the interest rate swap. This is basically a scheme where you exchange a payment stream at a fixed rate of interest, known as the swap rate, for a payment stream at a floating rate (typically a LIBOR rate). There are many versions of interest rate swaps, and we will study the

forward swap settled in arrears, which is defined as follows. We denote the principal by K, and the swap rate by R. By assumption we have a number of equally spaced dates $T_0, \ldots, T_n$, and payment occurs at the dates $T_1, \ldots, T_n$ (not at T_0). If you swap a fixed rate for a floating rate (in this case the LIBOR spot rate), then, at time T_i, you will receive the amount $K\delta L(T_{i-1}, T_i)$, which is exactly Kc_i, where C_i is the i-th coupon for the floating rate bond in the previous section. At T_i you will pay the amount $K\delta R$.
The net cash flow at T_i is thus given by

$$K\delta[L(T_{i-1}, Ti) - R],$$

and using the results from the floating rate bond, we can compute the value at $t < T_0$ of this cash flow as

$$KP(t, T_{i-1}) - K(1 + \delta R)P(t, T_i).$$

The total value $\Pi(t)$, at t, of the swap is thus given by

$$\Pi(t) = K \sum_{i=1}^{n} [P(t, T_{i-1}) - (1 + \delta R)P(t, T_i)],$$

and we can simplify this to obtain the following result.

Proposition 2.20 *The price, for $T_0 < t < T_1$, of the swap above is given by*

$$\Pi(t) = KP(t, T_0) - \sum_{i=1}^{n} d_i P(t, T_i),$$

where

$$d_i = R\delta, i = 1, \ldots, n - 1,$$

and

$$d_n = 1 + R\delta.$$

The remaining question is how the swap rate R is determined. By definition it is chosen such that the value of the swap equals zero at the time when the contract is made. We have the following easy result.

Proposition 2.21 *If, by convention, we assume that the contract is written at $t = 0$, the swap rate is given by*

$$R = \frac{P(0, T_0) - P(0, T_n)}{\delta \sum_{i=1}^{n}} P(0, T_i).$$

In the case that $T_0 = 0$ this formula reduces to

$$R = \frac{1 - P(0, T_n)}{\delta \sum_{i=1}^{n}} P(0, T_i).$$

Yield and duration

Consider a zero coupon T-bond with market price $P(t,T)$. We now look for the bond's "internal rate of interest", i.e. the constant short rate of interest which will give the same value to this bond as the value given by the market. Denoting this value of the short rate by y, we thus want to solve the equation

$$P(t,T) = e^{-y(T-t)} \cdot 1,$$

where the factor 1 indicates the face value of the bond. We are thus led to the following definition.

Definition 2.22 *The continuously compounded zero coupon yield, $y(t,T)$, is given by*

$$y(t,T) = -\frac{logP(t,T)}{T-t}.$$

For a fixed t, the function $T \mapsto y(t,T)$ is called the (zero coupon) yield curve.

We note that the yield $y(t,T)$ is nothing more than the spot rate for the interval $[t,T]$. Now let us consider a fixed coupon bond of the form discussed above where, for simplicity of notation, we include the face value in the coupon c_n. We denote its market value at t by $P(t)$. In the same spirit as above we now look for its internal rate of interest, i.e. the constant value of the short rate, which will give the market value of the coupon bond.

Definition 2.23 *The yield to maturity, y(t,T), of a fixed coupon bond at time t, with market price P, and payments c_i at T_i for $i = 1, \dots, n$, is defined as the value of y which solves the equation*

$$P(t) = \sum_{i=1}^{n} c_i e^{-y(T_i-t)}.$$

An important concept in bond portfolio management is the "Macaulay duration". Without loss of generality we may assume that $t = 0$.

Definition 2.24 *For the fixed coupon bond above, with price P at $t = 0$, and yield to maturity y, the duration D is defined as*

$$D = \frac{\sum_{i=1}^{n} T_i c_i e^{-yT_i}}{P}.$$

The duration is thus a weighted average of the coupon dates of the bond, where the discounted values of the coupon payments are used as weights, and it will in a sense provide you with the "mean time to coupon payment". As such it is an important concept, and it also acts a measure of the sensitivity of the bond price w.r.t. changes in the yield. This is shown by the following obvious result.

Proposition 2.25 *With notation as above we have*

$$\frac{dp}{dy} = \frac{d}{dy}\Big(\sum_{i=1}^{n} c_i e^{-yT_i}\Big) = -DP.$$

In a hedging language, we can see that duration is essentially for bonds (w.r.t. yield) what the Delta is for derivatives (w.r.t. the underlying price). The bond equivalent of the Gamma is convexity, which is defined as

$$C = \frac{\partial^2 P}{\partial y^2}.$$

2.3 Short rate models

2.3.1 Generalities

In this chapter we turn to the problem of how to model an arbitrage free family of zero coupon bond price processes $\{P(\cdot, T)|T \geq 0\}$. Since, at least intuitively, the price, $P(t, T)$, should in some sense depend upon the behavior of the short rate of interest over the interval $[t, T]$, a natural starting point is to give an a priori specification of the dynamics of the short rate of interest. This has in fact been the "classical" approach to interest rate theory, so let us model the short rate, under the objective probability measure $\mathcal{P}$, as the solution of an SDE of the form

$$dr(t) = \mu(t, r(t))dt + \sigma(t, r(t))dW(t). \tag{2.5}$$

The short rate of interest is the only object given a priori, so the only exogenously given asset is the money account, with price process B defined by the dynamics

$$dB(t) = r(t)B(t)dt. \tag{2.6}$$

As usual we interpret this as a model of a bank with the stochastic short rate of interest r. The dynamics of B can then be interpreted as the dynamics of the value of a bank account. To be quite clear let us formulate the above as a formalized assumption.

Assumption 2.26 *We assume the existence of one exogenously given (locally risk free) asset. The price B of this asset has dynamics given by (2.5), where the dynamics of r, under the objective probability measure $\mathcal{P}$, are given by (2.5).*

As in the previous chapter, we make an assumption to guarantee the existence of a sufficiently rich bond market.

Assumption 2.27 *We assume that there exists a market for zero coupon T-bonds for every value of T.*

We thus assume that our market contains all possible bonds (plus, of course, the risk free asset above). Consequently it is a market containing an infinite number of assets, but we again stress the fact that only the risk free asset is exogenously given. In other words, in this model the risk free asset is considered as the underlying asset whereas all bonds are regarded as derivatives of the "underlying" short rate r.

Our main goal is broadly to investigate the relationship which must hold in an arbitrage free market between the price processes of bonds with different maturities.

As a second step we also want to obtain arbitrage free prices for other interest rate derivatives such as bond options and interest rate swaps. Since we view bonds as interest rate derivatives it is natural to ask whether the bond prices are uniquely determined by the given r dynamics in (2.5) and the condition that the bond market shall be free of arbitrage. The answer to this question is fundamental:

bond prices are not uniquely determined
by the $\mathcal{P}$-dynamics of the short rate r.

In order to investigate this problem, let us start by viewing the bond market in the light of the meta-theorem 2.11. We see that in the present situation the number M of exogenously given traded assets excluding the risk free asset equals zero. The number R of random sources on the other hand equals one (we have one driving Wiener process). From the meta-theorem we may thus expect that the exogenously given market is arbitrage free but not complete. The lack of completeness is quite clear: since the only exogenously given asset is the risk free one we have no possibility of forming interesting portfolios. The only thing we can do on the a priori given market is simply to invest our initial capital in the bank and then sit down and wait while the portfolio value evolves according to the dynamics (2.6). It is thus impossible to replicate an interesting derivative, even such a simple one as a T-bond.

Another way of seeing this problem appears if we try to price a certain T-bond using a Black-Scholes' style technique. In order to imitate the classic argument we would assume that the price of a certain bond is of the form $F(t, r(t))$. Then we would like to form a risk free portfolio based on this bond and on the underlying asset. The rate of return of this risk free portfolio would then, by an arbitrage argument, have to equal the short rate of interest, thus giving us some kind of equation for the determination of the function F. Now, in the Black-Scholes model the underlying asset is the stock S, and at first glance this would correspond to r in the present situation. Here, however, we have the major difference between the Black-Scholes model and our present model. The short rate of interest r is not the price of a traded asset, i.e. there is no asset on the market whose price process is given by r. Thus it is meaningless to form a portfolio "based on r".

To sum up:

- The price of a particular bond will not be completely determined by the specification (2.5) of the r-dynamics and the requirement that the bond market is free of arbitrage.

- The reason for this fact is that arbitrage pricing is always a case of pricing a derivative in terms of the price of some underlying assets. In our market we do not have sufficiently many underlying assets.

We thus fail to determine a unique price of a particular bond. Fortunately this (perhaps disappointing) fact does not mean that bond prices can take any form whatsoever. On the contrary we have the following basic intuition.

Observation 2.28 *Prices of bonds with different maturities will have to satisfy certain internal consistency relations in order to avoid arbitrage possibilities on the bond market. If we take the price of one particular "benchmark" bond as given then the prices of all other bonds will be uniquely determined in terms of the price of the benchmark bond (and the r-dynamics).*

This fact is in complete agreement with the meta-theorem 2.11, since in the a priori given market consisting of one benchmark bond plus the risk free asset we will have $R = M = 1$ thus guaranteeing completeness.

2.3.2 The term structure equation

To make the ideas presented in the previous section more concrete we now begin our formal treatment.

Assumption 2.29 *We assume that there is a market for T-bonds for every choice of T and that the market is arbitrage free. We assume furthermore that, for every T, the price of a T-bond has the form*

$$P(t,T) = F(t, r(t), T),$$

where F is a smooth function of three real variables.

Conceptually it is perhaps easiest to think of F as a function of only two variables, namely r and t, whereas T is regarded as a parameter. Sometimes we will therefore write $F^T(t,r)$ instead of $F(t,r,T)$. The main problem now is to find out what F^T may look like on an arbitrage free market. Just as in the case of stock derivatives we have a simple boundary condition. At the time of maturity a T-bond is of course worth exactly 1 Euro, so we have the relation

$$F(T,r,T) = 1, \text{ for all } r.$$

Note that in the equation above the letter r denotes a real variable, while at the same time r is used as the name of the stochastic process for the short rate. To conform with our general notational principles we should really denote the stochastic process by a capital letter like R, and then denote an outcome of R by the letter r. Unfortunately the use of r as the name of the stochastic process seems to be so fixed that it cannot be changed. We will thus continue to use r as a name both for the process and for a generic outcome of the process. This is somewhat sloppy, but we hope that the meaning will be clear from the context. In order to implement the ideas above we will now form a portfolio consisting of bonds having different times of maturity. We thus fix two times of maturity S and T. From the last assumption and the Itô Formula we get the following price dynamics for the T-bond, with corresponding equations for the S-bond:

$$dF^T = F^T \alpha_T dt + F^T \sigma_T dW, \tag{2.7}$$

where, with subindexes r and t denoting partial derivatives,

$$\alpha_T = \frac{F_t^T + \mu F_r^T + \frac{1}{2}\sigma^2 F_{rr}^T}{F^T},$$

and

$$\sigma_T = \frac{\sigma F_r^T}{F^T}.$$

Denoting the relative portfolio by (u_S, u_T) we have the following value dynamics for our portfolio:

$$dV = V\left(u_T \frac{dF^T}{F^T} + u_S dF^S F^S\right),$$

and inserting the differential from (2.7), as well as the corresponding equation for the S-bond, gives us, after some reshuffling of terms,

$$dV = V(u_T\alpha_T + u_S\alpha_S)dt + V(u_T\sigma_T + u_S\sigma_S)dW. \tag{2.8}$$

We now define our portfolio in this way:

$$\begin{cases} u_T + u_S = 1, \\ u_T\sigma_T + u_S\sigma_S = 0. \end{cases}$$

With this portfolio the dW-term in (2.8) will vanish, so the value dynamics reduce to

$$dV = V u_T\alpha_T + u_S\alpha_S dt.$$

The system can easily be solved as

$$\begin{cases} u_T = -\frac{\sigma_S}{\sigma_T - \sigma_S}, \\ u_S = -\frac{\sigma_T}{\sigma_T - \sigma_S}, \end{cases}$$

and substituting this into dV we get

$$dV = V\Big(\frac{\alpha_S\sigma_T - \alpha_T\sigma_S}{\sigma_T - \sigma_S}\Big)dt.$$

Using the whole theory presented in [7], the assumption of no arbitrage implies that this portfolio must have a rate of return equal to the short rate of interest. Thus we have the condition

$$\frac{\alpha_S\sigma_T - \alpha_T\sigma_S}{\sigma_T - \sigma_S} = r(t), \text{ for all } t, \text{ with probability } 1,$$

or, written differently,

$$\frac{\alpha_S(t) - r(t)}{\sigma_S(t)} = \frac{\alpha_T(t) - r(t)}{\sigma_T(t)}.$$

The interesting fact about the last equation is that on the left hand side we have a stochastic process which does not depend on the choice of T, whereas on the right hand side we have a process which does not depend on the choice of S. The common quotient will thus not depend on the choice of either T or S, so we have thus proved the following fundamental result.

Proposition 2.30 *Assume that the bond market is free of arbitrage. Then there exists a process λ such that the relation*

$$\frac{\alpha_T(t) - r(t)}{\sigma_T(t)} = \lambda(t) \tag{2.9}$$

holds for all t and for every choice of maturity time T.

Observe that the process λ is universal in the sense that it is the same λ which occurs on the right hand side of (2.9) regardless of the choice of T. Let us now take a somewhat closer look at this process. In the numerator of (2.9) we have the term $\alpha_T(t) - r(t)$. By (2.7), $\alpha_T(t)$ is the local rate of return on the T-bond, whereas r is the rate of return of the risk free asset. The difference $\alpha_T(t) - r(t)$ is thus the risk premium of the T-bond. It measures the excess rate of return for the risky T-bond over the riskless rate of return which is required by the market in order to avoid arbitrage possibilities. In the denominator of (2.9) we have $\sigma_T(t)$, i.e. the local volatility of the T-bond.
Thus we see that the process λ has the dimension "risk premium per unit of volatility". The process λ is known as the market price of risk, and we can paraphrase the proposition in the same way:

In a no arbitrage market all bonds will, regardless of maturity time, have the same market price of risk.

Before we move on, a brief word of warning: the name "market price of risk" is in some sense rather appealing and reasonable, but it is important to realize that the market price of risk is not a price in the technical (general equilibrium) sense reserved for the word "price" in the rest of this work. The quantity λ is not something which we pay in order to obtain some commodity. Thus the usage of the word "price" in this context is that of informal everyday language, and one should be careful not to overinterpret the words "market price of risk" by assuming that properties holding for price processes in general equilibrium theory also automatically hold for the process λ.

We may obtain even more information from (2.9) by inserting our earlier formulas for α_T and σ_T. After some manipulation we then obtain one of the most important equations in the theory of interest rates, the so called "term structure equation". Since this equation is so fundamental we formulate it as a separate result.

Proposition 2.31 *In an arbitrage free bond market, F^T will satisfy the term structure equation*

$$\begin{cases} F_t^T + (\mu - \lambda\sigma)F_r^T + \frac{1}{2}\sigma^2 F_{rr}^T - rF^T &= 0, \\ \\ F^T(T,r) &= 1. \end{cases}$$

The term structure equation is obviously closely related to the Black-Scholes equation, but it is a more complicated object due to the appearance of the market price of risk λ. It follows from equations for α_T and σ_T and from (2.9) that λ is of the form $\lambda = \lambda(t,r)$, so the term structure equation is a standard PDE, but the problem is that λ is not determined within the model. In order to be able to solve the term structure equation we must specify λ exogenously just as we have to specify μ and σ. Despite this problem it is not hard to obtain a Feynman-Kac representation of F^T. This is done by fixing (t,r) and then using the process

$$exp\Big(-\int_t^s r(u)du\Big)F^T(s,r(s)).$$

If we apply the Itô Formula to this equation and use the fact that F^T satisfies the term structure equation, then we can obtain the following stochastic representation formula.

Proposition 2.32 *Bond prices are given by the formula $P(t,T) = F(t,r(t),T)$ where*

$$F(t,r,T) = \mathbb{E}^{\mathcal{Q}}_{t,r}\Big(e^{-\int_t^T r(s)ds}\Big). \tag{2.10}$$

Here the martingale measure $\mathcal{Q}$ and the subscripts t,r denote that the expectation shall be taken given the following dynamics for the short rate.

$$\begin{cases} dr(s) &= (\mu - \lambda\sigma)ds + \sigma dW(s), \\ \\ r(t) &= r. \end{cases}$$

The formula (2.10) has the usual natural economic interpretation, which is most easily seen if we write it as

$$F(t, r, T) = \mathbb{E}_{t,r}^{\mathcal{Q}}\Big(e^{-\int_t^T r(s)ds} \times 1\Big).$$

We see that the value of a T-bond at time t is given as the expected value of the final payoff of one Euro, discounted to present value. The deflator used is the natural one, namely $exp\Big(-\int_t^T r(s)ds\Big)$, but we observe that the expectation is not to be taken using the underlying objective probability measure $\mathcal{P}$. Instead we must, as usual, use the martingale measure $\mathcal{Q}$ and we see that we have different martingale measures for different choices of λ.

The main difference between the present situation and the Black-Scholes setting is that in the Black-Scholes model the martingale measure is uniquely determined. It can be shown that the uniqueness of the martingale measure is due to the fact that the Black-Scholes model is complete. In the present case our exogenously given market is not complete, so bond prices will not be uniquely determined by the given $\mathcal{P}$-dynamics of the short rate r. To express this fact more precisely, the various bond prices will be determined partly by the $\mathcal{P}$-dynamics of the short rate of interest, and partly by market forces. The fact that there are different possible choices of λ simply means that there are different conceivable bond markets all of which are consistent with the given r-dynamics. Precisely which set of bond price processes will be realized by an actual market will depend on the relations between supply and demand for bonds in this particular market, and these factors are in their turn determined by such things as the forms of risk aversion possessed by the various agents on the market. In particular this means that if we make an ad hoc choice of λ (e.g. such as λ=0) then we have implicitly made an assumption concerning the aggregate risk aversion on the market.

We can also turn the argument around and say that when the market has determined the dynamics of one bond price process, say with maturity T, then the market has indirectly specified λ by (2.9). When λ is thus determined, all other bond prices will be determined by the term structure equation. Expressed in another way: all bond prices will be determined in terms of the basic T-bond and the short rate of interest. Again we see that arbitrage pricing always is a case of determining prices of derivatives in terms of some a priori given price processes.

There remains one important and natural question, namely how we ought to choose λ in a concrete case. This question will be treated in some detail later , and the moral is that we must go to the actual market and, by using market data, infer the market's choice of λ.

The bonds treated above are of course contingent claims of a particularly simple type; they are deterministic. Let us close this section by looking at a more general type of contingent

T-claim of the form

$$\mathcal{X} = \Phi(r(T)),$$

where Φ is some real valued function. Using the same type of arguments as above it is easy to see that we have the following result.

Proposition 2.33 *Let $\mathcal{X}$ be a contingent T-claim of the form $\mathcal{X} = \Phi(r(T))$. In an arbitrage free market the price $\Pi(t, \Phi)$ will be given as*

$$\Pi(t, \Phi) = F(t, r(t)),$$

where F solves the boundary value problem

$$\begin{cases} F_t + (\mu - \lambda\sigma)F_r + \frac{1}{2}\sigma^2 F_{rr} - rF &= 0, \\ \\ F(T, r) &= \Phi(r). \end{cases}$$

Furthermore F has the stochastic representation

$$F(t, r, T) = \mathbb{E}^{\mathcal{Q}}_{t,r}\Big[exp\Big(- \int_t^T r(s)ds\Big) \times \Phi(r(T))\Big],$$

where the martingale measure $\mathcal{Q}$ and the subscripts t, r denote that the expectation shall be taken using the following dynamics:

$$\begin{cases} dr(s) &= (\mu - \lambda\sigma)ds + \sigma dW(s), \\ \\ r(t) &= r. \end{cases}$$

2.4 Martingale models for the short rate

2.4.1 $\mathcal{Q}$-dynamics

Let us again study an interest rate model where the $\mathcal{P}$-dynamics of the short rate of interest are given by

$$dr(t) = \mu(t, r(t))dt + \sigma(t, r(t))dW.$$

As we saw in the previous section, the term structure (i.e. the family of bond price processes) will, together with all other derivatives, be completely determined by the general term structure equation

$$\begin{cases} F_t + (\mu - \lambda\sigma)F_r + \frac{1}{2}\sigma^2 F_{rr} - rF &= 0, \\ \\ F(T, r) &= \Phi(r). \end{cases} \tag{2.11}$$

as soon as we have specified the following objects:

- the drift term μ;

- the diffusion term σ;

- the market price of risk λ.

Consider for a moment σ to be given a priori. Then it is clear from (2.11) that it is irrelevant exactly how we specify μ and λ per se. The object, apart from σ, that really determines the term structure (and all other derivatives) is the term $\mu - \lambda\sigma$ in (2.11). Now, from the proposition for the general case in the last section, we recall that the term $\mu - \lambda\sigma$ is precisely the drift term of the short rate of interest under the martingale measure $\mathcal{Q}$. This fact is so important that we stress it again.

The term structure is completely determined by specifying the r-dynamics under the martingale measure $\mathcal{Q}$.

Instead of specifying μ and λ under the objective probability measure $\mathcal{P}$ we will henceforth specify the dynamics of the short rate r directly under the martingale measure $\mathcal{Q}$. This procedure is known as martingale modelling, and the typical assumption will thus be that r under $\mathcal{Q}$ has dynamics given by

$$dr(t) = \mu(t, r(t))dt + \sigma(t, r(t))dW(t),$$

where μ and σ are given functions. From now on the letter μ will thus always denote the drift term of the short rate of interest under the martingale measure $\mathcal{Q}$.
In the literature there are a large number of proposals on how to specify the $\mathcal{Q}$-dynamics for r. We now present a (far from complete) list of the most popular models. If a parameter is time dependent this is written out explicitly. Otherwise all parameters are constant.

Vasicek:	$dr = (b - ar)dt + \sigma dW(t),$
Cox-Ingersoll-Ross (CIR):	$dr = a(b - r)dt + \sigma\sqrt{r}dW,$
Dothan:	$dr = ardt + \sigma rdW,$
Black-Derman-Toy:	$dr = \phi(t)rdt + \sigma(t)rdW,$
Ho-Lee:	$dr = \phi(t)dt + \sigma dW,$
Hull-White (extended Vasicek):	$dr = (\phi(t) - a(t)r)dt + \sigma(t)dW,$

Between all this classical approaches, the CIR model in particular will be deeply investigated in the next chapters.

2.4.2 Inversion of the yield curve

Let us now address the question of how we will estimate the various model parameters in the martingale models above. To take a specific case, assume that we have decided to use the Vasicek model. Then we have to get values for a, b and σ in some way, and a natural procedure would be to look in some textbook dealing with parameter estimation for SDEs (see the list at the very beginning of the chapter). This procedure, however, is unfortunately completely nonsensical and the reason is as follows. We have chosen to model our r-process by giving the $\mathcal{Q}$-dynamics, which means that a, b and σ are the parameters which hold under the martingale measure $\mathcal{Q}$. When we make observations in the real world we are not observing r under the martingale measure $\mathcal{Q}$, but under the objective measure $\mathcal{P}$. This means that if we apply standard statistical procedures to our observed data we will not get our $\mathcal{Q}$-parameters. What we get instead is pure nonsense. This looks extremely disturbing but the situation is not hopeless. It is in fact possible to show that the diffusion term is the same under $\mathcal{P}$ and under $\mathcal{Q}$, so "in principle" it may be possible to estimate diffusion parameters using $\mathcal{P}$-data. From a purely theoretical point of view this fact is obvious: a Girsanov transformation will only affect the drift term of a diffusion but not the diffusion term.
When it comes to the estimation of parameters affecting the drift term of r we have to use completely different methods, in fact it is well known that

The market itself chooses the martingale measure.

Thus, in order to obtain information about the $\mathcal{Q}$-drift parameters we have to collect price information from the market, and the typical approach is that of inverting the yield curve which works as follows.

1. Choose a particular model involving one or several parameters. Let us denote the entire parameter vector by α. Thus we write the r-dynamics (under $\mathcal{Q}$) as

$$dr(t) = \mu(t, r(t), \alpha)dt + \sigma(t, r(t), \alpha)dW(t).$$

2. Solve, for every conceivable time of maturity T, the term structure equation

$$\begin{cases} F_t^T + \mu F_r^T + \frac{1}{2}\sigma^2 F_{rr}^T - rF^T &= 0, \\ \\ F^T(T, r) &= 1. \end{cases}$$

   In this way we have computed the theoretical term structure as $P(t, T, \alpha) = F^T(t, r, \alpha)$. Note that the form of the term structure will depend upon our choice of parameter vector. We have not made this choice yet...

3. Collect price data from the bond market. In particular we may today (i.e. at $t = 0$) observe $P(0, T)$ for all values of T. Denote this empirical term structure by $\{P^*(0, T)|T \geq 0\}$.

4. Now choose the parameter vector α in such a way that the theoretical curve $\{P(0, T, \alpha)|T \geq 0\}$ fits the empirical curve $\{P^*(0, T)|T \geq 0\}$ as well as possible (according to some objective function). This gives us our estimated parameter vector α^*.

5. Insert α^* into μ and σ. Now we have pinned down exactly which martingale measure we are working with. Let us denote the result of inserting α^* into μ and σ by μ^* and σ^* respectively.

6. We have now pinned down our martingale measure $\mathcal{Q}$, and we can go on to compute prices of interest rate derivatives, like, say, $\mathcal{X} = \Phi(r(T))$. The price process is then given by $\Pi(t, \Phi) = G(t, r(t))$ where G solves the term structure equation

$$\begin{cases} G_t + \mu^* G_r + \frac{1}{2}(\sigma^*)^2 G_{rr} - rG & = & 0, \\ G(T, r) & = & \Phi(r). \end{cases}$$

If the above program is to be carried out within reasonable time limits it is of course of great importance that the PDEs involved are easy to salve. It turns out that some of the models above are much easier to deal with analytically than the others, and this leads us to the subject of so called affine term structures.

2.4.3 Affine term structures

Definition 2.34 *If the term structure $\{P(t, T)|0 \leq t \leq T, T > 0\}$ has the form*

$$P(t, T) = F(t, r(t), T),$$

where F has the form

$$F(t, r, T) = e^{A(t,T) - B(t,T)r}, \tag{2.12}$$

and where A and B are deterministic functions, then the model is said to possess an affine term structure (ATS).

The functions A and B above are functions of the two real variables t and T, but conceptually it is easier to think of A and B as being functions of t, while T serves as a parameter. It turns out that the existence of an affine term structure is extremely pleasing from an analytical and a computational point of view, so it is of considerable interest to understand when such a structure appears. In particular we would like to answer the following question: for which choices of μ and σ in the $\mathcal{Q}$-dynamics for r do we get an affine term structure? We will try to give at least a partial answer to this question, and we start by investigating some of the implications of an affine term structure. Assume then that we have the Q-dynamics

$$dr(t) = \mu(t, r(t))dt + \sigma(t, r(t))dW(t),$$

and assume that this model actually possesses an ATS. In other words we assume that the bond prices have the form (2.12)above. Using (2.12) we may easily compute the various partial derivatives of F, and since F must solve the term structure equation (2.11), we thus obtain

$$A_t(t,T)[1 + B_t(t,T)]r - \mu(t,r)B(t,T) + \frac{1}{2}\sigma^2(t,r)B^2(t,T) = 0. \tag{2.13}$$

The boundary value $F(T, r, T) = 1$ implies

$$A(T,T) = B(T,T) = 0.$$

Equation (2.13) gives us the relations which must hold between A, B, μ and σ in order for an ATS to exist, and for a certain choice of μ and σ there may or may not exist functions A and B such that (2.13) is satisfied. Our immediate task is thus to give conditions on μ and σ which guarantee the existence of functions A and B solving (2.13).

Generally speaking this is a fairly complex question, but we may give a very nice partial answer. We observe that if μ and σ^2 are both affine (i.e. linear plus a constant) functions of r, with possibly time dependent coefficients, then equation (2.13) becomes a separable differential equation for the unknown functions A and B. Assume thus that μ and σ have the form

$$\begin{cases} \mu(t,r) &= \alpha(t)r + \beta(t), \\ \\ \sigma(t,r) &= \sqrt{\gamma(t) + \delta(t)}. \end{cases}$$

Then, after collecting terms, (2.13) transforms into

$$A_t(t,T) - \beta(t)B(t,T) + \frac{1}{2}\delta(t)B^2(t,T) - [1 + B_t(t,T) + \alpha(t)B(t,T)\frac{1}{2}\gamma(t)B^2(t,T)]r = 0.$$

This equation holds for all t, T and r, so let us consider it for a fixed choice of T and t. Since the equation holds for all values of r the coefficient of r must be equal to zero. Thus we have the equation

$$B_t(t,T) + \alpha(t)B(t,T) - \frac{1}{2}\gamma(t)B^2(t,T) = -1.$$

Since the r-term is zero we see that the other term must also vanish, giving us the equation

$$A_t(t,T) = \beta(t)B(t,T) - \frac{1}{2}\delta(t)B^2(t,T).$$

We may thus formulate the main result.

Proposition 2.35 *Assume that μ and σ are of the form*

$$\begin{cases} \mu(t,r) &= \alpha(t)r + \beta(t), \\ \\ \sigma(t,r) &= \sqrt{\gamma(t) + \delta(t)}. \end{cases}$$

Then the model admits an ATS of the form (2.12), where A and B satisfy the system

$$\begin{cases} B_t(t,T) + \alpha(t)B(t,T) - \frac{1}{2}\gamma(t)B^2(t,T) &= -1, \\ \\ B(T,T) &= 0. \end{cases} \tag{2.14}$$

$$\begin{cases} \beta(t)B(t,T) - \frac{1}{2}\delta(t)B^2(t,T) &= A_t(t,T), \\ \\ A(T,T) &= 0. \end{cases} \tag{2.15}$$

We note that equation (2.14) is a Riccati equation for the determination of B which does not involve A. Having solved equation (2.14) we may then insert the solution B into equation (2.15) and simply integrate in order to obtain A. An interesting question is if it is only for an affine choice of μ and σ^2 that we get an ATS. This is not generally the case, but it can fairly easily be shown that if we demand that μ and σ^2 are time independent, then a necessary condition for the existence of an ATS is that μ and σ^2 are affine. Looking at the list of models in the previous section we see that all models except the Dothan and the Black-Derman-Toy models have an ATS.

2.4.4 A probabilistic discussion

There are good probabilistic reasons why some of the models in the previous list are easier to handle than others. We see that the models of Vasicek, Ho-Lee and Hull-White all describe the short rate using a linear SDE. Such SDEs are easy to solve and the corresponding r-processes can be shown to be normally distributed. Now, bond prices are given by expressions like

$$p(0,T) = \mathbb{E}\Big(exp^{-\int_0^T r(S)ds}\Big),$$

and the normal property of r is inherited by the integral $\int_0^T r(s)ds$. Thus we see that the computation of bond prices for a model with a normally distributed short rate boils down to the easy problem of computing the expected value of a log-normal stochastic variable.

This purely probabilistic program can in fact be carried out for all the linear models above, but it turns out that from a computational point of view it is easier to solve the system of equations (2.14)-(2.15).
In contrast with the linear models above, consider for a moment the Dothan model. This model for the short rate is the same as the Black-Scholes model for the underlying stock, so one is easily led to believe that computationally this is the nicest model conceivable. This is, however, not the case. For the Dothan model the short rate will be log-normally distributed, which means that in order to compute bond prices we are faced with determining the distribution of an integral for $\int_0^T r(s)ds$ of log-normal stochastic variables. It is, however, a sad fact that a sum (or an integral) of log-normally distributed variables is a particularly nasty object, so this model leads to great computational problems. It also has the unreasonable property that the expected value of the money account equals plus infinity. As for the CIR model and the Hull-White extension, these models for the short rate are roughly obtained by taking the square of the solution of a linear SDE, and can thus be handled analytically. They are, however, quite a bit messier to deal with than the normally distributed models. From a computational point of view there is thus a lot to be said in favor of a linear SDE describing the short rate. The price we have to pay for these models is again the Gaussian property. Since the short rate will be normally distributed this means that for every t there is a positive probability that $r(t)$ is negative, and this is unreasonable from an economic point of view. For the Dothan model on the other hand, the short rate is log-normal and thus positive with probability 1. It is also possible to show that the CIR model will produce a strictly positive short rate process, as we will se in the next chapters.
We end this section with a comment on the procedure of calibrating the model to data described in the previous section. If we want a complete fit between the theoretical and the observed bond prices this calibration procedure is formally that of solving the system of equations

$$p(0, T, \alpha) = p^*(0, T) \text{ for all } T > 0.$$

We observe that this is an infinite dimensional system of equations (one equation for each T) with α as the unknown, so if we work with a model containing a finite parameter vector α (like the Vasicek model) there is no hope of obtaining a perfect fit. Now, one of the main goals of interest rate theory is to compute prices of various derivatives, like, for example, bond options, and it is well known that the price of a derivative can be very sensitive with respect to the price of the underlying asset. For bond options the underlying asset is a bond, and it is thus disturbing if we have a model for derivative pricing which is not even able to correctly price the underlying asset. This leads to a natural demand for models which can be made to fit the observed bond data completely, and this is the reason why the Hull-White model has become so popular. In this model (and related ones) we

introduce an infinite dimensional parameter vector a by letting some or all parameters be time dependent. Whether it is possible to actually solve the system $P(0, T, \alpha) = P^*(0, T)$ for a concrete model such as the Hull-White extension of the Vasicek model, and how this is to be done in detail, is of course not clear a priori but has to be dealt with in a deeper study. As an example, we will carry out this study for the Hull-White model in the next section. It should, however, be noted that the introduction of an infinite parameter, in order to fit the entire initial term structure, has its dangers in terms of numerical instability of the parameter estimates.
There is also a completely different approach to the problem of obtaining a perfect fit between today's theoretical bond prices and today's observed bond prices. This is the Heath-Jarrow-Morton approach which roughly takes the observed term structure as an initial condition for the forward rate curve, thus automatically obtaining a perfect fit. This model will be studied in the next chapters.

2.4.5 Some standard models

In this section we will apply the ATS theory above, in order to study some affine one factor models.

The Vasicek model

To illustrate the technique we now compute the term structure for the Vasicek model (see [66]), which is based on this SDE:

$$dr = (b - ar)dt + \sigma dW.$$

Before starting the computations we note that this model has the property of being mean reverting (under $\mathcal{Q}$) in the sense that it will tend to revert to the mean level b/a. The equations (2.14)-(2.15) become

$$\begin{cases} B_t(t,T) - aB(t,T) &= -1, \\ B(T,T) &= 0. \end{cases}$$

$$\begin{cases} bB(t,T) - \frac{1}{2}\sigma^2 B^2(t,T) &= A_t(t,T), \\ A(T,T) &= 0. \end{cases}$$

The first equation is, for each fixed T, a simple linear ODE in the t-variable. It can easily be solved as

$$B(t,T) = \frac{1}{a}\Big(1 - e^{-a(T-t)}\Big).$$

Integrating the other equation we obtain

$$A(t,T) = \frac{\sigma^2}{2}\int_t^T B^2(s,T)ds - b\int_t^T B(s,T)ds,$$

and, substituting the expression for B above, we obtain the following result.

Proposition 2.36 *In the Vasicek model, bond prices are given by*

$$P(t,T) = e^{A(t,T)-B(t,T)r(t)},$$

where

$$B(t,T) = \frac{1}{a}\Big(1 - e^{-a(T-t)}\Big),$$

and

$$A(t,T) = \frac{(B(t,T) - T + t)(ab - \frac{1}{2}\sigma^2)}{a^2} - \frac{\sigma^2 B^2(t,T)}{4a}.$$

The CIR model

The CIR model (see [18] and [19]) is much more difficult to handle than the Vasicek model, since we have to solve a Riccati equation. We now simply cite the following result, leaving an in depth analysis to the next chapters.

Proposition 2.37 *The term structure for the CIR model is given by*

$$F^T(t,r) = A(T-t)e^{-B(T-t)r},$$

where

$$\begin{aligned} B(x) &= \tfrac{2(e^{\gamma x})}{(\gamma+a)(e^{\gamma x}-1)+2\gamma}, \\ A(x) &= \left(\tfrac{2\gamma e^{(a+\gamma)(x/2)}}{(\gamma+a)(e^{\gamma x}-1)+2\gamma}\right)^{2ab/\sigma^2}, \end{aligned}$$

and

$$\gamma = \sqrt{a^2 + 2\sigma^2}.$$

Closed form expressions for zero coupon bonds are rather complicated, so we refer to the detailed discussion in chapter 4.

The Hull-White model

In this section we will make a fairly detailed study of a simplified version of the Hull-White model (see [41]). We take care on this model because we will come back on this approach later, to justify our idea about consistency, in the last chapters.
The $\mathcal{Q}$-dynamics of the short rate are given by

$$dr = (\phi(t) - ar)dt + \sigma dW(t),$$

where a and σ are constants while ϕ is a deterministic function of time. In this model we typically choose a and σ in order to obtain a nice volatility structure whereas ϕ is chosen in order to fit the theoretical bond prices $\{P(0,T)|T>0\}$ to the observed curve $\{P^*(0,T)|T>0\}$.
We have an affine structure so bond prices are given by

$$P(t,T)=e^{A(t,T)-B(t,T)r(t)},$$

where A and B solve

$$\begin{cases} B_t(t,T)-aB(t,T) &= -1, \\ \\ B(T,T) &= 0. \end{cases}$$

$$\begin{cases} \phi(t)B(t,T)-\frac{1}{2}\sigma^2B^2(t,T) &= A_t(t,T), \\ \\ A(T,T) &= 0. \end{cases}$$

The solutions to these equations are given by

$$B(t,T)=\frac{1}{a}\Big(1-e^{-a(T-t)}\Big),$$

and

$$A(t,T)=\int_t^T\Big(\frac{1}{2}\sigma^2B^2(s,T)-\phi(s)B(s,T)\Big)ds.$$

Now we want to fit the theoretical prices above to the observed prices and it is convenient to do this using the forward rates. Since, as showed in (2.19), there is the one-to-one correspondence between forward rates and bond prices, we may just as well fit the theoretical forward rate curve $\{f(0,T)|T>0\}$ to the observed curve $\{f^*(0,T)|T>0\}$, where of course f^* is defined by $f^*(t,T)=\frac{\partial log P^*(t,T)}{\partial T}$. In any affine model the forward rates are given by

$$f(0,T)=B_T(0,T)r(0)-A_T(0,T),$$

which, substituting A and B, becomes

$$f(0,T)=e^{-aT}r(0)+\int_0^T e^{-a(T-s)}\phi(s)ds-\frac{\sigma^2}{2a^2}(1-e^{-aT})^2.$$

Given an observed forward rate structure f^* our problem is to find a function ϕ which solves the equation

$$f^*(0,T)=e^{-aT}r(0)+\int_0^T e^{-a(T-s)}\phi(s)ds-\frac{\sigma^2}{2a^2}(1-e^{-aT})^2, \forall T>0.$$

One way of solving this equation is to write it as

$$f^*(0,T)=x(T)-g(T),$$

where x and g are defined by

$$\begin{cases} \dot{x} &= -ax(t) + \phi(t), \\ x(0) &= r(0). \end{cases}$$

$$g(t) = \frac{\sigma^2}{2a^2}(1 - e{-}at)^2 = \frac{\sigma^2}{s}B^2(0,t).$$

Now we have

$$\phi(T) = \dot{x}(T) + ax(T) = f_T^*(0,T) + \dot{g}(T) + ax(T),$$

hence

$$\phi(T) = f_T^*(0,T) + \dot{g}(T) + a(f^*(0,T) - g(T)), \tag{2.16}$$

so we have in fact proved the following result.

Lemma 2.38 *Fix an arbitrary bond curve $\{P^*(0,T)|T > 0\}$, subject only to the condition that $P^*(0,T)$ is twice differentiable w.r.t. T. Choosing ϕ according to (2.16) will then produce a term structure $\{P(0,T)|T > 0\}$ such that $P(0,T) = P^*(0,T)$ for all $T > 0$.*

By choosing ϕ according to (2.16) we have, for a fixed choice of a and σ, determined our martingale measure. Now we would like to compute the theoretical bond prices under this martingale measure, and in order to do this we have to substitute our choice of ϕ into the precedent equations. The result is as follows.

Proposition 2.39 *Consider the Hull-White model with a and σ fixed. Having inverted the yield curve by choosing ϕ according to (2.16) we obtain the bond prices as*

$$P(t,T) = \frac{P(0,T)}{P(0,t)} exp\Big(B(t,T)f^*(0,t) - \frac{\sigma^2}{4a}B^2(t,T)(1 - e^{-2at}) - B(t,T)r(t)\Big),$$

where B is given by

$$B(t,T) = \frac{1}{a}\Big(1 - e^{-a(T-t)}\Big),$$

2.5 Forward rate models

2.5.1 The Heath-Jarrow-Morton framework

Up to this point we have studied interest models where the short rate r is the only explanatory variable. The main advantages with such models are as follows.

- Specifying r as the solution of an SDE allows us to use Markov process theory, so we may work within a PDE framework.

- In particular it is often possible to obtain analytical formulas for bond prices and derivatives.

The main drawbacks of short rate models are as follows.

- From an economic point of view it seems unreasonable to assume that the entire money market is governed by only one explanatory variable.

- It is hard to obtain a realistic volatility structure for the forward rates without introducing a very complicated short rate model.

- As the short rate model becomes more realistic, the inversion of the yield curve described above becomes increasingly more difficult.

These, and other considerations, have led various authors to propose models which use more than one state variable. One obvious idea would, for example, be to present an a priori model for the short rate as well as for some long rate, and one could of course also model one or several intermediary interest rates. The method proposed by Heath-Jarrow-Morton in [39] is at the far end of this spectrum:

in the HJM model the entire forward rate curve is the (infinite dimensional) state variable.

We now turn to the specification of the Heath-Jarrow-Morton framework. We start by specifying everything under a given objective measure $\mathcal{P}$.

Assumption 2.40 *We assume that, for every fixed $T > 0$, the forward rate $f(\cdot, T)$ has a stochastic differential which under the objective measure $\mathcal{P}$ is given by*

$$\begin{aligned} df(t,T) &= \alpha(t,T)dt + \sigma(t,T)dW(t), \\ f(0,T) &= f^*(0,T), \end{aligned} \tag{2.17}$$

where W is a (d-dimensional) $\mathcal{P}$-Wiener process whereas $\alpha(\cdot, T)$ and $\sigma(\cdot, T)$ are adapted processes.

Note that conceptually (2.17) is one stochastic differential in the t-variable for each fixed choice of T. The index T thus only serves as a "mark" or "parameter" in order to indicate which maturity we are looking at. Also note that we use the observed forward rate curve $\{f^*(0,T)|T \geq 0\}$ as the initial condition. This will automatically give us a perfect fit between observed and theoretical bond prices at $t = 0$, thus relieving us of the task of inverting the yield curve.

Observation 2.41 *It is important to observe that the HJM approach to interest rates is not a proposal of a specific model. It is instead a framework to be used for analyzing interest rate models. Every short rate model can be equivalently formulated in forward rate terms, and for every forward rate model, the arbitrage free price of a contingent T-claim $\mathcal{X}$ will still be given by the pricing formula*

$$\Pi(0, \mathcal{X}) = \mathbb{E}\Big[exp\Big(\int_0^T r(s)ds\Big) \cdot \mathcal{X}\Big],$$

where the short rate as usual is given by $r(s) = f(s,s)$.

Suppose now that we have specified α, σ and $\{f^*(0,T)|T \geq 0\}$. Then we have specified the entire forward rate structure and thus, by the relation

$$P(t,T) = exp\Big(-\int_t^T f(t,s)ds\Big),$$

we have in fact specified the entire term structure $\{P(t,T)|T > 0, 0 \leq t \leq T\}$. Since we have d sources of randomness (one for every Wiener process), and an infinite number of traded assets (one bond for each maturity T), we run a clear risk of having introduced arbitrage possibilities into the bond market. The first question we pose is thus very natural: how must the processes α and σ be related in order that the induced system of bond prices admits no arbitrage possibilities? The answer is given by the HJM drift condition below.

Theorem 2.42 *Assume that the family of forward rates is given by (2.17) and that the induced bond market is arbitrage free. Then there exists a d-dimensional column-vector process*

$$\lambda(t) = [\lambda_1(t), \ldots, \lambda_d(t)]^T$$

with the property that for all $T \geq 0$ and for all $t \leq T$, we have

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,s)^T ds - \sigma(t,T)\lambda(t). \tag{2.18}$$

In these formulas ^{T} denotes transpose.

Proof See [7]. □

2.5.2 Martingale modelling

We now turn to the question of martingale modelling, and thus assume that the forward rates are specified directly under a martingale measure $\mathcal{Q}$ as

$$\begin{aligned} df(t,T) &= \alpha(t,T)dt + \sigma(t,T)dW(t), \\ f(0,T) &= f^*(0,T), \end{aligned} \tag{2.19}$$

where W is a (d-dimensional) $\mathcal{Q}$-Wiener process.
Since a martingale measure automatically provides arbitrage free prices, we no longer have a problem of absence of arbitrage, but instead we have another problem. This is so because we now have the following two different formulas for bond prices

$$P(0,T) = exp\Big(- \int_0^T f(0,s)ds \Big),$$

$$P(0,T) = \mathbb{E}^{\mathcal{Q}}\Big[exp\Big(- \int_0^T r(s)ds \Big) \Big],$$

where the short rate r and the forward rates fore connected by $r(t) = f(t,t)$. In order for these formulas to hold simultaneously, we have to impose some sort of consistency relation between α and σ in the forward rate dynamics. The result is the famous Heath-Jarrow-Morton drift condition.

Theorem 2.43 *Under the martingale measure $\mathcal{Q}$, the processes α and σ must satisfy the following relation, for every t and every $T \geq t$.*

$$\alpha(t,T) = \sigma(t,T) \int_t^T \sigma(t,s)^T ds. \tag{2.20}$$

Proof See [7]. □

The moral of the HJM theorem is that when we specify the forward rate dynamics (under $\mathcal{Q}$) we may freely specify the volatility structure. The drift parameters are then uniquely determined. An "algorithm" for the use of an HJM model can be written schematically as follows.

1. Specify, by your own choice, the volatilities $\sigma(t,T)$.

2. The drift parameters of the forward rates are now given by

$$\alpha(t,T) = \sigma(t,T) \int_t^T \sigma(t,s)^T ds.$$

3. Go to the market and observe today's forward rate structure

$$\{f^*(0,T) | T \geq 0\}.$$

4. Integrate in order to get the forward rates as

$$f(t,T) = f^*(0,T) + \int_0^t \alpha(s,T)ds + \int_0^t \sigma(s,T)dW(s). \tag{2.21}$$

5. Compute bond prices using the formula

$$P(t,T) = exp\Big(- \int_t^T f(t,s)ds \Big).$$

6. Use the results above in order to compute prices for derivatives.

To see at least how part of this machinery works we now study the simplest example conceivable, which occurs when the process σ is a deterministic constant.
With a slight abuse of notation let us thus write $\sigma(t,T) \equiv \sigma$, where $\sigma > 0$. Equation (2.20) gives us the drift process as

$$\alpha(t,T) = \sigma \int_t^T \sigma ds = \sigma^2 (T-t),$$

so equation (2.21) becomes

$$f(t,T) = f^*(0,T) + \int_0^t \sigma^2(T-s)ds + \int_0^t \sigma dW(s),$$

that is

$$f(t,T) = f^*(0,T) + \sigma^2 t\Big(T - \frac{t}{2}\Big) + \sigma W(t).$$

In particular we see that r is given as

$$r(t) = f(t,t) = f^*(0,t) + \sigma^2 \frac{t^2}{2} + \sigma W(t),$$

so the short rate dynamics are

$$dr(t) = (f_T(0,t) + \sigma^2 t)dt + \sigma dW(t),$$

which is exactly the Ho-Lee model, fitted to the initial term structure. Observe in particular the ease with which we obtained a perfect fit to the initial term structure.

2.5.3 The Musiela parametrization

In many practical applications it is more natural to use time to maturity, rather than time of maturity, to parameterize bonds and forward rates. If we denote running time by t, time of maturity by T, and time to maturity by x, then we have $x = T - t$, and in terms of x the forward rates are defined as follows.

Definition 2.44 *For all $x \geq 0$ the forward rates $\tilde{f}(t,x)$ are defined by the relation*

$$\tilde{f}(t,x) = f(t, t+x).$$

Suppose now that we have the standard HJM-type model for the forward rates under a martingale measure $\mathcal{Q}$

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)dW(t).$$

The question is to find the $\mathcal{Q}$-dynamics for $\tilde{f}(t,x)$, and we have the following result, known as the Musiela equation.

Proposition 2.45 *Assume that the forward rate dynamics under Q are given by (2.19). Then*

$$d\tilde{f}(t,x) = (\mathcal{F}\tilde{f}(t,x) + D(t,x))dt + \tilde{\sigma}(t,x)dW(t), \tag{2.22}$$

where

$$\begin{aligned} \tilde{\sigma}(t,x) &= \sigma(t,t+x), \\ D(t,x) &= \tilde{\sigma}(t,x)\int_0^x \tilde{\sigma}(t,s)^T ds, \end{aligned}$$

and

$$\mathcal{F} = \frac{\partial}{\partial x}.$$

Proof See [7]. □

The point of the Musiela parametrization is that it highlights equation (2.22) as an infinite dimensional SDE. It has become an indispensable tool of modern interest rate theory.

Chapter 3

The problem

3.1 Public Debt Management

Debt strategy is defined as the manner in which a government finances an excess of government expenditures aver revenues and any maturing debt issued in previous periods. The question concerns the best way for the government to borrow these required funds. Should it, for example, use short-term debt, such as treasury bills or longer-term coupon bonds? Interestingly, an extensive academic literature on this subject does not exist.
Our analysis is based on the belief that a sustainable and prudent debt structure is critical for any sovereign nation. Moreover, we take the government's fiscal policy as given and attempt to characterize the set of financing strategies that have desirable risk-cost characteristics. Indeed, our primary objective is to learn more about the nature of the risk and cost trade-offs associated with different financing strategies.
The practitioner literature relating to better understanding this issue is found in publications from sovereign debt managers (see for example [12]).
Adopting this pragmatic perspective, we prove that one can conceptualize the government's borrowing decision as an optimal-control problem in a stochastic setting.
This problem has been extensively studied in the asset-pricing setting where an investor attempts to optimally select the proportion of risky and riskless assets that maximize their expected utility subject to appropriate wealth constraints (see, among the others, [46]).
In our situation, the government is attempting to optimally select the composition of its debt portfolio to minimize expected debt costs subject to risk and liquidity constraints.
Given practical complexities, however, it is not obvious how to use dynamic programming techniques to find a solution. Instead, we rely on simulation. We have also found that a simulation methodology termed dynamic financial analysis in the actuarial science literature is relevant for this task. Insurers are often faced with the problem of trying to set premiums and capital reserves, given stochastically evolving claims and investment

returns. Structurally, the techniques used in dynamic financial analysis are relevant for our work in debt strategy analysis.
Because our approach to debt strategy analysis involves stochastic simulation, another objective is to present the details of this simulation framework. Moreover, it is our view that management of the government's borrowing program is an important and difficult task requiring a combination of judgment and comprehensive analytical tools.

3.2 Description of the problem

We now introduce in a more formal way the problem studied in the scientific collaboration between the Italian Treasury Ministry and the Institute for the applications of Calculus of the Italian National Researches Council. We basically follows the working paper [1].

The Growth and Stability Pact (GSP), subscribed by the countries of the European Union (EU) in Maastricht, defines "sound and disciplined public finances" as an essential condition for strong and sustainable growth with improved employment creation. Since in Italy the expenses for interest payments on Public Debt is about 13% of the Budget Deficit (that is the difference between revenues and expenditures) the Public Debt Management Division of the Italian Ministry of Economy and Finance is deeply interested in studying which securities to issue, in order to achieve an optimal debt composition.

The goal is to determine the composition of the portfolio issued every month which minimizes a predefined cost function.

This can be, for instance, the width of fluctuations of deficit over a given time horizon or the interest expenses.

Mathematically speaking, this is a stochastic optimal control problem

with several constraints imposed by national and supranational regulations and by market practices. Among the former, for example, the Stability and Growth Pact rules require that:

- the Budget Deficit, has to be below 3% of Gross Domestic Product (GDP) (i.e. the total output of the economy).

- The Nominal Debt, that is the nominal amount of securities issued to finance the Budget Deficit, has to be less than 60% of the GDP.

- Countries should have an inflation rate within 1.5% of the three EU countries with the lowest rate.

- Long-term interest rates must be within 2% of the three lowest interest rates in EU.

Moreover, there are a number of other constraints such as the amount of money in the Treasury Cash Account. The complexity of the problem is further increased by the need for realistic solutions to take into account several side issues, like macroeconomic factors which are complicated as well, see [22].
The stochastic component of the problem is represented by the evolution of interest rates and Primary Budget Surplus (PBS) (Primary Balance is defined as domestic revenue minus domestic expenditure excluding interest payments).
Once a scenario for the evolution of these variables is set-up, the portfolio optimization can be formulated as a finite dimensional Linear Programming problem, neglecting some nonlinear effects of the bond issuances (for instance, a variation of the portfolio composition might trigger, by market reaction, a change in the term structure of the interest rate).
By means of standard methods (i.e. the simplex method, see [21]) we determine an optimal issuance strategy for each scenario.
The selection of the optimal strategy among the many optimal portfolios turns out to be a major problem. For example, it is likely that a combination of portfolios does not fulfill all the constraints (like the refunding of the expired securities).
Note that the Government announces the expected expenditure for the payment of interests in the yearly Financial Law (*Legge Finanziaria* in Italian) that is essentially the expected balance of the State for the following year. Therefore we need to provide strong probability estimates for our optimal control problem.
We thus turned the attention to iterative control algorithms to deal with scenario realizations. In engineering literature iterative control methods, called Model Predictive Control (MPC), have been successfully used in presence of disturbances, uncertainties and strict control and state constraints. The main difference of our framework is the presence of dominant stochastic behaviors, but the same techniques can be adapted to deal with that. The use of MPC allows us to obtain reliable probability estimates for the cost function opposed to predefined strategies that appear much less reliable.

3.3 The model

At present, the Italian Treasury Department issues ten different types of securities including one with floating rate. The securities differ in the maturity (or expiration date) m_k and in the rules for the payment of interests.

- The *Buoni Ordinari del Tesoro* (BOT) do not have coupons. From the accounting viewpoint the issuing price p is determined with a discount factor d: $p = 100 - d$, i.e. at the maturity date the nominal value 100 is reimbursed.

- The *Certificati del Tesoro Zero-coupon* (CTZ), like BOTs, do not have coupons. The issuing price is determined in such a way that the interests are comprised in the reimbursement $p(1 + r) = 100$.

- Both the *Buoni del Tesoro Poliennali* (BTP) and the *Certificati di Credito del Tesoro* (CCT) pay cash dividends by means of coupons corresponded every 6 months. The difference among them lies in the rate of interest (i.e. the value of the coupon) that is set at issuance time for BTPs whereas is variable for CCTs. More precisely, the interest rate for CCTs is determined by the interest rate for the 6-month BOTs.

For each of these four types of bonds we make a further distinction depending on the maturity. We order the bond types with an integer k taking values in $K = \{1, \ldots, 10\}$. Moreover we indicate by m_k the maturity in months of k. The issuance dates depend on the type of bond and we indicate them by a couple (d, m), where d is the day and m the month. In synthesis we have:

- k=1, BOT $m_1 = 3$ issuance dates: $(15, m),\ m = 1, \ldots, 12$;

- k=2, BOT $m_2 = 6$, issuance dates: $(30, m)$ or $(28, m),\ m = 1, \ldots, 12$;

- k=3, BOT $m_3 = 12$, issuance dates: $(15, m),\ m = 1, \ldots, 12$;

- k=4, CTZ $m_4 = 24$, issuance dates: $(15, m),\ m = 1, \ldots, 12$;

- k=5, BTP $m_5 = 36$, issuance dates: $(15, m),\ m = 1, \ldots, 12$;

- k=6, BTP $m_6 = 60$, issuance dates: $(15, m),\ m = 1, \ldots, 12$;

- k=7, BTP $m_7 = 120$, issuance dates: $(1, m),\ m = 1, \ldots, 12$;

- k=8, BTP $m_8 = 180$, issuance dates: $(15, m),\ m = 2, 3, 6, 7, 10, 11$;

- k=9, BTP $m_9 = 360$, issuance dates: $(15, m),\ m = 1, 3, 5, 7, 9, 11$;

- k=10, CCT $m_{10} = 84$, issuance dates: $(1, m),\ m = 1, \ldots, 12$;

By bonds' portfolio we mean the collection of bonds issued by the Italian Treasury that are still on the market, that is bonds that have not reached their maturity.

3.3.1 Cash flow for a single bond and for the portfolio

Let $u_k(t)$ be the money collected with the issuance, at time t, of bonds of k type, $p_k(t)$ the unit price and $c_k(s,t)$ the coupon percentage at time s for the same bond. For each bond there is an income of $p_k(t)$ at issuance time t, a payment of the nominal value that we set as equal to 100 at maturity $t + m_k$ and possibly payments of $100\, c_k(s,t)$ of coupons for all times s between the issuance date and maturity. Thus for a single bond we obtain the cash flow at time s:

$$R_k(s,t) = \delta_t(s) p_k(t) - 100 \left[\delta_{t+m_k}(s) + \sum_{\ell=1}^{m_k/6} \delta_{t+6\ell}(s) c_k(\ell, t) \right],$$

where the function $\delta_\tau(s) = 1$ if $s = \tau$ and 0 otherwise. Similarly we derive the cash flow for the whole portfolio:

$$\text{Flow}(s) = \sum_{k \in K} \sum_{t=s-m_k}^{s} \frac{u_k(t)}{100} R_k(s,t). \qquad (3.1)$$

3.3.2 Treasury Cash Account and Primary Budget Surplus

The cash flow of bonds' issuances and payments goes through a Bank of Italy account owned by the Treasury called Treasury Cash Account. There are some institutional positive lower bounds on the amount of money this account must have at the end of each

month (15 Euro billion). We indicate by TCA(s) the amount of money in the Treasury Cash Account at month s.
As to the PBS, any forecast is difficult due to many issues like seasonality and changes in the status of the economy. However, we assume that the PBS are defined every month and we indicate with PBS(s) the PBS at month s.

3.3.3 Constraints

Institutional constraints

A fundamental constraint is to guarantee the payment of coupons and the reimbursement of bonds at maturity:

$$\mathrm{TCA}(s) = \mathrm{TCA}(s-1) + \mathrm{Flow}(s) + \mathrm{PBS}(s) \geq \beta,$$

where β=15 Euro billion is fixed by the Italian law as explained in a previous paragraph. Note that PBS(s) may be negative. The Yearly Net Issuance (YNI) measures the difference between the volume of bonds issued during the year and the volume of bonds reimbursed during the same year. There is a constraint on the YNI indicated by the Government in the *Legge Finanziaria* (LF). In formula

$$\sum_{s=1}^{12} \sum_{k \in K} \left[p_k(t_0 + s) \frac{u_k(t_0 + s)}{100} - u_k(t_0 + s - m_k) \right] \leq \eta,$$

where t_0 is the first month of the year and η is fixed by the LF. More precisely the above formula must be corrected for BOT with a 100 nominal value instead of an issuance price p_k.

The growth and stability pact constraints

The Nominal Debt is defined as:

$$D(s) = \sum_{k \in K} \sum_{t=s-m_k+1}^{s} u_k(t),$$

and consists of all the money the State will reimburse in the future for bonds reaching maturity. Then the GSP imposes:

$$\frac{D(s)}{\mathrm{GDP}(s)} \leq \alpha,$$

where $\alpha = 0.6$ for the 60% constraint imposed by the Maastricht treaty that Italy is committed to reach at a satisfactory pace. [1]

[1] Note that Italy does not fulfill, at this time, such requirement.

The market practice constraints

The Treasury needs to consider also the problem of market stability. For instance, the amount of short-term bonds determines the behavior of the corresponding market. If a significant variation of the nominal amount of a short term bill offered was proposed in a single issuance, the market would react with a major change of the issuance price.
As a consequence, there are institutional constraints on the composition of portfolio which can be classified as dynamic constraints for short term securities, namely BOT, and static constraints for the medium and long term ones, namely CTZ, BTP and CCT.
Thus for $k = 1, 2$ and 3 the dynamic constraint can be modeled as:

$$\begin{aligned} \tfrac{u_k(t)-u_k(t-m_k)}{u_k(t-m_k)} &\leq \Gamma_k, \\ \tfrac{u_k(t-m_k)-u_k(t)}{u_k(t-m_k)} &\leq \gamma_k, \end{aligned} \tag{3.2}$$

where the values of Γ_k, γ_k are determined by the Ministry officers relying on their experience and market knowledge. The static constraints for $k \geq 4$ are stated as:

$$\lambda_k \leq u_k(t) \leq \Lambda_k, \tag{3.3}$$

where λ_k and Λ_k are the minimum and maximum amounts of long term bonds of each issuance.

The risk constraints

The last constraint is related to the possibility of operating changes in the issuance strategy in case of interest rates shocks. For each bond of type k issued at time t we define its Refixing Period as:

$$T_k(t, s) = m_k - (s - t),$$

that is the remaining time to maturity. The CCT is considered as a six month bond.
The Weighted Refixing Period (WRP) of the whole portfolio is an average time to maturity of the portfolio with weights proportional to the issued quantities:

$$\text{WRP}(s) = \frac{\sum_{k \in K} \sum_{t=s-m_k}^{s} u_k(t)\, T_k(t, s)}{D(s)}.$$

Since $T_k(t, s)$ is the time after which a bond has to be re-payed with a (probably) different interest rate, the WRP is an estimate of the averaged time period in which the Ministry is protected against changes of interest rates.
For the zero coupon bonds (BOT and CTZ) the WRP is equivalent to the duration, whereas for BTP is the weighted average time to maturity and for CCT is the weighted

average coupon refixing time.
A flexible management of the Public Debt requires that:

$$\tau_{min} \leq \text{WRP}(s) \leq \tau_{Max},$$

for some fixed values τ_{min} and τ_{Max}.

3.3.4 The cost function: ESA95 and other possible choices

A reasonable cost function is the yearly cost of the Public Debt calculated according to the European System of Accounts [43] (ESA95).
Roughly speaking, the ESA95 criteria consider for each bond its total cost (coupons plus the difference between nominal value and issuance price) distributed over its existence period, namely from issuance to maturity. Thus the cost over a set year is measured by the cost of bonds only for those days that fall inside the considered year. For instance, a 12-month BOT issued on July 1st 2000 counts for one half of its cost for the year 2000 according to ESA95 criteria.
In formula:

$$\text{ESA95}([t_1,t_2]) = \sum_{k\in K}\sum_{t=t_1-m_k}^{t_2} \frac{u_k(t)}{100}\Big((100-p_k(t))\frac{[t_1,t_2]\cap[t,t+m_k]}{[t,t+m_k]} + \sum_{\ell=1}^{m_k/6} c_k(t,\ell)\frac{[t_1,t_2]\cap[t+6(\ell-1),t+6\ell]}{[t+6(\ell-1),t+6\ell]}\Big) \tag{3.4}$$

is the cost for the time period $[t_1, t_2]$.
We are now ready to state our main goal:

Definition 3.1 *The Optimal Issuance Strategy (OIS), is the problem of determining a strategy for the selection of Public Debt securities that minimizes, within a given probability, the expenditure for interest payment (according to the ESA95 criteria) and satisfies, at the same time, the constraints on Debt management.*

A number of other possible cost functions can be chosen as an indicator of the Debt behavior. For instance, the discounted Debt which can be defined as follows.
Consider the total amount to be payed by the Treasury after some fixed time t_0, that is all the negative parts in the cash flows $R_k(s,t)$ for $k \in K$, issuance dates $t \leq t_0$ and times $s > t_0$. We denote such negative parts $Q_k(s,t)$. Let $a(t_0, s-t_0)$ be the annual interest rate of a bond with maturity $s - t_0$ (months) issued at time t_0 and $M = max_{k\in K}\, m_k$. In formula, the discounted Debt at time t_0 is:

$$\sum_{s=t_0+1}^{t_0+M}\sum_{k\in K}\sum_{t=s-m_k}^{s}\frac{u_k(t)}{100}Q_k(s,t)\left(\frac{1}{(1+a(t_0,s-t_0))^{\frac{s-t_0}{12}}}\right).$$

3.4 Optimization and Linear Programming

Since issuances happen at fixed dates, once per month, we use a discrete time model of evolution. For the sake of simplicity, the time step is one month. For the months in which some types of securities are not issued, the corresponding quantities are set equal to zero. We indicate by X_t the total amount of bonds that are not expired at time t. Thus X_t must contain, for every $k \in K$, one component for every $s \in \{t - m_k, \dots, t-1\}$. The evolution of X_t is determined at each step by cancelling bonds reaching maturity and adding the just issued ones. For example, for $k = 1$, one has to remove from X_t the quantity of 3 months BOT issued at time $t-3$ and insert that issued at time t. Clearly this can be done by shifting the components of X_t and adding the new issuances, thus we can write:

$$X_{t+1} = AX_t + B\,U_t, \tag{3.5}$$

where A is a shift matrix, $U_t = (\frac{u_k(t)}{100})_{k \in K}$ is the vector of the new issuances and B is a sparse matrix. Hence we get a linear discrete time control system.
Note that the stochastic behavior of interest rates (short rate and forward rates) influences the Flow (3.1), hence the Treasury Cash Account constraints, and the cost function ESA95 (3.4). The latter is influenced also by the PBS.

3.4.1 Input and output data

To specify completely the control problem it is necessary to set the input and output data and the optimization horizon.
The input data consist of:

- Past issuances.
- Issuance data.
- Gross Domestic Product and PBS forecasts.

Pastissuances. If the optimization horizon starts at time t_0, then for every $k \in K$ it is necessary to know the quantities issued at all dates $t_0 - m_k, \dots, t_0 - 1$.
Issuance data. The Italian Treasury sets the dates of issuance for each type of bonds. These dates are set in advance, usually for the next two or three years, and are not part of the control problem.
GDP forecasts. This point is quite critical, since it is difficult to have reliable GDP

forecasts. At least, the Treasury must take into account the forecasts reported in the *Legge Finanziaria*.
The output data are represented by the number of bonds that, for each issuance, fulfill all the constraints and, at the same time, minimize the cost function. From these data it is possible to derive:

- The Yearly Net Issuance.
- The Public Debt cost defined according to the ESA95 criteria.
- The *duration* and WRP of the portfolio.

The duration of a portfolio of bonds is, from the issuer viewpoint, the weighted average of the maturity of all the outcome cash flows. The duration describes the exposure to parallel shifts in the yield curve and is a widely used indicator of the risk associated with a particular choice of a fixed income securities portfolio [44].
The final goal is to provide an "optimal issuance strategy". There are, at least, two possible choices:

1. define the most probable scenario for the interest rates evolution, determine the corresponding optimal strategy, estimate the consequences of applying this strategy to a set of other scenarios (this step is necessary since the forecast on the interest rates can be wrong);

2. employ an "adaptive" strategy based on the available information on interest rates at issuance date (using interest rate models) and estimate the outcoming costs on a wide set of scenarios.

We call these approaches:

1. Probabilistic (fixed) Strategy.

2. Model Predictive Control Strategy (by similarity with engineering control problems).

For the purposes of the Ministry, a reasonable optimization horizon is 5 years.

3.4.2 Optimal control

Beside input and output data given at initial and final time respectively, there are some input and output variables evolving in the optimization horizon.
In control jargon Nominal Debt, Flow and Treasury Cash Account can be seen as output variables of the control system (3.5) and in formula can be indicated by:

$$Y_t = Y(X_t, U_t, PBS(t), p(t,T)). \tag{3.6}$$

In fact, all these quantities are computable since X_t, U_t and the exogenous stochastic parameters $PBS(t)$ and $p(t,T)$ are known. Finally, we get:

Proposition 3.2 *The OIS consists of an optimal control problem for the system (3.5) with constraints on the outputs (3.6) and with a cost function defined according to the ESA95 specs (3.4). Both constraints and cost function depend on the stochastic exogenous variables $PBS(t)$ and $p(t,T)$.*

A wide literature for stochastic optimal control problem is available, e.g. see [67]. However, the large number of variables (some hundreds components) and the needs for strict estimate in terms of probability prevent the applicability of most techniques.

3.4.3 Fixed scenario optimization

In [1] it is proved that:

Proposition 3.3 *For a fixed term structure evolution $t \mapsto p(t,T)$ and a PBS realization $t \mapsto PBS(t)$, the optimization problem becomes a linear programming problem with linear constraints.*

To solve the problem we resorted to the classic Simplex Method [21]. In figure 3.1 we report a block diagram of the software package that we realized to manage all the phases of the optimization.

3.5 Interest Rate Scenarios and Monte Carlo simulations

Once the single scenario optimization has been solved we can study the behavior of optimal controls and costs via Monte Carlo simulations. Some interesting parameters as the spread between maximum and minimum costs can be easily obtained.
For numerical results and empirical tests see [1].

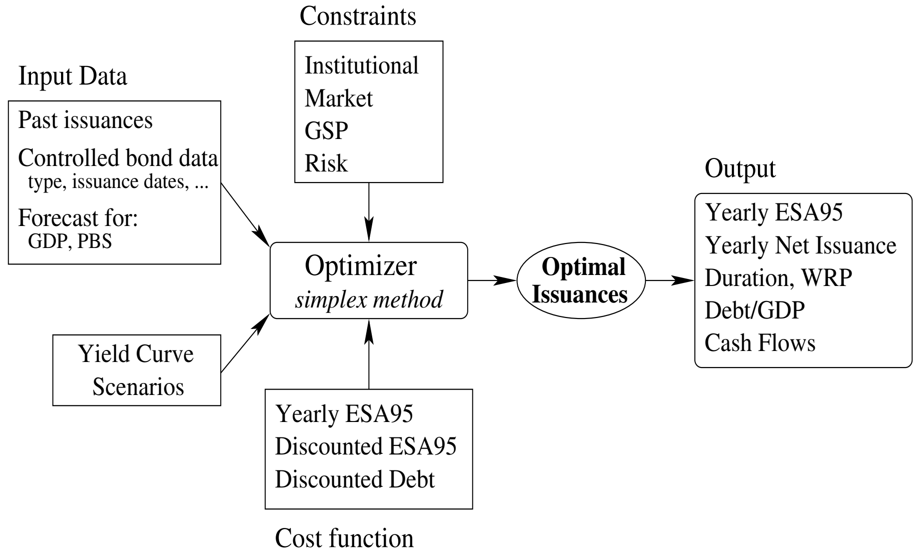

Figure 3.1: Block diagram of the optimization package.

3.6 Model Predictive Control strategies

It is well known that interest rate models have poor performance in forecasting rates behavior, thus we put the accent on advanced control techniques in order to reduce Debt risk.

In engineering literature an iterative strategy called Model Predictive Control (and/or Receding Horizon Control), briefly MPC, is often used in industrial applications for stabilization of systems under measurement uncertainties and disturbances, see [50]. This approach is particularly useful in case of hard constraints.

The basic idea of MPC is the following. In a discrete time setting, at step k, obtain an estimate up to an horizon $k + H$, $H > 0$, of the system behavior. Then choose a control according to some optimization criteria, such as optimal tracking of a benchmark trajectory. Finally, apply the obtained control and repeat the operation at next step.

Let us describe more precisely our application of MPC to the OIS problem. Fix a time window, say $[t_0, t_N]$ on which the evolution of the system is considered. Our procedure consists of the following steps:

Step 1. At a given issuance time t_j, in the time window $[t_0, t_N]$, we assume to know the term structure $p(t_j, T)$ for $T = t_j + m_k$ for $k \in K$, that is the rates for all bonds. Then we use some generator (predictor) $\tilde{p}(t, T)$ for the term structures at all times up to an optimization horizon H, i.e. up to $t_j + H$.

Step 2. We solve the OIS for the considered most probable scenario according to the generator $\tilde{p}(t,T)$. Alternatively, we can use a more sophisticated selection procedure for optimal portfolio but always based on the generator. This produces optimal issuance quantities $\tilde{u}_k(t_j,s)$ for all $s \geq t_j$ in the optimization window $[t_j, t_j+H]$.

Step3. We issue securities according to the found optimal values $\tilde{u}_k(t_j,t_j)$ and then we go back to Step 1 for the new issuance time t_{j+1}.

There are some key parameters as the optimization horizon H and the overall window $[t_0, t_N]$. However, the most interesting point is that the MPC strategy is more important than the choice of the generator $\tilde{p}$. More precisely, we show via simulations that the performance of an MPC strategy is much better, in probabilistic terms, than that of a fixed strategy for every choice of the interest rate forecast. Let us explain this in detail.
Fix the model described previously: let P_{Stat} be the portfolio selected by a strategy based on the most probable scenario. For each term structure scenario p_i, we indicate by $P^i_{MPC}(\tilde{p})$ the portfolio selected by the MPC strategy in case of scenario i. Notice that obviously $P^i_{MPC}(\tilde{p})$ does depend on the scenario because the procedure measures the actual rates at issuance date. Recalling the definition of P^i_{min} in the previous section, we evaluate the following quantities:

$$\frac{\mathrm{ESA95}\left(P_{Stat}\right) - \mathrm{ESA95}\left(P^i_{min}\right)}{\mathrm{ESA95}\left(P^i_{min}\right)},$$

and

$$\frac{\mathrm{ESA95}\left(P^i_{MPC}(\tilde{p})\right) - \mathrm{ESA95}\left(P^i_{min}\right)}{\mathrm{ESA95}\left(P^i_{min}\right)}.$$

Finally we consider a *reasonable* constant forecast for MPC strategy: at each issuance date t_j the generator simply replicates the actual term structure for all future times in the optimization horizon. The error is much smaller (see [1]).

3.6.1 Risk estimate

As explained at the very beginning, beside the mean value of the ESA95 cost, the Treasury must ensure the Debt performance with a very high probability.
Let us indicate by $\mathbb{P}$ the probability distribution over the set of scenarios for interest rates and by $\tilde{p}$ a fixed forecast.
Given a fixed probability level ℓ we can find a ESA95 cost level $C = C(\ell)$ such that

$$\mathbb{P}\left\{\mathrm{ESA95}\left(P^i_{min}\right) \leq C\right\} \geq \frac{1+\ell}{2}.$$

Then we find a percentage error level $\epsilon = \epsilon(\ell)$ such that

$$\mathbb{P}\left\{\frac{\text{ESA95}\left(P^i_{MPC}(\tilde{p})\right) - \text{ESA95}\left(P^i_{min}\right)}{\text{ESA95}\left(P^i_{min}\right)} \leq \epsilon\right\} \geq \frac{1+\ell}{2}.$$

Finally the ESA95 cost level can be set equal to $C \times (1+\epsilon)$, that is ensured with probability greater than or equal to ℓ. This method could, in principle, perform poorly, but for MPC strategies the error ϵ is extremely small, so such estimate is quite satisfactory.

Chapter 4

CIR model vs Nelson-Siegel model

This chapter is based on *Empirical Analysis of the Italian Treasury Fixed Income Securities Market*, a joint work with Luca Torosantucci.

4.1 Introduction

The paper [64] was born with the double aim **to perform an empirical analysis about the Italian Secondary Bond Market**, almost absent in literature, and **to give a practical scheme and suggestions to financial institutions**. Many financial institutions, in fact, find a lot of difficulties in the application of the interest rates models (for this kind of problems see, for example, the books [15] and [44]): these models are often characterized by some sophisticated mathematics and many times they were born for theoretical purposes. This work, instead, is an application of a theoretical scheme to an empirical problem so, it is an interesting example of how mathematics can help to make important decision in the real world.
The data, extracted by Datastream, are composed by the prices of bonds quoted from November 1999 to November 2000 in the Italian Treasury Fixed Income Securities Market.Datastream is one of the most famous financial database, very used by financial institutions. The main problem of Datastream is that it is only possible to download prices of bonds not expired yet. We will see that this fact produces some problems in the analysis of the evolution of the term structure.
In this chapter we implement from a practitioner point of view the Cox-Ingersoll-Ross (CIR) model. This model describes the evolution of the short rate with an only one factor of risk. Under this hypothesis the entire term structure can be determined by only one stochastic variable. In literature this kind of models are considered too poor to describe the term structure, but there are different reasons to consider the CIR model sufficient for the empirical purposes of this work.

Our results are compared with the results of a previous work of Barone-Cuoco-Zautzik (BCZ), dealing with the Italian Interest Rates Market in the period 1984-1989. To implement the term structure, BCZ chose to apply a procedure of data rejection for bonds with unreasonable returns. We believe that, in spite of the reasonableness of the BCZ approach, this procedure could be dangerous in the trading activity of a financial institution, and for this reason we choose to implement the term structure with all quoted bonds, without reject any data. We also prove that the choice of the method to implement the CIR model can influence the results.

It is important to stress that the analysis of the spreads between the market price and the theoretical price obtained by the CIR model can justify fundamentalist trading strategies. Moreover, the CIR model appears to be capable to record changes of the macroeconomic scenarios.

Finally, the CIR model is compared with the parsimonious model of Nelson and Siegel (NS), which is a model that implements the term structure postulating a particular functional form for the forward rates. The choice of the NS model is mainly due to the fact that it allows to the short, the medium and the long term to change in different ways, that is with different characteristic times. This property is very important because the CIR model, on the contrary, imposes a perfect correlation between the movements of the term structure at different maturities. So, the comparison between the NS model and the CIR model can stress which results depend on the method used to implement the term structure and which results are true characteristics of the market.

In our study we consider $(\Omega, (\mathcal{F}_t)_t, (W_t)_t, \mathbb{P})$ with Ω probability space, $(\mathcal{F}_t)_t$ natural filtration generated by the standard Brownian motion $(W_t)_t$ and $\mathbb{P}$ probability measure.

Definition 4.1 *$p(t,T)$ is the price at time t of a bond with maturity T.*

The relations between $p(t,T)$, the forward rate $f(t,T)$ and the short rate $r(t)$ are well known:

$$f(t,T) = -\frac{\partial \log(p(t,T))}{\partial T},$$

$$r(t) = f(t,t).$$

4.2 The CIR model

Several work has been done during the years in order to model the interest rates, as shown in the first chapter): the first approach is to give a stochastic evolution for the short term rate r_t, in order to derive the complete term structure from this only state variable. Between all models proposed in literature we choose the model proposed by Cox, Ingersoll and Ross in [18] and [19], which is the "natural" evolution of the standard Gaussian Vasicek model (see [66]).

The CIR model follows an "economical equilibrium" approach based on a stochastic differential equation of the type:

$$dr_t = \mu(t, r_t)dt + \sigma\sqrt{r_t}dW_t.$$

where μ is linear in r_t and σ is constant.
Equation (4.2) is joint with the so called Local Expectation Hypothesis:

$$\mathbb{E}(\frac{dP}{P}) = r_t + \lambda r_t \frac{P_r}{P}, \tag{4.1}$$

where $P(t, T)$ is the price at time t of a zero-coupon with maturity T. We slightly modify the drift coefficient choosing $\mu(t, r_t) = b - ar_t$ in order to obtain a mean-reverting structure, a typical feature of interest rates market. Ignoring the Brownian noise, the model would be purely deterministic and the resulting ordinary differential equation would admit an explicit solution, asymptotic to a particular value which represents the long term mean. To simplify, in the case of constants coefficients, the problem $r'(t) = b - ar(t)$ has this solution: $r(t) = \frac{b}{a} + \left(r_0 - \frac{b}{a}\right)e^{-at}$. Its asymptotic value, as $t \to +\infty$, is $\frac{b}{a}$; the standard interpretation of this value is the long period short rate.
Hence our equation is:

$$dr_t = k(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t. \tag{4.2}$$

It can be proved that the CIR model possesses an ATS (Affine Term Structure), that is the price of a pure discount bond can be written as:

$$p(t, T) = F(t, T)e^{G(t,T)r_t}, \tag{4.3}$$

where $F(t, T)$ and $G(t, T)$ are given by:

$$F(t, T) = [\frac{\phi_1 e^{\phi_2(T-t)}}{\phi_2(e^{\phi_1(T-t)} - 1) + \phi_1}]^{\phi_3},$$

$$G(t, T) = [\frac{(e^{\phi_1(T-t)} - 1)}{\phi_2(e^{\phi_1(T-t)} - 1) + \phi_1}].$$

Obviously ϕ_i, $_{i=1,2,3}$ depend on the parameters of the model:

$$\phi_1 = \sqrt{(k + \lambda)^2 + 2\sigma^2},$$

$$\phi_2 = \frac{k+\lambda+\phi_1}{2}, \tag{4.4}$$

$$\phi_3 = \frac{2k\mu}{\sigma^2}.$$

where $-\lambda$ is as usual the market price of risk.

4.2.1 Mathematical setting of the CIR model

In this paragraph we report two theorems about the CIR model. The first one ensures the existence and uniqueness of the solution of (4.2): we remark that this fact is not obvious, since the volatility $\sigma\sqrt{r_t}$ is not Lipschitz continuous, but only Hölder continuous.

Theorem 4.2 *For every $x \geq 0$ there exists a unique continuous adapted process r_t with values in $[0, +\infty)$, satisfying $r_0 = x$ and (4.2).*

Proof See [42]. □

The second theorem is crucial because it guarantees the positivity of the solution of (4.2). For this reason the CIR model substituted the classical Gaussian Vasicek model proposed in [66]. Let denote r_t^x the solution of (4.2) starting at x at $t = 0$ and define the stopping time $\tau_0^x = inf\{t \geq 0 | r_t^x = 0\}$.

Theorem 4.3 *If $k\mu \geq \frac{\sigma^2}{2}$ then $\mathbb{P}(\tau_0^x = \infty) = 1$ for all $x > 0$.*

Proof See [48]. □

To close this section, we want to remark an important classification result: in [16] and [23] it is proved an "inverse" result of the two theorems above.

Any nonnegative diffusion short rate process $r = r(t)_{t \geq 0}$
providing an ATS is necessarily of the CIR type.

This result explain the deep reason of the central role of this model in financial studies: in some sense, the CIR model is **"the"** model if one is in interested in simple stochastic processes...

4.2.2 Why the CIR model?

There are different reasons to use the CIR model in the Italian secondary bond market:

1. The CIR model is easy to implement and this property is very important for a financial institution: it desires easy, clear and quick instruments. For this reason, the CIR model is often used for practical purposes. As our aim is to help financial institutions, it is useful to start with instruments already used by banks and managers.

2. The CIR model is a one-factor model and this is sufficient for our purposes. One-factor models use a single variable (generally the short rate) to describe the entire yield curve. Usually, a one-factor model is written as $dr_t = \mu(r_t)dt + \sigma(r_t)dW_t$, where both drift $\mu(r_t)$ and volatility $\sigma(r_t)$ are functions of r_t but not of t. With this choice all maturities of the term structure are described by only one stochastic

variable. The assumption of a single factor of risk implies that all interest rates move in the same direction for every short time interval. In other words, one-factor models imply a perfect correlation between movements at different maturities. It is very important to understand that this characteristic does not mean that returns on every maturity change by the same amount each time, and it does not mean that the term structure must always have the same form. One-factor models can produce many different forms for the term structure curve so, in general, one factor of risk is not too restrictive. Nevertheless, the perfect correlation between bonds at different maturities (with the consequence that the long term rate is a deterministic function of the spot rate) is often criticized, so in literature two-factor models are generally preferred to one-factor models.

To explain why a one-factor model like the CIR is adequate for our aim, it is therefore fundamental to understand the necessities of a practitioner on the secondary bond market. Many financial papers examine very long historical series, even as long as 10 years. From a theoretical point of view it is necessary, because you must have many data to obtain reliable statistics, but from a practical point of view it could be a great problem. If a model is implemented over a period of 10 years, for instance, the resultant parameters of the model will describe an "average" of the different markets that follow each other. On the contrary, a practitioner needs a model that describes only the present situation, the present market, and not the mean of the market of the last years. So, it is very important, for practical purposes, to restrict our analysis to a shorter period. In general, a year, a year and a half of data, can be an excellent period, but also in this case it is necessary to be prudent.

When a model is implemented for practical purposes, it is necessary to use homogeneous historical data. We can not use, for example, a period that contains the crash of October 1987, or the crash of 1929, or the devaluation of the Italian Lira in 1992. During the period used to implement the CIR model, therefore, there must not be any shock in the market. If the period used to implement the model is shorter than 1 year and a half and if there is not any shock in that period, it would be reasonable to think that one-factor models are sufficient to describe the evolution of the term structure. This is also justified in two papers by Litterman and Scheinkman, see [49], and by Dybvig, see [28]: Litterman and Scheinkman showed that, historically, 80% of the yield curve movements are parallel shifts, while Dybvig explained that one-factor models offer an appropriate first-order approximation of the yield curve. For the Italian market situation, see also [35].

3. Pegoraro showed in [56] that the CIR model is a good model in determining the mis-pricing of bonds. In particular, the spread between the market price and the

theoretical CIR price is very useful to determine if a bond is under-priced or over-priced. This is a great information for a practitioner because, in this way, he can implement fundamentalist strategies on the secondary bond market. We will see that our work could provide further supports to Pegoraro's work.

4. The CIR model, in our form, has a mean-reversion behavior, a typical characteristic of interest rate movements. This means that for the short rate it is not possible to assume unreasonable high or low values. In particular, as t becomes very large, the mean reversion forces the short rate to fluctuate around the value of the parameter μ, named long-term average rate. As showed early, it is easy to see this fact analyzing the unconditioned mean of the short rate:

   $$\mathbb{E}(r_t) = \mu + e^{-kt}(r_0 - \mu).$$

   It is also possible to see that the speed of adjustment depends on k, so k is often called speed of adjustment.

5. The implied volatility of the CIR model depends on the square root of r_t. As seen before, under the condition $0 \leq k\mu < \frac{\sigma^2}{2}$, r_t is almost surely positive. Obviously, it is another essential characteristic of the interest rates.

The last two characteristics are not as important as what one would expect. If a practitioner wants to estimate the mis-pricing of a bond with the CIR model, he could implement the model every day and forecast the evolution of interest rates only for a few days. Pelliccia (see [57]), showed on our same dataset that the CIR model can follow the future evolution of the short rate only for a few days, less than a week.
On the other hand, Gerace in [36] showed analogous results also for models very different from the CIR model, as the HJM family. We believe substantially impossible to create a model capable of following perfectly the future evolution of interest rates; we believe that it is only possible to make statistical forecasts. Therefore, any model can be used for a few days only, but in a few days it is very difficult for the modelled short rate to reach unreasonable levels. For this reason, the mean reversion and the implied volatility are not so important in our analysis.

4.3 The Nelson-Siegel model

The NS model, see [54], is an example of a so called "parsimonious" model. The main idea behind these models is to postulate a unique functional form for the whole range of maturities for the discount function or, it is analogous, for particular rates as the forward or the spot rates. The NS model tries to capture some properties of the term structure thanks to the functional form imposed on the forward interest rates. The idea of Nelson and Siegel is trivial but ingenious:

the discount function is the derivative of the price and the forward rate is linked to the derivative of the discount function, hence the forward rate will be linked to the second derivative of the price.

Therefore, it is reasonable to think that the form of the forward rate will be similar to the solution of a second order differential equation:

$$f(t) = \beta_0 + \beta_1 e^{-\frac{t}{\tau}} + \beta_2 \frac{t}{\tau} e^{-\frac{t}{\tau}}. \tag{4.5}$$

The parameters of this model have a clear financial meaning, as can be seen here:

$$\begin{aligned} &\lim_{t\to+\infty} f(t) = \beta_0, \\ &\lim_{t\to 0} f(t) = \beta_0 + \beta_1, \\ &\lim_{t\to+\tau} f(t) = \beta_0 + \tfrac{\beta_1+\beta_2}{e}. \end{aligned} \tag{4.6}$$

Clearly β_0 is the long-term forward rate; the short-term forward rate is modulated by β_1 and the medium-term forward rate is modulated by β_2. Also the parameter τ has an interesting rule: the amount of time τ separates the short from the medium-long term. In a working paper on the Italian secondary bond market, Torosantucci shows that τ is generally included between 12 months and 18 months, and this result agrees with the practitioner's thought.

4.3.1 Why the Nelson-Siegel model?

There are different reasons to use the Nelson-Siegel model in the Italian secondary bond market:

1. The NS model is very easy to implement and, as we have seen, this property is very important for financial institutions.

2. The parameters of the model have a clear financial meaning and this is very useful for a practitioner.

3. The NS model determines a very stable term structure. It is possible to understand the sense of this sentence with an example: if there were not a shock between today

and tomorrow the bond prices would change, but it is reasonable to believe that the term structure of tomorrow will be very similar to the term structure of today. In other words, until it occurs a shock, the term structure will remain roughly constant. Hence, if a model is too sensitive to bond prices,i.e. if it generates completely different term structures one day after another even when there is not any shock, that model is not a good model for our purposes. The NS model generates stable term structure and so from a practical point of view it is a good model.

4. The yield curve generated by NS model does not "explode" in the long term. Many institutions generate the term structure with regression techniques like, for example, splines and bi-splines (see for example McCullogh in [52]). These techniques are very interesting and very easy to apply, but have a great problem: many times the curve explodes at long term, and it is absurd from a financial point of view. The NS model does not share this problem, because the forward rate, and as consequence the discount function $d(t)$, tends to an asymptotic constant, as can be seen in the following formula:

$$\lim_{t\to+\infty} d(t) = \lim_{t\to+\infty} e^{-t\left[\beta_0+(\beta_1+\beta_2)\frac{\tau}{t}\left(1-e^{-\frac{t}{\tau}}\right)-\beta_2 e^{-\frac{t}{\tau}}\right]} = e^{-(\beta_1+\beta_2)\tau}$$

   This consideration will be very useful at the end of this chapter.

5. For the Italian secondary bond market, Torosantucci showed that the characteristic times of the three parameters $\beta_i,\ _{i=1,2,3}$ are completely different from each other. In particular, the characteristic time of β_0 is greater than 1 year, the characteristic time of β_1 is about 2 days and the characteristic time of β_2 is about 6 months. Hence, different maturities of the NS term structure can have movements with different intensities and different directions. It is very interesting for our aim, because we will be able to compare the results of the CIR model, where the long rate is a deterministic function of the short rate, with the results of the NS model, where the long, the medium and the short term can move with different characteristic times.

4.4 Data Problems

We use an historical archive of 523 days composed by prices of the quoted bonds in Italian secondary market downloaded from Datastream, one of the most popular and used financial database. The time series starts at 11/20/1998 and finish at 11/8/2000. **The choice of Datastream is due to its relevant role in financial institutions** and for this reason we choose to front its data problems. The archive extracted by Datastream has in fact two problems:

- Rarely data can contain mistakes, probably due to transcription errors. This is a typical problem of all archives.

- It is possible to download from Datastream only prices of the "active bonds", which are bonds that have not expired yet. So, if the time series is 2 years long, it is not possible, for example, to have information about a BOT (Italian BOT are analogous to American Bills) expired one year ago. As a consequence, in the first part of our archive there are many "holes" in the short term. We will see that this lack of information causes some difficulties in interpreting the term structure, especially for some kinds of models like models that use regression techniques (Nelson-Siegel, McCulloch, etc...).

4.4.1 Hypothesis, choices and data rejection

We have exposed the problems related to our data. In order to solve them, our choices are:

1. We analyze only the last 274 days of the whole archive, from 11/4/1999 to 11/21/2000, in order to limit the lack of information on the short term. We choose to start from 11/4/1999 because it is the first day where we have a data on a BOT, that is a data on the short-term.

2. We define, for every day and for every bond:

$$\text{Spread} = \text{(Market Price) - (Theoretical CIR Price)}$$

3. Some bonds seem to give "unreasonable" returns, sometimes higher, sometimes lower than the market expectation. The problem is: is it right to reject all this kind of data? What method of rejection should we use? In the paper [4], BCZ adopt the following procedure of data rejection: if some bonds have spread on return greater than 2.57 times their standard deviation (we call these observations outliers), the observation with greater spread on return is rejected and the regression to determine the model parameters is repeated. The rejection procedure is repeated recursively, until all outliers are erased.
   There are different theoretical reasons to accept BCZ procedure, but is it justified if our aim is to compute the term structure to implement fundamentalist strategies? Obviously, outliers can produce a misleading term structure, but the market is also characterized by anomalous and significant data, due to liquidity problems. With their procedure, BCZ eliminate different bonds that are not benchmarks for

the market. For a practitioner it is probably not meaningful to follow this approach, because he also buys and sells illiquid bonds, and a term structure determined without outliers does not consider this kind of securities.
For these reasons, we decide to implement the term structure with all data of our archive (except, obviously, when there is a clear transcription mistake). We do not reject any data to implement the term structure, but as we will see later, we apply the Chauvenet principle (see [63]) to compute the moments of the distribution.

4.5 Computation of the term structure

For the reasons that will become clear at the end of this section, in order to implement the CIR model, we follow the procedure below:

1. For each day we consider the bond prices.

2. Since in the CIR model the price of a bond can be expressed as in (4.3), we compute the model parameters in 4.4 applying the non-linear least square method (by cross section). It should be remarked that it is possible to compute σ and $(k + \lambda)$, but not μ and separately k and λ. It is not a great problem for our aim.

3. We repeat the procedure for each day.

At the end of this procedure, we have three historical series: one for ϕ_1, one for ϕ_2 and one for ϕ_3. From these series it is possible to determine other two series: one for σ and one for $(k + \lambda)$. Therefore for each day we have a different value for each parameter. It is also important to note that the value of $\phi_i,\ _{i=1,2,3}$, and therefore of σ and $(k + \lambda)$, on a specific day depends only on the bond prices in that specific day.
Roughly speaking, we implement the model in a static way, and this choice might seem very strange. The CIR model is a dynamical model that describes the time evolution of the short rate: the "natural" method to implement it would be using a time series of a bond (or of a rate) chosen as benchmark for the short rate. In other words, the most logical procedure would be:

1. To take the time series of the bond or the rate chosen as a benchmark for the short rate, for example the evolution of the EURIBOR.

2. To compute the model parameters with the likelihood method or with the EMM method, or also with the GMM method.

3. At the end, there will be only one value for each model parameters and not a historical series, as in our procedure. In this case, the value of each parameter is the same for each day and depends on the evolution of the benchmark and not on the price of all quoted bonds on a specific day.

The question is: why, to implement the CIR model, do we chose a statical procedure instead of a dynamical one? There are three principal reasons:

1. In the first part of our archive we don't have any information about the short term. Therefore, it is impossible to determine a benchmark for it, and even if it is possible to have a benchmark, there would be two other problems.

2. If we use the historical series of a benchmark, we would obtain a unique value for each parameter describing the "average" market of that days. So, as in the case of a shock during the period of analysis, the model would not be able to introduce the necessary changes in the term structure.

3. Moreover, if we use the historical series of a benchmark, we would not be able to introduce all the information about all quoted bonds, and in particular about the illiquid bonds. This would be a worrying restriction for a practitioner.

We believe that for fund managers these reasons are sufficient to justify the use of a statical procedure. To sum up: for every day we compute the model parameters applying the non linear least squares method (by cross section) to the prices of bonds. We do not reject any data. So, for each day we obtained a value for each parameter.

4.6 Computation and analysis of spreads

Because our aim is to help a practitioner to implement fundamentalist strategies, that is to determine if a bond is mis-priced, we must analyze the spread between the market price and the theoretical CIR price. Now, we want to examine the statistical distribution of all spreads obtained from all bonds in every day. For this analysis we have to discuss further the problem of data rejection. If we want to compute only the statistical distribution, or better, the moments of the distribution, it is possible to reject all spreads which are too far from their mean (depending on their standard deviation). It is very important to understand that in this way the form of the term structure is not influenced by data rejection, because we reject the spreads after the computation of the term structure, and we do not reject historical data of our archive before to compute the term structures (like in BCZ procedure).

The difference between the two approaches is clear. If we analyze the moments of the distribution, we would have to study the general behavior of the market; otherwise, if we use the CIR model to implement a fundamentalist strategy, we would have to work on all bonds. So, in the first case it is useful to erase anomalous data, in the latter case this choice would not be correct. A priori we do not know which data are significant and which are not, but we can assume that almost all of the spreads which are too far from their mean are insignificant. What is the best procedure to reject data in the computation of the distribution moments of the spreads? There are different methods, but we chose the Chauvenet principle (see [63]).
The mean and the standard deviation of the spreads obtained applying the Chauvenet principle are:

$$\text{mean}_{TU} = -0.1412, \qquad \text{std}_{TU} = 0.4987. \tag{4.7}$$

The results of BCZ are:

$$\text{mean}_{BCZ} = -0.04, \qquad \text{std}_{BCZ} = 0.52. \tag{4.8}$$

BCZ suggest that their distribution of the spreads is normal and that their mean is non statistically different from zero.
We obtain completely different results, so we ask ourselves if this difference is due to the different market situations or to the analysis procedures. In order to solve this problem we apply the BCZ method on our same dataset, obtaining:

$$\text{mean}_{(\text{TU with BCZ method})} = -0.020, \qquad \text{std}_{(\text{TU with BCZ method})} = 0.247. \tag{4.9}$$

Our new results agree with BCZ ones, and therefore we prove that the differences in (4.7) and (4.8) are not due to different market situations, but to the methods used to implement the CIR model. In other words, **the choice of the method of implementation** is absolutely critical and **can heavily influence the results**.
Due to our empirical aim and for the reasons discussed in the previous paragraphs (especially the necessity to use all information), we thought it is better not to reject any data in the implementation of the model.

4.6.1 Computation of the spreads

The moments of our distribution are:

$$\text{mean}_{CIR} = -0.1412,$$

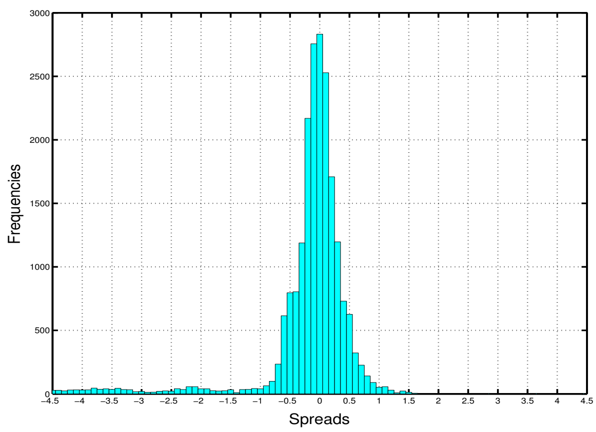

Figure 4.1: Distribution of the spreads with the CIR model.

$$\text{std}_{CIR} = 0.4987,$$
$$\text{skew}_{CIR} = -3.1371,$$
$$\text{kurt}_{CIR} = 16.1561$$

The observations on these results can be expressed in different points:

1. The skewness is significantly less than zero. As can be seen in figure 4.1, the distribution of the spreads shows a great asymmetry on the left of the curve.

2. The kurtosis is significantly higher than 1. So, the distribution of the spreads shows a higher peak and fat tails.

3. The number of total spreads is equal to 21601. Applying the Chauvenet principle we reject 1220 data, equal to 5.7% of the total.

4. A rough decomposition of the distribution of the spreads is: 27, 2% of the spreads $\in [-0.10; 0.10]$, and 81, 2% of the spreads $\in [-0.50; 0.50]$ (in BCZ respectively 25% and 30%).

5. The maximum positive spread is equal to $+4.033$ while the minimum negative spread is equal to -4.341. Also with this this simple information it is possible to note the asymmetry of the distribution.

6. The values of the kurtosis and of the skewness show a non normal behavior.

7. The hypothesis test of Kolmogorov-Smirnov rejects the normality hypothesis (this fact appears clearly also without the test, thanks to the value of the kurtosis and the skewness).

4.6.2 Consequences of the non-normality

Is the non-normality a problem? Is it a proof that the CIR model is not adequate to implement the term structure? The answer is no. Every bond has own life that depends on the entire market, but also on its characteristics. So, some bonds are greatly appreciated by investors and other bonds are not, that is there are very liquid bonds and other very illiquid bonds. In this case it is reasonable to think that some bonds are always over and other are always under the term structure. But, if it is true, the mean distance from the term structure, i.e. the mean of the spreads, could generally be different from 0, different from bond to bond, and consequently the distribution of the spreads of all bonds probably could be not normal.
Because we have proved that the distribution is not normal, we have also proved that some bonds are always over and some bonds are always under the term structure. This analysis has a direct consequence for a practitioner. If we suppose that the model used to implement the term structure is perfect, i.e. it can capture the real market evolution and therefore the right return for every maturity, a simplicistic and trivial fundamentalist strategy can be the following:

- when a bond has a market price greater (smaller) than the theoretical price, the bond is overpriced (under-priced).

- because it is reasonable to expect that under-priced bonds will go up and overpriced bonds will go down, the strategy is to buy underpriced and sell overpriced bonds.

But this strategy is wrong for the CIR model because in our analysis we have seen that some bonds are always over and other are always under the term structure. Consequently, it is not true that a market price that is less than the theoretical CIR price will increase in the future. In this environment, the approach must change and it is necessary to implement another strategy. For example, the practitioner could compute the mean distance between the market price and the theoretical CIR price of a particular bond (not of all bonds, as we have computed in the analysis of the distribution of the spreads). Only when the spread between a particular bond and the term structure is statistically different from its mean,

Short rate r_t					
Model and Period	Mean	St. Dev.	Rel. Var.	Min	Max
BCZ 1984-1989	11.519	3.156	27.4%	8.071	16.129
BCZ 1984	14.948	2.702	18.1%	13.295	16.129
TU 1999-2000	4.232	0.421	9.9%	3.255	5.114

Table 4.1: Short rate r_t

the bond is mis-priced. In particular: if the spread of the bond is statistically greater than its mean then the bond is overpriced, otherwise if the spread of the bond is statistically less than its mean the bond is under-priced. Pegoraro in [56] analyzed an analogous approach for the Italian secondary market between 1992 and 1993, and the results were very good. Our analysis enforces and encourages this type of approach, so we think that an in-depth study is necessary.

To conclude, we can say that the non-normality is not a problem for the CIR model: the non-normality also suggests to implement trading strategies based on the mean distance between the market price and the theoretical CIR price.

It is important to stress that the non-normality is true for the CIR model and for the Italian market in the historical period we analyzed (1999-2000). There is not any evidence that the non-normality is always present, also with different models, different markets and different historical periods. However, we think that it is possible that a strategy based on the mean distance between the market price and the theoretical CIR price could also be used in other markets and in other historical periods. There are some indications about these ideas in [56]. In his paper, in fact, Pegoraro applies with success a fundamentalist strategy in other historical periods and in particular before and after the devaluation of the Italian Lira, in 1992.

4.7 Analysis of the parameters of the model

In this paragraph we show the evolution of r_t, μ, k and σ. We also compare our results with the results in [4], obtained by BCZ for the years 1984-1989.

4.7.1 Evolution of the short rate r_t

In the second row of table 4.1 we report the results of BCZ about 1984, the year between 1984 and 1989 with the greatest mean of the short rate (in some sense the "worst" year). During the 1980s, the Italian macroeconomic situation was completely different from the

Implied volatility $\sigma\sqrt{r_t}$					
Model and Period	Mean	St. Dev.	Rel. Var.	Min	Max
BCZ 1984-1989	2.1004	0.3321	15.8%	1.1606	3.5880
BCZ 1984	1.9865	0.0397	2.0%	1.9092	2.0761
TU 1999-2000	0.8098	0.375	46.4%	0.2217	1.4807

Table 4.2: Implied volatility $\sigma\sqrt{r_t}$

present situation. In that period, in fact, the TUS of Banca d'Italia was greater than 10% while nowadays the Reference Rate of BCE is about 4%. It is very interesting to note that the CIR model can capture this change, in fact the mean of the short rate was 11.519% in 1984-1989 while it is 4.232% in 1999-2000 (see again table 4.1).
The relative variability, i.e the ratio between the standard deviation and the mean, in 1999-2000 is less than the relative variability in 1984-1989. This is a proof that today the Italian market is more stable, that is the investors economic expectations are more solid. Obviously, it is "necessary" that a term structure model captures the changes in the macroeconomic conditions. In other words, a good model have to capture the changes of scenarios but, on the contrary, if a model can capture the changes of macroeconomic scenarios it is not necessarily a good model! However, we think that these results are substantially a proof of the reasonableness of the CIR model.

4.7.2 Evolution of the implied volatility $\sigma\sqrt{r_t}$

In the second row of table 4.2 we report the results of BCZ about 1984, the year between 1984 and 1989 with the greatest mean of the short rate (in some sense the "worst" year). The implied volatility is very important because it describes the behavior of the stochastic part of the model. In a market where the expectations are stable, for example, this parameter is low. For this reason, the implied volatility of the period 1999-2000 is less than the implied volatility of the period 1984-1989. On the other hand, it is not definitively clear the reason why the relative variability of 1999-2000 is greater than the relative variability of 1984-1989. To understand this strange behavior it is probably necessary to repeat the analysis also on other markets and for other periods.

4.7.3 Evolution of the speed of adjustment k and of the average rate μ

If we choose the trivial case of null market price of risk, $\lambda = 0$, it is possible to explicate the parameter k and the parameter μ. In table 4.3 it can be seen that the average

Speed of adjustment k					
Model and Period	Mean	St. Dev.	Rel. Var.	Min	Max
BCZ 1984-1989	0.2433	0.0031	1.3%	0.2269	0.2524
TU 1999-2000	0.2804	0.1192	42.5%	0.0386	0.4653

Table 4.3: Speed of adjustment k

Long term average rate μ					
Model and Period	Mean	St. Dev.	Rel. Var.	Min	Max
BCZ 1984-1989	11.1280	1.8023	17.2%	4.261	16.837
TU 1999-2000	7.4701	1.2304	16.5%	6.192	14.454

Table 4.4: Long term average rate μ

of the speed of adjustment k for the period 1999-2000 is very similar to the speed of adjustment computed by BCZ for the period 1984-1989. This fact is very strange. We can not say that the Italian market (or even every market) has its own and constant speed of adjustment; we only want to remark the result and the necessity to carry out an in-depth study. However, if it were proved the constantness of the speed of adjustment k, it would be an important aspect characterizing the bond market. Finally, we want to stress the greater relative variability of k during 1999-2000: we don't know how to explain this value. The long-term average rate μ in the period 1999-2000 is less than the long-term

average rate in the period 1984-1989 (see table 4.4). As discussed before, obviously this is a consequence of the different Italian macroeconomic scenery.

4.8 Comparison between the CIR model and the NS model

As explained in the introduction, we also implement the term structure with the NS model. In the following four figures we report the evolution of the returns for four maturities: 3 months, 1 year, 5 years, 10 years. For example, in figure 4.2 we represent the evolution of the return for the three months maturity, extracted by the term structures of the NS model (dotted) and of the CIR model (solid).

The results are not always good...

For very short maturities the returns obtained by the NS model are completely different from the returns obtained by the CIR model. For the three months maturity this problem is clearly evident: not only the NS return is different from the CIR return (the NS return

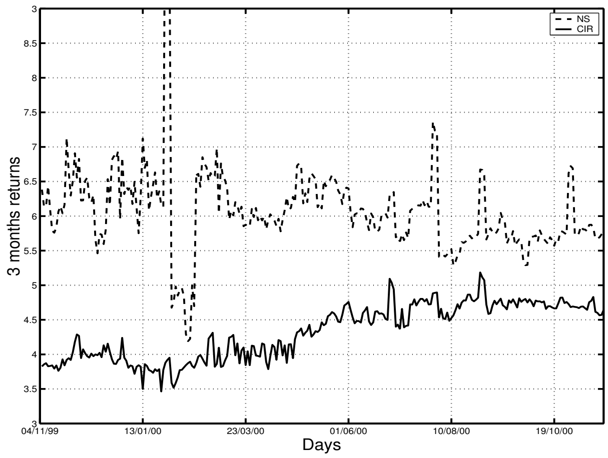

Figure 4.2: 3 months returns.

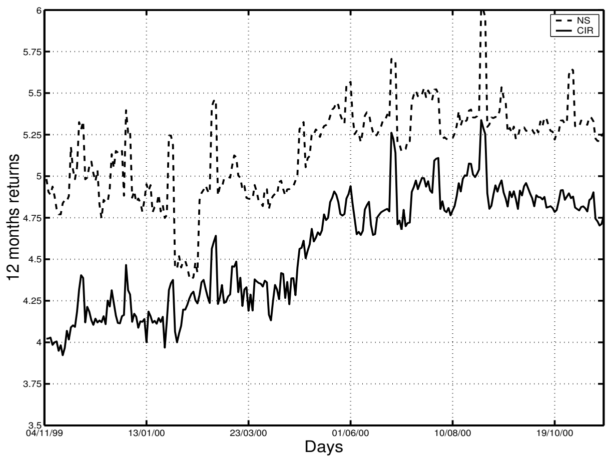

Figure 4.3: 12 months returns

fluctuates between 5% and 7%, except for two peaks, the CIR return fluctuates between 3.5% and 5%), but also everyone has own behavior and own trends. For example, in the first part (until about 03/23/00) the NS model shows a very nervous behavior with unreasonable peaks during February. Between March 2000 and May 2000, the NS return follows a downtrend while the CIR return follows an uptrend. These different behaviors force us to think that neither the NS nor the CIR model can really explain the real evolution of the term structure for the short maturities: nevertheless, comparing them with the real return obtained by BOT during 1999-2000, about 4%, the CIR model seems more adequate than the Nelson-Siegel model (which shows a mean return greater than 6%). So, we can suggest to a practitioner to be warn with the short time horizon.

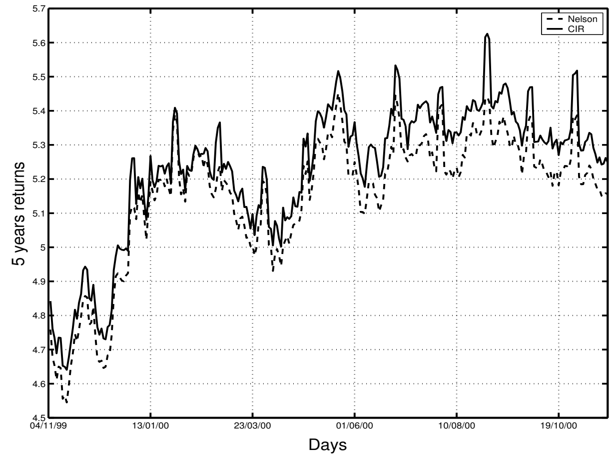

Figure 4.4: 5 years returns

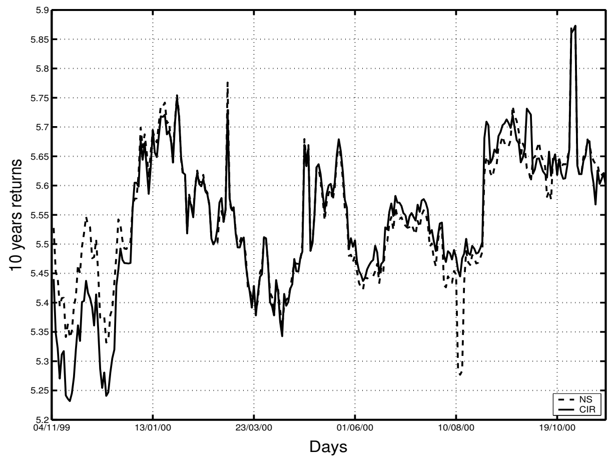

Figure 4.5: 10 years returns

We will come back later on this point, trying to explain that these bad results on short maturities are mainly due to the composition of the archive.
For medium maturities, from 1 to 5 years (see figure 4.3 and figure 4.4), the two models seem show different returns, but in this case there is an interesting news: both the NS and the CIR models show similar trends, in fact:

- during the period 11/99 - 03/00 both them show a trendless;

- during the period 03/00 - 06/00 both them show an uptrend;

- during the period 06/00 - 11/00 both them show another trendless.

It must be observed, however, that during the period 11/99 - 03/00 the NS model shows a behavior more nervous than the CIR one (see the peaks in February).
For long maturities the two models show the same behavior (see fig 4.5). For maturities greater than 10 years the evolutions of the NS model and of the CIR model become indistinguishable each other. Really, this result is less surprising than it would seem, because in the long term both models share an asymptotic behavior: their very strict similarity only tells us that the two calibration procedures are consistent. It is very interesting to note how, for maturities longer than 1 year, the evolution and the trends of the market are captured by both models. This is a good news for a practitioner interested to the evolution of the market: probably, he could use either the NS or the CIR model to follow the trends of the market. We stress that our analysis does not allow to forecast the evolution of the term structure, that is to compute the evolution of the term structure ex ante, but merely it allows to compute the evolution of the term structure ex post. Despite this, the computation ex post could be useful for the elaboration of the future scenarios.
It is well-known (see Mandelbrot in [51] and various analysis of famous practitioner, for example Zweig in [68]) that the trends are self-persistent: if until today the average price has increased (decreased), probably tomorrow it will continue to increase (decrease) until something will change significantly the market condition; so, knowing the present trend it is possible to manage the next evolution of the term structure, with an uncertainty margin that grow up with the time.
We must also remember that the problem on the short term of the term structure is not so dangerous for a fundamentalist fund manager. In fact, it is well known that an active management is reasonable in a long term bonds fund. Because of the characteristics of the fluctuations on the short term, in fact, it is very difficult to plan a trading strategy with the short term bonds. [1]

4.8.1 Why are there great differences between the NS and the CIR models on the short term?

We have explained that in our archive there is a significant lack of information on the short-term. Is it possible that the great difference between the results on the short term

[1] To manage the monetary funds, in fact, managers generally use to buy bonds with residual life less than 24 months and use to hold them until the expiration. It is interesting to remember that there are many types of bond funds, with different risk profile, that work on different market etc... but, in spite of this, it could be easier to separate all bond funds in only two macro-groups: the "real" bond funds that operate on every maturities, especially on long-term bonds, and monetary funds that operate only on short term maturities, generally less than 24 months. The first group is used as alternative to, or for a diversification with, the stock funds, the second group is used to invest capitals that could be necessary to investors in less than 18 months.

Distance between NS return and CIR return on 3 months maturity				
Period	Mean	St. Dev.	Max	Min
11/04/1999 - 11/21/2000	1.7492	0.8788	8.0569	3.5880
11/04/1999 - 08/30/2000	1.9446	0.8945	8.0569	0.3190
08/30/2000 - 11/21/2000	1.0668	0.2794	2.0397	0.4878

Table 4.5: Distance between NS return and CIR return on 3 months maturity

is due to this lack of information? To answer this question we divide our archive in two parts:
a) From 11/04/99 to 08/30/00.
b) From 08/30/00 to 11/21/00.
Hence we separately repeat our analysis. We choose 08/30/00 because in that day we have the first information about a bond with maturity less than 90 days. The results for the three months maturity are reported in table 4.5, respectively on the whole archive, on the first part and on the second part. For the whole archive, the average distance

between NS return and CIR return is 1.7492%. In the first period between 11/04/99 and 08/30/00 (in this period there is not any data on bond with maturities less than 90 days) the average distance is 1.9446%; in the second period between 08/30/00 and 11/21/00 (in this period there are data on bond with maturities less than 90 days, so in some sense it is the "good" period for us) the spread becomes roughly an half of the spread of the first period, precisely 1.0668%. Therefore the presence of data on very-short term reduces the spread between the two models and this is a "proof" of the importance to have information on all maturities, especially with regression models like NS model. However we can not conclude that the problem depends only on the lack of information: a mean difference of about 1% on return, in the second part of our archive, remains too high. As the CIR returns on the short term are closer to the real BOT returns, we believe that the CIR model is always better than the NS model to implement the term structure on the short term. Summing up:

1. The CIR model provides an evolution of the term structure closer to reality on the short term maturities.

2. The lack of data in the database is still a great problem, but not the only one.

3. In the medium term period both models can describe the evolution and the trends

Distance between NS model and CIR model				
Maturities	Mean	St. Dev.	Max	Min
3 months 11/04/1999 - 11/21/2000	1.7492	0.8788	8.0569	0.3190
3 months 11/04/1999 - 08/30/2000	1.9446	0.8945	8.0569	0.3190
3 months 08/30/2000 - 11/21/2000	1.0668	0.2794	2.0397	0.4878
6 months 11/04/1999 - 11/21/2000	1.2733	0.5317	4.4147	0.2221
6 months 11/04/1999 - 08/30/2000	1.3950	0.5339	4.4147	0.2221
6 months 08/30/2000 - 11/21/2000	0.8481	0.2071	1.5500	0.4291
12 months 11/04/1999 - 11/21/2000	0.5881	0.1810	1.0862	0.0856
12 months 11/04/1999 - 08/30/2000	0.6188	0.1874	1.0862	0.0856
12 months 08/30/2000 - 11/21/2000	0.4808	0.0998	0.7788	0.3140
3 years 11/04/1999 - 11/21/2000	-0.0940	0.0385	0.0439	-0.1877
3 years 11/04/1999 - 08/30/2000	-0.1035	0.0350	0.0439	-0.1877
3 years 08/30/2000 - 11/21/2000	-0.0607	0.0307	0.0332	-0.1412
5 years 11/04/1999 - 11/21/2000	-0.0805	0.0294	-0.0041	-0.1849
5 years 11/04/1999 - 08/30/2000	-0.0757	0.0284	-0.0041	-0.1449
5 years 08/30/2000 - 11/21/2000	-0.0975	0.0264	-0.0626	-0.1849
10 years 11/04/1999 - 11/21/2000	0.0056	0.0459	0.1270	-0.1921
10 years 11/04/1999 - 08/30/2000	0.0090	0.0481	0.1270	-0.1921
10 years 08/30/2000 - 11/21/2000	-0.0063	0.0350	0.0688	-0.1148

Table 4.6: Distance between NS model and CIR model

of the market.

4. If a financial institution is looking for a model to study the evolution of the yield curve, the CIR and the NS models could be good candidates.

In table 4.6 we present the distance between the CIR model and the NS model for different maturities and for different historical periods of our archive; as before, for every maturity we quote first the whole archive, then the first part and finally the second part.

4.8.2 Analysis of spreads obtained by the Nelson-Siegel model

We analyze the distribution of the spreads between the market prices and the theoretical NS prices. The method and the observations are identical to those seen during the analysis

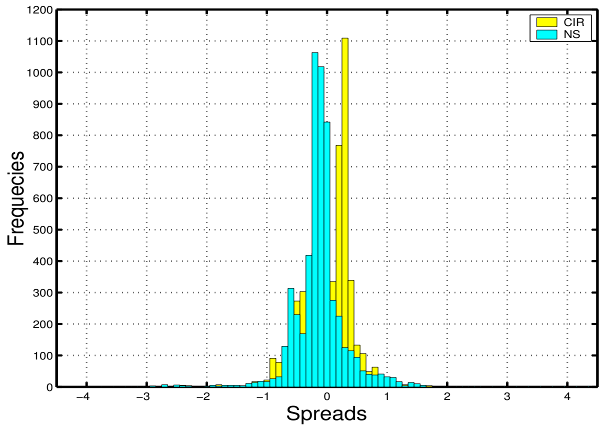

Figure 4.6: Distributions of the spreads in the first part of the dataset

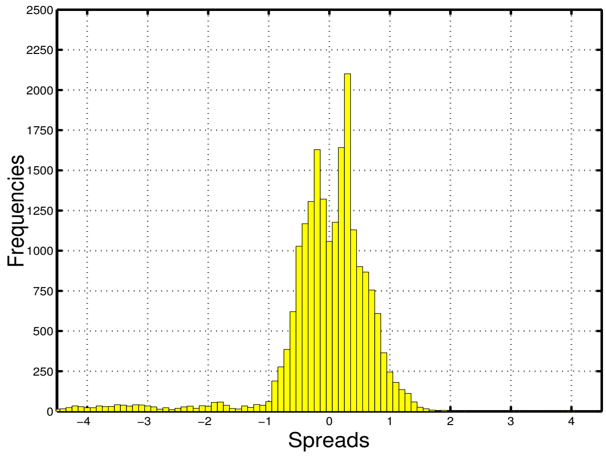

Figure 4.7: Distribution of the spreads with the NS method.

of the CIR prices. We report the distribution of the spreads in figure 4.7, while in figure 4.8.2 we report the two distributions, obtained in the second part of our archive. In table 4.7 we compare the results of BCZ for the years 1984-1989 with our results, obtained with the CIR model and with the NS model. The analysis is computed for two periods (as

in the previous section): for the whole archive, from 04/11/99 to 21/11/00, and for the period from 30/08/00 to 21/11/00 (the "good" period). We decide to repeat our analysis for the two models also for this sub-period, because in it there are data on the short term bonds. Some observations are fundamental: for every period and for both models, our analysis shows a skewness less than 0 and a kurtosis greater than 1, so it is possible to

Comparison between the distributions							
Period	Model	Data	Rejected	Mean	St. Dev.	Skewness	Kurtosis
1984-1989 (BCZ)	CIR			-0.04	0.52		
04/11/99 - 21/11/00	CIR	21601	5.7%	-0.1412	0.4987	-3.1371	16.1561
04/11/99 - 21/11/00	NS	21601	5.6%	-0.0352	0.6032	-2.2197	11.2847
30/08/00 - 21/11/00	CIR	5639	2.8%	-0.1293	0.1987	-1.2948	15.2099
30/08/00 - 21/11/00	NS	5639	2.8%	0.0179	0.2283	-1.4673	11.0870

Table 4.7: Comparison between the distributions

affirm that the asymmetry in the distribution is a characteristic of the market and not of the model used to implement the term structure. In clear disagreement respect to the CIR model, with the NS model we obtain a mean roughly equal to zero. Probably this result is due to the regression nature of the model. Therefore, we can not say that the distribution of the spreads has mean equal to zero independently from the model used to implement the term structure.

4.9 Conclusions and suggestions to a financial institution

Our work is an empirical analysis of the evolutions of the term structure, with the aim **to give suggestions to a financial institution only with available data**. Our historical archive is extracted from Datastream, one of the most used financial databases. The characteristic of Datastream is to provide data only on bonds not expired yet, so if the archive is longer than 6 months, there is a lack of information on the short term. We implement the CIR model and the NS model and we compare the results. From our analysis we can conclude with some suggestions:

1. As in every work in finance, our results are surely valid only for the Italian secondary bond market and for the period between 1999 and 2000. To generalize our results it is necessary to repeat our analysis also in other periods and for other markets, but in our opinion it is possible that some results could be always true.

2. The lack of information on the short term can be extremely dangerous, especially for regression models like the NS model.

3. For the medium and long-term maturities of the term structure both the NS model and the CIR model are good choices. On the other hand, approaches that use spline

and bi-spline, like McCulloch in [52], lead to an explosion of the curve.

4. For the short-term maturities of the term structure, the CIR model is better than the NS model, because it is less sensitive to the lack of information in the first part of the data set.

5. Both models can describe the trends of the returns on the different maturities of the term structure.

6. It seems reasonable to implement fundamentalist strategies based on the mean spread between the market price and the theoretical CIR price of a bond.

We think that our results can provide different fields of analysis, in which a practitioner can go deeper to improve his strategies.

Chapter 5

Cross-section vs time-series approach to the CIR model and empirical evaluation of the market price of risk

This chapter is based on *Static and Dynamic Approach to the CIR Model and Empirical Evaluation of the Market Price of Risk*, a joint work with Luca Torosantucci.

5.1 Introduction

In chapter 4 we applied the CIR model and the Nelson-Siegel model to the Italian Treasury fixed income securities market, using the price of bonds quoted from November 1999 to November 2000: we will denote this dataset as "Archive-A". We will refer to these results in chapter 4 denoting them as results of [64], since they are the same.
We extracted our dataset from Datastream, one of the most important financial database; the choice of Datastream was due to its relevant role in financial institutions. However, the historical archive showed two main problems:

1. Datastream allows to download only prices of the bonds not expired yet. So, for example, if the time series is 2 years long it is not possible to have information about a BOT (Italian BOT are analogous to American Bills) expired one year ago. This fact produces a lack of information on the short term, hence some difficulties in the implementation of the term structure, especially for the Nelson-Siegel model.

2. Not all quoted bonds are sufficiently liquid and this fact produces some noise that

makes difficult to implement the term structure by simple interpolation on data. Precisely, the illiquid bonds shift from the "real" term structure so, if we try to implement it using also this data, it is probable that we will obtain a bias term structure. We could avoid to use illiquid data in the implementation, but in general all the short term Italian bonds are very illiquid, so following this choice it could be necessary to give all the short term information up. Moreover, it is not so simple to identify all the illiquid quoted bonds and finally, is it meaningful, for a practitioner, to implement the term structure neglecting part of the information of the market, i.e. cutting bonds that he uses in his daily activities? It is not, in our opinion...

In spite of these problems, in [64] we proved that the CIR model can be useful for a financial institution, because it can explain very different market situations and could help to set up fundamentalist trading strategies.
In this paper we apply the CIR model to another historical archive with the main aim to have a confirm about the reliability of the model respect to the market. Our dataset is now composed by daily Euribor and swaps rates in the period from January 1999 to December 2002 (the whole archive is composed by 1042 days): precisely, for each day we obtain a complete term structure using Euribor rates for the maturities under 1 year (3, 6 months) and swap rates for maturities from 1 to 30 years (precisely from 1 to 10, 15, 20, 25 and 30 years), so we have a set of 16 maturities. These rates are very liquid and this is the main reason of the choice of this archive. However, for each day the new historical archive provides a few number of data and this fact makes delicate the implementation of the term structure. We will denote "Archive-B" this dataset.
In this paper we implement the CIR model with two different methods:

1. first method: we apply the non-linear least squares method by cross section for each day of our archive. We denote this method as **static**, to stress that data used to implement the model are composed by the rates of a single day.

2. second method: we consider the short rate obtained by the static procedure as a "proxy" of the real market short rate. With this new proxy we can compute the parameters of the CIR model by means of martingale estimation techniques. To study the reliability of this approach, we also analyze the asymptotic convergence of this method by means of a simulation of CIR-paths. We denote this method as **dynamic**, to stress that data used to implement the model are composed by a single time series.

In [64] the CIR model was implemented with the first method, so the comparison between Archive-A and Archive-B is only performed with the results of the static model. This comparison shows that parameters obtained by Archive-A are a little bit more volatile than parameters obtained by Archive-B, especially the volatility of σ. This effect is due to the higher liquidity of Archive-B respect to Archive-A. The long term average rate μ is more stable for Archive-B than for Archive-A, and closer to the expectation of the practitioners. This effect is due, for the most part, to the lack of information on the short term of Archive-A.

We also compare the results of Archive-B by static method with real data; in particular we prove that the short rate r_t obtained by Archive-B is very close to 3-months Euribor (the shorter maturity in archive), to 6-months Euribor and to 1-year Swap rates showing that the CIR model provides a good description of the market. Due to the theoretical characteristics of the CIR model, the description of the term structure is more precise for the medium and long maturities than for the short maturities, especially when there is an inversion of the curve on the very short term (for example, when the 3-months Euribor is greater than the 6-months Euribor).

The main drawback of the static method is that is not possible to separate the market price of risk $-\lambda$ from the speed of adjustment k (the equations are in the next section). Comparing this last value with the speed of adjustment k obtained by the dynamic method, we can isolate the market price of the risk $-\lambda$ and evaluate, under the Local Expectations Hypothesis, the risk premium of the market for different maturities. This is maybe the most interesting part of our analysis, because it allows to review the expectations of the market on the yield curve.

5.2 The CIR Model

The CIR model was detailed discussed in the chapter 4: we now simply recall that it is based on this Stochastic Differential Equation:

$$dr_t = k(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t. \tag{5.1}$$

Equation (5.1) is joint with the so called Local Expectation Hypothesis:

$$\mathbb{E}(\frac{dP}{P}) = r_t + \lambda r_t \frac{P_r}{P}, \tag{5.2}$$

In a CIR-based market, zero coupon prices are given by:

$$p(t,T) = F(t,T)e^{G(t,T)r_t}, \tag{5.3}$$

with

$$F(t,T) = [\frac{\phi_1 e^{\phi_2(T-t)}}{\phi_2(e^{\phi_1(T-t)}-1)+\phi_1}]^{\phi_3},$$
$$G(t,T) = [\frac{(e^{\phi_1(T-t)}-1)}{\phi_2(e^{\phi_1(T-t)}-1)+\phi_1}]. \tag{5.4}$$

Obviously $\phi_i,\ i = 1, 2, 3$ depends on the parameters of the model:

$$\phi_1 = \sqrt{(k + \lambda)^2 + 2\sigma^2},$$

$$\phi_2 = \tfrac{k+\lambda+\phi_1}{2}, \tag{5.5}$$

$$\phi_3 = \tfrac{2k\mu}{\sigma^2},$$

where $-\lambda$ is as usual the market price of risk.
We recall that

$$lim_{T\to 0}R(t, T) = r_t,$$

$$lim_{T\to +\infty}R(t, T) = (\phi_1 - \phi_2)\phi_3.$$

where $R(t, T)$ is the return at time t of a zero-coupon with maturity T.

5.3 Static implementation of the CIR model

The first method we consider to implement the CIR model is the original one (see [18] and [19]): it describes every pure discount bond in the form (5.3), where $F(t, T)$ and $G(t, T)$ depend on the parameters as in (5.4).
Using (5.4) and (5.5) we apply the non linear least squared method over the m maturities, that is we look for a vector

$$x = (\phi_1, \phi_2, \phi_3, r_t),$$

in order to obtain

$$M = \min_x f(x) = \min_x (f_1^2(x) + f_2^2(x) + \ldots + f_m^2(x)), \tag{5.6}$$

where

$$f_i(x) = [P_{\text{market}}(t, i) - P_{\text{CIR}}(t, i)].$$

5.4 Numerical problems in the static implementation: a practical solution

The problem of the Archive-B is the small number of data for each day, equal to 16. For this reason the function f that we want to minimize (the distance between the real market price and the CIR price) could have many local minimum, and this drawback makes difficult to find the global minimum. Precisely, this is the problem: the non-linear least squares method used to find the parameter of the model is a numerical method that find the minimum of a function in a recursive way, starting from a parameter vector arbitrary chosen. From a geometrical point of view, this method could be thought as a path that,

from a beginning point, chooses the way that brings to the nearest deep valley down. As the function f is similar to an environment where there are many hills and valleys, it could be possible that the nearest deep valley is not the deepest valley, that is the non-linear least squares method could bring only to a local minimum down. So, the choice of the starting point is very delicate.
In general it is not possible to know, neither in approximate way, where the global minimum lies, so in general we can not choose a reasonable starting point for the parameter vector sufficiently close to the minimum. To overcome this difficult, a rough but efficiency method is to build a net of starting point. Precisely, this is the procedure:

- Fix a large interval for each element of the parameter vector. This interval has to be large enough to assume that the global minimum belongs to this interval. Because the parameter vector is composed by 4 elements, we will obtain a 4-dimensional interval.

- Divide the 4-dimensional interval in n sub-intervals and for any sub-interval fix a generic point belonging to it. In this way it is possible to build a set S of n 4-dimensional points.

- Apply for n times the non-linear least squares method, choosing as starting point all the points in S.

In this way there will be two new sets that we call A and B. Each element of A is a local minimum for the function f, whereas each element of A is the vector parameter for which we obtain exactly that local minimum. It is important to stress that not all local minimum in A are necessary acceptable. The values of ϕ_1, ϕ_2, ϕ_3 have to ensure both a reasonable expectation of the market, and the mathematical hypothesis of the CIR model and, that is:

$$\begin{cases} \sigma > 0, \\ \mu > 0, \\ k\mu \geq \sigma^2. \end{cases}$$

Eventually, between all allowed local minima we can determine the vector parameter that minimizes the function f.
The main drawback of this method is the long time necessary to give us a solution. For this reason it could be a good choice to restrict, as much as possible, the number n of iterations. To perform this strategy, there are two empirical but not rigorous solutions:

1. To implement the CIR model as above described for some arbitrary chosen days in the archive; these days must be far one from another. If the minimum parameter

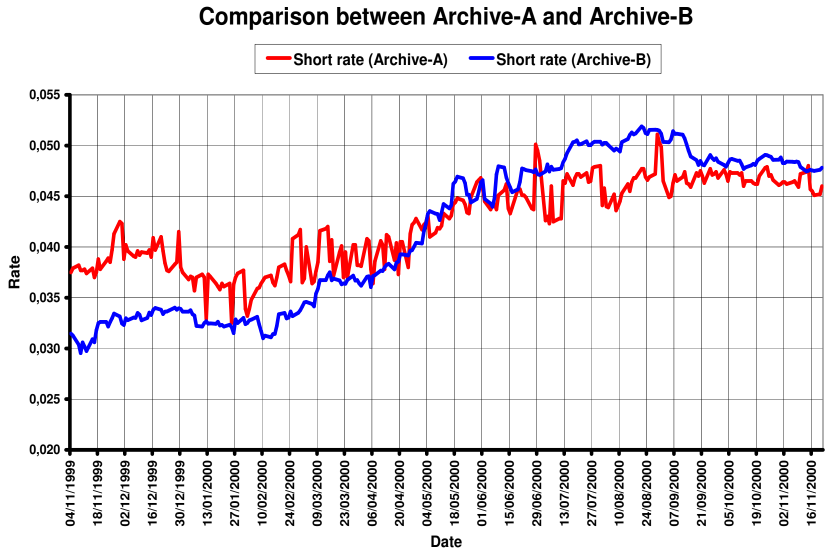

Figure 5.1:

vector determined by the non-linear procedure does not change dramatically from a day to another, for all the other days of the archive, it becomes reasonable to restrict the number of the starting point using a limited interval that contain all the obtained minimum parameter vectors.

2. To implement the CIR model as above described for the first day of the archive. For the following days of the archive fix as starting point the minimum parameter vectors of the previous day. The advantage is clear: from the second day of the archive it is used a unique starting point. The problem of this method is the following: if at the day t occurs a strange and particular market situation, it is reasonable to obtain an anomalous minimum parameter vector. Now, if that particular situation disappears at the day $(t+1)$ and if the starting point for the day $(t+1)$ is fixed in the anomalous minimum of the day t, the risk is that (caused by the great number of local minima) the minimum parameter vector for the day $(t+1)$ does not represent the real market situation.

So, if the aim is to obtain a quick analysis the second method is the best, otherwise it is advisable to use the first one method.

Short rate r_t					
Model and Period	Mean	St. Dev.	Rel. Var.	Min	Max
Archive-B 1999-2002	3.309	0.878	26.5%	1.838	5.189
Archive-B 11/99-11/00	4.162	0.738	17.7%	2.952	5.189
Archive-A	4.232	0.421	9.9%	3.255	5.114
BCZ 1984	14.948	2.702	18.1%	13.295	16.129
BCZ 1984-1989	11.519	3.156	27.4%	8.071	16.129

Table 5.1: Short rate r_t.

5.5 Analysis of the model parameters

In this section we show the evolution of r_t, m, k and σ obtained by the static implementation for the Archive-B. We then compare these results with the results obtained in [64] and [4].

5.5.1 Evolution of the short rate r_t

In the first row of table 5.1 we report the results of the static implementation for the Archive-B in the period between 1999 and 2002. In the second row we show the results in the shorter period between November 1999 and November 2000, the same historical period analyzed in [64]; in the third row we report the results of [64] obtained by the static implementation of Archive-A in the period between November 1999 and November 2000. In the fourth row we report the results of [4] about 1984, the year between 1984 and 1989 with the greatest mean of the short rate (in some sense their "worst" year).
During the 1980s, the Italian macroeconomic situation was completely different from the present situation. In that period, in fact, the TUS ("Tasso Ufficiale di sconto") of Banca d'Italia was greater than 10%, while nowadays the Reference Rate of BCE is about 4%.
It is very interesting to note that the CIR model can capture this change, in fact the mean of the short rate was 11.519% in 1984-1989 while it is 3.269% in 1999-2002 (see again table 5.1).
Obviously, it is necessary that a term structure model captures the changes in the macroeconomic conditions. In other words, a good model have to capture the changes of scenarios but, on the contrary, if a model can capture the changes of macroeconomic scenarios it is not necessarily a good model! However, we think that these results are substantially a proof of the reasonableness of the CIR model. For the period between 11/99 and 11/00, we must emphasize that both Archive-B and Archive-A provide short rates very close each

other. It is a necessary and an obvious check to verify the coherence of the archives (and of the model...) respect to the market situation of that period.
In spite of these considerations, from figure 5.4 it is clear the improvement of the implementation of the CIR model for Archive-B respect to the same implementation for Archive-A. In fact:

- In the last part of 1999 and in the first part of 2000, where the lack of information of Archive-A is more serious, the short rate from Archive-B is sufficiently stable, showing a more natural evolution. On the other hand, the short rate from Archive-A appears more nervous and higher than the short rate of Archive-B, showing how the lack of information on the short term maturities (that characterize Archive-A data) can introduce a bias in the implementation of the term structure.

- In the second part of figure 5.4, where also Archive-A has information about short maturities, the BOT short rate remains more nervous than the Euribor short rate. This fact can be explained remembering the higher liquidity of Archive-B data respect to Archive-A data. But, in spite of this, the short rate A becomes very close to the short rate B and this empirically proves that if there are data for all maturities, it is possible to obtain a reasonable term structure from the quoted bonds in the market.

- However, the choice of Swap and Euribor rates remains preferable, because these data provide a more stable term structure.

To conclude this analysis, it is useful to compare in figure 5.5.1 the behavior of the short rate from Archive-B with the behavior of Euribor rates for the whole period between January 1999 and December 2002. The short rate appears very close but lower than the Euribor rates. It agrees with the nature of "instantaneous maturity" of the short rate.

5.5.2 Evolution of the implied volatility $\sigma\sqrt{r_t}$

As in the preceding table, in the first row of table 5.2 we report the results of the static implementation for Archive-B for the period between 1999 and 2002. In the second row we show the results in the shorter period between November 1999 and November 2000, the same historical period analyzed in [64]; in the third row we report the results of [64] obtained by the static implementation of Archive-A in the period between November 1999 and November 2000; in the fourth row we show the results of [4] and, finally, in the fifth

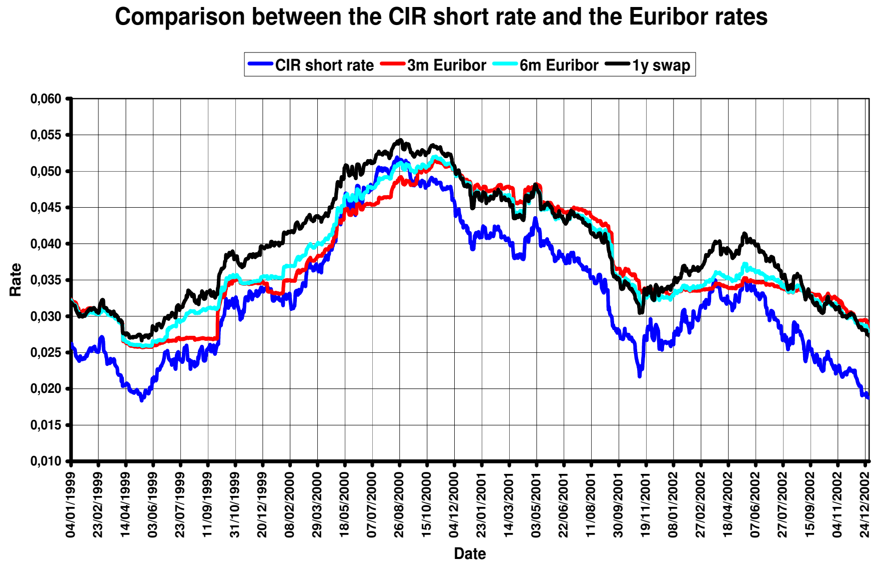

Figure 5.2:

row we show the results of [4] about 1984, the year between 1984 and 1989 with the greatest mean of the short rate (in some sense their "worst" year).
The implied volatility is very important because it describes the behavior of the stochastic part of the model. In a market where expectations are stable, for example, this parameter should be low. For this reason, the implied volatility of the period 1999-2000 or of the period 1999-2002 is less than the implied volatility of the period 1984-1989.
It is interesting to note that the implied volatility for Archive-A is very similar to the implied volatility for Archive-B and this is a check of the ability of the model to capture the real fluctuations of the market in spite of the lack of information on the short term maturities, and also in spite of the lower liquidity of the quoted bonds respect to the swap

Implied volatility $\sigma\sqrt{r_t}$					
Model and Period	Mean	St. Dev.	Rel. Var.	Min	Max
Archive-B 1999-2002	1.1522	0.2648	22.9%	0.6594	1.7872
Archive-B 11/99-11/00	1.4092	0.1704	12.1%	1.1579	1.7872
Archive-A	0.8098	0.3753	46.4%	0.2217	1.4807
BCZ 1984	1.9865	0.0397	2.0%	1.9092	2.0761
BCZ 1984-1989	2.1004	0.3321	15.8%	1.1606	3.5880

Table 5.2: Implied volatility $\sigma\sqrt{r_t}$.

Speed of adjustment k under the hypothesis $\lambda = 0$					
Model and Period	Mean	St. Dev.	Rel. Var.	Min	Max
Archive-B 1999-2002	0.3603	0.1301	36.1%	0.1791	0.6956
Archive-B 11/99-11/00	0.3913	0.1187	30.3%	0.2311	0.6956
Archive-A	0.2804	0.1192	42.5%	0.0386	0.4653
BCZ 1984-1989	0.2433	0.0031	1.3%	0.2269	0.2524

Table 5.3: Speed of adjustment k.

Long term average rate μ under the hypothesis $\lambda = 0$					
Model and Period	Mean	St. Dev.	Rel. Var.	Min	Max
Archive-B 1999-2002	5.910	0.353	5.9%	5.149	6.565
Archive-B 11/99-11/00	6.187	0.115	1.8%	5.923	6.565
Archive-A	7.470	1.230	16.5%	6.192	14.454
BCZ 1984-1989	11.128	1.802	17.2%	4.261	16.837

Table 5.4: Long term average rate μ.

rates. In fact, a generic quoted bond is not necessarily "on" the term structure, but up or down the curve with an average distance that in general depends on the type of the bond (that is it depends on its nature and on the appreciation of the market, in other words on its liquidity). So every noise in the market can generate a similar movement of both the curve and the bond position, but probably not of the average distance between the bond and the term structure. In this way it becomes intuitive that the CIR model could estimate the real volatility of the term structure also with the quoted bonds.

5.5.3 Evolution of the speed of adjustment k and of the average rate μ

If we choose the trivial case of null market price of risk, $\lambda = 0$, it is possible to study independently the parameters k and μ. This is a severe condition that we will try to eliminate in the next sections.
Looking at table 5.3 it is clear that the values obtained in [64] and [4] are very similar to the results for Archive-B. As we will see, the speed of adjustment k play a key role in the

definition of the expected returns for the Euribor and swaps rates.
In table 5.4 we report the long-term average rate μ. For this parameter the results are very similar to the results obtained in [64] and very different from the results of and [4]. As discussed before, obviously this is a consequence of the different Italian macroeconomic scenery.

5.6 Dynamic implementation of the CIR model

The CIR model is a dynamic model that defines a stochastic evolution of the short rate as in (5.1), so it could seem strange to use a static method to implement it.
The "natural" method to implement the CIR model would be based on the use of a time series of a bond (or of a rate) as benchmark for the short rate. In other words, the most logical procedure would be:

- Take the time series of the bond or of the rate chosen as benchmark for the short rate, for example the evolution of the 3 months Euribor.

- Compute the model parameters with the likelihood method or with the EMM method, or also with the GMM method.

- At the end, there will be only one value for each model parameter and not an historical series, as in our first procedure. In this case, the value of each parameter is the same for each day and depends on the evolution of the benchmark and not on the price of all quoted bonds on a specific day.

About the dynamic approach, there are three possible "critical points" to discuss:

1. In order to implement the CIR model using a dynamical approach, it is necessary to choose a benchmark for the short rate, but what is the best representative of the short rate? It could be the 3 months Euribor, but it could be too long as instantaneous rate. The behavior of other shorter rates, as for example the Eonia rate, is so particular that it is not possible to identify the real characteristics of the whole term structure by means of them.

2. We would obtain a unique value for each parameter describing the average market of the period considered. So, in the case of a shock during the period of analysis, the model would not be able to introduce the necessary changes in the term structure. We will come back later on this crucial point.

3. Moreover, if we use the historical series of a benchmark, we would not be able to introduce information about all quoted maturities. This would be a worrying restriction for a practitioner.

We believe that for fund managers these reasons are sufficient to justify the use of a static procedure. In spite of this considerations, for the static procedure there is a serious[1] drawback: **it is not possible to identify the real speed of adjustment of the model** because we can not separate k from the market price of risk $-\lambda$ (see (5.5)). In this section we will propose a method to implement the CIR model from a dynamical point of view, with the aim to estimate the market price of risk $-\lambda$.

5.6.1 The choice of the proxy

The natural candidates to the rule of proxy for the short rate are:

- The shortest maturity rate in the archive: in our case the 3 months Euribor rate;

- The short rate obtained by the static implementation of the CIR model (we also call it as "CIR short rate").

In figure 5.5.1 it is possible to note that the behavior of the CIR short rate is very close to the behavior of the Euribor rates and of the 1 year swap rate. On the other hand, the implied volatility of the CIR short rate is larger than the volatility of the 3 months and 6 months Euribor rates. However, the Euribor rates show anomalous characteristics. For example, during the period between 03/06/99 and 11/09/99 the 3 months Euribor rate is almost constant even though the 6 months Euribor shows a clear up-trend. After this period, the 3 months Euribor rate grows up immediately until to catch up with 6 months Euribor rate. In the same period, the CIR short rate shows a natural fluctuation without a trend. A similar problem occurs in the period between 19/11/01 and 18/04/02: the 3 months Euribor rate appears constant in spite of a growth of the 6 months Euribor rate. Last but not least, sometimes the curve is affected by the so called "inversion" on short term maturities (the 3-months Euribor rates is greater than 6-months Euribor rates and, sometimes, than the 1-year Swap rate).
Because of his strange periods, it is quite reasonable to think that the 3-months Euribor rate is unfit to describe the real short term. Besides, how we explained, the CIR short rate seems to represent a right compromise between the behavior of the Euribor rates and of the 1 year swap rate.

[1]From an pragmatical point of view it is not a great problem, because a practitioner can obtain the term structure directly from the parameters ϕ_1, ϕ_2, ϕ_3 and r, if he does not want to identify the values of the model parameters.

Finally, we can consider the behavior of the CIR short rate as the synthesis of all information inside the whole term structure, information that are extracted by the implementation of the model. For these reasons we believe that the CIR short rate is the better proxy of the real short rate.
We can compute the parameters of the CIR model by means of martingale estimations techniques (see [37] and [47]). We denote this method as dynamic, to stress that the data used to implement the model are composed by a single time series: the proxy of the short rate.

5.6.2 The Martingale Estimation

For the sake of simplicity, we redefine the CIR model (5.1) in this form:

$$dr_t = (a + br_t)dt + \sigma\sqrt{r_t}dW_t, \tag{5.7}$$

where obviously

$$\begin{aligned} k &= -b \\ \mu &= -\tfrac{a}{b} \end{aligned} \tag{5.8}$$

It is useful to remember that $b < 0, a > 0$ and $\sigma > 0$. To apply martingale estimations it is necessary to discretize the model (5.7): we add more terms from the Ito-Taylor expansion to the Milstein scheme. Hence we have this formulation (see [37]) and [47]):

$$\begin{aligned} r_{t+1} &= r_t + \Delta(a + br_t) + \sigma\sqrt{r_t}dW + \tfrac{\sigma^2}{4}(dW^2 - \Delta) \\ &\quad + b\sigma\sqrt{r_t}dZ + \tfrac{1}{2}\Delta^2 b(a + br_t) + \tfrac{\sigma}{2\sqrt{r_t}}(a + br_t - \tfrac{\sigma^2}{4})(dW\Delta - dZ), \end{aligned} \tag{5.9}$$

where:

$$dW = U_1\sqrt{\Delta},$$
$$dZ = \Delta^{3/2}(U_1 + \tfrac{U_2}{2}).$$

As usual U_1 and U_2 are independent $N(0, 1)$-distributed random variables and Δ is the discretized time step. We will set up the right value of Δ later, as in our study the time step is equal to one day. In particular this approach, via martingale estimating functions, guarantees closed formulas for a and b; in fact, in [37] and [5] it is proved that:

$$\begin{cases} b = \frac{1}{\Delta} ln\left[\frac{n\sum_t(\frac{r_t}{r_{t-1}}) - \sum_t r_t \sum_t(\frac{1}{r_{t-1}})}{n^2 - \sum_t r_{t-1}\sum_t(\frac{1}{r_{t-1}})}\right], \\ \\ a = \frac{b}{1-e^{b\Delta}} \frac{ne^{b\Delta} - \sum_t(\frac{r_t}{r_{t-1}})}{\frac{1}{r_{t-1}}}. \end{cases} \tag{5.10}$$

Besides, the estimator for σ^2 is:

$$\sigma^2 = \frac{\sum_t \left[r_t - \frac{(a+br_{t-1})e^{b\Delta} - a}{b}\right]^2 \frac{1}{r_{t-1}}}{\sum_t \left[\frac{(a+2br_{t-1})e^{2b\Delta} - 2(a+br_{t-1})e^{b\Delta} + a}{2r_{t-1}b^2}\right]}. \tag{5.11}$$

Before to apply the martingale estimators to our data, we analyze their convergence by means of simulations of CIR paths, to decide the reliability of (5.10) and (5.11) and to decide how many data are necessary to obtain a good estimation of the parameters.
Another problem is the choice of the time step parameter Δ: is it possible that this choice can influence the convergence of the martingale estimations? The answer is yes: the value of Δ influences the number of data necessary to ensure a reliable convergence. So, it is necessary to compare the number of data of the archive with the Δ chosen.
To test the convergence of the method we simulate CIR paths with these parameters:

$$\begin{cases} a = 2.60 \\ b = -0.80 \\ \sigma = 0.03 \end{cases} \text{or} \begin{cases} \mu = 3.24 \\ k = 0.80 \\ \sigma = 0.03 \end{cases} \tag{5.12}$$

respectively for (5.7) and (5.1)).
The choice of the parameters as in (5.12) is not so casual, but near to the mean obtained by the static implementation. We simulate paths with different number of steps N (50, 100, 200, 400, 500, 600, 700, 800, 1000 and 1500 steps) and with different value of Δ (1, 1/2, 1/100, 1/250); in particular, for each number of steps and for each value of Δ, we simulate 250 paths.
In table 5.5 we quote the mean value of the estimated parameters, whereas in table 5.6 we report the percentage relative errors, computed as the difference between "real" and "simulated" values over the "real" value.
With this convention, a positive errors means that the simulated parameter is lower than the "real" value showed in (5.12). For example: for Δ=1/250=0.004 and N=1500 we simulate 250 paths. For each path we apply the martingale estimation. The mean value of μ estimated for those 250 paths is 3.2430 (see table 5.5) with an error respect to the "real" value mu (equal to 3.24) of 0.02% (see table 5.6).

The results are in general very good but it is necessary to emphasize that:

- The estimates seem to provide a good evaluation of the parameters and this is the key answer for our purposes. The other observations are only necessary for a calibration of the optimum Δ and N, and not to judge the validity of the method.

- When Δ is very little it is necessary to use more data to obtain a convergent and reliable estimation of the parameters. In some sense it is an obvious consideration: roughly speaking, for Δ=0.004 we must consider 250 days to have a description of 1

Convergence of the martingale estimations: estimations of the parameters with "real" values $\mu = 3.24, k = 0.80, \sigma = 0.03$										
N	50	100	200	400	500	600	700	800	1000	1500
parameter μ										
Δ=1	3.2444	3.2445	3.2447	3.2454	3.2450	3.2449	3.2449	3.2449	3.2451	3.2451
Δ=0.5	3.2447	3.2461	3.2465	3.2450	3.2456	3.2448	3.2448	3.2452	3.2452	3.2452
Δ=0.01	2.8975	3.3976	3.2624	3.2483	3.2375	3.2472	3.2442	3.2482	3.2418	3.2437
Δ=0.004	2.7228	2.6144	3.2715	3.2591	3.2500	3.2609	3.2505	3.2504	3.2481	3.2430
parameter k										
Δ=1	0.6739	0.6763	0.6641	0.6617	0.6691	0.6579	0.6628	0.6612	0.6691	0.6600
Δ=0.5	0.7873	0.7910	0.7840	0.7854	0.7879	0.7837	0.7836	0.7764	0.7790	0.7748
Δ=0.01	1.0571	0.9234	0.8504	0.8293	0.8504	0.8272	0.8368	0.8292	0.8380	0.8274
Δ=0.004	3.0692	1.4747	1.0082	0.9011	0.8740	0.8187	0.8279	0.8451	0.8209	0.8290
parameter σ										
Δ=1	0.0258	0.0262	0.0264	0.0264	0.0265	0.0264	0.0264	0.0264	0.0265	0.0264
Δ=0.5	0.0289	0.0294	0.0297	0.0298	0.0298	0.0297	0.0298	0.0297	0.0298	0.0298
Δ=0.01	0.0299	0.0305	0.0305	0.0307	0.0306	0.0306	0.0307	0.0307	0.0307	0.0306
Δ=0.004	0.0303	0.0303	0.0305	0.0306	0.0307	0.0307	0.0307	0.0306	0.0307	0.0307

Table 5.5: Mean values.

Convergence of the martingale estimations:										
percentage relative error (with sign): ("real"-simulated)/"real"										
N	50	100	200	400	500	600	700	800	1000	1500
parameter μ										
Δ=1	-0.07	-0.07	-0.07	-0.10	-0.08	-0.08	-0.08	-0.08	-0.09	-0.09
Δ=0.5	-0.07	-0.12	-0.13	-0.08	-0.10	-0.08	-0.08	-0.09	-0.09	-0.09
Δ=0.01	10.64	-4.79	-0.62	-0.18	0.15	-0.15	-0.06	-0.18	0.02	-0.04
Δ=0.004	16.02	19.36	-0.90	-0.52	-0.24	-0.57	-0.25	-0.25	-0.18	-0.02
parameter k										
Δ=1	15.91	15.61	17.12	17.41	16.50	17.89	17.28	17.48	16.48	17.62
Δ=0.5	1.77	1.25	2.11	1.98	1.65	2.20	2.21	3.10	2.77	3.29
Δ=0.01	-26.83	-13.32	-5.32	-3.31	-5.76	-3.21	-4.31	-3.50	-4.42	-3.19
Δ=0.004	-231.34	-70.20	-22.20	-11.06	-8.16	-2.06	-3.00	-4.96	-2.29	-3.28
parameter σ										
Δ=1	16.02	14.60	14.18	14.23	13.76	14.03	14.01	13.97	13.88	13.96
Δ=0.5	6.13	4.45	3.49	2.92	3.00	3.29	2.95	3.27	2.98	3.09
Δ=0.01	2.77	0.63	0.67	0.06	0.32	0.37	0.10	-0.03	0.22	0.30
Δ=0.004	1.29	1.46	0.92	0.36	0.17	0.12	0.18	0.53	0.25	0.11

Table 5.6: Percentage relative errors (with sign).

year; for Δ=0.01 we must consider 100 days to have a description of 1 year, and so on. In the same way, to describe the stochastic behavior of the model it is necessary to have a number of data that growths with the reduction in Δ.

- The best results are obtained for low values of Δ. This is an interesting backtesting for our analysis. In fact, we must remember that in financial papers there are different conventions about the number of day in a year (360, 300, 250 days or also 220 days). However, in our dataset we have 1042 days for a period of 4 years, equal about to 250 days per year, so the more natural choice for our analysis is Δ=1/250=0.004.

For Δ=1/250 and for the number of the data in our archive (1042), thanks to table 5.5 and table 5.6, we have a direct empirical proof that in our conditions we can use the martingale estimations.
To sum up, in this subsection we have proved that:

1. The martingale estimations provide good results.

2. For the specific characteristic of our archive (1042 daily data of the short rate, with Δ=0.004) it is possible to calibrate adequately the model.

5.6.3 An empirical warning about the dynamic implementation

In the previous subsection we have seen that in order to obtain reliable parameters estimations it is better to use the largest as possible number of data. From a practical point of view this assertion is false: the market is not stable, so it can suddenly change its characteristics. Suppose that the market really follows a CIR evolution: this model can only describes the average characteristics of the market in the period used to calibrate the parameters. So, if in that period it occurs a shock, the model is unable to discriminate the two different market situation, before and after the shock. In that case, the resulting parameters of the implementation are a mix of two different market behavior, representative of none of these and, as consequence, substantially useless.
Summing up, we need to:

- Use the largest number of data as possible to obtain a reliable implementation.

- Use the smallest number of data as possible to limit the problem of the market instability.

Percentage errors	
Parameter	Error for N=600
Speed of adjustment k	$\sim$3%
Long term average rate μ	<1%
volatility σ	<0.5%

Table 5.7: Percentage errors for our final choice N=600.

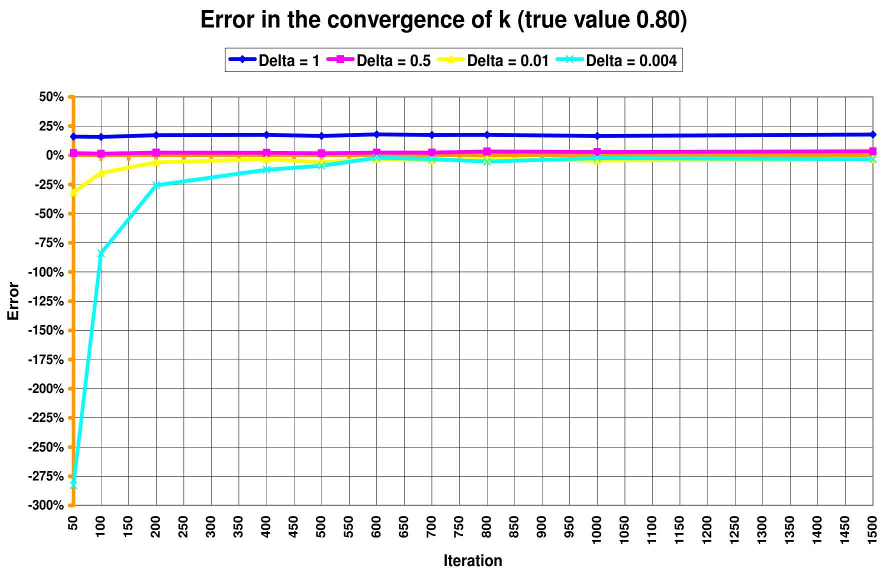

Figure 5.3:

So, the optimum number of data N for a dynamical implementation is a compromise between these two opposite requirements.

In our opinion, the natural compromise about the number of data is represented by the minimum number of data necessary to obtain a reliable implementation of the model. Once more, the study of the convergence can help us. In figure 5.6.3, 5.6.3 and 5.6.3 it is clear that over 600 data we have a very close position to the real value of the parameters. In particular, with 600 data we can roughly maintain the error of the estimates on the percentage levels showed table 5.7. We must note that, also with a number of 500 or 550 data is possible to have reliable statistics, but to limit the influence of occasional fluctuations we choose an slightly higher value for N.

Before to conclude this subsection we want to stress that, for the early observations, in the

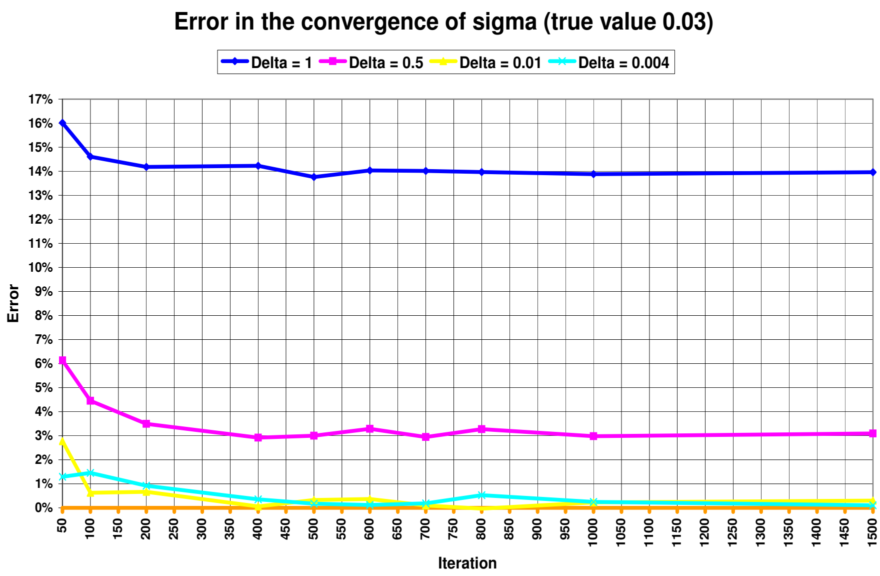

Figure 5.4:

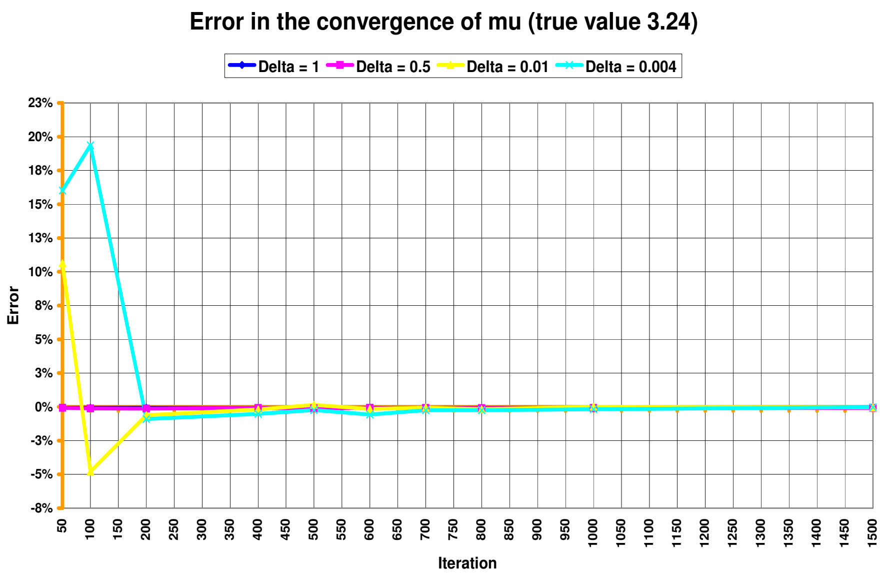

Figure 5.5:

case of a clear shock in the market, a practitioner should have to use a lower number of data. Anyway, he should also remember that the reliability of the implementation decrease naturally with the decreasing of N.

5.6.4 Results of the dynamic implementation

After the analysis of the convergence, we apply the dynamic procedure as follow:

1. We use the time series of the short rate proxy.

2. We start at 601-th day and we estimate the CIR parameters k, μ, σ with martingale estimation formulas, using the last 600 data.

3. Then, we continue with the 602-th day and we estimate the parameters using only the last 600 data.

4. And so on, until the last day of our archive. Due to anomalous and temporary fluctuations of the market, in rare case the dynamic implementation provides negative values of m and k, that are inconsistent from a financial point of view. To absorb this anomalous fluctuations we re-implement the model with a greater number of data: corrections are realized adding up sets of 10 data, equal to 10 days[2], until to achieve a reasonable value of the parameters. To be precise, this phenomenon occurs only for two days and in order to implement these two "strange" days 660 days are sufficient.

5. At the end of this procedure we obtain 3 new time series, respectively for the parameters k, μ, and σ. We want now to repeat that each triple $\{k_t, \mu_t, \sigma_t\}$ describes the average market of the last 600 days. The number of data for each time series is obviously equal to 442 (= 1042, the total number of days in our archive, - 600), precisely from 23/04/2001 to 31/12/2002.

These results are quoted in table 5.8.

It is interesting to compare the results between the dynamical and the static implementation:

- The mean values of the parameters obtained by the dynamic implementation and by the static implementation have the same magnitude.

[2]Note that 10 days are equal to 2 week so, in the bond market, this period is not too long to influence the results of the implementation with past market situations.

Dynamical Approach vs Static Approach From 2/1/2001 to 31/12/2002, time step Δ=1/250						
Parameter	Method	Mean	St. Dev.	Rel. Var.	Min	Max
Speed of adjustment	Dynamic	0.7449	0.2812	37.8%	0.1975	1.6179
k	Static	0.4078	0.1248	30.6%	0.1791	0.5937
Long term average rate	Dynamic	3.5882	1.0179	28.4%	0.9159	6.6734
μ	Static	5.7112	0.3441	6.0%	5.1494	6.4191
Implied volatility	Dynamic	0.7798	0.0569	7.3%	0.6462	0.9016
σ	Static	1.1552	0.2067	17.9%	0.8632	1.5650

Table 5.8: Dynamical Approach vs Static Approach.

- However, the estimated value of the dynamic implied volatility is less than the value of the static implied volatility. The reason is probably that, whereas the static method is based on the whole term structure and it could be influenced by the "noise" on every rate (both Euribor and Swap), the dynamic method could be influenced by only the noise on the short term. In other words, every gust of wind on the market that generates a temporary fluctuations on the medium or long term but not on the short term, can influence more significantly the static method than the dynamic one;

- The standard deviations of the parameters obtained by the dynamic implementation are bigger than the standard deviations obtained by the static implementation, except for the implied volatility;

- The fact that the variability of the dynamic implied volatility is less than the variability of the static implied volatility (see table 5.9) is not, in some sense, a surprise. In fact, in the analysis of the convergence of the martingale estimation, we have proved the ability of the dynamic method to capture the volatility of the CIR model (see table 5.6). As consequence, the dynamic result could be more reliable than the static result of the implied volatility;

- The implied volatility obtained by the dynamic implementation is very close to the explicit volatility of the short rate proxy.

We want to investigate deeper the last point. From (5.9) it is clear that the implied volatility measures the annual volatility and not the daily volatility. To view this interesting

Comparison between the volatilities				
From 23/04/2001 to 31/12/2002, time step Δ=1/250				
Parameter	Method	Mean	St. Dev.	Relative Error
Implied (annual) volatility	Static	1.1552	0.2066	17.9%
Implied (annual) volatility	Dynamic	0.7798	0.05688	7.3%
Annual volatility	Short rate proxy	0.6577	0.5872	89.3%

Table 5.9: Comparison between the volatilities.

aspect we can rewrite (5.9) up to the first order in Δ:

$$r_{t+1} = r_t + \Delta(a + br_t) + \sigma\sqrt{r_t}\sqrt{\Delta}dU, \tag{5.13}$$

where U is a $N(0,1)$-distributed random variable.
The discretization introduces a reduction of the implied volatility for a factor of $\sqrt{\Delta}$, so:

- We can compute the daily volatility multiplying the value of the implied volatility for $\sqrt{\Delta}$.

- Or, it is the same, we can compute the implied volatility dividing the daily volatility of the short rate for $\sqrt{\Delta}$.

To compare these quantities we choose the second approach. In table 5.9 it is possible to compare the implied volatility obtained by the static implementation, that one obtained by the dynamic implementation and the volatility of the short rate proxy. We underly that, not too surprisingly, they are very close to each other.

This result is very interesting. Obviously, it cannot alone justify the validity of the CIR model in the market, because we use as proxy the particular short rate determined by the static implementation of the term structure and not a particular rate exchanged in the market. However, it can be considered as **a proof of the coherence of the two different methods used to implement the CIR model**.

5.6.5 Comparison between the short rate proxy and the Euribor rates

Before to conclude the section of the dynamic implementation, we want to discuss again the delicate choice of the proxy for the short term. We have justified before the use of the static short rate as proxy of the real short rate with three observations:

Comparison between implementations with different short rate proxy From 02/07/2001 to 31/12/2002						
Parameter	Method	Mean	St. Dev.	Rel.Var.	Min	Max
Speed of adjustment k	Dynamic	0.7449	0.2812	37.8%	0.1975	1.6179
	Euribor 3m	0.4036	0.4019	99.5%	0.0227	1.5002
	Euribor 6m	0.3632	0.3082	84.9%	0.0275	1.1609
Long term average rate μ	Dynamic	3.5882	1.0179	28.4%	0.9159	6.6734
	Euribor 3m	4.4346	2.0360	45.9%	0.2744	13.2694
	Euribor 6m	3.7452	1.7658	47.1%	0.0162	7.8324
Implied volatility $\sigma\sqrt{r_t}$	Dynamic	0.7798	0.0569	7.3%	0.6462	0.9016
	Euribor 3m	0.4386	0.0572	13.0%	0.3491	0.5226
	Euribor 6m	0.4251	0.0197	4.6%	0.3761	0.4639

Table 5.10: Comparison between short rate proxy and Euribor rates.

1. The 6-months Euribor or the 1-year Swap rate are too long, considering the existence of a shorter Euribor rate and considering the implicit instantaneous nature of the short rate.

2. In some periods the 3 months Euribor rate shows anomalous behavior respect to the other maturities, so it is possible that it is unfit to describe the whole term structure.

3. Finally, the behavior of the CIR short rate is a quite good synthesis of all information inside the whole term structure.

Now we want to use other information to prove the reasonableness of our choice. To do this, we have also implemented the CIR model using before the 3-months Euribor rate and after the 6-months Euribor rate as a proxy of the short rate. The method used to implement the model is the dynamic approach. Finally, we have compared the results obtained for the Euribor rates and for the static short rate. The parameters obtained with the three different proxies are quoted in table 5.10.

- **The speed of adjustment k:** the speeds of adjustment are similar to each other: it is an evident positive result. However, the relative errors obtained by 3-months Euribor and 6-months Euribor rate are about the double amount of the relative error obtained by the static short rate. As we want to determine the market price of risk, it is better to have a variability of the parameter as small as possible so, for the speed of adjustment, the static short rate seems preferable respect to the others.

- **The long term average rate μ:** for the long term average rate μ each proxy provide the same amount. Consequently, the choice of the proxy seems indifferent for the parameter μ.

- **The implied volatility $\sigma\sqrt{r_t}$:** also if the three implied volatility show the same magnitude, the implied volatility obtained by the static short rate is greater than the implied volatility obtained by the others. In our opinion this is not a problem and the result is, in some ways, obvious. In the archive B, in fact, the Euribor rates (especially the 3-months) show periods where they remain almost constant; periods that, clearly, decrease their total volatility. However, the same behavior is not confirmed by the other rates (e.g. the behavior of the 1 year swap rate. See figure 5.4). Now, if the CIR model have to really represent the market, the real short rate have to consider the fluctuations of the whole term structures and it has to absorb the anomalous constant behavior of the Euribor rates; consequently, it is reasonable to expect that the real short rate shows a greater volatility than the Euribor rates.

For all these reasons we think that the choice of the static short rate as proxy of the real short rate is completely acceptable.

5.7 The market price of risk

5.7.1 On the meaning and on the usefulness market price of risk

We have seen that in the CIR environment the expected return of a bond is equal to (5.2), where the elasticity of the bond is given by:

$$el = r\frac{P_r}{P} = r\frac{1}{P}\frac{\partial P}{\partial r}.$$

In general, an investor wants a greater return for a longer investment, especially respect to the short rate (considering it as the virtual shortest maturity is possible to exchange). For these reasons, almost always, the second term of (5.2) is positive.
Inside the Local Expectation Hypothesis equation there is the derivative of the price P respect to the short rate: it is well known that when the rates go down the prices increase and viceversa so, in average, the derivative in (5.2) is negative. We want to focus on the term "in average": the market is not deterministic and it is possible that in a particular

day or in a particular situation both the prices and the short rate increase or decrease.
But now, if the derivative of P respect to r is negative, the second term of the Local Expectation Hypothesis equation can be positive only if λ is negative. That is the reason why the market price of risk is chosen as $-\lambda$ and not λ.
However, in some very short periods, the market price of risk can be negative. Considering the analysis just developed, the meaning of this situation would have to be clear: in that case the expected return for a generic bond is less than the short rate. It seems to be a non-sense, but this situation represents the moment in which the market strongly believes in a heavy decreasing of the rates.
So, from the market price of risk is possible to evaluate the expectation of the market about the evolution of the rates and of the whole term structure.

5.7.2 Computation of the market price of risk

We recall that in the static implementation we can only obtain $(k + \lambda)$, whereas in the dynamic implementation we can directly obtain the parameter k. So, comparing every day k with $(k + \lambda)$ we can only achieve an **empirical estimation of the market price of risk** $-\lambda$.
Because we have two time series obtained by our implementations, one for the static speed of adjustment and one (restricted to 442 days) for the dynamic speed of adjustment, it is possible to obtain the value of λ for every day from 02/07/2001 to 31/12/2002.
We insist on this point: the computation of the price of risk is exclusively empirical. We remember that the dynamic implementation provides a set of parameters that represents an average of the market, whereas the static implementation provides parameters that represent the actual situation of the market, i.e. the term structure of a single day. So, when we compute the market price of risk, we compare two variables not perfectly homogeneous.
In spite of this consideration, we believe that the approach is correct from a pragmatical point of view: in fact, it is the simplest approximations we can compute from the Archive-B. Moreover, we will see that this approach can help us to understand the tensions of the market. The values of the static $(k + \lambda)$, of the dynamic k and of the market price of risk are showed in table 5.11.

5.7.3 Consideration about the market price of risk

The behavior of the market price of risk showed in 5.7.2, seems to give a good representation of the evolution of the market between 23/04/01 and 31/12/02. In fact:

Estimation of the market price of risk $-\lambda$ From 02/07/2001 to 31/12/2002					
Parameter	Mean	St. Dev.	Rel.Var.	Min	Max
Static speed of adjustment k	0.7449	0.2812	37.8%	0.1975	1.6179
Dynamic speed of adjustment k	0.4078	0.1248	30.6%	0.1791	0.5937
Market price of risk $-\lambda$	0.3236	0.3397	105.0%	-0.2794	1.3493

Table 5.11: Comparison between short rate proxy and Euribor rates.

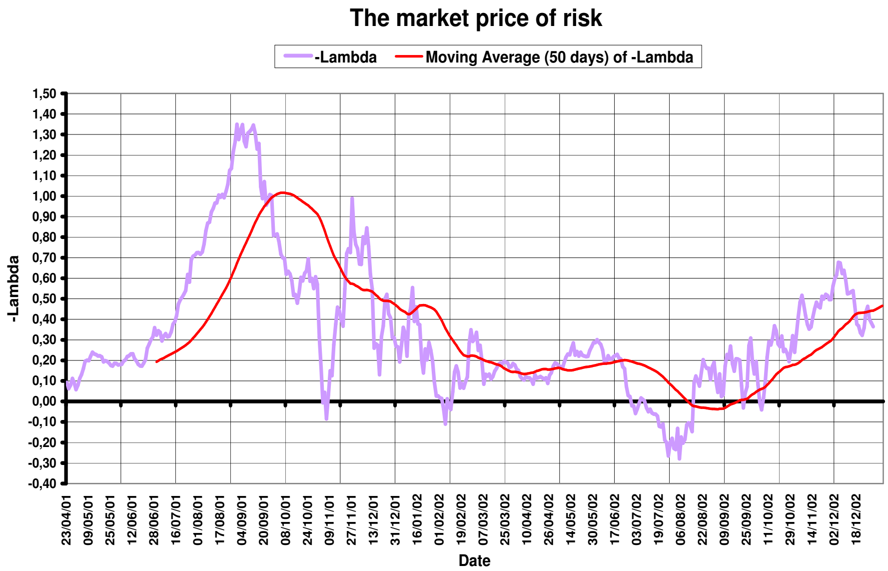

Figure 5.6:

- After the 11/09/01, $-\lambda$ goes down and shows a very nervous behavior for the next 2 or 3 months. These fact are probably due to the shock of the Twin Towers attempt that created new fears and the conviction of a necessary great reduction of the Reference Rate by BCE and by Federal Reserve, to help the economy of Europe and USA.

- After January 2002, $-\lambda$ remains almost stable and low according to the idea of a still long crisis.

- Between 06/02 and 08/02, $-\lambda$ remains under 0. In that very short period, there was the expectation of a new possible reduction of the rates actually happened in the subsequent year.

- After 08/02 the $-\lambda$ come back positive, but nervous, signal of indecision on the market.

In figure 5.7.2 it is also showed the moving average of 50 days of the market price of risk that confirm the above discussion. In some ways, the moving average, eliminating small fluctuations of $-\lambda$, makes clear the different moments of the market described in the previous points. Finally, we want to note that the moving average of the market price of risk remains almost always positive, absorbing the negative expectations between 06/02 and 08/02 and showing a behavior close to the idea of a market price of risk always positive.

5.8 The intrinsic term structure

The market price of risk is very important but, alone, it can not allow to have information about the intrinsic term structure of the market, i.e. the term structure that the market thinks more probable for the next future. In other words, determining the market price of risk and the elasticity of each single Euribor and Swap rate on a specific day, it is possible, by means of (5.2), to determine the expected returns of the market for each maturity and, as consequence, to obtain an estimation of an intrinsic term structure.
Obviously this can not be exactly equal to the real term structure, in part due to the method (the dynamic implementation uses a mean of the parameters of the last 600 days), and in part because the elasticity of a generic maturity computed for a single day, can be affected by tension on the market and can generate very high and non significant fluctuations respect to the short rate.
However, taking average on the time, the extreme fluctuations of the elasticities are absorbed and it becomes possible to obtain a reasonable mean of the return for each maturity (average of (5.2) on the time for each maturity); so, it also becomes possible to obtain an intrinsic term structure representative of the market. In the next two sub-sections we apply this approach to our dataset, in order to study the term structures generated by our implementation of the CIR model.

5.8.1 Computation of the elasticity

It is possible to calculate the elasticity of a generic bond or rate using the incremental ratio. In particular, we could discretize the partial derivative as:

$$\left.\frac{\partial P}{\partial r}\right|_t \sim \frac{P_{t+1} - P_t}{r_{t+1} - r_t},$$

or as:

$$\left.\frac{\partial P}{\partial r}\right|_t \sim \frac{P_t - P_{t-1}}{r_t - r_{t-1}}.$$

Instead, we use a more symmetric approximation, doing an average of the two estimates. In this way it is possible to bound the influence of a temporary fluctuations:

$$\left.\frac{\partial P}{\partial r}\right|_t \sim \frac{1}{2}\Big(\frac{P_{t+1} - P_t}{r_{t+1} - r_t} + \frac{P_t - P_{t-1}}{r_t - r_{t-1}}\Big).$$

This discretization of the derivative is a rude approximation but it is the typical method used to compute a derivative with discrete data (see [40]).

5.8.2 Computation of the expected risk and of expected return

With the elasticity of a particular rate, by means of (5.2) we can calculate, for each day, the expected risk (equal to $\lambda r(P_r/P)$). The results of all maturities are in table 5.12. Finally, adding the short rate to the expected risk for each rate and for each day, we can determine the expected returns. The results of all maturities are in table 5.13.

In column 4 and 5 of table 5.12 we show the average expected risk computed with all data of the archive used to estimate the market price of risk. As we have explained and justified at the beginning of this section, the expected risk can show anomalous fluctuations that can influence the average values showed in table 5.12. The question is: are this anomalous values significant or not? If a practitioner answers yes, then he has to use all data but if he answers no then he has to use only the significant values of them. The simplest method to eliminate the non significant data, or outliers, is to accept only those inside the interval:

Mean-2(St.dev)<(Acceptable Data)<Mean+2(St.dev).

In column 2 and 4 we show the average expected risk without and with outliers. In column 7 there is the number of data used to compute the mean of the expected risk; this number is equal to 442, i.e. equal to number of days between 23/04/2001 and 31/12/2002. In column 6 there is the number of data considered and in column 7 the percentage of outliers respect to the total number of data. It is possible to see that the number of outliers is always very low, independently from the maturities. Finally in table 5.13 we show the average expected return. In figure 5.8.2 and 5.8.2 we show the graphical representation of the expected risk and return.

We want finally to recall that, obviously, the intrinsic term structure of the market is represented by the curves of the expected returns.

Average expected risk of each maturity From 02/07/2001 to 31/12/2002							
	With no outliers		With outliers		Data and outliers		
Rate	Mean	St. Dev.	Mean	St. Dev.	Data	Total	Outliers
3 months Euribor	0.000459	0.003253	0.000994	0.009025	430	442	2.7%
6 months Euribor	0.001332	0.008602	0.003030	0.025693	429	442	2.9%
1 year Swap	0.004679	0.013470	0.005574	0.034333	431	442	2.5%
2 years Swap	0.011554	0.026505	0.013048	0.073381	431	442	2.5%
3 years Swap	0.017188	0.040071	0.020315	0.100709	431	442	2.5%
4 years Swap	0.020125	0.055104	0.024372	0.132335	431	442	2.5%
5 years Swap	0.020139	0.066437	0.026625	0.162625	430	442	2.7%
6 years Swap	0.023516	0.073377	0.022725	0.194423	427	442	3.4%
7 years Swap	0.022416	0.085951	0.021161	0.219579	425	442	3.8%
8 years Swap	0.023986	0.097875	0.028662	0.265035	426	442	3.6%
9 years Swap	0.025434	0.111767	0.028802	0.297041	427	442	3.4%
10 years Swap	0.022053	0.119953	0.027184	0.313965	426	442	3.6%
15 years Swap	0.022836	0.150791	0.026782	0.382543	428	442	3.2%
20 years Swap	0.020486	0.139287	0.021334	0.380468	424	442	4.1%
25 years Swap	0.019037	0.149390	0.022515	0.369489	428	442	3.2%
30 years Swap	0.016764	0.139913	0.017553	0.342910	427	442	3.4%

Table 5.12: Average expected risk of each maturity.

Average expected return of each maturity From 02/07/2001 - 31/12/2002							
	With no outliers		With outliers		Data and outliers		
Rate	Mean	St. Dev.	Mean	St. Dev.	Data	Total	Outliers
3 months Euribor	0.030649	0.006695	0.031108	0.010822	430	442	2.7%
6 months Euribor	0.031331	0.011011	0.033144	0.026497	429	442	2.9%
1 year Swap	0.034582	0.014953	0.035688	0.035252	431	442	2.5%
2 years Swap	0.041628	0.028063	0.043162	0.074044	431	442	2.5%
3 years Swap	0.047261	0.041265	0.050429	0.101318	431	442	2.5%
4 years Swap	0.050844	0.057178	0.054486	0.132789	431	442	2.5%
5 years Swap	0.050160	0.067242	0.056739	0.163049	430	442	2.7%
6 years Swap	0.053568	0.074046	0.052839	0.194676	427	442	3.4%
7 years Swap	0.052460	0.086695	0.051275	0.219894	425	442	3.8%
8 years Swap	0.054046	0.098655	0.058775	0.265446	426	442	3.6%
9 years Swap	0.055488	0.112451	0.058916	0.297417	427	442	3.4%
10 years Swap	0.052113	0.120506	0.057297	0.314280	426	442	3.6%
15 years Swap	0.052912	0.151349	0.056896	0.382846	428	442	3.2%
20 years Swap	0.050514	0.139842	0.051448	0.380695	424	442	4.1%
25 years Swap	0.049113	0.149896	0.052629	0.369784	428	442	3.2%
30 years Swap	0.046825	0.140448	0.047666	0.343163	427	442	3.4%

Table 5.13: Average expected risk of each maturity.

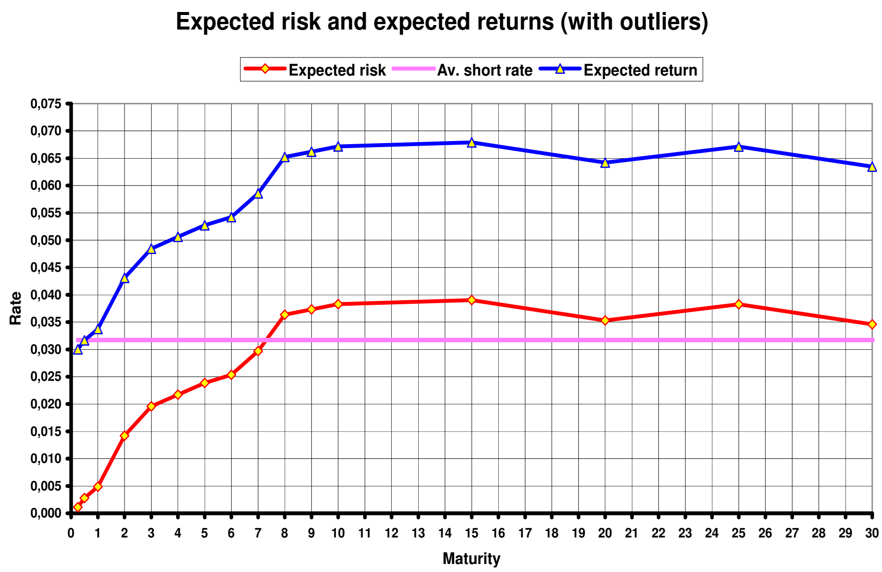

Figure 5.7:

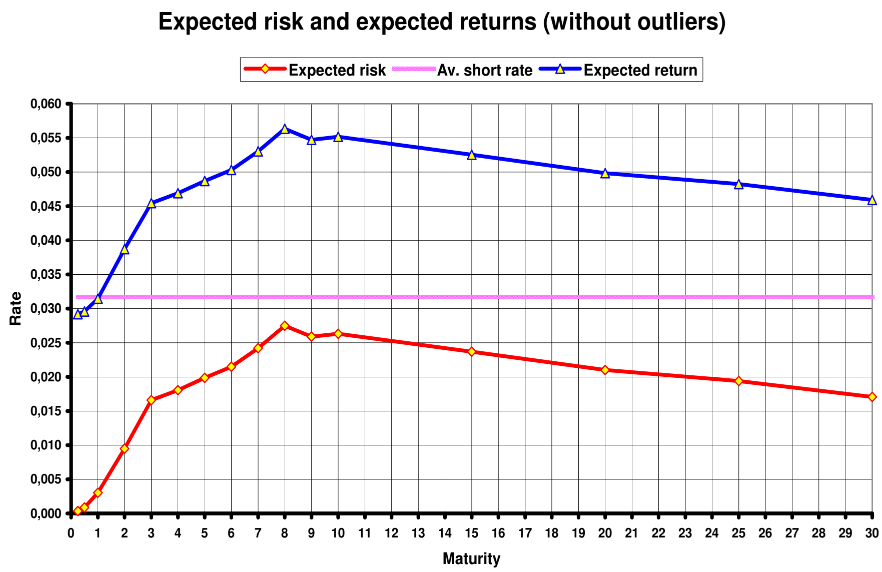

Figure 5.8:

Before to conclude we want to stress some interesting aspects:

- In figure 5.8.2 and 5.8.2 it is possible to see how the presence of the outliers, i.e. the anomalous fluctuations, provides a (reasonable) greater risk premium in maturities

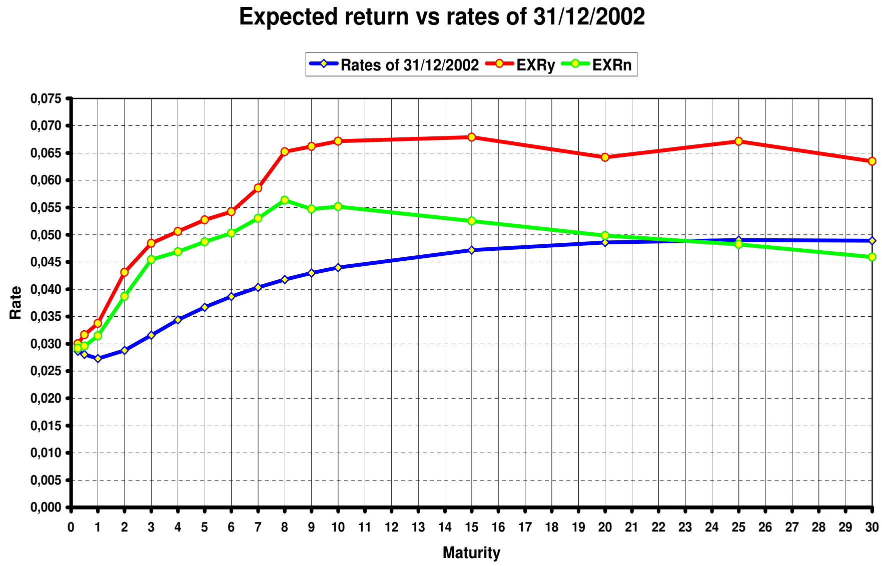

Figure 5.9:

between 3 and 20 years (with the exception of the 6 and 7 years). After 20 years and before 3 years the two "intrinsic" term structure remain very close each other. We can try to justify these facts with two observations:

1. For very long maturities the market seems to have a more stable idea of the risk's cost. In some sense, the long maturities are so far that they become less sensible to the fluctuations of the market[3].

2. For very short maturity the expected risk has to be small, so the expected returns can not be seriously influenced by the fluctuations of the market.

- The expected risk is almost zero for the Euribor rates. Because the short rate proxy and the Euribor rates have to be very close each other, this result is a indirect confirm of the coherence of the study (the expected risk have to go to zero when the maturity tends to zero).

- Both in the expected returns with outliers (we call them EXRy) and the expected returns without outliers (we call them EXRn) we obtain two reasonable term struc-

[3]This is a confirm of the results of L. Torosantucci inside "Progetto INA-IAC". In this work, applying the Nelson-Siegel model, it is proved that the characteristic time of the long term is longer than the characteristic time of the medium and short term

tures. Both the EXRy and the EXRn go a little bit down after the peak at the maturity of 8-9 years.

- It is still unsolved a problem: how to decide if to accept or not to accept the outliers in the computation of the intrinsic term structure? We believe that a deeper study is necessary. However, the differences in returns between EXRy and EXRn are sensitives but not dramatic and this observation let to think that the mistake to avoid the outliers should not be so serious. Anyway, for a practitioner it is probably better to consider the outliers, because this is in agree with the idea that all markets have fat tails that influence in a severe way the behavior of the financial variables.

- In figure 5.8.2 we plot EXRy, EXRn and the real returns of the 31/12/2002, the last day of our archive. We can see that both the EXRy and the EXRn decreases until to reach the present return for the long maturities. This is another confirm of the stability of the long maturities returns respect to the fluctuations above explained.

- Obviously, the expected returns are an average expectation of the market and not the real evolution of the market. The market is not forecastable so it is very difficult that the actual term structure of the last day coincides with the estimated expected return.

5.9 Conclusions

The static implementation of the term structure strongly depends on the quality of the dataset. So, it is better to choose a set of data along all maturities, from the shortest to the longest, being careful about their liquidity. For example, a practitioner could decide to implement the term structure using all the quoted bonds in an arbitrary day. As the quoted short term bonds in Italy (BOT and CTZ) are very illiquid, he could also decide to use only the price of the longer bonds (BTP). Table 5.1 and figure 5.4 prove that it is better to avoid this procedure because the lack on information on the short term could produce anomalous fluctuations in the short rate implemented from the model. For this reason, for a practitioner is better to implement the term structure using very liquid rates as, for example, Euribor and swaps rates.

Chapter 6

PCA based calibration of an HJM model

This chapter is based on *PCA based calibration of an HJM model*, a joint work with Roberto Renó.

6.1 Introduction

In every financial activity interest rates play a crucial role; for this reason the implementation of a reliable (and, possibly, simple) model is very important. In order to be used for practical purposes a good model must be at the same time:

- solid from a theoretical point of view

- flexible and as simple as possible

Usually, it can be very difficult to assess this trade-off. In this chapter, we propose a model which belongs to the HJM family, and we show how to calibrate and use it. Our aim is to provide a model which can be useful for scenario simulation, a common practice, for example, in insurance companies or public debt management. We implemented our model in a fast, light, user-friendly Matlab code. We fitted the model on observed yield curves in the period 1999-2001. Even if the last part of this period is very turbulent, we find a very satisfactory agreement between model assumption and observations.

6.2 The RU model

In this work, we focus on the well-known Heath-Jarrow-Morton framework, proposed in [39]. The HJM framework is a general arbitrage free framework for the term structure of forward rates. It is based on the following equation for the forward rates:

$$df(t,T) = \mu(t,T,f(t,T))dt + \sum_{i=1}^{N} \sigma_i(t,T,f(t,T))dW_i(t), \tag{6.1}$$

where t denotes time, T maturity, $f(t,T)$ is the forward rate and W_i, $i = 1,\ldots,N$ are independent Brownian motions under the risk neutral probability; σ_i are N general functions. In the original paper [39] it is shown that, in order to avoid arbitrage, one must impose the following relation between drift and volatility:

$$\mu(t,T,f(t,T)) = \sum_{i=1}^{N} \sigma_i(t,T,f(t,T)) \int_t^T \sigma_i(t,u,f(t,u))du. \tag{6.2}$$

It follows that the drift in equation (6.1) is completely specified when the number of factors and the corresponding volatility functions are.

In what follows, by an HJM model we mean a precise specification of the number and shape of volatility functions.

We will restrict our choice to a **Gaussian** HJM model, i.e. we drop the dependence from f in σ_i. With this choice of σ_i we can provide, as in [13] and [14], closed analytical formulas for the drift and for derivative prices, such as caps & floors and swaptions.

The main drawback of the Gaussian choice is that it allows negative rates, a feature which, however, is not ruled out by data and which is not as crucial as sometimes maintained in the financial literature. Indeed, interest rate data suggest that the probability of negative rates is very small, not zero. Moreover, there's overwhelming evidence of stochastic volatility in interest rates, see [24, 33, 35], which cannot be accounted for with deterministic volatility, especially for hedging purposes, see the discussion in [27].

In addition to the previous simplifications, we will adopt a constant maturity choice, letting the volatilities depend only on $T - t$. Summarizing we will choose:

$$\sigma_i(t,T,f(t,T)) = \sigma_i(T-t). \tag{6.3}$$

6.2.1 The choice of the volatility structure and the PCA approach

In this section, we sketch our PCA approach. Since the HJM model is factorial, it seems quite natural to extract the main factors driving the evolution of the term structure by principal component analysis (PCA); this approach goes back to [49], and has been largely used in the interest rate applications, see e.g. [58, 62, 26]. It is well known indeed that few factors (a number of three can always be regarded as an excellent approximation) can

explain a large component of the volatility structure across maturities, since the forward rates at different maturities are highly correlated.
In practice, we observe the forward rates on a set of K fixed maturities and the time step of observations is discrete, so that the model (6.1) can be rewritten, in its most general and non redundant form in the following fashion:

$$f_i(t_k) - f_i(t_{k-1}) = \mu_i(t_k - t_{k-1}) + \sum_{j=1}^{K} \sigma_{ij} dW_j(t_k), \; i = 1, \ldots, M. \tag{6.4}$$

We are taking as a reference measure for time $t_k - t_{k-1} = 1$. We remind here that μ is completely specified by the symmetric matrix σ_{ij}.
We set $y_i(t_k) = f_i(t_k) - f_i(t_{k-1})$. Since σ_{ij} is symmetric it can be diagonalized by an orthogonal matrix T, i.e. exist T such that

$$T^+ = T^{-1}, \; T^+ \sigma T = D = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_K \end{pmatrix}, \tag{6.5}$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_K$. Applying T^+ to equation (6.4) and using $T^+T = I$ we obtain:

$$T^+(y - \mu(t_k - t_{k-1})) = T^+ \sigma T T^+ dW = D T^+ dW. \tag{6.6}$$

What we do when we perform principal component analysis is to choose the largest F eigenvalues and replace the matrix D in (6.6) by D':

$$D' = \left(\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_F & \\ \hline & 0 & & 0 \end{array}\right), \tag{6.7}$$

Applying T to (6.6) with D replaced by D' we have

$$y - \mu(t_k - t_{k-1}) \simeq T D' T^+ dW \tag{6.8}$$

So, we easily obtain for the $i-th$ maturity:

$$y_i - \mu_i(t_k - t_{k-1}) = \sum_{m,n,j} T^+_{im} D'_{mn} T_{nj} dW_j = \sum_{k=1}^{F} \sigma_k T_{ik} \sum_{j=1}^{K} T_{jk} dW_j = \sum_{k=1}^{F} \sigma_k T_{ik} d\tilde{W}_k, \tag{6.9}$$

where $d\tilde{W}_k = \sum_{j=1}^{K} T_{jk} dW_j$ are F independent Brownian motions since $\sum_{j=1}^{K} T_{jk}^2 = 1$. We showed that the maturity-dependent weights in the volatility functions in (6.4) are given by the factor loadings of the F eigenvectors with the largest eigenvalues. We remark that in practice we compute the eigenvectors of the matrix Σ given by:

$$\Sigma_{ij} = \mathbb{E}[(y_i - \mathbb{E}[y_i]) \cdot (y_j - \mathbb{E}[y_j])],$$

which however are the same eigenvectors of σ since $\Sigma = \sigma^2$. In order to have a smooth dependence on maturity, we fit the eigenvectors with analytical functions which can be easily integrated. We found, almost generally, that the factor loadings of the volatility matrix eigenvectors can be fitted with the same parametric representation given, setting $x = T - t$, by:

$$\sigma_i(x) = (\alpha_i + \beta_i x)e^{\gamma_i x} + \delta_i. \tag{6.10}$$

Given such a function the integration in equation (6.2) is straightforward:

$$\mu_i = [(\alpha_i + \beta_i x)e^{\gamma_i x} + \delta_i] \cdot \left[\delta_i x + \tfrac{\beta_i}{\gamma_i} x e^{\gamma_i x} + (e^{\gamma_i x} - 1)\left(\tfrac{\alpha_i}{\gamma_i} - \tfrac{\beta_i}{\gamma_i^2}\right)\right]. \tag{6.11}$$

Moreover the functional form (6.10) allows a simple analytical determination of derivative prices. The coefficients $\alpha_i, \beta_i, \gamma_i, \delta_i$ can be calibrated on the historical eigenvectors' factor loadings, using common procedures of minimization, such as gradient descent. Since substituting the eigenvectors with an analytical fit and eliminating some factors may introduce some distortion in the resulting volatility function, we calibrate the weight of the selected factors solving:

$$\min_{\xi_1,\ldots,\xi_F} \sum_{i=1}^{K} \left(\Sigma_{ii} - \sum_{j=1}^{F} \xi_j^2 \sigma_j^2\right)^2 ,$$

since we are more interested in matching the term structure of volatility than eigenvectors themselves. Again this minimum can be found by gradient descent. Finally, our model will be:

$$df_i(t, T-t) = \mu_i(\alpha_i, \beta_i, \gamma_i, \delta_i)dt + \textstyle\sum_{i=1}^{F} \xi_i \sigma_i(T-t; \alpha_i, \beta_i, \gamma_i, \delta_i)dW_i(t), \tag{6.12}$$

with μ_i given by (6.11) and σ_i given by (6.10).

A model quite similar to this can be found in [3, 26, 29]. One-factor humped-volatility models are studied empirically in [53, 61].

	Daily	Weekly	Monthly
Factor 1	0.89151	0.88072	0.84821
Factor 2	0.95526	0.96452	0.96583
Factor 3	0.98009	0.98725	0.99084
Factor 4	0.9858	0.99323	0.99443
Factor 5	0.98983	0.99624	0.99712
Factor 6	0.99285	0.99762	0.99864

Table 6.1: Contribution of the first six factors to total volatility with different time intervals.

6.3 Results

When calibrating a model as described in the previous sections, one needs a data set of historical forward rates sampled with a given frequency (daily, weekly, monthly). The choice of the sampling interval $t_k - t_{k-1}$ is crucial, and it should be chosen according to the time scale one needs to simulate with the model he's going to fit. In life insurance applications, for example, a time difference of one month or even more is highly suggested, since the interesting time horizon can be as long as 30 years. In this regard, a good empirical rule could be to choose the interval $t_k - t_{k-1}$, measured in days, equal to the number of years one is interested in for simulation. For example if daily observations are used, the model is reliable on a time scale of the order of one year. In our analysis, we performed calibration on daily, weekly and monthly data. We tested the algorithm on a data set of daily forward Euro rates computed using deposit rates (Euribor) for maturities under one year and swap rates for maturities of one year and more.[1] The data set at our disposal ranges from January 1999 to December 2001, for a total of 781 yield curves; the maturities range from three months to thirty years. To be precise, our maturities for the Euribor are 3,6 and 9 months, while for the swap rates we have the well-known maturities: yearly from 1 to 10 years, then 12, 15, 20, 25 and 30 years.

Swap and deposit rates have been chosen since they are the most liquid instruments at our disposal.[2] For the common maturity of 1 year, the values of the 12 months Euribor and of 1 year swap are nearly indistinguishable.

For the bootstrapping procedure we need data on all maturities between 1 and 30 years, but historical data are available only for unevenly spaced maturities maturities. In order to have all discount factors, which are needed to run the bootstrap machinery, we adopt

[1]Before starting, we adjust data in order to take in account the conventions on the number of days in a month or in a year.

[2]Actually, for maturities under one year futures on interest rates are more liquid than their underlying.

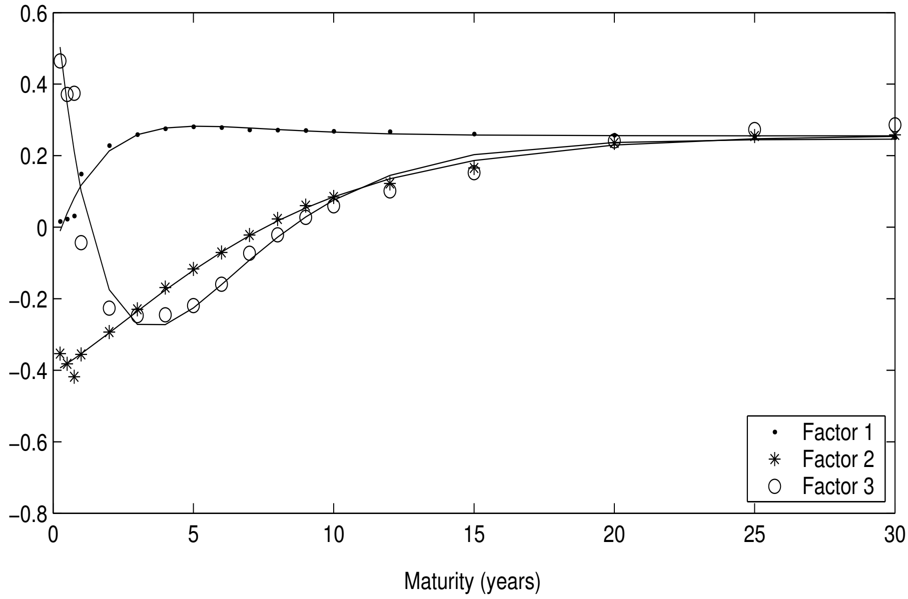

Figure 6.1: Eigenvectors and their fit with daily data

linear interpolation. Table 6.1 shows the result of PCA for three different time intervals: one trading day (daily), five trading days (weekly) and 21 trading days (monthly), on this data set. One factor explains nearly 89% of the variation of daily data, 88% of weekly data and 84% of monthly data. The second and third factor are more important for monthly data than for daily and weekly, indicating that while one-factor models could be adequate for the short-run, they could perform poorly on the long-run. Anyway, in our data set three factors explain more than 98% of the yield curve variation. Figures 6.1 to 6.3 show the fit of the first three eigenvectors with the function (6.10) for the three time intervals selected. The agreement of the fit with data is very good, and there are not shape differences in the factors among different time intervals. The first factor, as expected, is nearly constant: it represents almost equally the whole yield curve, with a small decline at low maturities. The second factor accounts for the difference among long and short maturity volatility, providing negative weights to short maturities and positive weights to long maturities. The third factor is humped, and it accounts for convexity changes.

In figure 6.4 we report the fit of the variance, which displays a well-known hump-shaped behavior, which has been extensively documented in the literature. Our fit is fairly reasonable and the peak is around the maturity of 4 years.

In Table 6.2 we report the estimates of the parameters: as in the previous table, we quote the results for all three different calibrations. Our results are fairly consistent among different time intervals. The biggest difference is for the third factor which, as previously noticed, is practically unimportant for daily changes: its relative weight ξ_3 is very close to

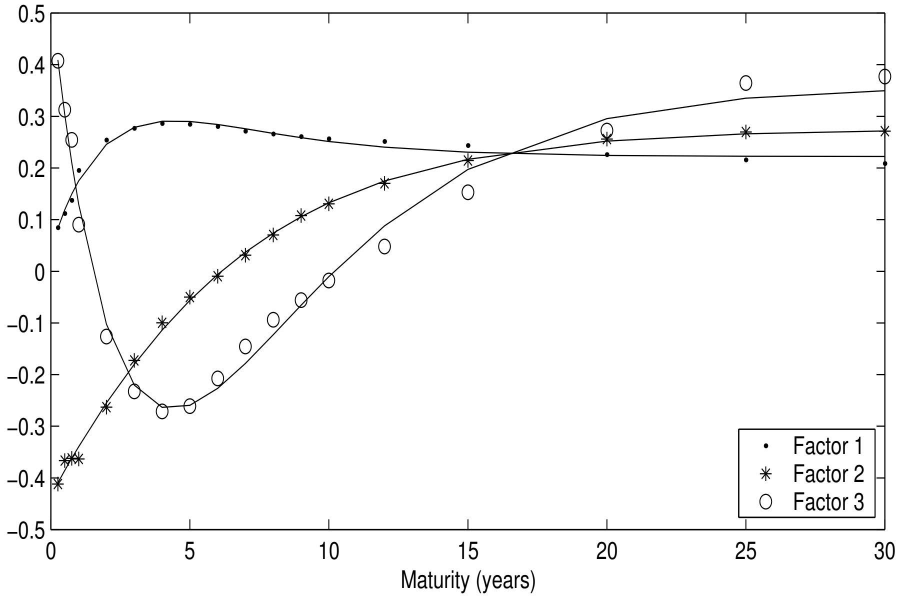

Figure 6.2: Eigenvectors and their fit with weekly data

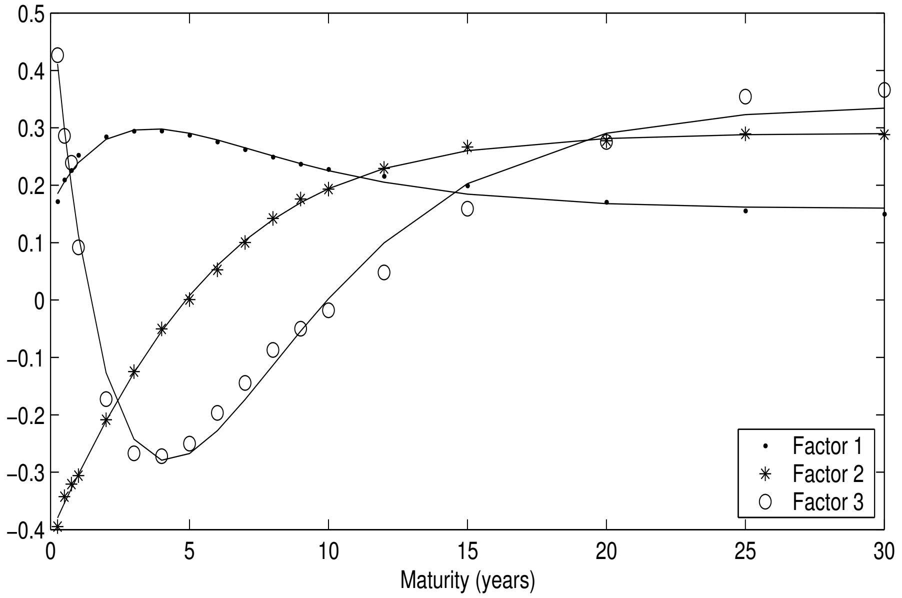

Figure 6.3: Eigenvectors and their fit with monthly data

zero. The coefficients γ, which represent the speed of convergence toward a flat curve, are estimated to be around -0.25, which corresponds to a time interval of 4 years, in agreement with Figure 6.4. The highest values of β, which represents the convexity change, is

Daily					weights
	α	β	γ	δ	ξ
Factor 1	-0.32511	0.11226	-0.43442	0.25544	0.024972
Factor 2	-0.66323	-0.10705	-0.23064	0.25776	0.012068
Factor 3	0.44395	-0.65227	-0.357	0.24681	4.0607×10^{-8}
Weekly					weights
	α	β	γ	δ	ξ
Factor 1	-0.1801	0.11382	-0.34901	0.22222	0.026464
Factor 2	-0.70891	-0.055989	-0.2187	0.27508	0.013335
Factor 3	0.17143	-0.46448	-0.24991	0.35724	0.00067072
Monthly					weights
	α	β	γ	δ	ξ
Factor 1	0.0017753	0.10376	-0.27544	0.15937	0.02386
Factor 2	-0.69607	-0.091575	-0.28125	0.29072	0.014302
Factor 3	0.20093	-0.49623	-0.26477	0.33946	0.0038445

Table 6.2: Parameter estimates.

for the third factor, and it is around -0.5. It is worth nothing that the estimates of the second and third factor are nearly the same for weekly and monthly data. After calibration, we tested the simulations obtained with this model, and we report some results for weekly data. A five-year evolution of the yield curve is shown in figure 6.5. In Figure 6.6 we report some yield curves, showing that the model can give rise to a rich variety of term structures: see also [58]. Two features of this model to be kept in mind for practical applications are:

- it happens with a given probability that the yield curve is a descending function of maturity, which is not very usual

- sometimes negative rates are observed: this effect is a direct consequence of the Gaussian approach. This drawback can be smoothed using the simulation for a reasonably low time horizon.

We give a last practical remark: when structural change are present in the historical data set (e.g. European Central Bank interventions on the discount rate), large changes in the forward rates (especially the ones with shortest maturity) are observed. These large changes could spuriously lead to an higher observed volatility, since the quadratic variation of a time series is strongly affected by large outliers. These observations with a large rate

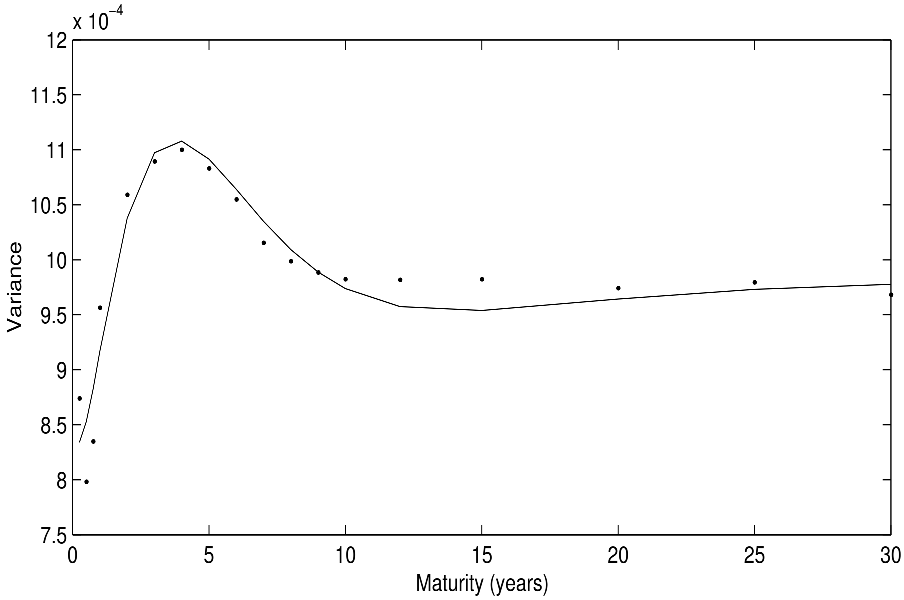

Figure 6.4: Fit of total variance as a function of maturity.

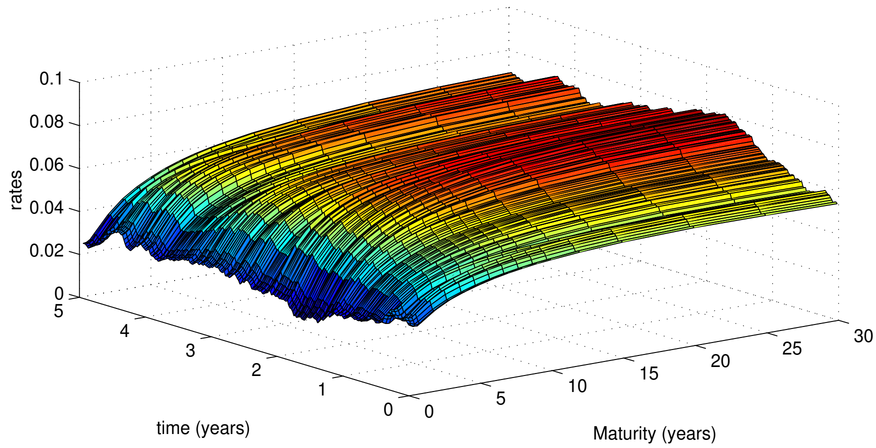

Figure 6.5: Simulated evolution of the yield curve (weekly calibration)

change should be eliminated by the historical sample. Taking into account such large changes is more a question of economics than of stochastic fluctuations.

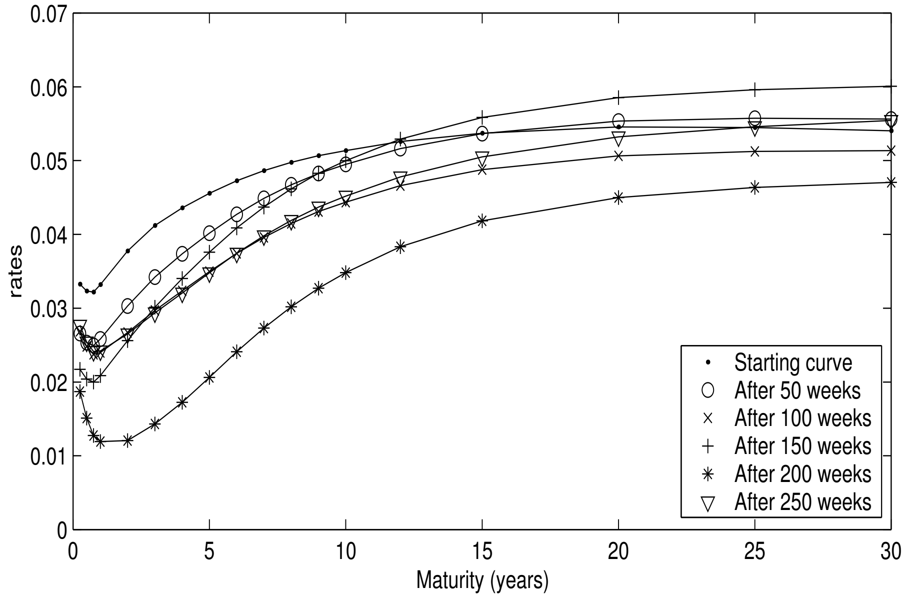

Figure 6.6: Some yield curves (weekly calibration)

6.4 Conclusions

We showed how to implement a model for the term structure of interest rates based on the HJM approach. The approach is simple enough to be easily and rapidly implemented in practice. In spite of its simplicity, the model gives a good description of the time evolution of the yield curve and can be reliably used in simulations. Some drawbacks of the model are reported: however, all these drawbacks are inherent in the choice of a Gaussian approach, and practical rules of thumb to avoid the main drawbacks are given. In this chapter we describe the model and its fit to European interest rate data; further research will consist in using the estimates to evaluate the pricing performance of interest rate derivatives, studying the role of the number of factors and time intervals.

Chapter 7

Consistency

The first paper on this subject is [8]. After this seminal paper, in a few years this research area grew rapidly, becoming up to now one of the most active sectors in Interest Rate Theory. For a detailed survey on the "state of the art" see [30].

7.1 Formal problem statement

We consider a default free bond market with $p(t,T)$ denoting the price at time $t=0$ of a zero coupon bond maturing at $T \geq t$. We assume frictionless markets, i.e. there are no transaction costs, taxes, or short sale restrictions, and the bonds are perfectly divisible. The forward rates $f(t,T)$ are defined as

$$f(t,T) = \frac{\partial logp(t,T)}{\partial T}, \tag{7.1}$$

and the short rate is $r(t) = f(t,t)$.
Suppose that we have a concrete term structure model, defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{Q}, \{\mathcal{F}_t\}_{t\geq 0})$, and that the model is free of arbitrage in the sense that the probability measure $\mathcal{Q}$ is a martingale measure. The physical probability measure $\mathcal{P}$ will play no role below, so all calculations are carried out under $\mathcal{Q}$.
More specifically, we assume that under the martingale measure $\mathcal{Q}$, the dynamics of the forward rates are of the form

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

where W is an m-dimensional $\mathcal{Q}$-Wiener process, while α and σ are as usual the drift and volatility functions. Thus, far each maturity date T, the evolution across calendar time t (where $t \leq T$) of the forward rate $f(t,T)$ is governed by this stochastic differential equation.
We furthermore assume that we have a parameterized family of forward rate curves

$$G : \mathcal{Z} \to [0, +\infty), \tag{7.2}$$

with $\mathcal{Z} \subseteq \mathbb{R}^d$ the parameter space, i.e. far each parameter value $z \in \mathcal{Z}$ we have a smooth curve $G(z)$. By slight abuse of notation we will sometimes write the curve as

$$x \mapsto G(x, z),$$

where the variable x is interpreted as the time to maturity, as opposed to the time of maturity T; clearly $x = T - t$. The main problem is to determine under which conditions the interest rate model (7.1) is **consistent** with the parameterized family of forward curves (7.2), in the following sense:

- Assume that, at an arbitrarily chosen time $t = s$, we have fitted a forward curve G to market data. Technically, this means that we have specified an initial forward curve, i.e. far some $z_0 \in \mathcal{Z}$ we have

$$f^*(s, s + x) = G(x, z_0), \forall x \geq 0.$$

- Is it then the case that the the subsequent forward curves produced by the interest rate model (7.1) always stay within the given forward curve family, i.e. does there at every fixed time $t \geq s$ exist some $z \in \mathcal{Z}$ such that

$$f(t, t + x) = G(x, z), \forall x \geq 0?$$

  Here, z may depend on t and on the elementary outcome $\omega \in \Omega$.

Observation 7.1 *Since we want to consider fairly general interest rate models, including those who are not time homogeneous, we are forced to consider an arbitrary initial time $t = s$, rather than just $t = 0$.*

Observation 7.2 *Note that we take the volatility structure $\sigma(t, T)$ as given. This structure is then kept fixed, so the calibration procedure only concerns the choice of initial forward rate curve. Put into measure theoretic terms this means that, by fixing the volatility structure, we have fixed a class of equivalent martingale measures, and the calibration step will then pin down a particular member of this class.*

To see more clearly what is going on in differential geometric terms, we define the forward curve manifold $\mathcal{G}$ as the set of all forward curves produced by the parameterized family.

Definition 7.3 *Given a smooth mapping $G : \mathcal{Z} \to C[0, +\infty)$, we define the forward curve manifold $\mathcal{G}$ by*

$$\mathcal{G} = ImG = \{G(\cdot, z) \in C[0, +\infty) | z \in \mathcal{Z}\}.$$

As we introduced in chapter 2 it is convenient to use the Musiela parametrization of forward rates

$$\tilde{f}(t,x) = f(t,t+x),$$

where the symbol x again denotes time to maturity as opposed to T, which denotes time of maturity. We denote the induced dynamics far the $\tilde{f}$-process by

$$d\tilde{f}(t,x) = \tilde{\mu}(t,x)dt + \tilde{\sigma}(t,x)dW(t). \tag{7.3}$$

We proved that is a one-to-one correspondence between the formulations (7.1) and (7.3):

$$\begin{aligned} \tilde{\mu} &= \tfrac{\partial \tilde{f}(t,x)}{\partial x} + \mu(t,t+x), \\ \tilde{\sigma}(t,x) &= \sigma(t,t+x). \end{aligned}$$

From now, with a clear abuse of notation we remove the symbol "$\sim$" from f,μ,σ, because in the whole section we will consider the HJM model under the Musiela parametrization. Thus the model $\mathcal{M}$ will be simply characterized by the particular volatility function $\sigma(t,x)$ in

$$df(t,x) = \Big(\frac{\partial f(t,x)}{\partial x} + \sigma(t,x)\int_o^x \sigma(t,u)^T du\Big)dt + \sigma(t,x)dW(t), \tag{7.4}$$

Definition 7.4 *Let be given a forward curve manifold $\mathcal{G}$ and a forward rate process $f(t,x)$. We say that $\mathcal{G}$ is invariant under the action of f if, for every fixed initial time s, the condition $f(s,\cdot) \in \mathcal{G}$ implies that $f(t,\cdot) \in \mathcal{G}, \forall t \geq s, \mathcal{G}$-a.s..*

We may now phrase the main problem in the following way:

- Suppose that we have a forward rate model $\mathcal{M}$ as in (7.4) and a forward curve manifold $\mathcal{G}$.

- Is $\mathcal{G}$ then invariant under the action of f?

Thus, the pair $(\mathcal{M},\mathcal{Q})$ is consistent if and only if the manifold $\mathcal{G}$ is invariant under the action of f, and the question we pursue is when this happens.
We now introduce a new tool, the Stratonovich integral.

Definition 7.5 *For given semimartingales X and Y, the Stratonovich integral of X with respect to Y is defined as:*

$$\int_0^t X(s) \circ dY(s) = \int_0^t X(s)dY(s) + \frac{1}{2}\langle X,Y\rangle_t, \tag{7.5}$$

where the first term on the right side is the Itô Integral.

In the entire work, based only on Wiener processes as driving noise, we can define the "quadratic variation process" $\langle X,Y\rangle_t$ in (7.5) by

$$\langle X,Y\rangle_t=\int_0^t dX(s)dY(s),$$

with the usual multiplication rules $dW\cdot dt=dt\cdot dt=0$ and $dW\cdot dW=dt$.

7.2 Invariant forward curve models

Assume that, by the Musiela parametrization of the HJM no arbitrage drift condition, the Itô dynamics for the forward rates are given by (7.4) under the martingale measure $\mathcal{Q}$. This implies that the Stratonovich dynamics under $\mathcal{Q}$ are given by

$$df(t,x)=\Big(\frac{\partial\tilde f(t,x)}{\partial x}+\sigma(t,x)\int_o^x\sigma(t,u)^T du\Big)dt-\frac{1}{2}d\langle\sigma(\cdot,x),W\rangle(t)+\sigma(t,x)dW(t), \tag{7.6}$$

This is the most general formulation of the infinite-dimensional stochastic case model $\mathcal{M}$. In fact, in many cases the quadratic variation process may be written in intensity form as

$$-\frac{1}{2}d\langle\sigma(\cdot,x),W\rangle(t)=\phi(t,x)dt. \tag{7.7}$$

In particular, in our case of a deterministic volatility function $\sigma(t,x)$, the Stratonovich formulation coincides with the Itô formulation, so we come back to (7.6).
We again take as given the forward curve manifold $\mathcal{G}$: the relevant invariance concept is the following.

Definition 7.6 *Consider a given interest rate model $\mathcal{M}$, specifying a forward rate process $f(t,x)$, as well as a forward curve manifold $\mathcal{G}$. We say that $\mathcal{G}$ is f-invariant under the action of the forward rate process $f(t,x)$ if there exists a stochastic process Z with state space $\mathcal{Z}$ and possessing a Stratonovich differential of the form*

$$dZ(t)=\gamma(t,Z(t))dt+\psi(t,Z(t))\circ dW(t), \tag{7.8}$$

such that, far every fixed choice of initial time s, whenever $y(s,.)\in\mathcal{G}$, the stochastic process defined by

$$y(t,x)=G(x;Z(t)),\forall t\geq s,x\geq 0,$$

solves the SDE (7.6) with initial condition $f(s,\cdot)=y(s,\cdot)$.

Observation 7.7 *The interpretation is that the Z-process describes the evolution of the z-parameters as the forward rate curve moves on the manifold $\mathcal{G}$. It is obvious that f-invariance implies invariance, and the conjecture is that invariance plus some minimal smoothness of the mapping G implies f-invariance.*

We may now state and prove the main invariance result. We assume that the forward rate Itô dynamics of $\mathcal{M}$ are given by (7.4), and that the quadratic variation process has the structure (7.7). We also assume that (7.8) has a solution that is unique in distribution, since in this case any solution y of (7.6) in the form $y(t,x) = G(x; Z(t))$ is unique in distribution, too, and hence y may be used in place of f.

Theorem 7.8 *The forward curve manifold $\mathcal{G}$ is f invariant for the forward rate process $f(t,x)$ in $\mathcal{M}$ if and only if*

$$G_x(\cdot,z) + \sigma(t,\cdot)\int_o^{\cdot} \sigma(t,u)^T du + \phi(t,\cdot) \in Im[G_z(\cdot,z)], \tag{7.9}$$

$$\sigma(t,\cdot) \in Im[G_z(\cdot,z)], \tag{7.10}$$

for all $(t,z) \in \mathbb{R}_+ \times \mathcal{Z}$. Here, G_z and G_x denote as usual the Frechet derivatives of G with respect to z and x, respectively, which are assumed to exist. Naturally, the condition (7.10) is interpreted columnwise for σ. Condition (7.9) is called "the consistent drift condition", and (7.10) is called "the consistent volatility condition".

Proof To prove necessity we assume f-invariance. Then we may take the differential in equation $y(t,x) = G(x; Z(t))$, producing

$$dy(t,x) = G_z(x, Z(t))\gamma(t, Z(t))dt + G_z(x, Z(t))\psi(t, Z(t)) \circ dW(t).$$

Comparing this equation with equation (7.6) and equating coefficients yields, (7.9) and (7.10), for every $t \geq s$, but with $z = Z(t)$. Since the initial time s as well as the initial point $f(s,\cdot)$ (and thus $Z(s)$) can be chosen arbitrarily, we finally obtain (7.9) and (7.10) in full generality.
To prove sufficiency, assume (7.9) and (7.10). Then we may select $\gamma : \mathbb{R}_+ \times \mathcal{Z} \to \mathbb{R}^d$ and $\psi : \mathbb{R}_+ \times \mathcal{Z} \to \mathbb{R}^{d\times m}$ satisfying

$$G_x(\cdot,z) + \sigma(t,\cdot)\int_o^{\cdot} \sigma(t,u)^T du + \phi(t,\cdot) = G_z(\cdot,z)\gamma(t,z),$$

and

$$\sigma(t,\cdot) = G_z(\cdot,z)\psi(t,z).$$

for all $(t,z) \in \mathbb{R}_+ \times \mathcal{Z}$.
Let $y(s,\cdot) \in \mathcal{G}$, i.e. $y(s,\cdot) = G(\cdot, z_0)$, for some $z_0 \in \mathcal{Z}$. Define Z as the solution to (7.8) with initial condition $Z(s) = z_0$, and define the infinite-dimensional process $y(x,t)$ by $y(t,x) = G(x, Z(t))$. Then

$$\begin{aligned} dy(t,x) &= G_z(x,Z(t))\gamma(t,Z(t))dt + G_z(x,Z(t))\psi(t,Z(t)) \circ dW(t) \\ &= \Big(G_x(x,Z(t)) + \sigma(t,x)\textstyle\int_o^x \sigma(t,u)^T du + \phi(t,x)\Big)dt + \sigma(t,x) \circ dW(t) \\ &= \Big(\tfrac{\partial}{\partial x} y(t,x) + \sigma(t,x)\textstyle\int_o^x \sigma(t,u)^T du + \phi(t,x)\Big)dt + \sigma(t,x) \circ dW(t). \end{aligned}$$

Thus, y solves the SDE (7.6). □

The moral is that, subject to regularity conditions, we have invariance if and only if conditions (7.9) and (7.9) hold. This leads us to the following definition.

Definition 7.9 *We say that the interest rate model $\mathcal{M}$ is consistent with the forward rate manifold $\mathcal{G}$ if the consistent drift and volatility conditions (7.9) and (7.9) hold.*

For all practical purposes, this implies f-consistency, as well, since the additional regularity condition on the coefficients γ and ψ usually is satisfied in the models of interest in finance. Summing up, we will have to deal with

$$\left\{ \begin{array}{rcl} G_x(\cdot, z) + \sigma(t, \cdot) \int_0^{\cdot} \sigma(t, u)^T du & \in & Im(G_z(\cdot, z)), \\ \\ \sigma(t, \cdot) & \in & Im(G_z(\cdot, z)). \end{array} \right. \tag{7.11}$$

for all $(t, z) \in \mathbb{R}_+ \times \mathcal{Z}$.

7.3 The Hull-White Model

The Hull-White model (from now briefly HW model) is the generalization to the time-dependent case of the celebrated Vasicek model, and was proposed in [41].

We showed the properties of the HW model in the first chapter, and we come back to this model in order to introduce the basic ideas that we will use in the main theorem of this chapter, about the new model presented in chapter 6.

The HW model is based on

$$dr(t) = (\phi(t) - \alpha r(t))dt + \sigma dW(t),$$

where $\alpha, \sigma > 0$. The drift function $\phi(t)$ is calibrated from the observed initial forward curve $\{f^*(0,T)|T \geq 0\}$, and has the closed form showed in the first chapter.

We now turn to the HJM formulation:

$$df(t) = (\mu(t,T)dt + \sigma e^{-\alpha(T-t)} dW(t),$$

so we have, setting $x = T - t$:

$$\sigma(t, x) = \sigma e^{-\alpha x}.$$

With this choice of the volatility function, the system (7.11) becomes:

$$\left\{ \begin{array}{rcl} G_x(x, z) + \frac{\sigma^2}{\alpha}(e^{-\alpha x} - e^{-2\alpha x}) & \in & Im(G_z(x, z)), \\ \\ \sigma e^{-\alpha x} & \in & Im(G_z(x, z)). \end{array} \right. \tag{7.12}$$

- **Step1** Let's start with the "natural" choice, the full NS family:

$$G(x,z) = z_1 + z_2 e^{-z_4 x} + z_3 x e^{-z_4 x}, \tag{7.13}$$

which gives us:

$$\begin{aligned} G_z(x,z) &= [1, e^{-z_4 x}, x e^{-z_4 x}, -(z_2 + z_3 x) x e^{-z_4 x}], \\ G_x(x,z) &= (z_3 - z_2 z_4 - z_3 z_4 x) x e^{-z_4 x}. \end{aligned}$$

We consider now the second equation, the simplest one: we are looking for A, B, C, D such that for all $x \geq 0$ we have

$$\sigma e^{-\alpha x} = A + B e^{-z_4 x} + C x e^{-z_4 x} - D(z_2 + z_3 x) x e^{-z_4 x}.$$

One can easily see that it is possible if and only if $z_4 = \alpha$. So,

First hint: fix $z_4 = \alpha$ in the parametrization.

- **Step2** Let's now consider the restricted NS family:

$$G(x,z) = z_1 + z_2 e^{-\alpha x} + z_3 x e^{-\alpha x}. \tag{7.14}$$

The second equation is easily verified, while in the first we look for A, B, C such that for all $x \geq 0$ we have:

$$(z_3 - \alpha z_2 - \alpha z_3 x) x e^{-\alpha x} + \frac{\sigma^2}{\alpha}(e^{-\alpha x} - e^{-2\alpha x}) = A + B e^{-\alpha x} + C x e^{-\alpha x}.$$

This equation can never be verified (unless $\alpha = 0$), due to the fact that there are two different exponential. So

Second hint: include $e^{-2\alpha x}$ in the parametrization.

- **Step3** Let's consider the extended NS family:

$$G(x,z) = z_1 + z_2 e^{-\alpha x} + z_3 x e^{-\alpha x} + z_4 e^{-2\alpha x}. \tag{7.15}$$

Theorem 7.10 *The HW model is consistent with the extended NS family (7.15).*

Proof The second equation is easily verified, while in the first one we are looking for A, B, C, D such that for all $x \geq 0$ we have

$$z_2 \alpha e^{-\alpha x} + z_3 e^{-\alpha x} + z_3 x \alpha e^{-\alpha x} + 2 z_4 \alpha e^{-2\alpha x} - \frac{\sigma^2}{\alpha}(e^{-\alpha x} - 1) e^{-\alpha x} = A - e^{-\alpha x} B - x e^{-\alpha x} C - e^{-2\alpha x} D.$$

We simply give the solution, as only routine calculation (boring and trivial) has to be done:

$$\begin{cases} A &= 0, \\ B &= z_3 - z_2\alpha + \frac{\sigma^2}{\alpha}, \\ C &= -\alpha z_3, \\ D &= -\frac{\sigma^2}{\alpha} - 2z_4\alpha. \end{cases}$$

□

7.4 The exponential-polynomial family

Our model presented in chapter (6) is in some sense an extension of the HW model, in fact if we take $\beta = 0$ we obtain a translated version of the HW model. We showed in the last section that the HW model is consistent only with the extended version of the NS family (7.15), so if we want to obtain consistency (which is our main goal in this chapter) we must extend another time our family of curves.
The choice naturally goes toward a generalization that include the NS family (7.13). We then restrict our study to a Gaussian framework, as our model has a deterministic volatility: the HJM formulation of the model will be of the type (7.4).

$$df(t,x) = \Big(\frac{\partial f(t,x)}{\partial x} + \sigma(t,x)\int_o^x \sigma(t,u)^T du\Big)dt + \sigma(t,x)dW(t), \tag{7.16}$$

In order to specify the forward curve manifold, let us fix a positive integer K and a vector $n = [n_1, \ldots, n_K]$ having non-negative integers as its components.

Definition 7.11 *The Exponential-Polynomial forward curve manifold $EP(K,n)$ is defined as the set of all curves of the form*

$$G(x) = \sum_{i=1}^{K} p_i(x)e^{-a_i x}, \tag{7.17}$$

where $a_i \in \mathbb{R}$ for all i, and where p_i is any polynomial with $deg(p_i) \leq n_i$ for all i.

If we write the polynomial p_i as

$$p_i(x) = \sum_{j=0}^{n_i} z_{i,j}x^j,$$

we see that a particular polynomial p_i is determined by its $(N_I + 1)$-dimensional vector of coefficients $z_i = [z_{i,0}, \ldots, z_{i,n_i}]$.
The family $EP(k, n)$ is thus defined as the mapping

$$G : \mathbb{R}^{|n|+K} \times \mathbb{R}^K \to C[0, +\infty),$$

where $|n| = \sum n_i$, and $G(x, z, a)$ is given by (7.17) with $a = [a_1, \ldots, a_K]$ and $z = [z_1, \ldots, z_K]$. It is so clear that the dimension of the manifold is $|n| + 2K$.
Unfortunately, the only known result for this family is negative.

Proposition 7.12 *No non trivial model of the form (7.16 is consistent with the forward curve family $EP(K, n)$.*

Hence, in order to obtain consistency, we must slightly modify our approach, fixing exponents in the family.

Definition 7.13 *Consider a fixed choice of (K, n) and a vector $a = [a_1, \ldots, a_K] \in \mathbb{R}^K$. The Fixed Exponential-Polynomial forward curve manifold $FEP(K, n, a)$ is defined as the set of all curves of the form*

$$G(x) = \sum_{i=1}^{K} p_i(x) e^{-a_i x}, \tag{7.18}$$

where p_i is any polynomial with $deg(p_i) \leq n_i$ for all i.

For the $FEP(K, n, a)$ manifold we have that

$$G : \mathbb{R}^{|n|+K} \to C[0, +\infty),$$

and the dimension of the manifold is now $|n| + K$.
We now turn to the last specification of the family: for reasons that we will explain in chapter 8, we are induced to introduce a **linear term** inside the family. This term can obviously be considered also in the family (7.18), taking $a_{\tilde{i}} = 0$ for some $\tilde{i}$ and choosing in the appropriate way the polynomial $p_{\tilde{i}}$. We want instead to stress the importance of the linear term writing it explicitly in the parametrization: without this term, we will have again a negative result, as we showed for the $EP(K, n)$ family.
At the end we can introduce the "right" family.

Definition 7.14 *Consider a fixed choice of (K, n) and a vector $a = [a_1, \ldots, a_K] \in \mathbb{R}^K$. The Linear Plus Fixed Exponential-Polynomial forward curve manifold $LPFEP(K, n, a)$ is defined as the set of all curves of the form*

$$G(x) = m + nx + \sum_{i=1}^{K} p_i(x) e^{-a_i x}, \tag{7.19}$$

where p_i is any polynomial with $deg(p_i) \leq n_i$ for all i.

In particular we will deal with a specific version of the family, precisely:

$$G(x) = z_1 + z_2 x + z_3 e^{ax} + z_4 x e^{ax} + z_5 x^2 e^{ax} + z_6 e^{2ax} + z_7 x e^{2ax} + z_8 x^2 e^{2ax}. \tag{7.20}$$

Again, the deep motivations of this choice will be cleansed in chapter 8.

7.5 Finite dimensional realizations

7.6 Extended HW = Restricted RU

We can now apply the whole machinery developed up to this section. We recall that the HW model has this choice of volatility:

$$\sigma(t,x) = \sigma e^{-\alpha x}.$$

Fixing the particular choice of $\beta = 0$ in the RU model, we can thus extend the HW model:

$$\sigma(x) = (\alpha + \beta x)e^{\gamma x} + \delta. \tag{7.21}$$

Theorem 7.15 *The restricted RU model obtained using the volatility function (7.21) is consistent with the $FEP(K,n,b)$ family.*

Proof Developing partial derivatives and calculations, equation (7.10 becomes

$$\alpha e^{\gamma x} + \delta = A + e^{\gamma x}B + xe^{\gamma x}C + e^{2\gamma x}D + xe^{2\gamma x}E + xF,$$

with the usual meaning that we are looking for A, B, C, D, E, F such that the equation hold for all $x \geq 0$. Collecting exponentials, we have

$$(-D - xE)e^{2\gamma x} + (\alpha - xC - B)e^{\gamma x} + \delta - A - xF = 0,$$

and we thus easily obtain the solution

$$A = \delta, B = \alpha, C = 0, D = 0, E = 0, F = 0.$$

For equation (7.9) a little bit of work has to be done: written in the form (7.11), we look for A, B, C, D, E, F such that

$$z_2\gamma e^{\gamma x} + z_3 e^{\gamma x} + z_3 x\gamma e^{\gamma x} + 2z_4\gamma e^{2\gamma x} + z_5 e^{2\gamma x} + 2z_5 x\gamma e^{2\gamma x} + z_6 + \frac{(\alpha e^{\gamma x} + \delta x\gamma - \alpha)(\alpha e^{\gamma x} + \delta)}{\gamma}$$

$$= A + e^{\gamma x}B + xe^{\gamma x}C + e^{2\gamma x}D + xe^{2\gamma x}E + xF.$$

holds for all $x \geq 0$.
Collecting exponentials we obtain

$$\begin{aligned} 0 = \quad & (-2z_4\gamma^2 - z_5\gamma - \alpha^2 + D\gamma + xE\gamma - 2z_5x\gamma^2)e^{2\gamma x} \\ & +(\alpha^2 - z_3\gamma + xC\gamma - \alpha\delta - \delta x\gamma\alpha - z_3x\gamma^2 - z_2\gamma^2 + B\gamma)e^{\gamma x} \\ & +A\gamma + xF\gamma - z_6\gamma + \alpha\delta - \delta^2 x\gamma. \end{aligned}$$

Imposing equal to 0 the coefficients of exponentials, we obtain the solution:

$$\begin{cases} A &= \delta^2 x + z_6 - \frac{\alpha\delta}{\gamma}, \\ B &= -\frac{\alpha^2}{\gamma} + z_3 + \frac{\alpha\delta}{\gamma} + \gamma z_2, \\ C &= \alpha\delta + z_3\gamma, \\ D &= 2z_4\gamma + z_5 + \frac{\alpha^2}{\gamma}, \\ E &= 2z_5\gamma, \\ F &= 0. \end{cases}$$

□

7.7 The RU model

Finally, we deal with the new model proposed in 6. We recall that in the RU model the volatility structure is given as in (7.21) by

$$\sigma(x) = (\alpha + \beta x)e^{\gamma x} + \delta.$$

As our model is a Gaussian model, because we chose a deterministic volatility function, equations (7.9) and (7.10) can be rewritten in the simple context of (7.11).
We thus have

$$G_x(\cdot, z) + \sigma(t, \cdot)\int_0^{\cdot} \sigma(t, u)^T du \in Im(G_z(\cdot, z)), \tag{7.22}$$

and

$$\sigma(t, \cdot) \in Im(G_z(\cdot, z)). \tag{7.23}$$

Theorem 7.16 (Main Theorem) *The RU model is consistent with the $LPFEP(K, n, b)$ family in the form (7.20).*

Proof Let's begin with the easier equation (7.10): we look for A, B, C, D, E, F, G, H such that for all $x \geq 0$ we have

$$(\alpha + \beta x)e^{\gamma x} + \delta = A + e^{\gamma x}B + xe^{\gamma x}C + x^2e^{\gamma x}D + e^{2\gamma x}E + xe^{2\gamma x}F + x^2e^{2\gamma x}G + xH.$$

Collecting exponentials we get

$$(-E - xF - x^2G)k^2 + (a + bx - B - xC - x^2D)k + d - A - xH = 0,$$

and we easily obtain the solution:

$$A = d, B = a, C = b, D = 0, E = 0, F = 0, G = 0, H = 0.$$

Now we turn to equation (7.9), written in the form (7.11).
We are looking for A, B, C, D, E, F, G, H such that for all $x \geq 0$ the following relation holds:

$$\begin{aligned} & z_3\gamma e^{\gamma x} + z_4e^{\gamma x} + z_4x\gamma e^{\gamma x} + 2z_5xe^{\gamma x} + z_5x^2\gamma e^{\gamma x} + 2z_6\gamma e^{2\gamma x} + z_7e^{2\gamma x} + 2z_7x\gamma e^{2\gamma x} + 2z_8xe^{2\gamma x} \\ & \quad +2z_8x^2\gamma e^{2\gamma x} + z_2 + \frac{(e^{\gamma x}\alpha\gamma + e^{\gamma x}\beta\gamma x - e^{\gamma x}\beta + \delta x\gamma^2 - \alpha\gamma + \beta)((\alpha + \beta x)e^{\gamma x} + \delta)}{\gamma^2} \\ & \quad = A + e^{\gamma x}B + xe^{\gamma x}C + x^2e^{\gamma x}D + e^{2\gamma x}E + xe^{2\gamma x}F + x^2e^{2\gamma x}G + xH. \end{aligned}$$

Collecting the exponentials we get

$$\begin{aligned} 0 = \; & \Big(2\alpha\gamma\beta x + \beta^2\gamma x^2 - \beta^2x - \beta\alpha + \alpha^2\gamma - E\gamma^2 + 2z_6\gamma^3 + z_7\gamma^2 \\ & +2z_8x^2\gamma^3 + 2z_8x\gamma^2 - xF\gamma^2 - x^2G\gamma^2 + 2z_7x\gamma^3\Big)e^{2\gamma x} \\ & +\Big(\beta\gamma x\delta + z_3\gamma^3 + \delta x^2\gamma^2\beta - \alpha\gamma\beta x + \beta^2x + z_4x\gamma^3 + \delta x\gamma^2\alpha + \beta\alpha - B\gamma^2 \\ & -xC\gamma^2 - \beta\delta - x^2D\gamma^2 + \alpha\gamma\delta - \alpha^2\gamma + z_4\gamma^2 + z_5x^2\gamma^3 + 2z_5x\gamma^2\Big)e^{\gamma x} \\ & -\alpha\gamma\delta + z_2\gamma^2 + \delta^2x\gamma^2 - xH\gamma^2 + \beta\delta - A\gamma^2 \end{aligned}$$

Imposing the coefficients equal to 0 we obtain the following system:

$$\begin{cases} 0 = & 2\alpha\gamma\beta x + \beta^2\gamma x^2 - \beta^2x - \beta\alpha + \alpha^2\gamma - E\gamma^2 + 2z_6\gamma^3 + z_7\gamma^2 + 2z_8x^2\gamma^3 + 2z_8x\gamma^2 \\ & -xF\gamma^2 - x^2G\gamma^2 + 2z_7x\gamma^3, \\ & \\ 0 = & \beta\gamma x\delta + z_3\gamma^3 + \delta x^2\gamma^2\beta - \alpha\gamma\beta x + \beta^2x + z_4x\gamma^3 + \delta x\gamma^2\alpha + \beta\alpha - B\gamma^2 - xC\gamma^2 \\ & -\beta\delta - x^2D\gamma^2 + \alpha\gamma\delta - \alpha^2\gamma + z_4\gamma^2 + z_5x^2\gamma^3 + 2z_5x\gamma^2, \\ & \\ 0 = & -\alpha\gamma\delta + z_2\gamma^2 + \delta^2x\gamma^2 - xH\gamma^2 + \beta\delta - A\gamma^2. \end{cases}$$

Since these relations must hold for all $x \geq 0$, we sort elements with respect to x:

$$\begin{cases} 0 = & (\beta^2\gamma + 2z_8\gamma^3 - G\gamma^2)x^2 + (2z_8\gamma^2 - \beta^2 + 2z_7\gamma^3 + 2\alpha\gamma\beta - F\gamma^2)x \\ & -E\gamma^2 + 2z_6\gamma^3 - \alpha\beta + z_7\gamma^2 + \alpha^2\gamma, \\ \\ 0 = & (-D\gamma^2 + \delta\gamma^2\beta + z_5\gamma^3)x^2 + (\delta\gamma^2\alpha + 2z_5\gamma^2 + \beta^2 + \beta\gamma\delta + z_4\gamma^3 - C\gamma^2 - \alpha\gamma\beta)x \\ & +\beta\alpha + z_4\gamma^2 - \beta\delta - \alpha^2\gamma + z_3\gamma^3 + \alpha\gamma\delta - B\gamma^2, \\ \\ 0 = & (H\gamma^2 - \delta^2\gamma^2)x + A\gamma^2 - \beta\delta + \alpha\gamma\delta - z_2\gamma^2. \end{cases}$$

Finally, by Polynomial Identity Principle, we obtain the linear system:

$$\begin{cases} 0 = & \beta^2\gamma + 2z_8\gamma^3 - G\gamma^2, \\ \\ 0 = & 2z_8\gamma^2 - \beta^2 + 2z_7\gamma^3 + 2\alpha\gamma\beta - F\gamma^2, \\ \\ 0 = & -E\gamma^2 + 2z_6\gamma^3 - \alpha\beta + z_7\gamma^2 + \alpha^2\gamma, \\ \\ 0 = & -D\gamma^2 + \delta\gamma^2\beta + z_5\gamma^3, \\ \\ 0 = & \delta\gamma^2\alpha + 2z_5\gamma^2 + \beta^2 + \beta\gamma\delta + z_4\gamma^3 - C\gamma^2 - \alpha\gamma\beta, \\ \\ 0 = & \beta\alpha + z_4\gamma^2 - \beta\delta - \alpha^2\gamma + z_3\gamma^3 + \alpha\gamma\delta - B\gamma^2, \\ \\ 0 = & H\gamma^2 - \delta^2\gamma^2, \\ \\ 0 = & A\gamma^2 - \beta\delta + \alpha\gamma\delta - z_2\gamma^2. \end{cases}$$

The solution is given by

$$\begin{cases} A = & z_2 - \frac{\alpha\delta}{\gamma} + \frac{\beta\delta}{\gamma^2}, \\ B = & z_3\gamma + z_4 + \frac{-\alpha^2+\alpha\delta}{\gamma} + \frac{-\beta\delta+\alpha\beta}{\gamma^2}, \\ C = & z_4\gamma + 2z_5 + \alpha\delta + \frac{\beta\delta-\alpha\beta}{\gamma} + \frac{\beta^2}{\gamma^2}, \\ D = & z_5\gamma + \beta\delta, \\ E = & 2z_6\gamma + z_7 + \frac{\alpha^2}{\gamma} - \frac{\beta\alpha}{\gamma^2}, \\ F = & 2z_7\gamma + 2z_8 + 2\frac{\alpha\beta}{\gamma} - \frac{\beta^2}{\gamma^2}, \\ G = & 2z_8\gamma + \frac{\beta^2}{\gamma}, \\ H = & \delta^2. \end{cases}$$

□

Chapter 8

Finite dimensional realizations

The first paper on this subject is [9], soon extended by [11] and [10]. In particular the paper [9] has been of great interest for our aim, since it deals with deterministic volatilities: from this work we took inspiration to investigate in the direction of finite dimensional realization.
This area is now deeply investigated from a theoretical point of view (see [31] and [32]) but (up to now) there is not empirical evidence of the usability of the appealing and sophisticated framework in "real" interest rates management. The only work known to us oriented to applications is [2].

8.1 Background

In this section we give the necessary (and sufficient) background on infinite dimensional differential geometry. In the context of interest rate theory below, we will only be working in a Hilbert space setting, but the result are just as easy to prove in the more general framework of Banach spaces.
Let us thus consider a real Banach space X. By an n-dimensional distribution, we mean a mapping F, which to each x in an open subset V of X associates an n-dimensional subspace $F(x) \subseteq X$. A mapping (vector field) $f : U \to X$, where U is an open subset of X, is said to lie in F (on U), if $U \subseteq V$ and $f(x) \in F(x)$ for every $x \in U$. A collection $f_1, \dots, f_n$ in of vector fields lying in F on U generate (or span) F on U if $span\{f_1(x), \dots, f_n(x)\} = F(x)$ for every x in U, where *span* denotes the linear hull aver the real field. The distribution is smooth if, to every x in V, there exist an open set U such that $x \in U \cap V$, and smooth vector fields $f_1, \dots, f_n$ spanning F on U. If F and G are distributions and $G(x) \subseteq F(x)$ for all x we say that F contains G, and we write $G \subseteq F$. The dimension of a distribution F is defined pointwise as $dimF(x)$.

Observation 8.1 *The prefix smooth above is interpreted as C^∞. Note, however, that* *in many situations below we only need C^k for some k which depends upon the particular context.*

Let f and g be smooth vector fields on U. Their Lie bracket is the vector field

$$[f,g](x) = f'(x)g(x) - g'(x)f(x),$$

where $f'(x)$ denotes the Frechet derivative of f at x, and similarly for g. For clarity we will sometimes write $f'(x)[g(x)]$ instead of $f'(x)g(x)$. A distribution F is called involutive if, for every smooth f and g lying in F on U, their Lie bracket also lies in F, i.e.

$$[f,g](x) \in F(x), \forall x \in U.$$

Let the distribution F be as above, and let $\phi : V \to W$ be a diffeomorphism between the open subsets V and W of X. Then we can define a new distribution $\phi_* F$ on W by

$$(\phi_* F)(\phi(x)) = \phi'(x)F(x).$$

Similarly, for any smooth vector field $f \in C^\infty(U,X)$, we define the field $\phi_* f$ by

$$\phi_* f = (\phi' \circ \phi^{-1})(f \circ \phi^{-l}).$$

By a straightforward calculation one verifies easily that

$$\phi_*[f,g] = [\phi_* f, \phi_* g].$$

From this we conclude that if F is generated by $f_1, \ldots, f_n$, then $\phi_* F$ is generated by $\phi_* f_1, \ldots, \phi_* f_n$, and we see that F is involutive if and only if $\phi_* F$ is involutive.
The main result of this section is the following infinite dimensional extension of the standard Frobenius Theorem. The particular formulation of the theorem and the proof can be found in [11].

Theorem 8.2 *Let F be a smooth distribution on the open set V in the Banach space X. Let furthermore x be an arbitrary point in V. Then there exists a diffeomorphism $\phi : U \to X$ on some neighborhood $U \subseteq V$ of x, such that $\phi_* F$ is constant on $\phi(U)$ if and only if F is involutive.*

We will use the Frobenius theorem in order to prove existence of so called tangential manifolds.

Definition 8.3 *Let F be a smooth distribution, and let x_0 be a fixed point in X. A submanifold $\mathcal{G} \subseteq X$ with $x_0 \in \mathcal{G}$ is called a tangential manifold through x for F, if $F(x) \subseteq T\mathcal{G}(x)$ for each x in a neighborhood of x_0 in $\mathcal{G}$ (where $T\mathcal{G}(x)$ denotes the tangent space to $\mathcal{G}$ at x).*

Observation 8.4 *We note that the definition of a tangential manifold is similar to the definition of an integral manifold, but that for an integral manifold we have the inclusion* $T\mathcal{G}(x) \subseteq F(x)$. *Just as one typically is looking for maximal integral manifolds, we will be looking for minimal tangential manifolds.*

We now have the following result which we will use below.

Theorem 8.5 *Let* F *be an* n*-dimensional distribution, and let* x_0 *be a fixed point in* X. *There exists an* n*-dimensional tangential manifold through* x *for all* x *in a neighborhood of* x_0 *if and only if* F *is involutive.*

Suppose now that we are given an n-dimensional involutive distribution F and a point $x_0 \in X$. From the result above we know that there exists a tangential manifold for F passing through x_0, and a natural question is if it possible to parameterize this manifold in some constructive way. This is in fact possible but we need some new notation.

Definition 8.6 *Let* f *be a smooth vector field on* X, *and let* x *be a fixed point in* X. *Consider the ODE*

$$\begin{cases} \frac{dx_t}{dt} & = & f(x_t), \\ \\ x_0 & = x. \end{cases}$$

We denote the solution x_t *as* $x_t = e^{ft}x$.

We have thus defined a group of operators $\{e^{ft} | t \in \mathbb{R}\}$, and we note that the set $\{e^{ft}x | t \in \mathbb{R}\} \subseteq X$ is nothing else than the integral curve of the vector field f, with initial point x, i.e. the curve obtained by starting at x and then gluing together infinitesimally small pieces of the vector field f.

Proposition 8.7 *Take as given an* n*-dimensional involutive distribution spanned by* $f_1, \ldots, f_n$, *and a point* $x_0 \in X$. *Denote the tangential manifold through* x_0 *by* $\mathcal{G}$. *Define the mapping* $G : \mathbb{R}^n \to X$ *by*

$$G(z_1, \ldots, z_n) = e^{f_n z_n} \ldots e^{f_1 z_1} x_0.$$

Then G *is a local parametrization of* $\mathcal{G}$ *in the sense that there exist open neighborhoods* $U \subseteq \mathbb{R}^n$ *and* $V \subset \mathcal{G}$, *of* 0 *and* x_0 *respectively, such that* $V = G(U)$.
Furthermore, the inverse of G *restricted to* V *is a local coordinate system for* $\mathcal{G}$ *at* x_0.

8.2 Realizations and invariant manifolds

8.2.1 Homogenous systems

We begin by considering here only time invariant systems. For the time varying case, see the next paragraph.

Take as given an m-dimensional standard Wiener process W with components $W^l, \ldots, W^m$, and a Hilbert space H, where a generic point will be denoted by r [1].
Let furthermore $\mu, \sigma_1, \ldots, \sigma_m$ be smooth vector fields on H. For a given initial point $r^o \in H$ we can then consider the following Stratonovich SDE on H:

$$\left\{ \begin{array}{rcl} dr_t & = & \mu(r_t)dt + \sigma(r_t) \circ dW_t, \\ \\ r_0 & = & r^0, \end{array} \right. \tag{8.1}$$

where

$$\sigma(r_t) \circ dW_t = \sum_{i=1}^{m} \sigma_i(rt) \circ dW_t^i$$

and where $\circ$ denotes the Stratonovich integral. We note that, because of the assumed smoothness, the SDE (8.1) will locally (up to a positive stopping time), always have a unique strong solution. Let us emphasize that in the sequel we are dealing exclusively with local strong (as opposed to weak or mild) solutions to all SDEs. For information on SDEs in Hilbert space see [20], which is the standard text on the subject.
The process r above is inherently an infinite dimensional process, but we will investigate under what conditions it can be realized by means of a finite dimensional SDE: a more detailed conceptual discussion can be found in [8].
We now adapt some material from chapter 7.

Definition 8.8 *We say that the SDE (8.1) has a (local) d-dimensional realization if there exists a point $z_0 \in \mathbb{R}^d$, smooth vector fields $a, b_1, \ldots, b_m$ on some open subset $\mathcal{Z}$ of $\mathbb{R}^d$ and a smooth (submanifold) map $G : \mathcal{Z} \to H$, such that r has the local representation*

$$r_t = G(Z_t),$$

where Z is the solution of the d-dimensional Stratonovich SDE

$$\left\{ \begin{array}{rcl} dZ_t & = & a(Z_t)dt + b(Z_t) \circ dW_t, \\ \\ Z_0 & = & z_0, \end{array} \right.$$

The prefix "local" above means that the representation is assumed to hold for all t with $O \leq t < \tau(r^0)$, $\mathbb{P}$-a.s. where, for each $r^0 \in H$, $\tau(r^0)$ is a a strictly positive stopping time.

We want to give conditions for the existence of a finite realization in terms of the local characteristics μ, σ and in this context a local realization is the best one can hope for.
We will thus often suppress the prefix "local", so in the sequel the word realization should

[1] We choose r to denote a point (and above forward rates), to avoid misunderstandings, because of the presence of f as vector fields.

always be interpreted as local realization. The realization concept is closely connected to the concept of an invariant sub-manifold.

Definition 8.9 *A submanifold $\mathcal{G}$ in H is said to be (locally) invariant under the action of the SDE (8.1), if for every choice of $r^0 \in \mathcal{G}$ we have $r_t \in \mathcal{G}$ for $0 \leq t < \tau(r^0)$, where τ is a strictly positive stopping time.*

The first step towards a solution of the realization problem lies in the following result.

Proposition 8.10 *There exists a local d-dimensional realization to (8.1) if and only if there exists an invariant submanifold $\mathcal{G}$ with $r^0 \in \mathcal{G}$.*

The problem of finding a realization is thus reduced to the problem of finding a finite dimensional invariant submanifold.
We can now connect this problem to the Frobenius theory exposed before.

Theorem 8.11 *A submanifold $\mathcal{G}$ is invariant under the action of the SDE (8.1) if and only if, for every point $r \in \mathcal{G}$, the vectors $\mu(r), \sigma_1(r), \ldots, \sigma_m(r)$ belong to the tangent space of $\mathcal{G}$ at r. Thus $\mathcal{G}$ is invariant if and only if it is an tangential manifold for the distribution generated by $\mu(r), \sigma_1(r), \ldots, \sigma_m(r)$.*

Putting these results together we immediately have the following.

Proposition 8.12 *The SDE (8.1) has a finite dimensional realization if and only if there exists a finite dimensional tangential manifold for $\mu(r), \sigma_1(r), \ldots, \sigma_m(r)$ containing the initial point r^0.*
The dimension of a minimal (w.r.t. the dimension d) realization coincides with the dimension of the minimal tangential manifold.

Before going to our main result we need a new concept.

Definition 8.13 *Let F be a smooth distribution on H. The Lie algebra generated by F, denoted by $\{F\}_{LA}$, is defined as the minimal (under inclusion) involutive distribution containing F.*

If, for example, the distribution F is spanned by the vector fields $f_1, \ldots, f_n$ then, to construct the Lie algebra $\{f_1, \ldots, f_n\}_{LA}$, you simply form all possible brackets, and brackets of brackets, etc. of the fields $f_1, \ldots, f_n$ and adjoin these to the original distribution until the dimension of the distribution is no longer increased.
We can now formulate the main result on the existence of finite dimensional realizations.

Theorem 8.14 *Take as given the vector fields μ, σ and a point $\hat{r} \in H$. The following statements are equivalent:*

- *For each choice of initial point r^0 near $\hat{r} \in H$, there exists a d-dimensional realization of the infinite dimensional SDE (8.1).*
- *The Lie algebra $\{\mu, \sigma_1, \ldots, \sigma_m\}_{LA}$ has dimension d near $\hat{r}$.*

Note that, as for as finite realizations are concerned, we are only considering pure existence results. We do not now treat the problem of how to construct a concrete realization. "In principle" this is simple: it is sufficient to fix a coordinate system on the invariant manifold and to write down the coordinate dynamics of the r-process, but this is easier said than done. Note also that there is no such thing as a unique realization, since any diffeomorphic transformation of the state space will give a new equivalent realization. The problem of finding canonical realizations is developed in [10].

8.2.2 Time varying systems

We now extend the analysis to cover time varying systems of the form

$$(8.2) \qquad \begin{cases} dr_t &= \mu(r_t, t)dt + \sigma(r_t, t) \circ dW_t, \\ r_0 &= r^0, \end{cases}$$

Note that we no longer initiate the system at $t = 0$ but at an arbitrary point s in time. The relevant definitions of realizations and invariant manifolds are the natural ones with the obvious extension of the prefix "local".

Definition 8.15 *We say that the SDE (8.2) has a (local) d-dimensional realization at (s, r^0), if there exists a point $z^s \in \mathbb{R}^d$, smooth vector fields $a, b_1, \ldots, b_m$ on some open subset $\mathcal{Z}$ of $\mathbb{R}^d$ and a smooth (submanifold) map $G : \mathcal{Z} \to H$, such that r has the local representation*

$$r_t = G(Z_t) \text{ for } t \geq s,$$

where Z is the solution of the d-dimensional Stratonovich SDE

$$\begin{cases} dZ_t &= a(Z_t)dt + b(Z_t) \circ dW_t, \\ Z_s &= z^s, \end{cases}$$

Definition 8.16 *A submanifold $\mathcal{G}$ in H is said to be (locally) invariant under the action of the SDE (8.2), if for every choice of $(s, r^0) \in \mathbb{R}_+ \times \mathcal{G}$ we have $r_t \in \mathcal{G}$ locally for $s \leq t$.*

As before we are investigating the existence of finite dimensional realizations and invariant manifolds. We handle this new situation by enlarging the state space and introducing running time as a new state. Defining, far each fixed s, the Wiener process W^s as $W_t^s = W_t - W_s$, we can write (8.2) as

$$\begin{cases} dr_u &= \mu(r_u, t)du + \sigma(r_u, t) \circ dW_u^s, \\ dt &= 1 \cdot du + 0 \cdot dW_u^s, \\ r_0 &= r^0, \\ t_0 &= s. \end{cases}$$

We are led to the following natural notation.

Definition 8.17 *Define the following extended objects.*

$$\hat{H} = H \times \mathbb{R},$$

$$\hat{r} = \begin{bmatrix} r \\ t \end{bmatrix},$$

$$\hat{\mu}(\hat{r}) = \begin{bmatrix} \mu(r, t) \\ 1 \end{bmatrix},$$

$$\hat{\sigma}(\hat{r}) = \begin{bmatrix} \sigma(r, t) \\ 0 \end{bmatrix},$$

We can now write (8.2) as the following SDE on $\hat{H}$ where, in order to save space, we suppress the superindex in W^s.

$$\begin{cases} d\hat{r}_u &= \hat{\mu}(\hat{r}_u)du + \hat{\sigma}(\hat{r}_u) \circ dW_u, \\ \hat{r}_0 &= \hat{r}^0. \end{cases} \tag{8.3}$$

Here we have used the notation

$$\hat{r}^0 = \begin{bmatrix} r^0 \\ s \end{bmatrix}.$$

It is now easy to see that a manifold $\mathcal{G}$ is invariant in H for the time varying system (8.2) if and only if $\mathcal{G} \times \mathbb{R}$ is invariant for the homogenous system (8.3).
Furthermore it is clear that the problem of finding a realization for (8.2) is equivalent to that of finding a realization for (8.3); so we have the main result, which follows immediately from the previous theory.

Theorem 8.18 *The time varying system (8.2) has a finite dimensional realization if and only if*

$$dim\{\hat{\mu}, \hat{\sigma}\}_{LA} < +\infty.$$

8.3 Back to interest rates

We are now looking for a suitable Hilbert space to set our problem. We consider the forward rates dynamics in the Musiela parametrization (see chapter 2):

$$dr(t,x) = \Big(\frac{\partial}{\partial x} r(t,x) + \sigma(t,x) \int_0^x \sigma(t,u)^T du\Big) dt + \sigma(t,x) dW_t. \tag{8.4}$$

The problem with equation (8.4) is the operator $\frac{\partial}{\partial x}$ in the drift term. In most "naturally chosen" spaces, like a weighted Sobolev space, the operator is unbounded, which means that we have no existence results concerning strong solutions. Thus, we need a very regular space to work in.

Definition 8.19 *Consider fixed real numbers $\beta > 1$ and $\gamma > 0$. The space $H_{\beta,\gamma}$ is defined as the space of all infinitely differentiable functions*

$$r : \mathbb{R}_+ \to \mathbb{R}$$

satisfying the norm condition $\|r\|_{\beta,\gamma} < +\infty$. Here the norm is defined as

$$\|r\|^2_{\beta,\gamma} = \sum_{n=0}^{+\infty} \beta^{-n} \int_0^{+\infty} \Big(\frac{d^n r}{dx^n}(x)\Big)^2 e^{-\gamma x} dx.$$

Note that $H_{\beta,\gamma}$ is not a space of distributions, but a space of functions. The reason for the exponential weighting is to include all constant functions in the space. Most of the results below are uniform w.r.t. (β, γ), so in the sequel we will often suppress the subindices. With the obvious inner product H is a pre-Hilbert space, and we have the following result.

Proposition 8.20 *The space H is a Hilbert space (i.e. complete), and if $f_n \to f$ in H then $f_n^{(m)} \to f^{(m)}$ uniformly on compacts for every $m \geq 0$, where $f^{(m)} = \frac{d^m f}{dx^m}$. Thus in particular, for any fixed $x \in \mathbb{R}_+$, the point evaluation mapping $r \mapsto r(x)$ is a bounded linear functional. Furthermore, every function in H is in fact real analytic, and can thus be uniquely extended to a holomorphic function in the entire complex plane.*

We take as given a volatility σ of the form

$$\sigma : H \times \mathbb{R}_+ \to \mathbb{R}^m,$$

i.e. each component of $\sigma(r,x) = [\sigma_1(r,x), \ldots, \sigma_m(r,x)]$ is a functional of the infinite dimensional r-variable and of the real variable x.

We will also consider σ_i as a mapping from H to a space of functions, and we will in fact assume that each component σ_i is a smooth vector field on H.

Observation 8.21 *Note that we are primarily interested in homogenous systems, i.e. we consider forward rate volatilities without an explicit time-dependence. This is partly due to the fact that it (to us) seems "not so meaningful" to assume time-varying volatilities. In our experience, the introduction of a time dependent volatility seems to add very little of interest, and time varying models are easily handled by the methods in the previous section. Our opinion is shared by the authors of [11].*

We need to introduce some more compact notation.

Definition 8.22 *In the sequel, we will denote integration w.r.t. time to maturity by* $\mathbf{H}$. *Thus, we will write*

$$\mathbf{H}\sigma(r,x) = \int_0^x \sigma(r,s)ds.$$

With the same operator notation we will write

$$\mathbf{F} = \frac{\partial}{\partial x}$$

Suppressing the x-variable, the Itô's dynamics of the forward rates (8.4) under the martingale measure $\mathcal{Q}$ can thus be written as

$$dr_t = \left(\mathbf{F}r_t + \sigma(r_t)\mathbf{H}\sigma(r_t)^T\right)dt + \sigma(r_t)dW_t.$$

In chapter 7) we proved that the forward rate model on Stratonovich form is given by

$$dr_t = \mu(r_t)dt + \sigma(r_t) \circ dW_t,$$

where

$$\mu(r) = \mathbf{F}r + \sigma(r)\mathbf{H}\sigma^T - \frac{1}{2}\sigma_r'(r)(\sigma(r)),$$

where as usual $\sigma_r'(r)(\sigma(r))$ denotes the Frechet derivative $\sigma_r'(r)$ operating on $\sigma(r)$.
We need some regularity assumptions.

Assumption 8.23 *From now on we assume that a has the following properties:*

- *The mappings* $\sigma_1, \ldots, \sigma_m$ *are smooth vector fields on* H.
- *The mapping* $r \mapsto \sigma(r)\mathbf{H}\sigma(r)^T$ *is a smooth vector field on* H.

As justified before, the reason for our choice of H as the underlying space is that the linear operator $\mathbf{F} = \frac{\partial}{\partial x}$ is bounded in this space. Together with the assumptions above, this implies that both μ and σ are smooth vector fields on H, thus ensuring the local existence of a strong solution to the forward rate equation for every initial point $r^0 \in H$.

8.4 Main Results

We now go on to apply the general results from previous sections to the class of interest rate models in the form (8.4).

Theorem 8.24 *Take as given the volatility mapping* $\sigma = (\sigma_1, \ldots, \sigma_m)$ *as well as an initial forward rate curve* $r^0 \in H$. *Then the forward rate model generated by* σ *generically admits a finite dimensional realization at* r^0, *if and only if*

$$dim\{\mu, \sigma_1, \ldots, \sigma_m\}_{LA} < +\infty,$$

in a neighborhood of r^0, *where* μ *is given by*

$$\mu(r) = \mathbf{F}r + \sigma(r)\mathbf{H}\sigma^T - \frac{1}{2}\sigma_r'(r)(\sigma(r)), \tag{8.5}$$

and $\mathbf{H}\sigma$ *is defined by*

$$\mathbf{H}\sigma(r,x) = \int_0^x \sigma(r,s)ds.$$

If the assumptions of the theorem are in force, we can furthermore use the previous propositions to give a description of the minimal invariant (tangential) manifold, containing r^0, which is generated by the interest rate dynamics. In other words, we can describe the structure of all possible forward curves that can be produced by the model.

Theorem 8.25 *Assume that the Lie algebra* $\{\mu, \sigma\}_{LA}$ *is spanned by the smooth vector fields* $f_1, \ldots, f_d$. *Then, for the initial point* r^0, *all forward rate curves produced by the model will belong to the induced tangential manifold* $\mathcal{G}$, *which can be parameterized as* $\mathcal{G} = Im(G)$, *where*

$$G(z_1, \ldots, z_d) = e^{f_d s_d} \ldots e^{f_1 s_1} r^0,$$

and where the operator $e^{f_d s_d}$ *is defined as before.*

With this machinery we can also very easily give the solution found in [11]to a related problem, which was left open in [8].
Consider a fixed interest rate model, specified by the volatility σ, and also a fixed family of forward rate curves parameterized by the mapping $G_0 : \mathbb{R}^k \to H$. Now, if $\mathcal{G}_0 = Im(Go)$ is invariant, then the interest rate model will, given any initial point r^0 in $\mathcal{G}_0$, only produce forward rate curves belonging to $\mathcal{G}_0$, in which case we say that the model and the family $\mathcal{G}_0$ are consistent, as defined in chapter 7.
If the family is not consistent, then an initial forward rate curve in $\mathcal{G}_0$ may produce future forward rate curves outside $\mathcal{G}_0$, and this question arises: how to construct the smallest possible family of forward rate curves which contains the initial family $\mathcal{G}_0$, and is consistent (i.e. invariant) w.r.t the interest rate model?

As a concrete example, one may want to find the minimal extension of the Nelson-Siegel family of forward rate curves which is consistent with the Hull-White model. An explicit solution to this problem is given in chapter 7. In particular one would like to know under what conditions this minimal extension of $\mathcal{G}_0$ is finite dimensional.
In geometrical terms we thus want to construct the minimal manifold containing $\mathcal{G}_0$, which is tangential w.r.t. the vector fields $\mu, \sigma_1, \dots, \sigma_m$. The solution is obvious: for very point on $\mathcal{G}_0$ we construct the minimal tangential manifold through that point, and then we define the extension $\mathcal{G}$ as the union of all these fibers. Thus we have the following result.

Proposition 8.26 *Consider a fixed volatility mapping σ, and lei $\mathcal{G}_0$ be a k-dimensional submanifold parameterized by $G_0 : \mathbb{R}^k \to H$. Then $\mathcal{G}_0$ can be extended to a finite dimensional invariant submanifold $\mathcal{G}$, if and only if*

$$dim\{\mu, \sigma_1, \dots, \sigma_m\}_{LA} < +\infty.$$

Moreover, if $\mathcal{G}_0$ is transversal to $\{\mu, \sigma\}_{LA}$ and if the Lie algebra is spanned by the independent vector fields $f_1, \dots, f_d$, then $dim\mathcal{G} = k + d$ and a parametrization of $\mathcal{G}$ is given by the map $G : \mathbb{R}^{k+d} \to H$, defined by

$$G(z_l, \dots, z_k, y_1, \dots, y_d) = e^{f_d y_d} \dots e^{f_1 y_1} G_0(z_1, \dots, z_k).$$

Observation 8.27 *The term "transversal" above means that no vector in the Lie algebra $\{\mu, \sigma\}_{LA}$ is contained in the tangent space of $\mathcal{G}_0$ at any point of $\mathcal{G}_0$. This prohibits an integral curve of the algebra to be contained in $\mathcal{G}_0$, which otherwise would lead to an extension with lower dimension than $d + k$. In such a case the parametrization above would amount to an over parametrization in the sense that G would not be injective.*

We end this section by remarking that it is always possible to choose the state variables as a set of benchmark forward rates. From a pragmatical point of view, this assertion is a crucial point in order to obtain applications to interest rates market, as in chapter 9.

Proposition 8.28 *Suppose that $dim\{\mu, \sigma_1, \dots, \sigma_m\}_{LA} = d$. Then, for almost every choice of distinct benchmark maturities $x_1, \dots, x_d$, the realization can be chosen such that the state process Z_t is given by $Z_t = (r_t(x_1), \dots, r_t(x_d))$. The expression "almost for every choice" means that, apart from a discrete set of forbidden values, $x_1, \dots, x_d$ can be chosen freely.*

8.5 Deterministic volatility

We now turn to present some applications of the theory developed above. We start with the simplest case, which is when the volatility $\sigma(r, x)$ does not depend on r, i.e. when we can write $\sigma(r, x) = \sigma(x)$. This case was first studied and more or less completely analyzed in [9]. For simplicity we a assume (for the moment) that we have only one driving Wiener process, so the volatility is a constant vector field in H.

8.5.1 Finite dimensional realizations

In order to use the main theorems we now compute the Lie algebra $\{\mu, \sigma\}_{LA}$. Since the vector field σ is constant we have $\sigma' = 0$ so, from (8.5) we have

$$\mu(r) = \mathbf{F}r + D,$$

where

$$D(x) = \sigma(x) \int_0^x \sigma(u) du.$$

The Frechet derivatives are trivial in this case: since $\mathbf{F}$ is linear (and bounded in our space), we obtain:

$$\mu' = \mathbf{F},$$

$$\sigma' = 0.$$

Thus the Lie bracket $[\mu, \sigma]$ is given by

$$[\mu, \sigma] = \mathbf{F}\sigma,$$

and in the same way we have

$$[\mu, [\mu, \sigma]] = \mathbf{F}^2\sigma.$$

Continuing in the same way it is easily seen that the relevant Lie algebra is given by

$$\{\mu, \sigma\}_{LA} = span\{\mu, \sigma, \mathbf{F}\sigma, \mathbf{F}^2\sigma, \ldots\}.$$

It is thus clear that $\{\mu, \sigma\}_{LA}$ is finite dimensional (at each point r) if and only if function space

$$span\{\mathbf{F}^n\sigma | n = 0, 1, 2, \ldots\}$$

is finite dimensional. This argument easily carries over to the case of several driving Wiener processes, and we have the following result (originally proved in [9]).

Proposition 8.29 *Assume that the volatility components $\sigma_1, \ldots, \sigma_m$ are deterministic. Then the model possesses a finite dimensional realization if and only if the function space*

$$span\{\mathbf{F}^k\sigma_i | i = 1, \ldots, m; k = 0, 1, \ldots\}$$

is finite dimensional. If the dimension of the function space above equals d, then the dimension of a minimal realization is $d + 1$.

We thus have a finite dimensional realization if and only if the components of σ solve, as functions of x, a multidimensional linear ODE with constant coefficients. Using standard results from ODE theory, a more concrete characterization is given as follows.

Proposition 8.30 *Given the assumptions above there exists a finite dimensional realization if and only if the volatility* $\sigma(x) = (\sigma_1(x), \ldots, \sigma(x))$ *can be written*

$$\sigma(x) = ce^{Ax}B,$$

where c is a row vector, whereas A *and* B *are matrices. In particular, if every component of* σ *is of the form*

$$\sigma_i(x) = p_i(x)e^{-\lambda_i x},$$

for some polynomial p_i *and some real positive number* λ_i*, then there exists a finite dimensional realization.*

8.5.2 Invariant manifolds

We now turn to the construction of invariant manifolds, and to this end we assume that the Lie algebra above is finite dimensional. Thus it is spanned by a finite number of vector fields as

$$\{\mu, \sigma\}_{LA} = span\{\mu, \sigma_i^{(k)} | i = 1, \ldots, m; k = 0, 1, \ldots\},$$

where

$$\sigma_i^{(k)}(x) = \frac{\partial^k \sigma_i}{\partial x^k}(x).$$

In order to apply the previous machinery, we have to compute the operators $exp\Big(\mu t\Big)$ and $exp\Big(\sigma_i^{(k)}\Big)$, i.e. we have to solve H-valued ODEs. We recall that

$$\mu(r) = \mathbf{F}r + D,$$

where the constant field D is given

$$D(x) = \sum_{i=1}^{m} \sigma_i(x) \int_0^x \sigma_i(s)ds$$

which can be written as

$$D(x) = \frac{1}{2}\frac{\partial}{\partial x}\|S(x)\|^2,$$

where, as before, $S(x) = \int_0^x \sigma(s)ds$. Thus $e^{\mu t}$ is obtained by solving

$$\frac{dr}{dt} = \mathbf{F}r + D.$$

This is a linear equation with solution

$$r_t = e^{\mathbf{F}t} r_0 + \int_0^t e^{\mathbf{F}(t-s)} D ds$$

so

$$(e^{\mu t} r_0)(x) = r_0(x+t) + \frac{1}{2}(\|S(x+t)\|^2 - \|S(x)\|^2).$$

The vector fields $\sigma_i^{(k)}$ are constant, so the corresponding ODEs are trivial. We have

$$e^{\sigma_i^{(k)} t} r_0 = r_0 + \sigma_i^{(k)} t.$$

We thus have the following results on the parametrization of invariant manifolds. For a given mapping $G : \mathbb{R}^n \to H$, we write $G(z)(x)$ or $G(z, x)$ to denote the function $G(z) \in H$ evaluated at $x \in |R_+$.

Proposition 8.31 *The invariant manifold generated by the initial forward rate curve r_0 is parameterized as*

$$G(z_0, z_i^k | i = 1, \ldots, m; k = 0, \ldots, n_i)(x)$$
$$= r_0(x + z_0) + \frac{1}{2}(\|S(x+z_0)\|^2 - \|S(x)\|^2) + \sum_{i=1}^m \sum_{k=0}^{n_i} \sigma_i^{(k)}(x) z_i^k.$$

If the k-dimensional manifold $\mathcal{G}_0$ is transversal to $\{\mu, \sigma\}_{LA}$ and parameterized by $G_0(y_l, \ldots, y_k)$, then the minimal consistent (i.e. invariant) extension is parameterized as

$$G(y_1, \ldots, y_k, z_0, z_i^k | i = 1, \ldots, m; k = 0, \ldots, n_i)(x)$$

$$= G_0(y_1, \ldots, y_k)(x + z_0) + \frac{1}{2}(\|S(x+z_0)\|^2 - \|S(x)\|^2) + \sum_{i=1}^m \sum_{k=0}^{n_i} \sigma_i^{(k)}(x) z_i^k.$$

Note that if $\mathcal{G}_0$ is invariant under shift in the x-variable (this is in fact the typical case), then a simpler parametrization of $\mathcal{G}$ is given by

$$G(y_1, \ldots, y_k, z_0, z_i^k | i = 1, \ldots, m; k = 0, \ldots, n_i)(x)$$

$$= G_0(y_1, \ldots, y_k)(x) + \frac{1}{2}(\|S(x+z_0)\|^2 - \|S(x)\|^2) + \sum_{i=1}^m \sum_{k=0}^{n_i} \sigma_i^{(k)}(x) z_i^k.$$

8.6 Short rate realizations

We close this chapter with an appealing probabilistic properties of our model.
One of the classical problems concerning the HJM approach to interest rate modelling is that of determining when a given forward rate model is realized by a short rate model, i.e. when the short rate induced by the forward rate model is a one-dimensional Markov

process. There are several results in this area providing partial answers (see for example [17]).
As usual, we restrict ourselves to the case of time homogenous forward rate volatilities: using previous results, we immediately have the following general necessary condition.

Proposition 8.32 *The forward rate model generated by $\sigma(r, x)$ is a generic short rate model, i.e. the short rate is generically a Markov process, only if*

$$dim\{\mu, \sigma\}_{LA} \leq 2. \tag{8.6}$$

Observation 8.33 *The natural case is of course $dim\{\mu, \sigma\}_{LA} = 2$, since we are basically looking far a short rate realization where the coefficients are time dependent (in order to make it possible to fit any initial forward rate curve). It seems presently an open question whether there exists a nontrivial generic short rate model with dim $dim\{\mu, \sigma\}_{LA} = 1$.*

Let us start by observing that, in this context, the case of a scalar Wiener process is the only non pathological case.

Lemma 8.34 *Assume that there are m driving Wiener processes, and that the volatilities are of the general form $\sigma_1(r, x), \ldots, \sigma_m(r, x)$. Then, a necessary condition far the Lie algebra $\{\mu, \sigma\}_{LA}$ to be two dimensional is that there exists a vector field $\sigma_0(r, x)$ and scalar fields $\phi(r), i = 1, \ldots, m$ such that*

$$\sigma_i(r, x) = \phi_i(r)\sigma_0(r, x)$$

To avoid degenerate cases like this we make a standing assumption.

Assumption 8.35 *We assume that we have only one scalar driving Wiener process, i.e. that $m = 1$.*

Note that condition (8.6) is only a sufficient condition for the existence of a short rate realization. It guarantees that there exists a two-dimensional realization but the question remains whether the realization can chosen in such a way that the short rate and running time are the state variables. This question is completely solved by the following result.

Theorem 8.36 *Assume that the model is not deterministic, and take as given a time invariant volatility $\sigma(r, x)$. Then there exists a short rate realization if and only if the vector fields $[\mu, \sigma]$ and σ are parallel, i.e. if and only if there exits a scalar field $a(r)$ such that the following relation holds (locally) for all r.*

$$[\mu, \sigma](r) = a(r)\sigma(r). \tag{8.7}$$

This can also be written

$$[\mathbf{F}r + D, \sigma](r) - \frac{1}{2}\sigma''(r)[\sigma(r), \sigma(r)] = a(r)\sigma(r).$$

where

$$D(r,x) = \sigma(r,x)\int_0^x \sigma(r,s)ds.$$

If the volatility is of the form $\sigma(r,t,x)$ then the condition (8.7) is replaced by

$$\mu_r(r,t)[\sigma(r,t)] - \sigma_r(r,t)[\mu(r,t)] - \sigma_t(r,t) = a(r,t)\sigma(r,t),$$

where subindex denotes the Frechet derivative w.r.t the indicated variable.

We now turn to applications of the theorem: we will study now the simplest non-trivial example, which is the case studied in [17]. We have m driving Wiener processes, and in our notation the volatilities are assumed to be deterministic, i.e. (with a slight misuse of notation) they have the form $\sigma_i(t,r,x) = \sigma_i(t,x)$. Adapting the last lemma to our present setting, we see that there must exist functions $g(t)$ and $\sigma_0(t,x)$ such that

$$\sigma_i(t,x) = g_i(t)\sigma_0(t,x), i = 1,\ldots,m.$$

This however, implies that we can replace the m driving Wiener processes with just one Wiener process so, as was expected, we may as well assume that $m = 1$. Denoting σ_1 by σ, we obtain

$$\mu(r,t,x) = \frac{\partial}{\partial x}r(x) + \sigma(t,x)\int_0^x \sigma(t,s)ds.$$

and we see that $\mu_r = \mathbf{F}r$ and $\sigma_r = 0$. Since we have a time varying system, we use the second part of the theorem, which says that there exists a short rate realization if and only if there exists a scalar field $a(r,t)$, such that for all r, t, and x we have

$$\sigma_x(t,x) - \sigma_t(t,x) = a(r,t)\sigma(t,x).$$

From this relation it is clear that $a(r,t) = a(t)$. Dividing by σ and defining g by $g(t,x) = ln\sigma(t,x)$ we have the equation

$$g_x(t,x) - g_t(t,x) = a(t).$$

Taking x-derivatives, and defining h by $h = g_x$ we get

$$h_x(t,x) - h_t(t,x) = 0.$$

This is the simplest possible example of the wave equation, with solution $h(t,x) = \lambda(t+x)$ for some arbitrary function λ.
Thus

$$g(t,x) = \int_0^x \lambda(t+s)ds + \beta(t),$$

for some function β and, going back all the steps to σ, we have the following result, which was proved in [17].

Proposition 8.37 *The deterministic volatility structure* $\sigma(t,x)$ *induces a Markovian short rate if and only if it can be written as*

$$\sigma(t,x) = c(t)exp^{\int_t^{t+s}\lambda(s)ds}.$$

for some functions c *and* λ*, where* $c > 0$.

8.7 More in depth

In this section we come back to the family of curves introduced in chapter 7:

$$(8.8)\quad G(x) = z_1 + z_2x + z_3e^{ax} + z_4xe^{ax} + z_5x^2e^{ax} + z_6e^{2ax} + z_7xe^{2ax} + z_8x^2e^{2ax}.$$

We want to investigate the "lucky choice" of this parametrization, showing the deep reasons that justify such a choice. Consider again our volatility form (for sake of simplicity we restrict to a one dimensional σ)

$$\sigma(x) = (\alpha + \beta x)e^{\gamma x} + \delta.$$

The coefficient μ is easily obtained by the HJM condition on the drift:

$$\mu = [(\alpha + \beta x)e^{\gamma x} + \delta] \cdot \left[\delta x + \frac{\beta}{\gamma}xe^{\gamma x} + (e^{\gamma x} - 1)\left(\frac{\alpha}{\gamma} - \frac{\beta}{\gamma^2}\right)\right].$$

We now consider the Lie algebra $\{\mu,\sigma\}_{LA}$: it is clearly finite dimensional, so as a consequence of Theorem 8.18 the forward rate model generated by σ admits a finite dimensional realization.
We now show how to construct it: from Proposition 8.31 we recall that the realization is posed in this form

$$G(z_0, z_i^k | i = 1,\ldots,m; k = 0,\ldots,n_i)(x) = r_0(x+z_0) + \frac{1}{2}(\|S(x+z_0)\|^2 - \|S(x)\|^2) + \sum_{i=1}^{m}\sum_{k=0}^{n_i}\sigma_i^{(k)}(x)z_i^k.$$

We now use the fact that

$$D(x) = \sigma(x)\int_0^x \sigma(u)du,$$

or, in a equivalent way

$$D(x) = \frac{1}{2}\frac{\partial}{\partial x}\|S(x)\|^2.$$

A simple calculation of the crucial term of the realization gives us that

$$\frac{1}{2}\Big(\|S(x+z_0)\|^2 - \|S(x)\|^2\Big) = \int_x^{x+z_o} D(u)du.$$

In our case, with σ as before, we obtain that

$$D(x) = \frac{\Big((\alpha+\beta x)e^{\gamma x} + \delta\Big)(e^{\gamma x}\alpha\gamma + e^{\gamma x}\beta\gamma x - e^{\gamma x}\beta + \delta x\gamma^2 - \alpha\gamma + \beta)}{\gamma^2}.$$

and that

$$2\gamma^4 \int_x^{x+z_o} D(u)du = \Big((\beta\gamma e^{\gamma z_0} z_0 + \beta - \alpha\gamma - \beta e^{\gamma z_0} + \alpha\gamma e^{\gamma z_0} - \beta\gamma x + \beta\gamma e^{\gamma z_0} x)$$
$$(\beta\gamma e^{\gamma z_0} z_0 - \beta + \alpha\gamma - \beta e^{\gamma z_0} + \alpha\gamma e^{\gamma z_0} + \beta\gamma x + \beta\gamma e^{\gamma z_0} x)\Big) e^{2\gamma x}$$

$$-2\Big(-2\beta\alpha e^{\gamma z_0}\gamma + x\alpha\delta\gamma^3 + 2\alpha\beta\gamma + x\beta e^{\gamma z_0}\delta\gamma^2 + x\beta^2\gamma$$
$$+\gamma^2\alpha\beta e^{\gamma z_0} x + \alpha^2 e^{\gamma z_0}\gamma^2 + e^{\gamma z_0}\beta^2 - x^2\beta e^{\gamma z_0}\delta\gamma^3 - \gamma\beta^2 e^{\gamma z_0} z_0$$
$$-z_0^2\beta e^{\gamma z_0}\delta\gamma^3 + \gamma^2\beta e^{\gamma z_0} z_0\alpha - \alpha^2\gamma^2 - x\alpha e^{\gamma z_0}\delta\gamma^3 - z_0\alpha e^{\gamma z_0}\delta\gamma^3$$
$$+x^2\beta\delta\gamma^3 - \gamma\beta^2 e^{\gamma z_0} x + z_0\beta e^{\gamma z_0}\delta\gamma^2 - x\beta\alpha\gamma^2 - 2z_0 x\beta e^{\gamma z_0}\delta\gamma^3$$
$$-\beta^2 - x\beta\delta\gamma^2\Big) e^{\gamma x}$$

$$+\delta z_0(\delta z_0\gamma^2 + 2\beta + 2\delta x\gamma^2 - 2\alpha\gamma)\gamma^2$$

Collecting ordered powers of x in the second term we get

$$\Big((-\beta^2\gamma^2 + e^{2\gamma z_0}\beta^2\gamma^2)\mathbf{x^2} + (-2e^{2\gamma z_0}\beta^2\gamma + 2\beta\alpha e^{2\gamma z_0}\gamma^2 + 2\beta^2\gamma + 2z_0 e^{2\gamma z_0}\beta^2\gamma^2 - 2\beta\alpha\gamma^2)\mathbf{x}$$
$$-2\beta\alpha e^{2\gamma z_0}\gamma + e^{2\gamma z_0}\beta^2 - \beta^2 - 2z_0 e^{2\gamma z_0}\beta^2\gamma - \alpha^2\gamma^2 + 2\alpha\beta\gamma + \alpha^2 e^{2\gamma z_0}\gamma^2 + z_0^2 e^{2\gamma z_0}\beta^2\gamma^2$$
$$+2z_0\beta\alpha e^{2\gamma z_0}\gamma^2\Big)\mathbf{e^{2\gamma x}}$$
$$+\Big(2(\beta e^{\gamma z_0}\delta\gamma^3 - \beta\delta\gamma^3)\mathbf{x^2} + 2(\beta\delta\gamma^2 + \beta\alpha\gamma^2 - \beta\gamma^2 e^{\gamma z_0}\alpha + 2z_0\beta e^{\gamma z_0}\delta\gamma^3 + \alpha e^{\gamma z_0}\delta\gamma^3 + e^{\gamma z_0}\gamma\beta^2$$
$$-\alpha\delta\gamma^3 - \beta^2\gamma - \beta e^{\gamma z_0}\delta\gamma^2)\mathbf{x} - 4\alpha\beta\gamma + 2\beta^2\gamma e^{\gamma z_0} z_0 + 2\beta^2 - 2\alpha^2\gamma^2 e^{\gamma z_0} + 2\alpha^2\gamma^2 - 2\beta\gamma^2 e^{\gamma z_0} z_0\alpha$$
$$+2z_0\alpha e^{\gamma z_0}\delta\gamma^3 + 4\beta\gamma e^{\gamma z_0}\alpha - 2z_0\beta e^{\gamma z_0}\delta\gamma^2 + 2z_0^2\beta e^{\gamma z_0}\delta\gamma^3 - 2\beta^2 e^{\gamma z_0}\Big)\mathbf{e^{\gamma x}}$$
$$+(2\delta^2 z_0\gamma^4)\mathbf{x} + \delta^2 z_0^2\gamma^4 + 2\delta z_0\gamma^2\beta - 2\delta z_0\gamma^3\alpha$$

Ignoring constants, it easy to see that this expression is generated by the fundamental terms in (7.20), that is

$$\frac{1}{2}\Big(\|S(x+z_0)\|^2 - \|S(x)\|^2\Big) \in span\{1, x, e^{\gamma x}, xe^{\gamma x}, x^2 e^{\gamma x}, e^{2\gamma x}, xe^{2\gamma x}, x^2 e^{2\gamma x}\}.$$

We thus obtain an identical version of the family of curves chosen.
More precisely, we have just proved that every finite dimensional realization of the model must contain at least all the terms in (8.8). In chapter 7 we proved the consistency between our model and this family and so, with a "necessary and sufficient" reasoning we obtain this important **minimality** result.

Theorem 8.38 *The family of curves 8.8 is the* **smallest** *family available to obtain the consistency with our model.*

.

Chapter 9

Pricing Caps&Floors

This chapter is based on *Pricing Caps&Floors with a consistent HJM model*, a joint work with Roberto Renó.

9.1 Introduction

One of the main goals of financial mathematics is to determine analytically the prices of derivatives. In this chapter we apply the sophisticated machinery introduced in literature in the last years, in order to study our special market constituted by Fixed Income Securities. We will study the behavior of the new model introduced in chapter 6, pricing the most liquid and traded derivatives, caps & floors.
We recall that our choice of volatility function is

$$\sigma(x) = (\alpha + \beta x)e^{\gamma x} + \delta. \tag{9.1}$$

With this shape for every factor i the model can be written as

$$df_i(t, T-t) = \mu_i(\alpha_i, \beta_i, \gamma_i, \delta_i)dt + \sum_{i=1}^{F} \xi_i \sigma_i(T-t; \alpha_i, \beta_i, \gamma_i, \delta_i)dW_i(t), \tag{9.2}$$

where ξ_i are the factors' weights (see chapter 6) and μ given by

$$\mu_i = [(\alpha_i + \beta_i x)e^{\gamma_i x} + \delta_i] \cdot \left[\delta_i x + \frac{\beta_i}{\gamma_i} x e^{\gamma_i x} + (e^{\gamma_i x} - 1)\left(\frac{\alpha_i}{\gamma_i} - \frac{\beta_i}{\gamma_i^2}\right)\right]. \tag{9.3}$$

9.2 Caps&Floors

Caps&Floors are the most traded interest rates derivatives, so they are the "natural" candidate to test our model on the the real market.
As well known, an interest rate cap is a financial insurance which protects you from having

to pay more than a predetermined rate, the cap rate, even though you have a loan at a floating rate of interest. On the contrary, the floor contracts guarantee that the interest rate paid on a floating rate loan will never be below some predetermined floor rate.
Let suppose that we are standing at time $t = 0$, and the cap is to be in force over the interval $[0, T]$. Technically speaking, a cap is the sum of a number of basic contracts, known as caplets, which are defined as follows.

- The interval $[0, T]$ is subdivided by the equidistant points $0 = T_0, T_1, \ldots, T_n = T$. We use the notation δ for the length of an elementary interval, i.e. $\delta = T_i - T_{i-1}$. Typically, δ is a quarter of year, or half a year.

- The cap is working on some principal amount of money, denoted by K, and the cap rate is denoted by R.

- The floating rate of interest underlying the cap is not the short rate r, but rather some market rate, and we will assume that over the interval $[T_{i-1}, T_i]$ it is the LIBOR spot rate $L(T_{i-1}, T_i)$.

- Caplet i is now defined as the following contingent claim, paid at T_i

$$\mathcal{X}_i = K\delta max[L(T_{i-1}, T_i) - R, 0].$$

We now turn to the problem of pricing the caplet, and without loss of generality we may assume that $K = 1$. We will also use the notation $x^+ = max[x, 0]$, so the caplet can be written as

$$\mathcal{X} = \delta(L - R, 0)^+,$$

where $L = L(T_{i-1}, T_i)$. Denoting $p(T_{i-1}, T_i)$ by p, and recalling that

$$L = \frac{1-p}{p\delta},$$

we have

$$\mathcal{X} = \delta(L - R)^+ = \delta\Big(\frac{1-p}{p\delta} - R\Big)^+ = \Big(\frac{1}{p} - (1 + \delta R)\Big)^+ = \Big(\frac{1}{p} - R^*\Big) = \frac{R^*}{p}\Big(\frac{1}{R^*} - p\Big)^+,$$

where $R^* = 1 + \delta R$.
It is then clear that a payment of $\frac{R^*}{p}\Big(\frac{1}{R^*} - p\Big)^+$ at time T_i is equivalent to a payment of $R^*\Big(\frac{1}{R^*} - p\Big)^+$ at time T_{i-1}.
Consequently we see that a caplet is equivalent to R^* put options on an underlying T_i-bond, where the exercise date of the option is at T_{i-1} and the exercise price is $1/R^*$.
Hence we have proved that

an entire cap contract is a portfolio of put options,

and we may use the earlier results of chapter 2 to compute the theoretical price.

9.3 Pricing Caps&Floors

In order to obtain a pricing formula, we introduce the cumulative distribution function of the $N[0,1]$ distribution:

$$N(x) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{x} exp(-z^2/2)dz.$$

We now consider an European call option with expiration date T_0 and exercise price K, on an a security with price process $S(t)$. Assume now that the discounted price process

$$Z_{S,T}(t) = \frac{S(t)}{p(t,T)}$$

has a stochastic differential of the form

$$dZ_{S,T}(t) = Z_{S,T}(t)\mu_{S,T}(t)dt + Z_{S,T}(t)\sigma_{S,T}(t)dW_t,$$

where the volatility process $\sigma_{S,T}(t)$ is *deterministic*. Under this assumption, we have this generalized version of the celebrated Black&Scholes formula.

Theorem 9.1 *The price at $t = 0$ of the call option $\mathcal{X}$ is given by*

$$\Pi(0,\mathcal{X}) = S(0)N(d_1) - Kp(0,T)N(d_2),$$

where

$$d_1 = d_2 + \sqrt{\Sigma^2_{S,T}(T)},$$

$$d_2 = \frac{ln(\frac{S(0)}{Kp(0,T)}) - \frac{1}{2}\Sigma^2_{S,T}(T)}{\sqrt{\Sigma^2_{S,T}(T)}},$$

$$\Sigma^2_{S,T} = \int_0^T \|\sigma_{S,T}(s)\|^2 ds.$$

Proof See [34]. □

We now turn to bond options: given a specified model

$$df(t,T) = \mu(t,T,f(t,T))dt + \sigma(t,T)dW_t,$$

we know that the dynamics of $p(t,T)$ is given by

$$dp(t,T) = p(t,T)r(t)dt + p(t,T)v(t,T)dW_t,$$

where $v(t,T) = -\int_t^T \sigma(t,s)ds$.
Finally, we consider a European call option, with expiration date T_0 and exercise price K, on an underlying bond with maturity T_1 (of course $T_0 < T_1$).

Theorem 9.2 *The price at $t = 0$ of the option*

$$\mathcal{X} = max[p(T_0, T_1) - K, 0]$$

is given by

$$\Pi(0, \mathcal{X}) = p(0, T_1)N(d_1) - Kp(0, T_0)N(d_2),$$

where

$$d_1 = \frac{ln(\frac{p(0,T_1)}{Kp(0,T_0)}) + \frac{1}{2}\Sigma^2_{T_1,T_0}}{\sqrt{\Sigma^2_{T_1,T_0}}},$$

$$d_2 = d_1 - \sqrt{\Sigma^2_{T_1,T_0}},$$

$$\Sigma^2_{T_1,T_0} = \int_0^{T_0} \|\sigma_{T_1,T_0}(s)\|^2 ds.$$

Proof In order to use the previous theorem, we consider the process

$$Z(t) = \frac{p(t, T_1)}{p(t, T_0)}.$$

It's easy to show that

$$\sigma_{T_1,T_0} = v(t, T_1) - v(t, T_0) = -\int_{T_0}^{T_1} \sigma(t, s)ds;$$

this is again deterministic, so we can apply theorem 9.1 and we have the thesis. □

Corollary 9.3 *With our functional form of the volatility in (9.1) we have that*
$\Sigma^2_{T_1,T_0} = 1/4(2\alpha^2\gamma^2\tau_0^4 - 4\alpha^2\gamma^2\tau_0^3\tau_1 + 2\alpha^2\gamma^2\tau_0^2\tau_1^2 - 2\alpha^2\gamma^2\tau_0^2 + 4\alpha^2\gamma^2\tau_0\tau_1 - 2\alpha^2\gamma^2\tau_1^2 + 4\alpha\beta\gamma^2\tau_0^4T_0 - 4\alpha\beta\gamma^2\tau_0^3\tau_1T_0 - 4\alpha\beta\gamma^2\tau_0^3\tau_1T_1 + 4\alpha\beta\gamma^2\tau_0^2\tau_1^2T_1 - 4\alpha\beta\gamma^2\tau_0\tau_1T_0 + 4\alpha\beta\gamma^2\tau_0\tau_1T_1 + 4\alpha\beta\gamma^2\tau_1^2T_0 - 4\alpha\beta\gamma^2\tau_1^2T_1 - 6\alpha\beta\gamma\tau_0^4 + 12\alpha\beta\gamma\tau_0^3\tau_1 - 6\alpha\beta\gamma\tau_0^2\tau_1^2 + 6\alpha\beta\gamma\tau_0^2 - 12\alpha\beta\gamma\tau_0\tau_1 + 6\alpha\beta\gamma\tau_1^2 + 8\alpha\gamma^3\delta\tau_0^3T_0 - 8\alpha\gamma^3\delta\tau_0^3T_1 - 8\alpha\gamma^3\delta\tau_0^2\tau_1T_0 + 8\alpha\gamma^3\delta\tau_0^2\tau_1T_1 - 8\alpha\gamma^3\delta\tau_0^2T_0 + 8\alpha\gamma^3\delta\tau_0^2T_1 + 8\alpha\gamma^3\delta\tau_0\tau_1T_0 - 8\alpha\gamma^3\delta\tau_0\tau_1T_1 + 2\beta^2\gamma^2\tau_0^4T_0^2 - 4\beta^2\gamma^2\tau_0^3\tau_1T_0T_1 + 2\beta^2\gamma^2\tau_0^2\tau_1^2T_1^2 - 2\beta^2\gamma^2\tau_1^2T_0^2 + 4\beta^2\gamma^2\tau_1^2T_0T_1 - 2\beta^2\gamma^2\tau_1^2T_1^2 - 6\beta^2\gamma\tau_0^4T_0 + 6\beta^2\gamma\tau_0^3\tau_1T_0 + 6\beta^2\gamma\tau_0^3\tau_1T_1 - 6\beta^2\gamma\tau_0^2\tau_1^2T_1 + 6\beta^2\gamma\tau_0\tau_1T_0 - 6\beta^2\gamma\tau_0\tau_1T_1 - 6\beta^2\gamma\tau_1^2T_0 + 6\beta^2\gamma\tau_1^2T_1 + 5\beta^2\tau_0^4 - 10\beta^2\tau_0^3\tau_1 + 5\beta^2\tau_0^2\tau_1^2 - 5\beta^2\tau_0^2 + 10\beta^2\tau_0\tau_1 - 5\beta^2\tau_1^2 + 8\beta\gamma^3\delta\tau_0^3T_0^2 - 8\beta\gamma^3\delta\tau_0^3T_0T_1 - 8\beta\gamma^3\delta\tau_0^2\tau_1T_0T_1 + 8\beta\gamma^3\delta\tau_0^2\tau_1T_1^2 - 8\beta\gamma^3\delta\tau_0\tau_1T_0^2 + 16\beta\gamma^3\delta\tau_0\tau_1T_0T_1 - 8\beta\gamma^3\delta\tau_0\tau_1T_1^2 - 16\beta\gamma^2\delta\tau_0^3T_0 + 16\beta\gamma^2\delta\tau_0^3T_1 + 16\beta\gamma^2\delta\tau_0^2\tau_1T_0 - 16\beta\gamma^2\delta\tau_0^2\tau_1T_1 + 16\beta\gamma^2\delta\tau_0^2T_0 - 16\beta\gamma^2\delta\tau_0^2T_1 - 16\beta\gamma^2\delta\tau_0\tau_1T_0 + 16\beta\gamma^2\delta\tau_0\tau_1T_1 + 4\gamma^5\delta^2\tau_0^2T_0^3 - 8\gamma^5\delta^2\tau_0^2T_0^2T_1 + 4\gamma^5\delta^2\tau_0^2T_0T_1^2)/(\gamma^5\tau_0^2)$,
where $\tau_0 = e^{T_0\gamma}$ and $\tau_1 = e^{T_1\gamma}$.

Proof Simple but boring calculation are straightforward. □

9.4 Results

Considerations on data and variance estimation have been already exposed in chapter 6, so we can directly introduce results on simulations.

9.4.1 In-sample results

Some comments are in order. We start from the impact of the number of factors. It is clear that there's a great improvement passing from one factor to two factors, while there's no improvement at all passing from two factors to three factors. This is not surprising, looking at the very small weights of the third factor in the variance estimation, see Table 6.2. This result is of great practical importance.

Once the number of factors is fixed, we find that our model fits better caps of longer maturity. This is not surprising as well. Our model completely ignores stochastic volatility effects, as well as the risk premium associated with it. For index and equity options, it is well known that deterministic volatility functions are not capable to fully describe the term structure of observed volatility, see e.g. [27]. Thus we should expect our prices to be incorrect. The fact that this bias is more severe for shorter maturity options is a signature of mean reversion in volatility.

Looking at the time scale (daily, weekly, monthly) the evidence is mixed. Looking at the one-factor models, variance estimated with monthly data performs better than variance estimated with weekly data, which in turn outperforms that estimated with daily data. With two-factor and three-factor models, monthly variance still achieves the best performance, but daily data perform better than weekly data and in a very similar fashion that monthly data. We conclude that our analysis points out a very weak evidence for increasing the time scale, which anyway could be an artifact due to the particular period selected.

Finally, and most importantly, we obtain that the correlation between model prices and observed prices is very close to 1 in any case. Now, since volatility is the same across the sample and varies only as a function of maturity, than the daily movements in the prices predicted by the model are entirely due to the daily movements of the term structures. We then conclude that the movements in the term structure of cap prices is entirely due to the movements of the term structure. This result contrasts some current opinions, principally related to stock market evidence.

Table 9.1: This table shows the average pricing error on the whole sample for the model considered. We fit volatility in the 1999-2002 sample, then we compute cap prices in the same sample. Results are displayed for the model estimated with daily, weekly and monthly data, with one, two and three factor. The column labelled by μ reports the average pricing error (percentage), the column labelled by σ reports the pricing error standard deviation (percentage), while the column labelled ρ reports the correlation between variation in model prices and variation in observed prices.

Daily data								
1 factor			2 factor			3 factor		
μ	σ	ρ	μ	σ	ρ	μ	σ	ρ
-43.12	33.61	0.96	-4.53	15.99	0.98	-4.53	15.99	0.98
-26.78	22.43	0.99	-4.77	7.85	0.99	-4.77	7.85	0.99
-19.42	14.75	0.99	-6.07	5.92	0.99	-6.07	5.92	0.99
-14.83	10.66	0.99	-5.45	4.89	0.99	-5.45	4.89	0.99
-11.94	8.31	0.99	-4.89	4.41	0.99	-4.89	4.41	0.99
-7.90	5.60	0.99	-3.39	3.65	0.99	-3.39	3.65	0.99
-4.92	4.22	0.99	-2.05	3.42	0.99	-2.05	3.42	0.99
Weekly data								
1 factor			2 factor			3 factor		
μ	σ	ρ	μ	σ	ρ	μ	σ	ρ
-19.78	19.86	0.98	28.63	24.50	0.98	30.36	25.30	0.98
-9.22	9.33	0.99	16.55	14.88	0.99	17.15	15.21	0.99
-6.35	5.88	0.99	8.75	8.66	0.99	8.97	8.75	0.99
-3.96	4.26	0.99	6.23	6.60	0.99	6.34	6.65	0.99
-2.46	3.69	0.99	4.87	5.55	0.99	4.96	5.58	0.99
-0.08	3.18	0.99	4.19	4.54	0.99	4.27	4.57	0.99
1.94	3.30	0.99	4.31	4.16	0.99	4.35	4.17	0.99
Monthly data								
1 factor			2 factor			3 factor		
μ	σ	ρ	μ	σ	ρ	μ	σ	ρ
-33.00	26.53	0.98	9.65	17.24	0.98	9.69	17.25	0.98
-17.30	14.39	0.99	5.18	9.06	0.99	5.20	9.06	0.99
-11.94	8.90	0.99	1.25	5.56	0.99	1.25	5.56	0.99
-8.44	6.12	0.99	0.49	4.47	0.99	0.50	4.47	0.99
-6.30	4.77	0.99	0.16	4.03	0.99	0.16	4.03	0.99
-3.23	3.36	0.99	0.57	3.49	0.99	0.57	3.49	0.99
-0.86	3.02	0.99	1.24	3.38	0.99	1.24	3.39	0.99

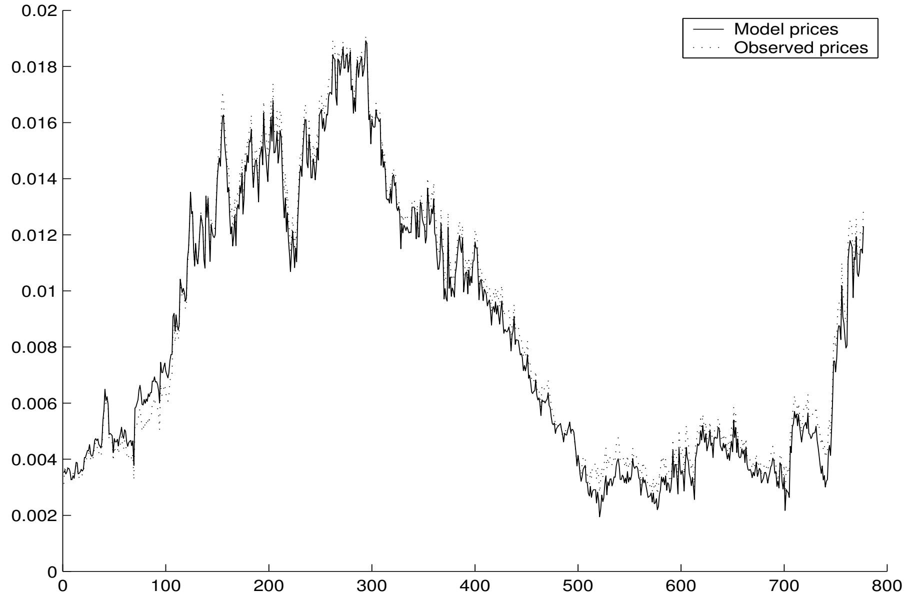

Figure 9.1: Two years cap, one factor.

Figure 9.2: Ten years cap, one factor.

Bibliography

[1] Adamo, M. &, Amadori, A.L. & Bernaschi, M. & La Chioma, C. & Marigo, A., & Piccoli, B. & Sbaraglia, S. & Uboldi, A. & Vergni, D. (2003) Optimal Strategies for the Issuances of Public Debt Securities. *IAC Report 2003.*

[2] Angelini, F. & Herzel, S. (2002). Consistent initial curves for interest rate models. *Journal of Derivatives*, 9, 8-18.

[3] Avellaneda, M. & Laurence, P. (2000). Quantitative modelling of derivative securities. *Chapman & Hall/CRC.*

[4] Barone, E. & Cuoco, D. & Zautzik, E. (1989). La struttura dei rendimenti per scadenza secondo il modello CIR: una verifica empirica. *Banca d'Italia, Temi di discussione*, 128.

[5] Bibby, B.M & Sorensen, M. (1995). Martingale estimating functions for discretely observed diffusion processes. *Bernoulli*, vol.1 , 17-39.

[6] Björk, T. (1996). Interest rate theory. *CIME Lectures.*

[7] Björk, T. (1998). Arbitrage theory in continuous time. *Oxford University Press.*

[8] Björk, T. & Christensen, B.J. (1997). Interest rate dynamics and consistent forward rate curves. *Working Paper Series in Economics and Finance*, 209.

[9] Björk, T. & Gombani, A. (1999). Minimal realizations of interest rate models. *Finance and Stochastics*, 3, 413432.

[10] Björk, T. & Landen C. (2000). On the construction of finite dimensional realizations for nonlinear forward rate models. *Working Paper Series in Economics and Finance*, 420.

[11] Björk, T. & Svensson, L. (1999). On the existence of finite dimensional realizations for nonlinear forward rate models. *Working Paper Series in Economics and Finance*, 338.

[12] Bolder, D.J. (2002). Towards a complete debt strategy simulation framework. *Bank of Canada Workin Paper*, 2002-13.

[13] Brace, A. & Musiela, M. (1994). A multi-factor Gauss Markov implementation of Heath, Jarrow and Morton. *Mathematical Finance*, 4, 259-283.

[14] Brace, A. & Musiela, M. (1997). Swap Derivatives in a Gaussian HJM framework. *Mathematics of Derivative Securities*, Michael A. H. Dempster and Stanley R. Pliska. Cambridge: Cambridge University Press, 336-368.

[15] Brigo, D. & Mercurio, F. (2001). Interest rate models: theory and practice. *Springer Finance.*

[16] Brown, R.G. & Schaefer, S.M. (1994). Interest rate volatility and the shape of the term structure. *Phil. Trans. R. Soc. Lond.*, 347, 563-576.

[17] Carverhill A. (1994). When is the spot rate Markovian?. *Mathematical Finance*, 4, 305-312.

[18] Cox, J. & Ingersoll, J. & Ross, S. (1985a). An inter-temporal general equilibrium model of asset prices. *Econometrica*, vol. 53, n.2, 363-384.

[19] Cox, J. & Ingersoll, J. & Ross, S. (1985b). A theory of the term structure of interest rates. *Econometrica*, vol. 53, n.2, 385-408.

[20] Da Prato, G. & Zabczyk, J. (1992). Stochastic Equations in Infinite Dimensions. *Cambridge University Press.*

[21] Dantzig, G.B. (1963). Linear Programming and Extensions, *Princeton University Press, Princeton N.J..*

[22] Demmel, R. (1999). Fiscal Policy, Public Debt and the Term Structure of Interest Rates. Lecture Notes in Economics and Mathematical Sciences 476, *Springer-Verlag, Berlin-Heidelberg DE.*

[23] Duffie, D. & Kan, R. (1996). A yield-factor model of interest rates. *Math. Finance*, 6, 379-406.

[24] Dai, Q. & Singleton, K. (2000). Specification analysis of affine term-structure models. *Journal of Finance*, 5, 1943-1978.

[25] Delbaen, F. & Schachermayer, W. (1994). A general version of the fundamental theorem on asset pricing. *Mathematishe Annalen*, 300, 463-520.

[26] Driessen, J. & Klaassen, P & Melenberg, B. (2000). The performance of multi-factor term structure models for pricing and hedging caps and swaptions. *CER Working Papers*, 2000-93.

[27] Dumas, B. & Fleming, J. & Whaley, R. (1998). Implied volatility functions: empirical tests. *Journal of Finance*, vol. 53, n. 6, 2059-2106.

[28] Dybvig, P.H. (1997). Bond and bond option pricing based on the current term structure (Mathematics of derivative securities). *Cambridge Press.*

[29] Fan, R. & Gupta, A. & Ritchken, P. (2001). On pricing and hedging in the swaption market: How many factors really? *Working Paper Case Western Reserve University.*

[30] Filipovic, D. (2001). Consistency problems for Heath-Jarrow-Morton interest rate models. *Springer Verlag.*

[31] Filipovic, D. & Teichmann J. (2002). On Finite-dimensional Term Structure models. *Working paper of Princeton University.*

[32] Filipovic, D. & Teichmann J. (2003). On Geometry of the term structure. *Working paper of Princeton University.*

[33] Gallant, R. & Tauchen, G. (1998). Reprojecting partially observed systems with application to interest rate diffusions. *Journal of the American Statistical Association*, 93, 93, 10-24.

[34] Geman, H. & El Karoui, N. & Rochet, J.C. (1995). Changes of numeraire, changes of probability measures and option pricing. *Journal of Applied Probability*, 32, 443-458.

[35] Gentile, M. & Renó, R. (2002). Which model for the Italian interest rates? *LEM Working Paper Series.*

[36] Gerace, A. (2001). Tesi di Laurea.

[37] Going-Jaeschke, A. (1998). Parameter estimation and Bessel processes in financial models and numerical analysis in Hamiltonian dynamics. *Diss. Math. Wiss. ETH Zurich*, n. 12566.

[38] Harrison, J.M. & Pliska, S. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes & Appl.*, 11, 215-260.

[39] Heath, D. & Jarrow, R. & Morton, A. (1992). Bond pricing and the term structure of interest rates: a new methodology. *Econometrica*, vol. 60, n.1, 77-106.

[40] Hull, J. (2002). Options, futures and other derivative securities. *Prentice & Hall*, 5th edition.

[41] Hull, J. & White, A. (1990). Pricing interest rate derivative securities. *The Review of Financial Studies*, 3, n.1, 573-592.

[42] Ikeda, N. & Watanabe, S. (1981). Stochastic differential equations and diffusion processes. *North Holland.*

[43] Jackson, D. (2000). The New National Accounts: An Introduction to the System of National Accounts 1993 and the European System of Accounts 1995. *Edward Elgar Pub, Camberley UK.*

[44] James, J. & Webber N. (2000). Interest rate modeling. *John Wiley & Sons.*

[45] John, F. (1982). Partial Differential Equations. *Springer-Verlag.*

[46] Karatzas, I. & Shreve, S. (1988). Brownian Motion and Stochastic Calculus. *Springer-Verlag.*

[47] Kloeden, P.E. & Platen, E. (1992). Numerical solution of stochastic differential equations. *Springer-Verlag.*

[48] Lamberton, D. & Lapeyre, B. (1996). Introduction to stochastic calculus applied to finance. *Chapman & Hall.*

[49] Litterman, R. & Scheinkman, E. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, June, 51-61.

[50] Maciejowski, J.M. (2002). Predictive Control with Constraints. *Prentice Hall, London UK*

[51] Mandelbrot, B. (1997). Fractals and scaling in finance. *Springer.*

[52] Mc Cullogh, J.H. (1971). Measuring the term structure of interest rates. *Journal of Business*, vol. XLIV (January), 19-31.

[53] Moraleda, J. & Vorst, T. (1997). Pricing American interest rate claims with humped volatility models. *Journal of Banking and Finance*, 21, 1131-1157.

[54] Nelson, C. & Siegel, A. (1987). Parsimonious modeling of yield curves. *Journal of Business*, 60, 473-489.

[55] Øksendal, B. (1995). Stochastic Differential Equations. *Springer.*

[56] Pegoraro, F. La struttura a termine dei tassi di interesse e trading su titoli: un approccio con il modello CIR al mercato italiano dei BTP. Preprint.

[57] Pelliccia, A. (2002). Tesi di Laurea.

[58] Rebonato, R. (1998). Interest-rate option models. *John Wiley & Sons.*

[59] Renó, R. & Uboldi, A. (2002). PCA based calibration of an HJM model. *Quaderno IAC*, Q28-002.

[60] Renó, R. & Uboldi, A. (2002). Pricing Caps&Floors with a consistent HJM model. *IAC Report.*

[61] Ritchken, P. & Chuang, I. (1999). Interest rate option pricing with volatility humps. *Review of Derivatives Research*, 3, 237-262.

[62] Scherer, K.P. & Avellaneda, M. (2002). All for one... one for all? A principal component analysis of latin american brady bond debt from 1994 to 2000. *International Journal of Theoretical and Applied Finance*, vol. 5, n.1, 79-106.

[63] Taylor, J.R. (1990). Introduzione all'analisi degli errori. *Zanichelli.*

[64] Torosantucci, L. & Uboldi, A. (2002). Empirical analysis of the Italian Treasury fixed income securities market. *Quaderno IAC*, Q27-002.

[65] Torosantucci, L. & Uboldi, A. (2003). Static and dynamic approach to the CIR model and empirical evaluation of the market price of risk. *IAC Report.*

[66] Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5, 177-188.

[67] Yong, J. & Zhou, X.Y. (1999). Stochastic controls. *Springer Verlag, New York NY.*

[68] Zweig, M. Winning on Wall Street. *Time Warner International.*