Universidad Cardenal Herrera-**CEU**

Departamento de Ciencias Físicas, Matemáticas y de la Computación

Efficient Methods for Calibrating and Pricing Interest Rate Options

TESIS DOCTORAL

Presentada por: Lluís Navarro Girbés

Dirigida por:
Dr. Antonio Falcó Montesinos
Dr. Juan Miguel Nave Pineda

Valencia, 2012

Prediction, not narration, is the real test of our understanding of the world.

NASSIM NICHOLAS TALEB

Agradecimientos

Me gustaría expresar en estas líneas mi agradecimiento a todas aquellas personas que con su ayuda han colaborado en la realización del presente trabajo, en especial al Dr. D. Antonio Falcó Montesinos y al Dr. D. Juan M. Nave Pineda, directores de esta investigación, por la orientación, el seguimiento y la supervisión continua de la misma, pero sobre todo por la motivación y el apoyo moral recibido a lo largo de estos años.

Especial reconocimiento merece el interés mostrado por mi trabajo y las sugerencias recibidas del profesional del sector bancario y amigo Óscar Bayona, con el que me encuentro en deuda por el ánimo infundido y la confianza en mí depositada hasta el final.

Quisiera hacer extensiva mi gratitud a mis antiguos compañeros de la Escuela Superior de Enseñanzas Técnicas de la Universidad Cardenal Herrera CEU por su amistad. También a diversos profesionales del sector financiero como Marcos Pérez de `Especular.com`, Alberto Montero y Carlos Salinas de *Mora Banc*, Jon Garavilla de *SIAG Consulting* y Marta Díaz de *Coterie Trading* por los buenos ratos pasados al compartir la experiencia, sin duda enriquecedora, de la creación de una nueva empresa del sector, así como por el interés manifiesto por el trabajo de investigación que ahora concluye.

Un agradecimiento muy especial merece la comprensión, paciencia y el ánimo recibidos de mi familia y amigos.

A todos ellos, muchas gracias.

Contents

1 Foundations of Interest Rate Theory **17**

1.1 Definitions and Notation . . . 17

1.2 Interest-Rate Curves . . . 21

1.2.1 The Short Rate and the Money-Market Account . . . 22

1.3 A Brief Note on Martingale Modeling . . . 23

1.3.1 Martingale Measures, Derivative Securities and Arbitrage . 24

2 Forward Rate Models **29**

2.1 The Heath-Jarrow-Morton Framework . . . 29

2.1.1 Absence of Arbitrage . . . 31

2.2 From HJM to Short-Rate Models . . . 32

2.3 Forward Measures . . . 33

2.4 The General HJM Gaussian Model . . . 34

3 Interest Rate Caps **39**

3.1 The Market Practice for Plain Vanilla Caps . . . 39

3.1.1 IRS and At-The-Money Plain Vanilla Caps . . . 41

3.2 Caps under The General HJM Gaussian Model . . . 44

4 Geometric Interest Rate Theory **47**

4.1 The Problem . . . 47

4.2 Setup . . . 49

4.3 The Formalized Problem . . . 50

4.3.1 The Forward Curve Manifold . . . 50

4.3.2 The Space . . . 51

4.3.3 The Interest Rate Model . . . 52

4.4 The Invariance Conditions . . . 52

4.4.1 Simple Invariance . . . 54

5 The Hull-White Model and multiobjective calibration **59**
5.1 The Hull-White Model 59
5.1.1 Markovianity of the HW model 60
5.2 The Nelson and Siegel family and Invariance 62
5.2.1 The NS family 63
5.3 The Minimal Consistent family and Realizations of Gaussian Models 64
5.3.1 The Minimal Consistent family 66
5.4 Calibration to Market Data Approaches 67
5.4.1 The Two-Step Traditional Method 69
5.4.2 The Joint Calibration to Cap and Bond Prices 70
5.5 Empirical Results 72
5.6 Concluding Remarks 78

6 Multiojective Calibration: More Empirical Evidence **81**
6.1 Introduction 82
6.2 Consistent Curves with The Model 84
6.2.1 The Minimal Consistent family 85
6.3 Empirical Results 86
6.4 Conclusions 91

7 Pricing Options with a Consistent HJM Model **95**
7.1 Introduction 95
7.2 Partial Differential Equation of the HW Model 96
7.2.1 Bypassing Forward Induction 97
7.3 Finite-Difference Implementation 98
7.3.1 A Stable Explicit Scheme 100
7.3.2 A Crank-Nicolson Scheme 105
7.4 Numerical Examples 107
7.5 Monte Carlo Simulation for Consistent HJM Models 114
7.5.1 Basic Monte Carlo 114

A Forward Rate Models **119**
A.1 The HJM Framework 119
A.2 From HJM to Short-Rate Models 120

B Geometric Interest Rate Theory **123**
B.1 Setup 123
B.2 The Invariance Conditions 125

C Pricing Options with a Consistent HJM Model **127**

List of Figures

1.1 Zero-rate curves on July 1, 2003. The normal line corresponds to US dollar rates and the dashed one to the euro rates. 21

1.2 Term structure of discount bonds on July 1, 2003. The normal line belongs to the US dollar discount bond curve and the dashed one to the euro. 22

4.1 The vector field from the the system (4.21) with $\mathcal{S}^1$ and a parabollic noninvariant manifold. 57

5.1 Average of the US market TSIR and TSV with 99% confidence levels. 73

5.2 Some daily calibration results for the minimal consistent family. . 75

5.3 Efficient points in the $RPE_B - RPE_C$ space using the method of convex combinations for two different days in sample. 76

5.4 Daily empirical distribution of weights with the best RPE_C for both consistent families as produced by the multi-objective calibration. 77

5.5 Time Series Comparison. 78

5.6 Summary statistics for the calibration results. In-sample descriptive statistics are carried out using the daily Pareto points with the best derivative fit outcomes. 79

6.1 Market TSIR and TSV data in the two different market scenarios. 83

6.2 Discrete data for initial yield-curve estimation. 87

6.3 Summary statistics for calibration results with simulated data. . 88

6.4 Daily estimates of parameters a and α for data simulated from the model with $\alpha = 0.002$ and $a = 0.35$ and starting forward curve $f_0(x) = f^o_{ANS+}(x)$. 89

6.5 Summary statistics for calibration results with US data on both periods. . . . 90
6.6 Not normalized daily empirical distribution of weights with the best RPE_C for both sample periods as produced by the multi-objective calibration. . . . 91
6.7 In-sample fitting time series for the first period (left) and the second period (right) with real market data. . . . 93

7.1 Downward Branching 102
7.2 Upward Branching 103
7.3 An example of stable explicit grid for x with $l = 10a = 1$. . . . 104
7.4 Discrete data for initial discount bond estimation. All rates are expressed with continuous compounding. . . . 107
7.5 Prices (in basispoints of the notional) for put options on 5yr discount bond. . . . 109
7.6 Relative pricing errors of vanilla 2yr put on 5yr discount bond. . . 110
7.7 Calculation times in milliseconds, running MATLAB on an Intel Core 2 Duo P8600 @ 2.39GHz computer for a $2yr$ option on a $5yr$ discount bond, with $a = 0.1$ and $\sigma = 0.01$. . . . 110
7.8 Prices of vanilla/digital European-style caps with semi-annual tenor. . . . 113
7.9 Relative valuation errors of 10yr vanilla/digital caps with semi-annual tenor. . . . 113
7.10 Prices (in percent of the notional) for a 1yr put option on 3yr discount bond. . . . 117

Introduction

This Ph.D. thesis is devoted to the application of dynamic consistent families to problems arising in interest rate modelling.

A self contained introduction to the theoretical framework is presented in Chapter 1. We report just those fundamental definitions and results, like the fundamental arbitrage-free equation or the LIBOR rate definiton among others, that are required in the following chapters. A detailed survey on the mathematical settings of interest rate models is presented in Chapter 2 where we present the Heath-Jarrow-Morton (HJM) framework for the forward rates. Finally, in Chapter 3 we expose the standard market practice and pricing techniques for interest rates derivatives like caps and bond options, that will be fully developed later.

A first aim of the present work is to study consistent families of the HJM models existing in mathematical and financial literature, in order to evaluate their applicability to specific financial engineering problems like calibration or valuation. Therefore, in Chapter 4 we review the general features of the geometric view of HJM models as seminally introduced by Björk and Christensen in [11], introducing the concept of consistent families with this class of models.

A second part of the thesis is oriented to propose new techniques for the application of existing consistent families. In particular, in Chapter 5 and 6, by means of a new multiobjective extension of the calibration techniques proposed by Herzel and Angelini in [2, 3], we develop a consistent framework for the calibration of vanilla derivatives to consistent families.

First results obtained by the implementation of the method suggest that this extended technique is quite robust and shows that the choice of consistent families are really relevant in the quality of joint calibration outcomes. At this point, 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 7 we show that the models empirically analyzed in Chapters 5 and 6 admit numerical implementions which preserve wide open the use of the consistent families introduced before by means of minor modifications of the standard numerical schemes introduced in the literature. In this chapter we face one of the most important problem in Mathematical Finance, that is the pricing of derivative securities. We apply several discretization and simulation techniques to the pricing of vanilla caps, the most important derivative product in fixed income markets, bond options and digital caps. The computational results confirms that the Crank-Nicolson method outperforms the other numerical schemes considered, and it encourages the search for more efficient implementations of this specific finite difference approach.

It is crucial to remark that although our choice of the models is quite restrictive, the results seems to be good, and, from a theoretical and computational point of view, support the use of an entire consistent framework.

Chapter 1

Foundations of Interest Rate Theory

1.1 Definitions and Notation

The primary objects of our investigation are pure discount bonds, of various maturities. All payments are assumed to be made in a fixed currency. Moreover, we need some formal definitions.

Definition 1 (Discount Bond.) *A $T-$maturity pure discount bond is a contract that guarantees its holder the payment of one unit of currency at time T, with no intermediate payments. The contract value at time $t < T$ is denoted by $P(t,T)$. Clearly, $P(T,T) = 1$ for all T.*

Definition 2 (Time to maturity.) *The time to maturity $x = T - t$ is the amount of time expressed in years from the present time t to the maturity time $T > t$.*

Coupon bonds give the owner a payment stream during the interval $[0, T]$. 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.

Pure discount bond prices are the basic quantities in interest-rate theory, and all interest rates can be defined in terms of discount bond prices, as we shall see now. Therefore, they are often used as basic auxiliary quantities from which all rates can be recovered, and in turn discount bond prices can be defined in terms of any given family of interest rates. Notice, however, that interest rates are what is usually quoted in (interbank) financial markets, whereas zero-coupon bonds are theoretical instruments that, as such, are not directly observable in the market. In moving from discount bond prices to interest rates, and vice versa, we need to know two fundamental features of the rates themselves: the compounding type and the day-count convention to be applied in the rate definition. What we mean by "compounding type" will be clear from the definitions below.

Definition 3 (Anually compounded spot interest rate.) *The annually compounded spot interest rate prevailing at time t for the maturity T is denoted by $Y(t,T)$ and is the constant rate at which an investment has to be made to produce an amount of one unit of currency at maturity, starting from $P(t,T)$ units of currency at time t, when reinvesting the obtained amounts once a year. In formulas*

$$Y(t,T) := P(t,T)^{-\frac{1}{T-t}} - 1 \tag{1.1}$$

which implies that bond prices can be expressed in terms of annually compounded rates as

$$P(t,T) = \frac{1}{(1+Y(t,T))^{T-t}} \tag{1.2}$$

Definition 4 (Continuously compounded spot interest rate.) *The continuously compounded spot interest rate prevailing at time t for the maturity T is denoted by $R(t,T)$ and is the constant rate at which an investment of $P(t,T)$ units of currency at time t accrues continously to yield a unit amount of currency at maturity T*

$$R(t,T) := -\frac{\log P(t,T)}{T-t} \tag{1.3}$$

The continuously compounded interest rate is therefore a constant rate that is consistent with the discount bond prices in that

$$e^{R(t,T)(T-t)} P(t,T) = 1 \tag{1.4}$$

from which we can express the bond price in terms of the continuously compounded rate R:

$$P(t,T) = e^{-R(t,T)(T-t)} \tag{1.5}$$

where $T-t$, the time difference expressed in years. An alternative to continuous compounding is simple compounding, which applies when accruing occurs proportionally to the time of the investment.

Definition 5 (Simply compounded spot interest rate.) *The simply compounded spot interest rate prevailing at time t for the maturity T is denoted by $L(t,T)$ and is the constant rate at which an investment has to be made to produce an amount of one unit of currency at maturity, starting from $P(t,T)$ units of currency at time t, when accruing occurs proportionally to the investment time.*

$$L(t,T) := \frac{1 - P(t,T)}{(T-t)P(t,T)} \tag{1.6}$$

We denote by L such rates because the market LIBOR rates are simply compounded. These are the most important interbank rates and they are considered as a reference for contracts, fixing daily in London (London InterBank Offered Rate). There exist equivalent interbank rates fixing in other markets (e.g. the EURIBOR rate, fixing in Brussels by the European Banking Federation).

Suppose that we are standing at time t, and let us fix two other points in time S and T with $t < S < T$. Let us consider now the project of writing a forward rate agreement at time t which allows us to make an investment of one unit of currency at time S, and have a *deterministic* rate of return, determined at the contract time t, over the interval $[S,T]$. This agreement can be achieved with the following replicating strategy

1. At time t we sell one S-bond. This will give us $P(t,S)$ units of our base currency.

2. With this money we may buy exactly a $\frac{P(t,S)}{P(t,T)}$ amount of T-bonds.

   $$P(t,S) - \frac{P(t,S)}{P(t,T)}P(t,T) = 0 \quad \text{in} \quad t$$

   Note that our net investment at initial time t is zero.

3. At time S the S-bond expires, so we must to pay out one monetary unit of our currency.

4. At time T each T-bond expires paying one unit of currency, so we will receive the payoff $P(t,S)/P(t,T) \cdot 1$.

5. The real effect of this strategy is that, based on the contract agreed at t, for an investment of one unit of currency we have received in turn $P(t,S)/P(t,T)$ at time T.

Now the following crutial definition is well motivated by the implementation of the financial strategy above.

Definition 6 (Simply compounded forward interest rate.) *The simple forward rate for the period $[S, T]$ contracted at $t < S < T$, is defined as*

$$L(t; S, T) := \frac{1}{T - S}\left(\frac{P(t, S)}{P(t, T)} - 1\right)$$

Or, in other words, the simple forward rate L, is the solution to the equation

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

Moreover, it is straightforward to recover the spot definition making the assignment $t = S$, i.e. the spot rates are forward when the time of the agreement coincides with the start of the interval over which the interest rate is effective.

The simple forward rate $L(t; T, S)$ may be viewed as an estimate of the future spot rate $L(T, S)$.

When the maturity of the forward rate collapses towards its expiry, we have the notion of *instantaneous forward rate*. Let us consider the limit

$$\begin{aligned} \lim_{\Delta T \to 0^+} L(t; T, T + \Delta T) &= -\lim_{\Delta T \to 0^+} \frac{P(t, T + \Delta T) - P(t, T)}{P(t, T + \Delta T)\Delta T} \\ &= -\frac{1}{P(t, T)} \frac{\partial P(t, T)}{\partial T} \\ &= -\frac{\partial \log P(t, T)}{\partial T} \end{aligned} \tag{1.7}$$

This leads to the following.

Definition 7 (Instantaneous forward interest rate.) *The instaneous forward interest rate prevailing at time t for the maturity $T > t$ is denoted by $F(t, T)$ and is defined as*

$$F(t, T) := \lim_{\Delta T \to 0^+} L(t; T, T + \Delta T) = -\frac{\partial \log P(t, T)}{\partial T}, \tag{1.8}$$

so that we also have

$$P(t, T) = \exp\left(-\int_t^T F(t, u)\, du\right) \tag{1.9}$$

Clearly for this notion to make sense, we need to asume smoothness of the discount bond price function $T \mapsto P(t,T)$ for all T's.

Intuitively, the instantaneous forward rate $F(t,T)$ is a forward interest rate at time t whose maturity is very close to its expiry T, say $F(t,T) \approx L(t;T,T+\Delta T)$ with ΔT small.

1.2 Interest-Rate Curves

A fundamental curve that can be obtained from the market data of interest rates is the zero-copupon curve at a given date t. This curve is the graph of the function mapping maturities into rates at times t. More precisely:

Definition 8 (Zero-rate curve.) *The zero-rate curve at time t is the graph of the function*

$$T \mapsto \begin{cases} L(t,T) & t < T \leq t+1 \\ Y(t,T) & T > t+1 \end{cases} \tag{1.10}$$

Such a zero-coupon curve is also called the *term structure of interest rates* (TSIR) at time t. By definition (1.10), it is a plot at time t of simply-compounded interest rates for all maturities T up to one year and of annually-compounded rates for maturities T larger than one year. Recall that at times it may be considered the

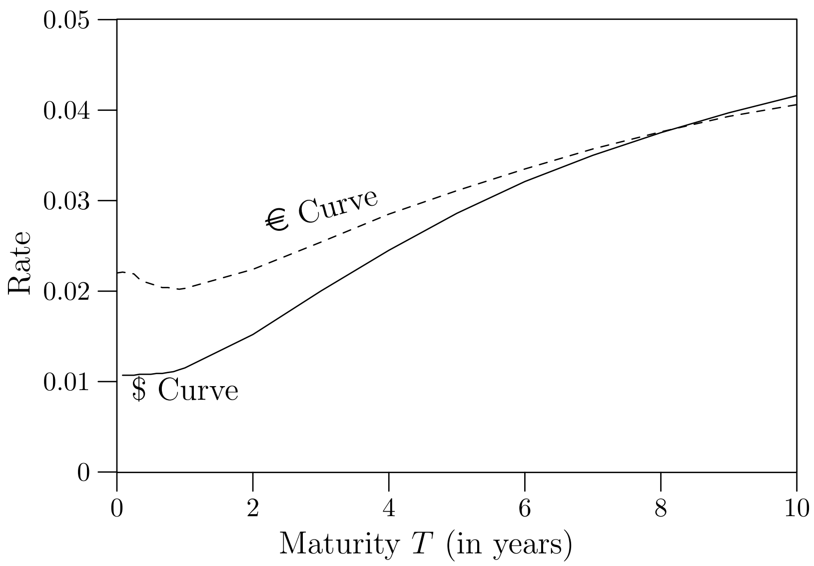


Figure 1.1: Zero-rate curves on July 1, 2003. The normal line corresponds to US dollar rates and the dashed one to the euro rates.

sample for rates with different compounding conventions, such as for example

$$T \mapsto R(t,T), \quad T > t$$

Definition 9 (Discount bond curve.) *The Discount bond curve at time t is the graph of the function*

$$T \mapsto P(t,T), \quad T > t \tag{1.11}$$

which, because of the positivity of interest rates, is a T-decreasing function starting from $P(t,t) = 1$. Two examples of such a curve ara shown in 1.2

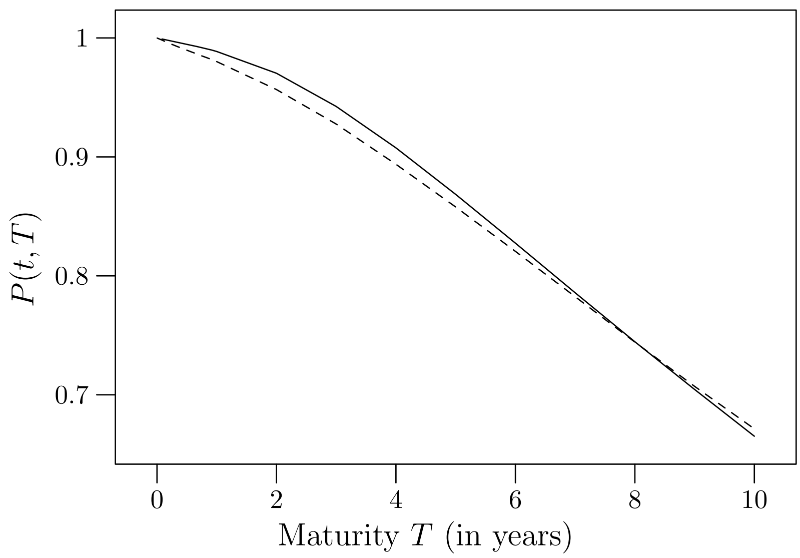


Figure 1.2: Term structure of discount bonds on July 1, 2003. The normal line belongs to the US dollar discount bond curve and the dashed one to the euro.

1.2.1 The Short Rate and the Money-Market Account

Definition 10 (Short rate.) *The instantaneous spot interest rate, also referred as the short rate, is the continuously-compounded insterest rate when time to maturity collapses to zero:*

$$r(t) = \lim_{\Delta t \to 0} R(t, t + \Delta t) \tag{1.12}$$

Let us work out this limit

$$\begin{aligned}\lim_{\Delta t\to 0} R(t,t+\Delta t) &= -\lim_{\Delta t\to 0}\frac{\log P(t,t+\Delta t)}{(t+\Delta t)-t}\\ &= -\lim_{\Delta t\to 0}\frac{\log P(t,t+\Delta t)-\log P(t,t)}{(t+\Delta t)-t}\\ &= -\left.\frac{\partial \log P(t,\theta)}{\partial\theta}\right|_{\theta=t}\\ &= F(t,t)\end{aligned} \tag{1.13}$$

The next definition we consider is the definition of a money-market account. A money-market account represents a locally riskless investment, where profit is accrued continuously at the short rate prevailing in the market at every instant.

Definition 11 (Money-market account.) *We define $B(t)$ to be the value of a money-market account at time $t \geq 0$. Assume that $B(0) = 1$, and that the money-market account evolves according to the following differential equation:*

$$dB(t) = r(t)B(t)dt, \qquad B(0) = 1, \tag{1.14}$$

where $r(t)$ is a positive stochastic process, i.e.,

$$B(t) = \exp\left(\int_0^t r(s)\,ds\right). \tag{1.15}$$

1.3 A Brief Note on Martingale Modeling

Throughout this work we consider a continuous trading economy, with a finite trading interval given by $[0,\Theta]$. The uncertainty is modelled by the filtered probability space $(\Omega,\mathcal{F},\mathbf{F},\mathbb{P})$ where Ω denotes a sample space, with elements $\omega\in\Omega$; $\mathcal{F}$ denotes a σ-algebra on Ω; and $\mathbb{P}$ denotes a probability measure in $(\Omega,\mathcal{F})$. The uncertainty is resolved over $[0,\Theta]$ according to the filtration $\mathbf{F}=\{\mathcal{F}_t\}_{t\geq 0}$.

We consider a financial market $S = [\ S_0\ S_1\ \dots\ S_n\ ]^T$ with a riskless investment, S_0, or money market account given by (1.14), and n risky assets which all follow Itô processes driven by a q-dimensional Wiener-Einstein process, W,

$$dS_i = S_i(\mu_i dt + \sigma_i\cdot dW), \quad S_i(0) > 0, \quad i = 1,\dots,n.$$

the appreciation rates μ_i and the volatility row vectors $\sigma_i = [\ \sigma_{i1}\ \dots\ \sigma_{iq}\ ]$ are assumed to be $\mathcal{F}_t$-adapted, intuitively, this means that they all depend on past values but not on future. They also satisfy the integrability conditions

$$\int_0^\Theta |\mu_i| dt < \infty, \int_0^\Theta ||\sigma_i||^2 dt < \infty \quad i = 1, \ldots, n \tag{1.16}$$

almost surely.

A continuous time *trading strategy* is any $\mathbb{R}^{n+1}$-valued $\mathcal{F}_t$-adapted stochastic process

$$\phi(t) = [\ \phi_0(t) \ \ldots \ \phi_n(t)\ ]$$

where $\phi_i(t)$ denotes the holdings in the asset i at time t. The asset holdings $\phi_i(t)$ are furthermore assumed to satisfy similar regularity conditions as the presented in (1.16).

Its corresponding *value process* is

$$V(\phi, t) = \phi(t) \cdot S(t) = \sum_{i=0}^{n} \phi_i(t) S_i(t)$$

The *portfolio* or trading strategy ϕ, is called *self-financing* when

$$V(\phi, t) = V(\phi, 0) + \sum_{i=0}^{n} \int_0^t \phi_i(s) dS_i(s), \quad t \in [0, \Theta], \tag{1.17}$$

where $\int \phi_i(s) dS_i(s)$ denote Itô integrals. Hence, a self-financing trading strategy is a trading strategy that requires nor generates funds between time 0 and time Θ.

1.3.1 Martingale Measures, Derivative Securities and Arbitrage

All prices above are interpreted as being given in terms of some a priori given *numeraire*, or monetary basis. Tipically this *numeraire* is the domestic currency like €, but we may, of course, equally express all prices denominated in some other *numeraire*. In fact, any asset which has strictly positive prices for all $t \in [0, \Theta]$ is a *numeraire*.

Suppose that, for some $p \leq n$, the p-asset is a *numeraire* . The prices of other assets $i \neq p$ denominated in S_p are called the *relative prices* or *discounted prices* and we denote them by

$$\widetilde{S}_i := S_i / S_p.$$

We denote the *relative value process* as well by

$$\widetilde{V} := \frac{V}{S_p} = \sum_{i \neq p}^{n} \phi_i \widetilde{S}_i$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote the probability space from the beginning of this section. Consider now the set that contains all probability measures $\mathbb{Q}^*$ such that:

1. $\mathbb{Q}^* \sim \mathbb{P}$, i.e. both measures have the same null-sets;
2. the relative processes $\widetilde{S}_i$ are martingales under $\mathbb{Q}^*$ for all i, i.e. for $t \leq s$

$$\widetilde{S}_i(t) = \mathbb{E}^{\mathbb{Q}^*}[\widetilde{S}_i(s)|\mathcal{F}_t].$$

The measures $\mathbb{Q}^*$ are called *equivalent martingale measures*. Suppose we pick one particular equivalent martingale measure $\mathbb{Q}^*$.

Definition 12 (Derivative security.) *Is any $\mathcal{F}_t$-measurable random variable $h(T)$ such that*

$$\mathbb{E}^*(|h(T)|) < \infty,$$

where $\mathbb{E}^*$ denotes expectation under the equivalent martingale measure $\mathbb{Q}^*$.

Hence, derivative securities are those assets for which the expectation of the payoff is well defined. If we can find a self-financing trading strategy ϕ such that $\widetilde{V}(\phi, T) = h(T)$ with probability one, the derivative is said to be *attainable*. The self-financing trading strategy is then called a *replicating strategy*. If in an economy *all* derivative securities are attainable, the economy is called *complete*.

An *arbitrage portfolio* is a self-financing trading strategy ϕ, with

$$\mathbb{P}[\widetilde{V}(\phi, T) \geq 0] = 1, \quad \text{with} \quad \widetilde{V}(\phi, 0) < 0,$$

thus, an arbitrage trading strategy is capable to produce a “free lunch”, because with initial negative costs we obtain at terminal time a non-negative value of the portfolio denominated in the chosen numeraire.

Theorem 1 (Unique Equivalent Martingale Measure.) *A continuous trading economy is free of arbitrage trading strategies and every derivative security is attainable, i.e. the market is complete, if for every choice of numeraire there exists a unique martingale measure.*

Proof. See [30]. □

Thus for a given numeraire M with unique martingale measure $\mathbb{Q}^M$, the value of a self-financing trading strategy

$$\widetilde{V}(\phi,t) = \frac{V(\phi,t)}{M(t)}$$

is a $\mathbb{Q}^M$-martingale. Hence, for a replicating strategy ϕ_h that replicates the derivative security $h(T)$ we obtain

$$\mathbb{E}^M\left[\frac{h(T)}{M(T)}\middle|\mathcal{F}_t\right] = \mathbb{E}^M\left[\frac{V(\phi_h,T)}{M(T)}\middle|\mathcal{F}_t\right] = \frac{V(\phi_h,t)}{M(t)}$$

where the last equality follows from the definition of a martingale. Combining the first and the last expression yields

$$V(\phi_h,t) = M(t)\mathbb{E}^M\left[\frac{h(T)}{M(T)}\middle|\mathcal{F}_t\right] \tag{1.18}$$

This formula can be used to determine the value at time $t < T$ for any derivative security $h(T)$. In particular, absence of arbitrage and market completeness implies the existence of the unique probability measure $\mathbb{Q}^B$, equivalent to the physical $\mathbb{P}$, under which the price of any discount bond or T-bond, appropiately discounted by the money-market account $S_0(t) = B(t)$, is a $\mathbb{Q}^B$-martingale.

$$\widetilde{P}(t,T) := \frac{P(t,T)}{B(t)} = \mathbb{E}^B\left[\frac{P(T,T)}{B(T)}\middle|\mathcal{F}_t\right] = \mathbb{E}^B\left[e^{-\int_0^T r(u)du}P(T,T)\middle|\mathcal{F}_t\right]$$

Combining this fact with the fact that a T-bond is a derivative security which has price 1 at its maturity we can write the well-known *arbitrage-free pricing* formula

$$P(t,T) = \mathbb{E}^B\left[e^{-\int_t^T r(s)ds}\middle|\mathcal{F}_t\right], \tag{1.19}$$

where we have used that $B(t)$ is $\mathcal{F}_t$-measurable. Let us now introduce a convenient definition.

Definition 13 (Stochastic discount factor.) *The stochastic discount factor* $D(t,T)$ *is given by*

$$D(t,T) = \frac{B(t)}{B(T)} = \exp\left(-\int_t^T r(u)\,du\right) \tag{1.20}$$

Chapter 2

Forward Rate Models

In the nineties, Heath, Jarrow and Morton (henceforth HJM) [31] proposed a new framework for modeling the entire forward curve directly.

2.1 The Heath-Jarrow-Morton Framework

The stochastic setup is as shown at the end of the Sect. 1.1.3. We consider that there exists a unique equivalent martingale measure $\mathbb{Q}$, associated to the money-market account. Therefore, the T-bond market is complete and there are no arbitrage strategies. Let W be a q-dimensional $\mathbb{Q}$-Wiener process.

We assume that we are given an $\mathbb{R}$-valued and $\mathbb{R}^q$-valued stochastic process $\alpha = \alpha(\omega, t, T)$ and $\sigma = [\ \sigma_1(\omega, t, T)\ \ldots\ \sigma_q(\omega, t, T)\ ]$, respectively, with $\alpha(\cdot, T)$ and $\sigma(\cdot, T)$ $\mathcal{F}_T$-adapted processes. We also assume that for $0 \leq t < T < \infty$, the forward rate $F(\cdot, T)$ has a stochastic differential which under $\mathbb{Q}$ is given by

$$\begin{cases} dF(t,T) &= \alpha(t,T)dt + \sum_{j=1}^{q} \sigma_j(t,T)dW_j(t) \\ F(0,T) &= F^o(0,T). \end{cases} \tag{2.1}$$

Note that conceptually equation (2.1) is one stochastic differential in the t-variable for every choice of T. Also note that we use the observed forward rate curve $T \mapsto F^o(0, T)$ as the initial condition. This will automatically give us a perfect fit between observed and theoretical T-bond prices at $t = 0$, thus relieving us of the task of inverting the term structure of discount bonds.

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

$$P(t,T) = e^{-Z(t,T)} \tag{2.2}$$

where

$$Z(t,T) = \int_t^T F(t,s)ds \tag{2.3}$$

we have in fact specified the entire term structure of discount bonds

$$\{P(t,T);\ T \geq 0,\ 0 \leq t \leq T\}.$$

We now show how bond price dynamics are induced by a given specification of the forward rate dynamics. By using Itô Lemma in (2.2), we have

$$dP(t,T) = -P(t,T)dZ(t,T) + \frac{1}{2}P(t,T)\left(dZ(t,T)\right)^2, \tag{2.4}$$

and it remains to compute $dZ(t,T)$. We have

$$dZ(t,T) = d\left(\int_t^T F(t,s)ds\right)$$

and this is a situation that is not covered by the standard Itô formula. Let us guess the answer.

Proposition 1 *Consider for $s \in [t_0, T]$ the Itô process defined by*

$$dF(t,s) = \alpha(t,s)dt + \sum_{j=1}^{q} \sigma_j(t,s)dW_j(t) \tag{2.5}$$

with $t \in [t_0, s]$. Then, the dynamics for the stochastic process (2.3) is

$$dZ(t,T) = \left[\left(\int_t^T \alpha(t,s)ds\right) - F(t,t)\right]dt + \sum_{j=1}^{q}\int_t^T \sigma_j(t,s)ds\ dW_j(t). \tag{2.6}$$

Proof. See Appendix A. □

Therefore, by substituting in equation (2.4)

$$\begin{aligned} dP(t,T) &= P(t,T)\Big\{\Big[r(t) - \textstyle\int_t^T \alpha(t,s) + \frac{1}{2}\sum_{j=1}^{q}\left(\int_t^T \sigma_j(t,s)ds\right)^2\Big]dt \\ &- \textstyle\sum_{j=1}^{q}\left(\int_t^T \sigma_j(t,s)ds\right)dW_j(t)\Big\} \end{aligned} \tag{2.7}$$

which may be summarized as the the following:

Corollary 1 *The $\mathbb{Q}$-dynamics for the T-bond price, $P(t,T)$, follows the stochastic differential equation*

$$dP(t,T) = P(t,T)\left[\left(r(t) + A(t,T) + \frac{1}{2}\|S(t,T)\|^2\right)dt + S(t,T)dW(t)\right], \quad (2.8)$$

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

$$\begin{array}{rcl} A(t,T) & := & -\int_t^T \alpha(t,s)\,ds \\ S_j(t,T) & := & -\int_t^T \sigma_j(t,s)\,ds, \end{array}$$

and we have used the matrix notations

$$\begin{array}{rcl} S(t,T) & = & [\ S_1(t,T)\quad S_2(t,T)\quad \dots\quad S_q(t,T)\ ] \\ W(t) & = & [\ W_1(t)\quad W_2(t)\quad \dots\quad W_q(t)\ ]^T. \end{array}$$

2.1.1 Absence of Arbitrage

Theorem 2 (HJM Drift Condition) *Assume that the family of forward rates is given by (2.1) and that the induced bond market is arbitrage free. Under the martingale measure $\mathbb{Q}$, the process α and σ must satisfy the following relation, for every t and every $T \geq t$.*

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,s)^T\,ds. \quad (2.9)$$

Proof. Since we are modeling the market under the equivalent martingale measure, $\mathbb{Q}$, the discounted T-bond price $\widetilde{P}(t,T)$ have to be a local $\mathbb{Q}$-martingale satisfying the following differential:

$$d\widetilde{P}(t,T) = \widetilde{P}(t,T)S(t,T)dW(t)$$

We now look for the corresponding differential of the ordinary discount bond price, $P(t,T) = B(t)\widetilde{P}(t,T)$. From the Itô Lemma we know that

$$dP(t,T) = P(t,T)\left(r(t)dt + S(t,T)dW(t)\right), \quad (2.10)$$

in other words, as $\mathbb{Q}$ is a martingale measure with the money account B as numeraire, the local rate of return of every asset price under $\mathbb{Q}$ equals the short rate. We thus have

$$A(t,T) + \frac{1}{2}\|S(t,T)\|^2 = 0.$$

Taking the T-derivative of this equation gives us the relation (2.9) □

2.2 From HJM to Short-Rate Models

What is the interplay between the short-rate dynamics and the present HJM framework? Let us consider the simplest one-dimensional HJM model: a constant $\sigma(t,T) \equiv \sigma > 0$. Then, we have under the risk-neutral measure $\mathbb{Q}$

$$dF(t,T) = \sigma^2(T-t)dt + \sigma dW(t),$$

which implies by direct integration

$$F(t,T) = F(0,T) + \frac{\sigma^2}{2}T^2 + \sigma W(T).$$

Hence for the short rates we obtain

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

and taking the differentials

$$dr(t) = (\partial_t F(0,t) + \sigma^2 t)dt + \sigma dW(t).$$

The observant reader may identify it with the Ho and Lee model [32]. The main inputs into the HJM framework are the forward rate volatility processes $\sigma_j(t,T)$, and as we have shown the Ho and Lee model is a special case of the general 1-factor HJM framework, corresponding to a particular choice of the volatility process. However, it has remained unclear whether other short-rate models could be derived within the HJM framework, and whether there exists a systematic approach for generating the short-rate models. In general, we have the following:

Proposition 2 *Suppose that $F(0,T)$, $\alpha(t,T)$ and $\sigma(t,T)$ are differentiable in T with $\int_0^T |\partial_u F(0,u)|\, du < \infty$.*

Then the short-rate process is an Itô process of the form

$$dr(t) = \zeta(t)dt + \sigma(t,t)dW(t), \tag{2.11}$$

where

$$\zeta(t) = \alpha(t,t) + \partial_t F(0,t) + \int_0^t \partial_t \alpha(s,t)\, ds + \int_0^t \partial_t \sigma(s,t) dW(s) \tag{2.12}$$

Proof. See Appendix A. □

Remark 1 *For every forward rate model, the arbitrage free price of a derivative security, with T-payoff $h(T)$, will still be given by the general pricing formula*

$$V(h,t) = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^T r(u)\, du} h(T) \Big| \mathcal{F}_t \right],$$

where the short-rate as usual is given by $r(t) = F(t,t)$.

2.3 Forward Measures

Equation (1.18) shows to calculate the arbitrage free price $V(t)$, of a derivative security. The value calculated must, of course, be independent of the choice of numeraire. Consider two numeraires M and N with martingale measures $\mathbb{Q}^M$ and $\mathbb{Q}^N$. Combining the result (1.18) applied to both numeraires yields

$$M(t)\mathbb{E}^M \left[\frac{h(T)}{M(T)} \Big| \mathcal{F}_t \right] = N(t)\mathbb{E}^N \left[\frac{h(T)}{N(T)} \Big| \mathcal{F}_t \right]$$

This expression can be rewritten as

$$\mathbb{E}^M \left[g(T) \Big| \mathcal{F}_t \right] = \mathbb{E}^N \left[g(T) \frac{M(T)/M(t)}{N(T)/N(t)} \Big| \mathcal{F}_t \right] \tag{2.13}$$

where $g(T) = h(T)/M(T)$. Since, h, M and N are general, this result holds for all random payoffs g and all numeraires M and N.

We have now derived a way to express the expectation $g(T)$ under the measure $\mathbb{Q}^M$ in terms of an expectation under the measure $\mathbb{Q}^N$.

Theorem 3 (Change of Numeraire) *Let $\mathbb{Q}^M$ be the equivalent martingale measure with respect to the numeraire $M(t)$. Let $\mathbb{Q}^N$ be the equivalent martingale measure with respect to the numeraire $N(t)$. The Radon-Nikodym derivative that changes the equivalent martingale measure $\mathbb{Q}^M$ into $\mathbb{Q}^N$ is given by*

$$\lambda(t) = \frac{d\mathbb{Q}^M}{d\mathbb{Q}^N} = \frac{M(T)/M(t)}{N(T)/N(t)}.$$

The *Change of Numeraire Theorem* is very powerful in the context of pricing interest rate derivatives. Instead of using the value of the money-market account $B(t)$ as a numeraire, the prices of T-bonds can also be used as a numeraire. A very convenient choice is to use the discount bond with maturity T as a numeraire for derivatives which have a payoff $h(T)$ at time T. Assume, on the other hand, that the probability measure $\mathbb{Q}^T$ associated to the numeraire $P(t,T)$ actually exists. Hence, we can apply the *Change of Numeraire Theorem* as follows. Under the measure $\mathbb{Q}^T$ the prices $V(h,t)/P(t,T)$ are martingales for $t < T$. Therefore, applying the definition of a martingale and taking into account that $P(T,T) = 1$, we obtain

$$V(h,t) = P(t,T)\mathbb{E}^T\left[h(T)\middle|\mathcal{F}_t\right] \tag{2.14}$$

The measure $\mathbb{Q}^T$ has another very interesting property, which virtually gave the name T-forward measure. Under the T-forward measure, the instantaneous forward rate, $F(t,T)$ is equal to the expected of the spot interest rate at time T. In formulas

$$F(t,T) = \mathbb{E}^T\left[r(T)\middle|\mathcal{F}_t\right],$$

e.g., see the straightforward arguments followed by Filipovič in [27, Sect. 7.1] or Björk in [5, Sect. 19.4.2]. Note that in this case, the corresponding Radon-Nikodym derivative that changes the T-forward measure $\mathbb{Q}^T$ into the risk-neutral measure (or money-market measure) $\mathbb{Q}$, is

$$\frac{d\mathbb{Q}^T}{d\mathbb{Q}} = \frac{P(T,T)/P(t,T)}{B(T)/B(t)} = \frac{D(t,T)}{P(t,T)} = \frac{e^{-\int_t^T r(s)\,ds}}{P(t,T)}. \tag{2.15}$$

2.4 The General HJM Gaussian Model

Consider a general HJM model under the risk-neutral measure $\mathbb{Q}$ specified by (2.1). We also assume that

$$\sigma(t,T) = [\,\sigma_1(t,T)\ \ldots\ \sigma_q(t,T)\,]$$

are deterministic functions of (t,T), and hence forward rates $F(t,T)$ are Gaussian distributed. We consider now a European call option, with expiration date T and exercise price K, on an underlying bond with maturity S (where of course $T < S$). The following general pricing formula may be derived:

Proposition 3 (Bond Option Pricing for Gaussian Forward Rates) *The price, at $t = 0$ of the bond option*

$$h(T) = (P(T,S) - K)^+$$

is given by

$$V(h,0) = P(0,S)N(d_+) - KP(0,T)N(d_-) \tag{2.16}$$

where

$$\begin{aligned} d_\pm &:= \frac{\log\left(\frac{P(0,S)}{KP(0,T)}\right) \pm \frac{1}{2}\vartheta^2(T,S)}{\vartheta(T,S)}, \\ \vartheta^2(T,S) &:= \int_0^T \|\varsigma(u;T,S)\|^2 \, du \,; \end{aligned} \tag{2.17}$$

and,

$$\varsigma(t;T,S) := S(t,S) - S(t,T) = -\int_T^S \sigma(t,s)\, ds. \tag{2.18}$$

Proof. Let us start with the fundamental arbitrage-free equation

$$V(h,0) = \mathbb{E}\left[D(0,T)(P(T,S) - K)^+\right],$$

where we are taking the expectations with respect the equivalent martingale measure $\mathbb{Q}$ associated to the money-market numeraire $B(\cdot)$. We decompose it as follows

$$V = \mathbb{E}\left[D(0,T)P(T,S)\mathbb{1}_{\{P(T,S)\geq K\}}\right] - K\mathbb{E}\left[D(0,T)\mathbb{1}_{\{P(T,S)\geq K\}}\right] \tag{2.19}$$

In this case, the Radon-Nikodym derivative that changes S-forward measure $\mathbb{Q}^S$ into the money-market measure $\mathbb{Q}$ will be given by

$$\lambda^S(T) = \frac{d\mathbb{Q}^S}{d\mathbb{Q}} = \frac{P(T,S)/P(0,S)}{B(T)/B(0)} = \frac{D(0,T)P(T,S)}{P(0,S)}.$$

In a similar way note that

$$\lambda^T(T) = \frac{d\mathbb{Q}^T}{d\mathbb{Q}} = \frac{P(T,T)/P(0,T)}{B(T)/B(0)} = \frac{D(0,T)}{P(0,T)},$$

is the conversion factor responsible for changing the T-forward measure $\mathbb{Q}^T$ into the risk-neutral world $\mathbb{Q}$. Substituting into decomposition (2.19), and combining with the measurability at $t=0$ of $P(0,S)$ and $P(0,T)$ we have

$$V = \mathbb{E}\left[P(0,S)\lambda^S(T)\mathbb{1}_{\{P(T,S)\geq K\}}\right] - K\mathbb{E}\left[P(0,T)\lambda^T(T)\mathbb{1}_{\{P(T,S)\geq K\}}\right]$$

$$V = P(0,S)\mathbb{Q}^S\left(P(T,S)\geq K\right) - KP(0,T)\mathbb{Q}^T\left(P(T,S)\geq K\right)$$

Now we have the value V for the call option in terms of the forward measures $\mathbb{Q}^S$ and $\mathbb{Q}^T$. Let us start with the probability computations referred to the T-forward measure. Note that the probability may be written as

$$\mathbb{Q}^T(P(T,S)\geq K) = \mathbb{Q}^T\left(\frac{P(T,S)}{P(T,T)}\geq K\right) = \mathbb{Q}^T\left(\log\frac{P(T,S)}{P(T,T)}\geq \log K\right)$$

Consider the “discounted” process

$$X_{S,T}(t) := \frac{P(t,S)}{P(t,T)},$$

with terminal value $X_{S,T}(T) = P(T,S)/P(T,T)$. By taking differentials under the risk-neutral measure $\mathbb{Q}$ we have

$$\begin{aligned}
d\left(\frac{P(t,S)}{P(t,T)}\right) &= \frac{1}{P(t,T)}dP(t,S) - \frac{P(t,S)}{(P(t,T))^2}dP(t,T) + dP(t,S)\cdot d\left(\frac{1}{P(t,T)}\right) = \\
&= \{\dots\}dt + X\left\{(S(t,S)-S(t,T))\,dW(t)\right\} \\
dX &= \{\dots\}dt + X\varsigma(t;T,S)dW(t).
\end{aligned}$$

For the second stage we have used equation (2.10) applied to the discount bonds $P(t,S)$ and $P(t,T)$. Recall that $\mathbb{Q}^T$ is a martingale measure and the multidimensional Girsanov's Theorem[1] which locally induces the change into this T-forward measure, does not affect the difussion coefficient of the initially taken differential. Therefore we have

$$dX_{T,S}(t) = X_{T,S}(t)\varsigma(t;T,S)dW^T(t).$$

Let us introduce the auxiliary process:

$$Y_{T,S}(t) = \log X_{T,S}(t)$$

[1] See [43] for a detailed discussion of it.

By means of the multidimensional Itô Lemma, it is not difficult to prove that the random variable $Y_{T,S}(T)$ distributes like

$$Y_{T,S}(T) \sim \mathcal{N}\left(\log \frac{P(0,S)}{P(0,T)} - \frac{1}{2}\vartheta^2(T,S), \vartheta^2(T,S)\right),$$

where $\vartheta^2(T,S) = \int_0^T \|\varsigma(u;T,S)\|^2 \, du$. Now the computation of the probability under the T-forward measure is straightforward:

$$\mathbb{Q}^T(P(T,S) \geq K) = \mathbb{Q}^T(Y_{T,S}(T) \geq \log K) = N(d_-)$$

For the pending probability $\mathbb{Q}^S$, first, note the following:

$$\mathbb{Q}^S(P(T,S) \geq K) = \mathbb{Q}^S\left(\frac{P(T,T)}{P(T,S)} \leq \frac{1}{K}\right) = \mathbb{Q}^S\left(\log \frac{P(T,T)}{P(T,S)} \leq -\log K\right).$$

It is enough to introduce the auxiliary processes,

$$W_{T,S}(t) := \frac{P(t,T)}{P(t,S)},$$

and,

$$Z_{T,S}(t) := \log W_{T,S}(t),$$

for concluding that $Z_{T,S}(T)$ distributes like

$$Z_{T,S}(T) \sim \mathcal{N}\left(\log \frac{P(0,T)}{P(0,S)} - \frac{1}{2}\vartheta^2(T,S), \vartheta^2(T,S)\right),$$

and then

$$\mathbb{Q}^S(Z_{T,S}(T) \leq -\log K) = N(d_+).$$

□

Corollary 2 *The price at $t = 0$ of the put option*

$$h(T) = (K - P(T,S))^+$$

is given by

$$\Pi(h,0) = KP(0,T)N(-d_-) - P(0,S)N(-d_+) \tag{2.20}$$

where the quantities $d_\pm$ are completely determined by the identities (2.17) to

(2.18).

Proof. First consider the difference between the call and the put option at time $t = 0$. Under the risk-neutral martingale measure we know

$$\begin{aligned} V - \Pi &= \mathbb{E}\left[D(0,T)\left\{(P(T,S)-K)^+ - (K-P(T,S))^+\right\}\right] \\ &= \mathbb{E}\left[D(0,T)(P(T,S)-K)\right] \end{aligned}$$

By equation (1.19) we have

$$\mathbb{E}\left[D(0,T)K\right] = KP(0,T).$$

However, we have the problem of the correlation between the discounting factor and the payoff factor for the first term

$$\mathbb{E}\left[D(0,T)P(T,S)\right].$$

We can circumvent this problem by using the *Change of Numeraire Theorem* in an identical way to that shown in Proposition 3. Recall first that the likelihood

$$\lambda^S(T) = \frac{d\mathbb{Q}^S}{d\mathbb{Q}} = \frac{P(T,S)/P(0,S)}{B(T)/B(0)} = \frac{D(0,T)P(T,S)}{P(0,S)},$$

induces the change of the S-forward measure into the risk-neutral measure. Thus we have:

$$\mathbb{E}\left[D(0,T)P(T,S)\right] = \mathbb{E}\left[\lambda^S(T)P(0,S)\right] = \mathbb{E}^S\left[P(0,S)\right],$$

and then the *Put-Call Parity Relation*:

$$V - \Pi = P(0,S) - KP(0,T), \tag{2.21}$$

is finally inferred. □

Chapter 3

Interest Rate Caps

Interest rate caps are widely traded OTC interest rate derivatives. An interest rate cap is a financial insurance which protects you from having to pay more than a predetermined rate, therefore, a cap is insurance against rising interest rates.

3.1 The Market Practice for Plain Vanilla Caps

In this section we discuss vanilla interest rate caps and the market practice for quoting these instruments. For concreteness suppose the underlying rate is the simply-compounded forward LIBOR rate $L(t;\cdot,\cdot)$ introduced in Sect. 1.1. Let suppose that we are standing at time $t = 0$. We consider a fixed set of increasing maturities $x_0,\ x_1,\ldots,x_n$ and we define τ_j, by

$$\tau_j = x_j - x_{j-1}, \quad j = 1, ..., N.$$

The number τ_j is known as the **tenor**.

Definition 14 *We let $P_j(t)$ denote the discount bond price $P(t,x_j)$ and let $L_j(t)$ summarize the notation for a forward LIBOR rate of the type $L(t;x_{j-1},x_j)$, i.e.*

$$L_j(t) = \frac{1}{\tau_j}\left(\frac{P_{j-1}(t)}{P_j(t)} - 1\right) \quad j = 1,\ldots,n.$$

Recall that a *vanilla cap* with *cap rate* K and *resettlement dates* $x_0,\ldots,x_n$ is a contrat which each time x_j gives the holder of the contract the amount

$$h_{\gamma_j}(x_j) = \tau_j(L_j(x_{j-1}) - K)^+, \tag{3.1}$$

where $j = 1, \ldots, n$. In fact, the cap is a strip of *caplets*. Note that the forward LIBOR rate $L_j(x_{j-1})$ above is in fact the simply-compounded spot LIBOR interest rate. By definition:

$$L_j(x_{j-1}) := L(x_{j-1}; x_{j-1}, x_j) = L(x_{j-1}, x_j),$$

which is observed already at time x_{j-1}. The payoff h_{γ_j} is determined at the reset date x_{j-1} but not payed out until the settlement date x_j. We also note that the caplet γ_j is a call option on the on the underlying spot rate.

For a very long time, the market practice has been to value caps by using a formal extension of the Black model [14]. This extension is typically obtained by an approximation argument where the short rate at one point in the argument is assumed to be deterministic, while later on in the argument the LIBOR rate is assumed to be stochastic. This is of course logically inconsistent.

Definition 15 (Black's Formula for Caplets.) *The Black-76 formula for the j-caplet with payoff:*

$$h_{\gamma_j}(x_j) = \tau_j (L_j(x_{j-1}) - K)^+,$$

at time $t = 0$ is given by the expression

$$\gamma_j(h_{\gamma_j}, 0) = \tau_j P_j(0) \left\{ L_j(0) N(d_1) - K N(d_2) \right\}, \quad j = 1, \ldots, n, \tag{3.2}$$

where

$$d_1 = \frac{\log\left(\frac{L_j(0)}{K}\right) + \frac{1}{2}\sigma_j^2 x_j}{\sigma_j \sqrt{x_j}}, \tag{3.3}$$

$$d_2 = d_1 - \sigma_j \sqrt{x_j}. \tag{3.4}$$

The constant σ_j is known as the *Black volatility* for the j-caplet, γ_j. In the market, cap prices are not quoted in monetary terms but instead in terms of *implied Black volatilities* and these volatilities can furthermore be quoted as *flat volatilities* or as *forward rate volatilities* but, in this work, we confine ourselves to *flat volatilities*. Suppose we are standing at time $t = 0$ and consider the fixed set of dates $x_0,\ x_1, \ldots, x_n$ where $x_0 \geq 0$, and a fixed cap rate K. We assume that, for each $j = 1, \ldots, n$, there is a traded cap with resettlement dates $x_0, x_1, \ldots, x_j$, and we denote the corresponding observed market price by C_j^o. From this data we can easily compute the market prices for the corresponding caplets by means of the recursion formula

$$\gamma_j^o = C_j^o - C_{j-1}^o, \quad j = 1, \dots, n \tag{3.5}$$

with the convention $C_0^o = 0$. Moreover, note that given market data for caplets we can easily compute the corresponding market data for caps by solving the previous recursion formula (3.5):

$$C_j = \sum_{j=1}^{n} \gamma_j, \quad j = 1, \dots, n \tag{3.6}$$

Given market price data as above, the implied Black flat volatilities are defined as follows.

Definition 16 *The implied flat volatilities $\bar{\sigma}_1, \dots, \bar{\sigma}_n$ are defined as the solutions of the equations:*

$$C_j^o = \sum_{k=1}^{j} \gamma_k^o(\bar{\sigma}_j), \quad j = 1, \dots, n, \tag{3.7}$$

In other words, the flat volatility $\bar{\sigma}_j$ is the volatility implied by the Black formula if you use the same volatility for each caplet, in the cap with maturity x_j.

3.1.1 IRS and At-The-Money Plain Vanilla Caps

An interest rate swap (henceforth IRS) is a scheme where you exchange a payment stream at a fixed rate of interest for a payment stream at a floating rate (e.g. LIBOR). A payer IRS settled in arrears is specified by:

- a number of future dates $x_0 < x_1 < \cdots < x_n$ where $x_j - x_{j-1} \equiv \tau_j$ are the settlement periods and x_n is called the maturity of the swap,
- a fixed rate K; and,
- a nominal value N.

Moreover, plain vanilla IRS satisfy the equidistance condition for the settlement periods; i.e., $\tau \equiv \tau_j$. We recall that cash flows take place just at the settlement dates $x_1, x_2, \dots, x_n$. At this dates, the holder of such an IRS, pays a predetermined amount

$$K\tau N$$

and receives in turn the floating payout

$$L_j(x_{j-1})\tau N.$$

The net cash at x_j is therefore

$$\{L_j(x_{j-1}) - K\}\,\tau N.$$

Without loss of generality we set the notional $N = 1$ and $t = 0$ with $x_0 > 0$. By means of the fundamental arbitrage free pricing formula, starting from the money-market martingale measure $\mathbb{Q}$ we can compute the value of this contract as:

$$\begin{aligned}\Pi_{sw} &= \sum_{j=1}^{n} \mathbb{E}\left[D(0,x_j)\left\{L_j(x_{j-1}) - K\right\}\tau\right] \\ &= \tau\left(\sum_{j=1}^{n} \mathbb{E}\left[D(0,x_j)L_j(x_{j-1})\right] - K\sum_{j=1}^{n} P_j(0)\right)\end{aligned} \tag{3.8}$$

where we have used the well-known definition

$$P_j(0) = \mathbb{E}\left[D(0,x_j)\right].$$

Let us consider the first term

$$\mathbb{E}\left[D(0,x_j)L_j(x_{j-1})\right].$$

Fist of all, we may change the risk-neutral measure $\mathbb{Q}$ by means of the *Change of Numeraire Theorem* into the more suitable x_j-forward measure $\mathbb{Q}^{x_j}$. Therefore, we have

$$\Pi_{sw} = \tau\left(\sum_{j=1}^{n} P_j(0)\mathbb{E}^{x_j}\left[L_j(x_{j-1})\right] - K\sum_{j=1}^{n} P_j(0)\right) \tag{3.9}$$

The following result has a crutial role for concluding.

Lemma 1 *For every $j = 1,\ldots,n$, the LIBOR process $L_j(t)$ is a martingale under the corresponding forward measure $\mathbb{Q}^{x_j}$, on the interval $[0, x_{j-1}]$.*

Proof. From Definition 6[1] for the simply-compounded LIBOR forward interest rate $L_j(t)$, we have:

$$\tau L_j(t) := \frac{P_{j-1}(t)}{P_j(t)} - 1.$$

We recall that the process

$$P_{j-1}(t)/P_j(t)$$

is the price of the x_{j-1}-bond in terms of the strictly positive asset $P_j(t)$, which is, by definition, the numeraire for the forward measure $\mathbb{Q}^{x_j}$. The process $P_{j-1}(t)/P_j(t)$ is thus trivially a $\mathbb{Q}^{x_j}$-martingale on the interval $[0, x_{j-1}]$, where the normalized process is well defined. Therefore, $L_j(t)$ is also a $\mathbb{Q}^{x_j}$-martingale on the same interval. □

By using the previous Lemma, we have

$$L_j(t) = \mathbb{E}^{x_j}\left[L_j(s)\middle|\mathcal{F}_t\right] \quad 0 \le t \le s \le x_{j-1};$$

and, in particular,

$$L_j(0) = \mathbb{E}^{x_j}\left[L_j(x_{j-1})\right].$$

By substituting into (3.9):

$$\begin{aligned}
\Pi_{sw} &= \tau\left(\sum_{j=1}^{n} P_j(0)L_j(0) - K\sum_{j=1}^{n} P_j(0)\right) = \\
&= \tau\left[\frac{1}{\tau}\sum_{j=1}^{n} P_j(0)\left(\frac{P_{j-1}(0)}{P_j(0)} - 1\right) - K\sum_{j=1}^{n} P_j(0)\right] = \\
&= \left[\sum_{j=1}^{n}\left(P_{j-1}(0) - P_j(0)\right) - K\tau\sum_{j=1}^{n} P_j(0)\right].
\end{aligned} \tag{3.10}$$

Finally, the total value Π_{sw} at time $t = 0$ is therefore

$$\Pi_{sw} = P_0(0) - P_n(0) - K\tau\sum_{j=1}^{n} P_j(0). \tag{3.11}$$

Proposition 4 (General Closed-Formula for Plain Vanilla IRS.) *The total value $\Pi_{sw}(t)$ of a plain vanilla IRS settled in arrears at time time $t \le x_0$ and*

[1]See Sect. 1.1, p. 7

notional N is

$$\Pi_{sw}(t) = N\left(P_0(t) - P_n(t) - K\tau\sum_{j=1}^{n} P_j(t)\right). \tag{3.12}$$

In contrast to the *interest rate caps* pricing, which depends on the particular choice of the volatility vector process $\sigma(t, T)$ within the HJM framework, the IRS closed-formula is generic. However, note that the plain vanilla interest rate swaps remain dependent on the term structure of discount bonds $T \mapsto P(t, T)$.

Definition 17 (Forward Swap Rate.) *The forward swap rate (also called par swap rate) is the rate $K_{sw}(t)$ at time $t \leq x_0$ which gives the "fair value" $\Pi_{sw}(t) = 0$:*

$$K_{sw}(t) = \frac{P_0(t) - P_n(t)}{\tau\sum_{j=1}^{n} P_j(t)}.$$

Let $t = 0$ again for simplicity and suppose, as above, that $x_0 > 0$, then:

Remark 2 *A plain vanilla cap is said to be at-the-money (ATM henceforth) if*

$$K = K_{sw}(0) = \frac{P_0(0) - P_n(0)}{\tau\sum\limits_{j=1}^{n} P_j(0)}. \tag{3.13}$$

3.2 Caps under The General HJM Gaussian Model

Let us now turn to the problem of rigorously pricing the caplet. Remember that the payoffs on settlement dates $x_1, \ldots, x_n$ are:

$$h_{\gamma_j}(x_j) = \tau_j(L_j(x_{j-1}) - K)^+.$$

Note that such an stream of payoffs for the corresponding strip of caplets γ_j, is equivalent in terms of pricing to those

$$h_{\gamma_j}(x_{j-1}) = \tau_j\mathbb{E}\left[D(x_{j-1}, x_j)(L_j(x_{j-1}) - K)^+\middle|\mathcal{F}_{x_{j-1}}\right],$$

which are received at fixing dates $x_0, \ldots, x_{n-1}$. Because of $\mathcal{F}_{x_{j-1}}$-mesurability of $L_j(x_{j-1})$, they can also be expressed as

$$\begin{aligned} h_{\gamma_j}(x_{j-1}) &= \tau_j(L_j(x_{j-1}) - K)^+\mathbb{E}\left[D(x_{j-1}, x_j)\middle|\mathcal{F}_{x_{j-1}}\right] \\ &= \tau_j P_j(x_{j-1})(L_j(x_{j-1}) - K)^+. \end{aligned} \tag{3.14}$$

From Definition 5 for the simply-compounded LIBOR spot rate $L_j(x_{j-1})$, we know:

$$L_j(x_{j-1}) := \frac{1}{\tau_j}\left(\frac{1}{P_j(x_{j-1})} - 1\right),$$

and after some trivial algebra

$$\begin{aligned} h_{\gamma_j}(x_{j-1}) &= \tau_j P_j(x_{j-1})\left(\frac{1}{\tau_j}\left(\frac{1}{P_j(x_{j-1})} - 1\right) - K\right)^+ \\ &= \left(P_j(x_{j-1})\left(\frac{1}{P_j(x_{j-1})} - 1\right) - K\tau_j P_j(x_{j-1})\right)^+ \\ &= \left(1 - (1+\tau_j K)P_j(x_{j-1})\right)^+ \end{aligned} \tag{3.15}$$

we may finally write the following representation for the stream of payoffs:

$$h_{\gamma_j}(x_{j-1}) = (1+\tau_j K)\left(\kappa - P_j(x_{j-1})\right)^+ \tag{3.16}$$

where $\kappa = (1+\tau_j K)^{-1}$.

Consequently we see that a j-caplet is equivalent to $(1+\tau_j K)$ put options on an underlying x_j-bond, where the exercise date of the option is at x_{j-1} and the exercise price is κ. An entire cap contract can thus be viewed as a portfolio of put options, and we may use the results on Corollary 2 of Sect. 2.4 to compute the theoretical price and, in particular, to price it under the General Gaussian HJM model.

Proposition 5 (Caplet Pricing for Gaussian Forward Rates.) *The price at $t = 0$ of the j-caplet, γ_j, with payoff:*

$$h_{\gamma_j}(x_j) = (L_j(x_{j-1}) - K)^+$$

is given by

$$\gamma_j(h_{\gamma_j}, 0) = (1+\tau_j K)\left\{\kappa P_{j-1}(0)N(-d_-) - P_j(0)N(-d_+)\right\} \tag{3.17}$$

where

$$d_{\pm} := \frac{\log\left(\frac{P_j(0)}{\kappa P_{j-1}(0)}\right) \pm \frac{1}{2}\vartheta^2(0, x_{j-1})}{\vartheta(0, x_{j-1})}\,, \tag{3.18}$$

$$\vartheta^2(0, x_{j-1}) := \int_0^{x_{j-1}} \|\varsigma(u; x_{j-1}, x_j)\|^2\, du\,; \tag{3.19}$$

$$\varsigma(t; x_{j-1}, x_j) := -\int_{x_{j-1}}^{x_j} \sigma(t, s)\, ds. \tag{3.20}$$

and $\kappa = (1 + \tau_j K)^{-1}$.

Chapter 4

Geometric Interest Rate Theory

4.1 The Problem

Any acceptable model which prices interest rate derivatives must fit the observed term structure. This idea pioneered by Ho and Lee [32], has been explored in the past by many other researchers like Black and Karasinski [13] and Hull and White [33].

The contemporary models are more complex because they consider the evolution of the whole forward curve as an infinite system of stochastic differential equations (Heath, Jarrow and Morton [31]). In particular, they use a continuous forward rate curve as initial input. In reality, one only observes a discrete set composed either by bond prices or swap rates. So, in practice, the usual approach is to interpolate the forward curve by using splines or other parametrized families of functions.

A very plausible question arises at this point: Choose a specific parametric family, $\mathcal{G}$, of functions that represent the forward curve, and also an arbitrage free interest rate model $\mathcal{M}$. Assume that we use an initial curve that lay within as input for model $\mathcal{M}$. Will this interest rate model evolve through forward curves that lay within the family? Motivated by this question, Björk and Christensen [11] define the so–called consistent pairs ($\mathcal{M}$, $\mathcal{G}$) as ones whose answer to the above question is positive. In particular, they studied the problem of consistency between the family of curves proposed by Nelson and Siegel [45] and any HJM interest rate model with deterministic volatility, obtaining that there is no such interest model consistent with it.

We remark that the Nelson and Siegel interpolating scheme is an important example of a parametric family of forward curves, because it is widely adopted by central banks (see for instance BIS [4]). Its forward curve shape, $G_{NS}(z,\cdot)$ is given by the expression

$$G_{NS}(z,x) = z_1 + z_2 e^{-z_4 x} + z_3 x e^{-z_4 x},$$

where x denotes time to maturity and z the parameter vector

$$z = [z_1 \; z_2 \; \ldots]^T.$$

Despite all the positive empirical features and general acceptance by the financial community, Filipovič [29] has shown that there is no Itô process that is consistent with the Nelson-Siegel family. In a recent study De Rossi [20] applies consistency results to propose a consistent exponential dynamic model, and estimates it using data on LIBOR and UK swap rates. On the other hand, Buraschi and Corielli [16] add results to theoretical framework indicating that the use of inconsistent parametric families to obtain smooth interest rate curves, violates the standard self financing arguments of replicating strategies, with direct consequences in risk management procedures.

In order to illustrate this situation, we describe a very common fixed-income market procedure. In the real world, practitioners usually re-estimate yield curve and HJM model parameters on a daily basis. This procedure consists of two steps:

- They fit the initial yield curve from discrete market data (bond prices, swap rates, short-term zero rates), and
- They obtain an estimate of the parameters of the HJM model, minimizing the pricing error of some actively traded (plain vanilla) interest rate derivatives (commonly swap options or caps).

In contrast with the parsimonious assumption that model parameters are constant, an unstable HJM model parameter estimation it is often observed. Perhaps, this fact is not relevant for mark to market, but it could have practical consequences on the hedging portfolios associated with these financial instruments. Recall that such dynamic strategies depend on the model assumptions. Thus, re-calibration is conceivable because the practitioners are aware of *model risk*. A particular HJM model is not a perfect description of reality, and they are forced to re-estimate day to day model parameters in order to include new information that arrives from the market. On the other hand, unstable estimates may be caused by reasons that are more theoretical, because the above mentioned set-up does not take into account that HJM model parameters are linked, in general, to the initial yield curve fit parameters. If a practitioner uses an interpolation scheme which is not consistent with the model, then the parameters will be artificially forced to change. Thus, it seems to be that there are a plethora of motivations for the study of the empirical evidence and the practical implications that are predicted by a consistent HJM build model.

4.2 Setup

We consider, as earlier in chapter 2, a given forward rate model under a risk neutral martingale measure $\mathbb{Q}$. We will adopt the Musiela parameterization [42] and use the notation

$$f(t,x) := F(t,t+x).$$

The reasons why this parameterization is better suited to address the main problems that are the subject of the present work have been accurately explained, for instance, by Filipoviȁ [28]. As we will see, this parametrization makes changes in the specification of the stochastic process originally proposed by HJM for the instantaneous forward rate process, $F(t,t+x)$, since now x is not considered any more a fixed variable which parametrizes the dynamics of these rates.

Proposition 6 (The Musiela HJM formulation.) *Under the martingale measure $\mathbb{Q}$ the f-dynamics are given by*

$$\begin{cases} df(t,x) &= \left(\frac{\partial f(t,x)}{\partial x} + \widetilde{\sigma}(t,x)\int_0^x \widetilde{\sigma}(t,u)^T\,du\right)dt + \widetilde{\sigma}(t,x)dW(t) \\ f(0,x) &= f^o(0,x). \end{cases} \tag{4.1}$$

where $\widetilde{\sigma}(t,x) := \sigma(t,t+x)$.

Proof. See Appendix B. □

From now, with a clear abuse of notation we remove the symbol $\tilde{}$ from σ because in the whole chapter we will consider the HJM model under the Musiela parametrization. Thus the interest rate model $\mathcal{M}$ will be characterized by the particular volatility function $\sigma(t,x)$ used in the following f-dynamics:

$$df(t,x) = \left(\partial_x f(t,x) + \sigma(t,x)\int_0^x \sigma(t,u)^T\,du\right)dt + \sigma(t,x)dW(t). \tag{4.2}$$

4.3 The Formalized Problem

4.3.1 The Forward Curve Manifold

Assume that we have a parametrized family of forward rate curves

$$G : \mathcal{Z} \longrightarrow \mathcal{H}, \tag{4.3}$$

with $\mathcal{Z} \subseteq \mathbb{R}^d$ the parameter space. For each parameter value $z \in \mathcal{Z}$ we have a smooth curve $G(z)$. The value of this curve at the point $x \in \mathbb{R}_+$ will be written as $G(z,x)$, so we see that G can also be viewed as the mapping

$$G : \mathcal{Z} \times \mathbb{R}_+ \longrightarrow \mathbb{R}. \tag{4.4}$$

The main problem is to determine under which conditions the f-dynamics given by (4.2) is *consistent* with the parametrized family of forward rate curves (4.3) as follows:

- Assume that, at an arbitrarily chosen time $t = s$, we have fitted a forward curve G to market data, i.e. for some $z^o \in \mathcal{Z}$ we have

$$f^o(s, s+x) = G(z^o, x), \quad \forall x \geq 0\ ,$$

- the *future* forward curves produced by the interest rate model (4.2) always stay within the given forward curve family? In other words, does there exist at every fixed time $t \geq s$ some $z \in \mathcal{Z}$ such that

$$f(t, t+x) = G(z, x), \quad \forall x \geq 0\ ?$$

First, 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 parametrized family.

Definition 18 *The* **forward curve manifold** $\mathcal{G} \subseteq \mathcal{H}$ *is defined as*

$$\mathcal{G} = Im(G).$$

We now move on to give precise mathematical definition of the consistency property discussed above.

Definition 19 (Invariant manifold.) *Take as given the f-dynamics (4.2). Consider also the forward curve manifold $\mathcal{G}$. We say that $\mathcal{G}$ is invariant under the action of f if, for each point $(s, f) \in R_+ \times \mathcal{G}$, the condition $f_s \in \mathcal{G}$ implies that $f_t \in \mathcal{G}$ on a time interval $t - s > 0$.*

The purpose of the following section will be to characterize invariance in terms of local characteristics of both $\mathcal{G}$ and $\mathcal{M}$.

4.3.2 The Space

As the space of forward rate curves we will use a weighted Sobolev space where a generic point will be denoted by f.

Definition 20 *Consider a fixed real number $\gamma > 0$. The space $\mathcal{H}_\gamma$ is defined as the space of all differentiable (in the distributional sense) functions*

$$f : \mathbb{R}_+ \longrightarrow \mathbb{R}$$

satisfying the norm condition $\|f\|_\gamma < \infty$. Here the norm is defined as

$$\|f\|_\gamma^2 = \int_0^\infty f^2(x)e^{-\gamma x}\, dx + \int_0^\infty \left(\frac{df}{dx}(x)\right)^2 e^{-\gamma x}\, dx$$

Intuitively, as a specific Sobolev space, $\mathcal{H}_\gamma$ is a vector space of functions equipped with a norm that is a combination of L^2-norms of the function itself as well as its first derivative. Recall that x is the time to maturity, as defined in Sect. 1.1.

In fact, if we introduce the inner product

$$(f, g) = \int_0^\infty f(x)g(x)e^{-\gamma x}\, dx + \int_0^\infty \left(\frac{df}{dx}(x)\right)\left(\frac{dg}{dx}(x)\right) e^{-\gamma x}\, dx,$$

the space $\mathcal{H}_\gamma$ becomes a Hilbert space as proved by Björk and Landen [7].

4.3.3 The Interest Rate Model

Finally, let us consider as given a volatility function σ of the form

$$\sigma : \mathcal{H}_\gamma \times \mathbb{R}_+ \to \mathbb{R}^q.$$

$\sigma(f, x)$ is thus a functional of the infinite dimensional f-variable, and a function of the real variable x. Denoting the forward curve at time t by f_t we then have the following forward rate equation.

$$df_t(x) = \left\{ \frac{\partial}{\partial x} f_t(x) + \sigma(f_t, x) \int_0^x \sigma(f_t, u)^T \, du \right\} dt + \sigma(f_t, x) \, dW_t. \tag{4.5}$$

4.4 The Invariance Conditions

As we see before, the pair $(\mathcal{M}, \mathcal{G})$ is consistent if and only if the forward curve manifold $\mathcal{G}$ is invariant under the action f, and the question we pursue from now is when it happens. In order to guess the precise answer we have to rewrite the analysis in terms of Stratonovich integrals instead of Itô integrals.

Definition 21 *For given semimartingales X and Y driven by a multidimensional Wiener process, the* **Stratonovich integral** *of X w.r.t Y,*

$$\int_0^t X_s \circ dY(s) \ ,$$

is defined as

$$\int_0^t X_s \circ dY_s := \int_0^t X_s \cdot dY_s + \frac{1}{2} dX_t \cdot dY_t$$

Remark 3 *For computing the "quadratic variation process" $dX_t \cdot dY_t$ the usual "multiplication rules" $dW \cdot dt = dt \cdot dt$, $dW \cdot dW = dt$ must be applied.*

Proposition 7 (Chain rule.) *Assume that the function $F(t, y)$ is smooth. Then we have*

$$dF(t, Y_t) = \frac{\partial F}{\partial t}(t, Y_t) dt + \frac{\partial F}{\partial y}(t, Y_t) \circ dY_t.$$

Note that under the Stratonovich formulation of the stochastic integral, the Itô formula takes the form of the standard chain rule of ordinary calculus. Now, returning to the above f-dynamics (4.5), we can write it in terms of Stratonovich calculus as the following

$$df_t(x) = \left\{ \frac{\partial}{\partial x} f_t(x) + \sigma(f_t, x) \int_0^x \sigma(f_t, u)^T \, du \right\} dt - \frac{1}{2} d\sigma(f_t, x) \cdot dW_t + \sigma(f_t, x) \circ dW_t \ , \tag{4.6}$$

and, as it can be seen, it appears a *quadratic variation term* commonly known as the *Stratonovich correction*. Note that σ is not a function but a functional, however in practical terms we can work with the ordinary Itô formula which is still correct as in the finite dimensional case. Then,

$$\begin{aligned} d\sigma \cdot dW(t) &= \Big(\{\dots\} dt + \sigma'_f(f_t)\sigma(f_t)^T dW_t \Big) \cdot dW_t \\ &= \sigma'_f(f_t)\sigma(f_t)^T dt. \end{aligned} \tag{4.7}$$

where σ'_f denotes the Frechet derivative of σ w.r.t the f-variable. This derivative extends the concept of Jacobian matrix to the infinite dimensional case. Its formal definition, which is somewhat technical, is left out. See [19].

Finally, we may write the Stratonovich formulation of the Musiela equation (4.6) as

$$df_t = \mu(f_t) + \sigma(f_t) \circ dW_t \tag{4.8}$$

where

$$\mu(f_t, x) = \partial_x f_t(x) + \sigma(f_t, x) \int_0^x \sigma(f_t, u)^T \, du - \frac{1}{2} \left[\sigma'_f(f_t) \sigma_f(f_t)^T \right](x). \tag{4.9}$$

Let us consider as given the forward curve manifold $\mathcal{G}$: the relevant concept is the following.

Definition 22 *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 $Z(t)$ with state space $\mathcal{Z}$ and possessing a differential of the form*

$$dZ(t) = \gamma(t, Z(t))dt + \psi(t, Z(t)) \circ dW_t, \tag{4.10}$$

such that, for every fixed choice of initial time s, whenever $y_s(\cdot) \in \mathcal{G}$, the stochastic process defined by

$$y_t(x) = G(Z(t), x), \ \forall t \geq s, \ x \geq 0, \tag{4.11}$$

solves the SDE (4.6) with initial condition $f_s(\cdot) = y_s(\cdot)$.

In fact, the stochastic Z-process is describing how evolves the vector parameter z as the forward rate curve moves on the manifold $\mathcal{G}$.

We assume that the forward rate Itô dynamics of $\mathcal{M}$ are given by (4.5), and that the quadratic variation process may be written in intensity form:

$$-\frac{1}{2}\left[\sigma'_f(f_t)\sigma(f_t)^T\right](x)dt = \phi(t,x)dt$$

Now we can state and prove the main invariance result.

Theorem 4 (Consistency Conditions.) *The forward curve manifold $\mathcal{G}$ is f invariant for the forward rate process $f(t,x)$ in $\mathcal{M}$ iff*

$$G_x(z,\cdot) + \sigma(t,\cdot)\int_0^{\cdot} \sigma(t,u)^T\, du + \phi(t,\cdot) \in Im\left[G_z(z,\cdot)\right], \tag{4.12}$$

$$\sigma(t,\cdot) \in Im\left[G_z(z,\cdot)\right]. \tag{4.13}$$

$\forall (t,z) \in \mathbb{R}_+ \times \mathcal{Z}$. *Here, G_z and G_x denote the Jacobian of G w.r.t. to z and x, provided some minimal smoothness of the mapping G.*

Proof. See Appendix B. □

Condition (4.12) is called *the consistent drift condition* (henceforth CDC) and condition (4.13) is called *the consistent volatility condition* (henceforth CVC). It is said that we have invariance if and only if the latter conditions hold which brings us the following definition.

Definition 23 (Consistency.) *We say that the interest rate model $\mathcal{M}$ is* **consistent** *with the forward rate manifold $\mathcal{G}$ if the CDC and CVC conditions (4.12)-(4.13) prevail.*

4.4.1 Simple Invariance

In order to obtain some geometric intuition, we will analyze a bidimensional deterministic version of our problem as a motivational example. Let us consider:

- first, a deterministic vector function $Q : \mathbb{R}_+ \to \mathbb{R}^2$

$$Q(t) = \begin{bmatrix} Q_1(t) \\ Q_2(t) \end{bmatrix}, \tag{4.14}$$

with differential given by

$$\dot{Q} = \mu(t, Q(t)), \tag{4.15}$$

where $\mu : \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}^2$ is some smooth vector field.

- Next, a smooth mapping $G : \mathcal{Z} \to \mathbb{R}^2$,

$$G(z) = \begin{bmatrix} G_1(z) \\ G_2(z) \end{bmatrix}. \tag{4.16}$$

It could be interpreted that the process Q corresponds to two specific coordinates of the infinite-dimensional object $f(t, x)$ in a purely deterministic world. Say, for instance, picking out the 3-month and 10-year key rates at any time t. Thus, this is our toy model $\mathcal{M}$.

Now define the manifold $\mathcal{G}$ as

$$\mathcal{G} = \{G(z) : z \in \mathcal{Z}\}, \tag{4.17}$$

and assume that $Q(s) \in \mathcal{G}$ for some initial value $z^o \in \mathcal{Z}$, that is

$$q^o = Q(s) = G(Z(s) = z^o), \tag{4.18}$$

where Z is also another deterministic d-dimensional process $Z : \mathbb{R}_+ \to \mathbb{R}^d$. When the relation for the future times $t \geq s$,

$$Q(t) \in \mathcal{G}, \quad \forall\, t \geq s, \tag{4.19}$$

prevails? The answer is geometrically obvious. We have the relation (4.19) iff the velocity vector

$$\frac{dQ}{dt},$$

belongs to the tangent space $T_{Q(t)}(\mathcal{G})$ for each $t \geq s$. Note, that a generic point of $\mathcal{G}$ is written as $q = G(z)$, and the tangent space at this point is given as the span of the tangent vectors

$$\frac{\partial G(z)}{\partial z_i}, \quad i = 1, \ldots, d. \tag{4.20}$$

Let us return to the first order deterministic system written on normal form

$$\dot{Q} = \mu(t, Q(t)).$$

Physically, we may interpret this system as a condition set for the velocity of a particle at position q at time t. Indeed, geometrically, we may think in this context that $\mu : U \to \mathbb{R}^2$, where U is an open set $U \subset \mathbb{R}_+ \times \mathbb{R}^2$, is a velocity field defined over the plane $\mathbb{R}^2$. By definition, the function $Q : I \to \mathbb{R}^2$ is a solution of the aforementioned linear dynamical system whenever the relation

$$\dot{Q} = \mu(t, Q(t)), \quad \forall t \in I \subset \mathbb{R}_+$$

holds. That is, $Q(t)$ is a solution for the differential system iff the trajectory of the particle is tangent to the vector field μ in all its points. Thus, we have related the problem of finding the solution for the differential system with the geometric problem of finding tangent trajectories for the velocity field μ. In fact, the theoretical problem for the infinite-dimensional stochastic case is the same, and the use of the Stratonovich form for the differential is the convenience trick in order to bridge the gap between stochastic differential calculus and ordinary calculus for real variable functions. Consider for instance the concrete bidimensional system in $\mathbb{R}^2$:

$$\begin{cases} \dot{Q}(t) &= [-Q_2(t)\ Q_1(t)]^T \\ Q(0) &= [1\ 0]^T \end{cases} \tag{4.21}$$

For this system the unit circle manifold

$$\mathcal{S}^1 = \{G(z) = [\cos z\ \ \sin z]^T : z \in \mathbb{R}\}$$

is invariant. The deterministic system (4.21) has the exact solution

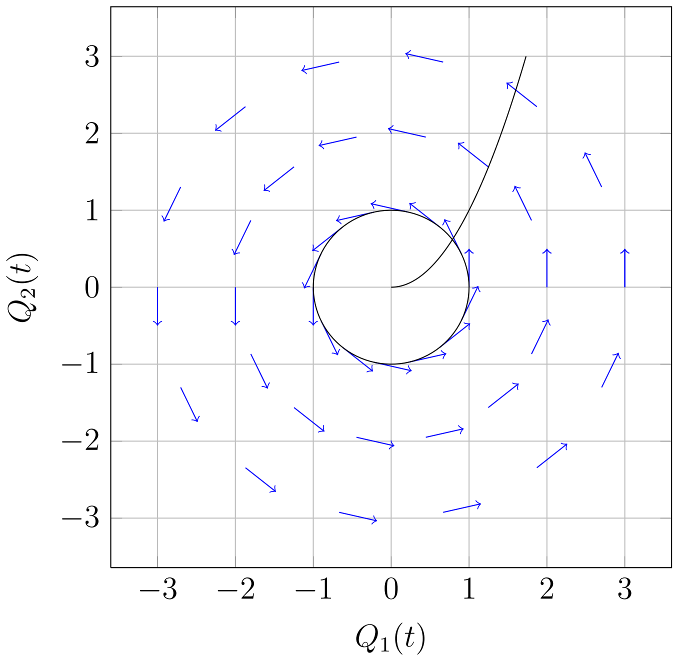
$$Q(t) = [\cos t\ \ \sin t]^T \ ,$$

and if we start the system on $\mathcal{S}^1$ it will stay forever on $\mathcal{S}^1$. In fact, the latter may be easily seen by introducing the trivial one-dimensional deterministic realization $Z(t) = t$, because making such a choice

$$Q(t) = G(Z(t)), \quad \forall t \geq s.$$

Now, it is easy to check the corresponding invariance condition first introduced

Figure 4.1: The vector field from the the system (4.21) with $\mathcal{S}^1$ and a parabollic noninvariant manifold.



in (4.20). Let $G_z(z)$ denote the Frechet derivative (which in turns is the Jacobian in this deterministic and finite-dimensional case) of G at z. The columns of the matrix representation of G_z are the tangent vectors above (4.20), so the tangent space $T_q(\mathcal{S}^1)$ to $\mathcal{S}^1$ at $q = G(z)$ coincides with the image $Im[G_z(z)]$:

$$G_z(z) = [-\sin(z)\ \cos(z)]^T.$$

Recall that the vector field μ at $q = G(z)$ is in turns:

$$\mu(G(z)) = [-\sin(z)\ \cos(z)]^T.$$

We thus trivially have

$$\mu(G(z)) \in Im[G_z(z)],$$

and, in fact, this is the consistency condition we have to check out for pairs ($\mathcal{M}$,$\mathcal{G}$) when $\mathcal{M}$ is an autonomous and deterministic differential system of finite dimension.

Chapter 5

The Hull-White Model and multiobjective calibration

This chapter is the first chapter with empirical and numerical applications[1] of this work. We have shown the foundations of interest rate theory and we have derived the general conditions of consistency. However, only little attention has been devoted to the empirical applications of these concepts. In this chapter we address these applications.

5.1 The Hull-White Model

Our test case is the following model, studied by Hull and White [33] (henceforth HW):

$$dr(t) = [\Phi(t) - ar(t)]\,dt + \sigma dW(t). \tag{5.1}$$

The Hull-White model improve Ho-Lee model incorporating mean-reversion and providing closed formulas for liquid options like interes rate caps. This model is one of the simplest Gaussian HJM models which preserves the Markov property, allowing very efficient numerical methods for the pricing of any kind of options. On the negative side, it does not capture large humps of the term structure of volatilities (TSV hereafter). The model instantaneous-forward volatility curve $T \to \sigma_f(t,T)$, as we will prove later, is monotonically decreasing and this fact often allows only small humps in the caplet curve. However, as pointed out by Brigo and Mercurio [15, pp. 91–92], in case of decreasing TSIR curves, large

[1] Some of the results presented in this chapter are also reported in [25].

humps can be produce even by this model.

Summarizing, it exhibits a relative good performance when it is chosen as a parsimonious solution for bussiness cycles with monotonically decreasing TSV, as it is shown by [3].

5.1.1 Markovianity of the HW model

The short-rate differential process (2.11) in Proposition 2 is not a Markov process in general. Notice in fact that time t appears in the expression (2.12) for the drift both as extreme of integration and inside the integrand function. However, the HW model as a particular Gaussian HJM have a suitable specification of σ for which the short-rate r is indeed a Markov process. This happens because we can write the model volatility function $\sigma(t,T)$ as a separable specification:

$$\sigma(t,T) = \varsigma(t)\varrho(T) \tag{5.2}$$

with $\varsigma(t)$ and $\varrho(T)$ strictly positive and deterministic functions of time. Under such a separable specification, the short-rate process becomes

$$\begin{aligned} r(t) := F(t,t) &= F(0,t) + \textstyle\int_0^t \left(\varsigma(u)\varrho(t) \int_u^t \varsigma(u)\varrho(s)\, ds\right)\, du + \int_0^t \varsigma(s)\varrho(t)\, dW(s) \\ &= F(0,t) + \varrho(t) \textstyle\int_0^t \left(\varsigma^2(u) \int_u^t \varrho(s)\, ds\right)\, du + \varrho(t) \int_0^t \varsigma(s)\, dW(s) \end{aligned} \tag{5.3}$$

Notice that, if we introduce the deterministic function $A(\cdot)$

$$A(t) := F(0,t) + \varrho(t) \int_0^t \left(\varsigma^2(u) \int_u^t \varrho(s) ds\right) du,$$

by taken differentials in (5.3) we can write

$$\begin{aligned} dr(t) &= A'(t) + \varrho'(t) \textstyle\int_0^t \varsigma(s) dW(s) + \varsigma(t)\varrho(t)\, dW(t) \\ &= \left[A'(t) + \varrho'(t)\frac{r(t) - A(t)}{\varrho(t)}\right] dt + \varsigma(t)\varrho(t)\, dW(t) \\ &= [a(t) + b(t)r(t)]\; dt + c(t)\, dW(t) \end{aligned} \tag{5.4}$$

where

$$\begin{aligned} a(t) &:= A'(t) - \frac{\varrho'(t)}{\varrho(t)} A(t), \\ b(t) &:= \frac{\varrho'(t)}{\varrho(t)}\ ; \text{ and,} \\ c(t) &:= \varsigma(t)\varrho(t) = \sigma(t,t). \end{aligned} \tag{5.5}$$

We can finally derive the HJM forward-rate dynamics that is equivalent to the original short-rate dynamics (5.1). To this end, let us set

$$\sigma(t,T) = \sigma e^{-a(T-t)},$$

where a and σ are real constants, so that

$$\begin{aligned} \varsigma(t) &= \sigma e^{at}, \\ \varrho(T) &= e^{-aT}, \\ A(t) &= F(0,t) + \tfrac{\sigma^2}{2a^2}(1-e^{-at})^2; \end{aligned} \tag{5.6}$$

and after some tedious but trivial algebra we have for the HW model:

$$\begin{aligned} a(t) &:= F_t(0,t) + aF(0,t) + \frac{\sigma^2}{2a}(1-e^{-2at}), \\ b(t) &:= -a \\ c(t) &:= \sigma. \end{aligned} \tag{5.7}$$

The resulting short-rate dynamics is then given by

$$dr(t) = \left[F_t(0,t) + aF(0,t) + \frac{\sigma^2}{2a}(1-e^{-2at}) - ar(t) \right] dt + \sigma\, dW(t), \tag{5.8}$$

which is equivalent to (5.1) when combined with the identity

$$\Phi(t) := F_t(0,t) + aF(0,t) + \frac{\sigma^2}{2a}(1-e^{-2at}).$$

In conclusion, we then have that the HW model is a short-rate markovian model that admits a one-factor HJM formulation. Turning back to the Musiela parametrization, we have that the volatility specification for this model is

$$\widetilde{\sigma}(t,x) := \sigma(t,t+x) = \sigma e^{-ax}, \tag{5.9}$$

which falls into the class of HJM models with deterministic volatility

$$\widetilde{\sigma}(f_t,x) = \widetilde{\sigma}(x).$$

5.2 The Nelson and Siegel family and Invariance

In order to illustrate the theoretical ideas shown in the previous chapter, we now move from abstract theory to the investigation of a number of concrete forward curve families and the Hull-White model. More fundamental results can be found in Björk and Christensen [11] in detail. The leading example is the popular forward curve family introduced by Nelson and Siegel [45], first introduced informally in previous chapter. We analyze the consistency of this family and several variations of it with the Hull-White model. We adapt some of the theoretical results to this Gaussian case study without no further technical discussion for the general case. From now, we again remove the symbol ~ as in Chap. 4., i.e. we will consider the HJM model under the Musiela parametrization.

Consider the space $\mathcal{H}_\gamma$ introduced in Sect. 4.3.2

Corollary 3 *Consider as given the mapping*

$$G : \mathcal{Z} \to \mathcal{H}_\gamma$$

where the parameter space $\mathcal{Z}$ is an open connected subset of R^d, $\mathcal{H}_\gamma$ a Hilbert space and the forward curve manifold $\mathcal{G} \subseteq H_\gamma$ is defined as $\mathcal{G} = \mathrm{Im}(G)$. The family $\mathcal{G}$ is consistent with the one-factor model $\mathcal{M}$ with deterministic volatility function $\sigma(\cdot)$, if and only if

$$G_x(z,x) + \sigma(x)\int_0^x \sigma(s)ds \in Im\left[G_z(z,x)\right], \tag{5.10}$$

$$\sigma(x) \in Im\left[G_z(z,x)\right], \tag{5.11}$$

for all $z \in Z$.

Proof. See Theorem 4 in Sect. 4.3.2 for the particular case of deterministic volatility $\sigma(f_t, x) = \sigma(x)$. □

The statements (5.10) and (5.11) are the particular CDC and CVC for the deterministic volatility case. These are easy to apply in concrete cases as shown Björk and Christensen [11] or De Rossi [20], among others.

5.2.1 The NS family

The NS forward curve manifold $\mathcal{G}$ is parametrized by $z \in \mathcal{Z} = \mathbb{R}^4$. Recall that the curve shape $G(z,x)$ is given by the expression

$$G(z,x) = z_1 + z_2 e^{z_4 x} + z_3 x e^{-z_4 x}. \tag{5.12}$$

For $z_4 = 0$ the *Frechet* derivatives $G_z(z,x)$ and $G_x(z,x)$ are easily obtained as

$$\begin{aligned} G_z(z,x) &= [1 \quad e^{-z_4 x} \quad x e^{-z_4 x}]^T, \\ G_x(z,x) &= (z_3 - z_2 z_4 - z_3 z_4 x) e^{-z_4 x}. \end{aligned} \tag{5.13}$$

Henceforth, we write $\mathcal{Z}_{NS} = \{[z_1\ z_2\ z_3\ z_4]^T\ : z_4 \neq 0\}$ for the NS parameter space and $\mathcal{G}_{NS} = G(\mathcal{Z}_{NS})$ for the associated manifold. For the HW model characterized by the volatility function (5.9), the consistency conditions of Theorem 4 become

$$\begin{cases} G_x(z,x) + \frac{\sigma^2}{a}\left[e^{-ax} - e^{-2ax}\right] & \in \ Im[G_z(z,x)], \\ \sigma e^{-ax} & \in \ Im[G_z(z,x)]. \end{cases} \tag{5.14}$$

To investigate whether $\mathcal{G}_{NS}$ is invariant under HW dynamics, we consider now simplest consistency condition, the CVC: let us consider for constants α_i, such $\forall x \geq 0$ we have

$$\sigma e^{-ax} = \alpha_1 + \alpha_2 e^{-z_4 x} + \alpha_3 x e^{-z_4 x} - \alpha_4 (z_2 + z_3 x) x e^{-z_4 x}. \tag{5.15}$$

One can easily see that is is possible iff $z_4 = a$. So as a first hint, let us fix $z_4 = a$ in the parametrization and introduce the restricted NS family:

$$G(z,x) = z_1 + z_2 e^{-ax} + z_3 x e^{-ax}.$$

Now the CVC is verified, while in the CDC we look for β_i such that $\forall x \geq 0$ we have:

$$(z_3 - a z_2 - a z_3 x) e^{-ax} + \frac{\sigma^2}{a}\left(e^{-ax} - e^{-2ax}\right) = \beta_1 + \beta_2 e^{-ax} + \beta_3 x e^{-ax}. \tag{5.16}$$

This equation can never be verified, due to the extra exponential e^{-2ax}, so we have proved the following.

Proposition 8 (Nelson-Siegel and Hull-White.) *The Hull-White model is* *inconsistent with the Nelson-Siegel family.*

Let us include this extra exponential in the parametrization, thus, we finally introduce the augmented NS family:

$$G_{ANS}(z,x) = z_1 + z_2 e^{-ax} + z_3 x e^{-ax} + z_4 e^{-2ax}.$$

Now, both CDC and CVC are verified for this family.

Proposition 9 (Augmented Nelson-Siegel and Hull-White.) *The augmented Nelson-Siegel family is consistent with the Hull-White model.*

Proof. The Frechet derivatives are in this case

$$\begin{aligned} \frac{\partial G_{ANS}}{\partial z}(z,x) &= [1 \quad e^{-ax} \quad xe^{-ax} \quad e^{-2ax}]^T, \\ \frac{\partial G_{ANS}}{\partial x}(z,x) &= [z_3 - a\,(z_2 + z_3 x)]\, e^{-ax} - 2az_4 e^{-2ax}. \end{aligned} \tag{5.17}$$

Now the set $Im[\partial_z G_{ANS}]$ is "large" enough to trivially satisfy CVC and CDC due to the extra component e^{-2ax}. The derivation of the balancing equations analogous to (5.15) and (5.16) are left to the meticulous reader. □

5.3 The Minimal Consistent family and Realizations of Gaussian Models

It should be also noted that $\sigma(x)$ is a one dimension *quasi-exponential* function (QE for short), because is of the form

$$f(x) = \sum_i e^{\lambda_i x} + \sum_i e^{\alpha_i x}[p_i(x)\cos(\omega_i x) + q_i(x)\sin(\omega_i x)], \tag{5.18}$$

with $\lambda_i, \alpha_i, \omega_i$ being real numbers and p_i, q_i are real polynomials.

If $f(x)$ is a q-dimensional QE function, then it admits the following matrix representation

$$f(x) = ce^{Ax}B, \tag{5.19}$$

where A is a $(n \times n)$-matrix, B is a $(n \times q)$-matrix and c is a n–dimensional row vector, see Björk [6, Lemma 2.1, p. 13]. Thus, $\sigma(x)$ can be written as

$$\begin{aligned} \sigma(x) &= ce^{Ax}b, \text{ where} \\ c &= 1, \\ A &= -a, \\ b &= \sigma. \end{aligned} \tag{5.20}$$

We can write the forward rate equation (4.2) following Björk [6, Proposition 2.1, pp. 8–9]

$$dq_t(x) = \mathbf{F}q_t(x)\,dt + \sigma(x)\,dW_t, \quad q_0(x) = 0 \tag{5.21}$$

$$f_t(x) = q_t(x) + \delta_t(x), \tag{5.22}$$

here $\mathbf{F}$ is a linear operator that is defined by

$$\mathbf{F} = \frac{\partial}{\partial x},$$

and $\delta_t(x)$ is the deterministic process given by

$$\delta_t(x) = f^o(x+t) + \int_0^t \Sigma(x+t-s)\,ds,$$

with

$$\Sigma(x) = \sigma(x)\int_0^x \sigma(s)\,ds.$$

Moreover, $q_t(x)$ has the concrete *finite dimensional* realization

$$dZ_t = -aZ_t\,dt + \sigma\,dW_t, \quad Z_0 = 0, \tag{5.23}$$

$$q_t(x) = e^{-ax}Z_t, \tag{5.24}$$

as a particular result from [9, Definition 2.1, p. 7] with the fundamental concluding remark derived by Björk for QE deterministic volatilities in [9, Proposition 2.3, p. 13]. The SDE (5.23) is linear in the narrow sense [39], with explicit solution

$$Z_t = \sigma e^{-at}\int_0^t e^{as}\,dW_s, \tag{5.25}$$

Now, with the definition of $S(x) = \int_0^x \sigma(u)du$, it is easy to obtain that

$$\int_0^t \Sigma(t+x-s)\, ds = \frac{1}{2}\left[S^2(t+x) - S^2(x)\right],$$

and, therefore, combining these explicit results with decomposition (5.22) we arrive to the forward rate dynamics

$$f_t(x) = f^o(x+t) + \frac{1}{2}\left[S^2(t+x) - S^2(x)\right] + e^{-ax}Z_t. \tag{5.26}$$

Equation (5.26) may be used for building initial forward rate curves $f^o(x)$ time-consistent with the model.

5.3.1 The Minimal Consistent family

Proposition 10 *The family*

$$G_{MIN}(z,x) = z_1 e^{-ax} + z_2 e^{-2ax}, \tag{5.27}$$

is the minimal dimension consistent family with the model characterized by

$$\sigma(x) = \sigma e^{-ax}.$$

Proof. As we mentionead earlier, there is a way to justify (5.27) focusing on forward rate evolution deduced at (5.26) we describe it next. By the definition of $S(x)$, we have that $S'(x) = \sigma(x)$. Then it is easy to derive that deterministic term $\frac{1}{2}\left[S^2(t+x) - S^2(x)\right]$ is of the form

$$g(t)e^{-ax} + h(t)e^{-2ax}.$$

Thus, the forward rate evolution becomes

$$f_t(x) = f^o(x+t) + (g(t) + Z_t)\, e^{-ax} + h(t)e^{-2ax}. \tag{5.28}$$

From (5.28) we see that a family which is invariant under time translation is consistent with the model if and only if it contains the linear space $\{e^{-ax}, e^{-2ax}\}$.
□

It should be also noted that the map

$$G(z, x) = G_{MIN}(z, x) + \phi(z, x),$$

where $\phi(\cdot)$, is an arbitrary function, is also consistent with this model.

Finally, we list some concluding remarks about the families analyzed.

Lemma 2 *The following hold for the Hull-White model*

- *The Nelson-Siegel family (henceforth NS)*

$$G_{NS}(z, x) = z_1 + z_2 e^{-z_4 x} + z_3 x e^{-z_4 x},$$

is not consistent with the model.

- *The family*

$$G_{MIN}(z, x) = z_1 e^{-ax} + z_2 e^{-2ax},$$

is the lowest dimension family consistent with the model (hereafter MIN).

- *The family*

$$G_{ANS}(z, x) = z_1 + z_2 e^{-ax} + z_3 x e^{-ax} + z_4 e^{-2ax},$$

is the simplest adjustment based on restricted NS family that allows model consistency (hereafter ANS).

5.4 Calibration to Market Data Approaches

To calibrate the model by means of real data, we actually need to determine the vector of parameters $\boldsymbol{p} = [\ \sigma\ a\ ]^T$. In order to estimate the forward rate volatility, the statistical analysis of past data can be a possible approach, but the practitioners usually prefer implied volatility, laying within some derivative market prices, based techniques. This way involves a minimization problem where the loss function can be taken as

$$l_C(\boldsymbol{p}) = \sum_{i=1}^{n} (C_i^o - C_i(\boldsymbol{p}))^2,$$

where $C_i(\boldsymbol{p})$ are the i–th theoretical derivative price and C_i^o is the i–th market price one. As we proved on Sect. 3.2, the model price, at $t = 0$, of the cap with equidistant settlement periods is given by

$$C = \sum_{j=1}^{n} \gamma_j = (1 + \tau K) \left(\sum_{j=1}^{n} \kappa P_{j-1}(0) N(-d_-) - P_j(0) N(-d_+) \right), \tag{5.29}$$

where the $d_\pm$ are given by (3.18). Moreover, recall that $P_j(0)$ is the initial x_j-maturity discount bond price $P(0, x_j)$, κ equals to $(1 + \tau K)^{-1}$ with K denoting the *cap rate*. The volatility function $\vartheta(0, \cdot)$ defined by the expressions (3.19) to (3.20) take the particular form for the HW model:

$$\vartheta(0, x_j) = \frac{\sigma}{a} \left(1 - e^{-a\tau}\right) \sqrt{\frac{1 - e^{-2ax_j}}{2a}}.$$

The equations (3.18) and (5.29), also express the effective influence of *ab initio* discount bond curve estimation on cap pricing.

The calibration procedures can be described formally as follows.

Let $\boldsymbol{p}$ be the parameter vector

$$[\sigma \; a]^T$$

for the model under consideration. Assume that we have time series observations of the flat volatilities $\bar{\sigma}_i$ of N at-the-money caps C_i which mature at T_i-times where $i = 1, \ldots, N$. Suppose we are also equipped with the discount bond curve estimation, $P(0, x)$, at time $t = 0$. As we introduce in Sect. 3.1, market participants translate volatility quotes to cash quotes adopting Black model [14]. Thus, according to the Definition 15 in Sect. 3.1., the market price, at $t = 0$, of the cap with regular payment periods is given by

$$C^o = \sum_{j=1}^{n} \gamma_j^o = \tau \sum_{j=1}^{n} P_j(0) \left(L_j(0) N(d_1) - K N(d_2) \right), \tag{5.30}$$

where d_1 and d_2 are given by the identities (3.3) to (3.4).

In addition, recall that according to Remark 2 in Sect. 3.1.1 the cap rates for the ATM plain vanilla caps must fulfill (3.13). By direct inspection, it is clear that this market convention makes the rates K depend on the discount bond curve estimation $P(0,x)$. Let us denote the market prices of caps by

$$C^o\left(T_i, P(0,x), K_i(P(0,x)), \bar{\sigma}_i\right).$$

Notice that this expression emphasizes explicit and implicit dependence (through ATM *strikes*) on discount bond curve estimation even for market prices. Let

$$C\left(T_i, P(0,x), K_i(P(0,x)), \boldsymbol{p}\right),$$

be the corresponding theoretical price under our particular model.

5.4.1 The Two-Step Traditional Method

Suppose that we are standing at time $t = 0$, the fixed time of calibration under study. For simplicity from now, we remove the dependency on initial time $t = 0$ from discount bond curve $P(0,x) \equiv P(x)$. First, we choose a non-consistent parametrized family of forward rate curves $G(z,x)$.

Let $P(z,x)$ be the zero-coupon bond prices produced by $G(z,x)$.

$$P(z) = [\, P_1(z) \ \ldots \ P_M(z)\,]$$

Let P_k^o be the corresponding discount x_k-bond observations with x_k-times running from $k = 1, \ldots, M$

$$P^o = [\, P_1^o \ \ldots \ P_M^o\,].$$

For each zero-coupon bond denoted with subscript k, the logarithmic pricing error[2] is written as follows

$$\epsilon_k(z) = \log P_k^o - \log P_k(z).$$

Then, we have chosen in this work the sum of squared logarithmic pricing errors, l_P, as the objective loss function to minimize:

$$\min_z l_P(z) = \min_z \|\log P^o - \log P(z)\|_2^2 = \min_z \sum_{k=1}^{M} \epsilon_k^2(z), \tag{5.31}$$

[2]Recall that, for small ϵ_k, it is also the relative pricing error $\frac{P_k^o - P_k(z)}{P_k^o}$.

with

$$\log P_k(z) = -\int_0^{x_k} G(z,u)\,du.$$

Now, via the least squares estimators $\hat{z}$, an entire discount bond curve estimation allows the pricing of caps using the market practice or the HW model. Following a similar scheme for the derivatives fitting as in the preceeding lines, we have

$$\eta_i(\boldsymbol{p}) = \log C_i^o - \log C_i(\boldsymbol{p}),$$

and

$$\min_{\boldsymbol{p}} l_C(\boldsymbol{p}) = \min_{\boldsymbol{p}} \|\log C^o - \log C(\boldsymbol{p})\|_2^2 = \min_{\boldsymbol{p}} \sum_{i=1}^{N} \eta_i^2(\boldsymbol{p}), \tag{5.32}$$

with the vector definitions:

$$C^o = [\, C_1^o \,\ldots\, C_N^o \,] \tag{5.33}$$

$$C(\boldsymbol{p}) = [\, C_1(\boldsymbol{p}) \,\ldots\, C_N(\boldsymbol{p}) \,]\,. \tag{5.34}$$

Note that here we have dropped the dependencies ($P(x)$, K, T, $\bar{\sigma}$) for simplicity. Moreover, notice that the discount bond curve estimation is external to the model in the sense that there is no need to know first any of the model parameters $\boldsymbol{p}$ for solving non-linear program (5.31).

5.4.2 The Joint Calibration to Cap and Bond Prices

Let us now describe in detail the joint cap-bond calibration procedure which has sense in a consistent family framework. We note that in this situation the parameters of the model are determined together with the initial forward rate curve.

This is different from the traditional fitting of the Hull-White model, where the two steps are separate, as we discussed before. From the expression (5.27), we notice the dependency of the family from the parameter a. Let $G(z, x, a)$ be a family consistent with the HW model (for instance, G_{MIN} and G_{ANS}) and define least-squares estimators, $\hat{z}(a)$

$$\hat{z}(a) = \arg\min_{z} \sum_{k=1}^{M} (\log P_k^o - \log P_k(z,a))^2. \tag{5.35}$$

From the expression

$$\log P_k(z,a) = -\int_0^{x_k} G(z,a,u)\,du = \sum_{j=1}^{n_p} M_{kj}(a) z_j, \tag{5.36}$$

we note that, for consistent families and for a fixed a the problem (5.35) is linear in z-parameters (for the G_{MIN} family $n_p = 2$, and for the G_{ANS} family $n_p = 4$). Thus, $\hat{z}$ is an explicit and continuous function of a.

Strictly speaking, joint calibration must be formalized as a multiobjetive optimization problem (MOO) of the form:

$$\min_{\boldsymbol{p}} \boldsymbol{l}(\boldsymbol{p}) \tag{5.37}$$

where $\boldsymbol{l} = [l_1(\boldsymbol{p})\, l_2(\boldsymbol{p})]^T$ is an objective function vector and $\boldsymbol{p}$ is the *design* vector $[\,\sigma\, a\,]^T$. Notice that in this case there are two objectives and two design variables. The *partial loss functions* $l_i(\sigma, a)$ are defined as

$$\begin{aligned} l_1(\boldsymbol{p}) &= \|\log C^o\left[P(\hat{z}(\boldsymbol{p}),\boldsymbol{p})\right] - \log C\left[P(\hat{z}(\boldsymbol{p}),\boldsymbol{p}),\boldsymbol{p}\right]\|_2^2 \\ l_2(\boldsymbol{p}) &= \|\log P^o - M(\boldsymbol{p})\hat{z}(\boldsymbol{p})\|_2^2 \end{aligned}$$

where

$$\hat{z}(\boldsymbol{p}) = R(a) Q^{-1}(a) \log P^o \tag{5.38}$$

being Q, R the matrices of the reduced QR decomposition of $M(\boldsymbol{p})$ which is defined by the relation (5.36).

Note that it is highly probable that these objectives would both be conflicting, in general, and no single $\hat{\boldsymbol{p}} = (\hat{\sigma}, \hat{a})$ would generally minimize simultaneously the pair of objective functions l_i. Tipically here, there is no single, global solution, and often, it is necessary to determine a *set* of points that all fit a predetermined definition for an optimum. Thus, the predominant concept in defining is that of Pareto optimality [47].

One of the most common and basic approach for multiobjective optimization requires to build a weighted sum of the objectives (see for instance [23, 40]). The result is the following scalarized utility function, which is minimized:

$$\tilde{l} = \omega_1 l_1 + \omega_2 l_2. \tag{5.39}$$

When an appropriate set of solutions is obtained by the single-objective optimization of $\tilde{l}$, the solutions can approximate a Pareto front. The weighted-sum method parametrically changes the weights among objective functions l_1 and l_2 to obtain this Pareto front. If the two weights are positive then minimizing the utility provides a sufficient condition for Pareto optimality, which means the minimum of (5.39) is always Pareto optimal [40, Sect. 4.1.1, pp. 41–42]. Thus, consistent calibration carried out with consistent families involves the entire Pareto optimal set, in contrast to the unique solution appearing in the two-step scalar problem.

At this point, note that the program used by Angelini and Herzel [2, 3] in their works, uses a different goal attainment

$$\min_{\boldsymbol{p}} l_1(\boldsymbol{p}) \tag{5.40}$$

where $l_1(\boldsymbol{p})$, and $\hat{z}(\boldsymbol{p})$ are defined trough the identities (5.32) and (5.35). As a consequence, the program used by these authors is a degenerate case of (5.39) with ω_1 fixed equal to 1 and ω_2 to 0, so it just allows to obtain one point of the implied trade-off front, which would be potentially only a weak Pareto optimum[3] (WPO) not a standard Pareto optimum [40, Sect. 4.1.1, p. 42].

5.5 Empirical Results

In this context the main goal is to analyze the impact that an alternative interpolation scheme has on the fitting capabilities of the model. To this end, we use as a measure, the daily (on average) relative pricing errors, hereafter RPE_C:

$$RPE_C = \frac{1}{N}\sum_{i=1}^{N}\frac{|C_i^o - C_i(\hat{\boldsymbol{p}})|}{C_i^o}$$

The same kind of measure is used for the zero-coupon bond prices and we denote it with RPE_B:

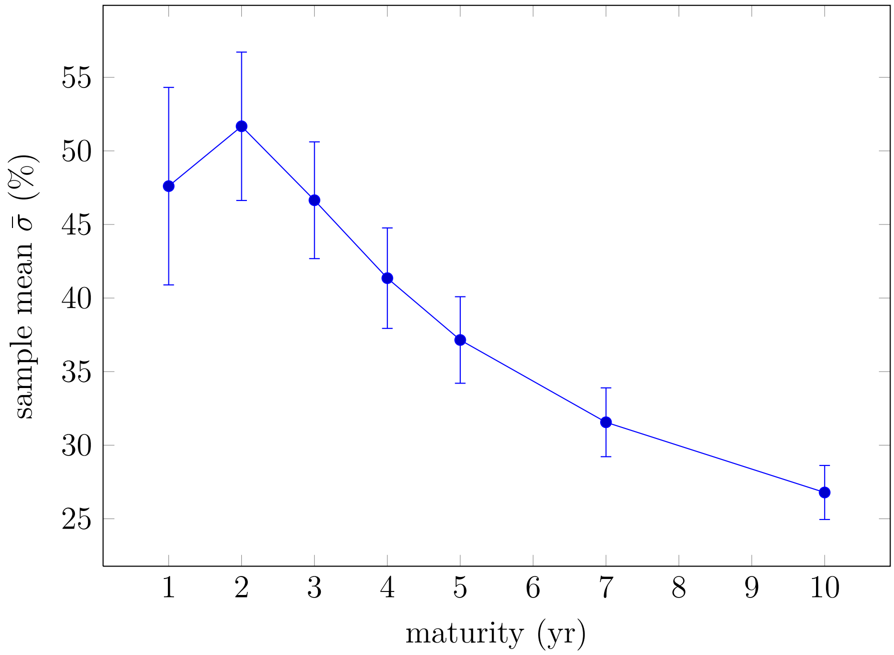
$$RPE_B = \frac{1}{M}\sum_{k=1}^{M}\frac{|P_k^o - P_k(\hat{z}(\hat{\boldsymbol{p}}), \hat{\boldsymbol{p}})|}{P_k^o}$$

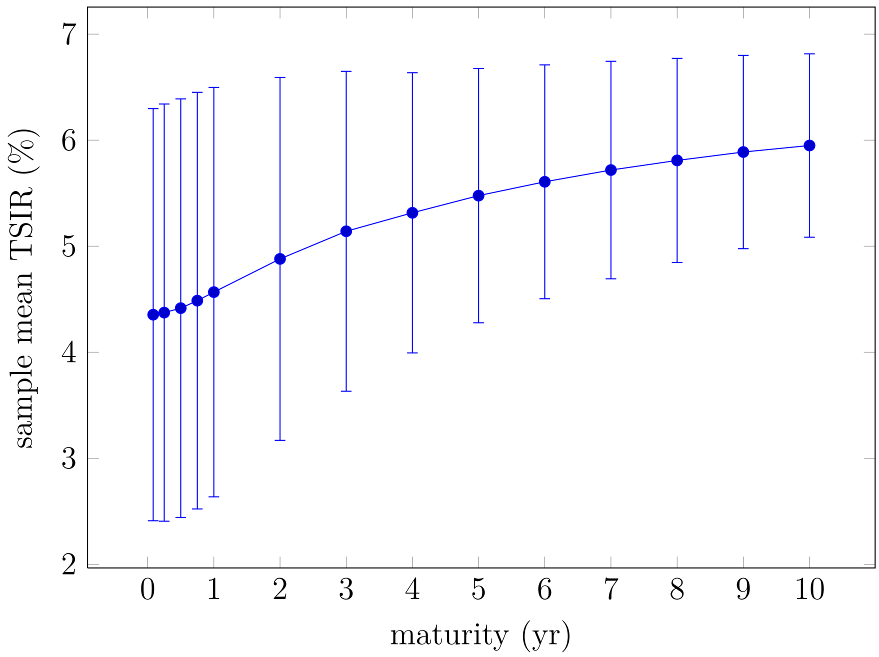
We perform such analysis focusing on US market. The real date consists of 282 daily observations, between 2/09/2002 and 30/09/2003. The data set is composed of US discount factors for fourteen maturities (1, 3, 6, 9 months and from 1 to 10 years) and of implied volatilities of at-the-money interest rate caps with maturities

[3]Pareto optimal points are WPO, but WPO are not Pareto optimal.

1,2,3,4,5,7,10 years. This database is provided by Thomson Reuters Datastream.

Figure 5.1: Average of the US market TSIR and TSV with 99% confidence levels.





As it have been explored before, the daily joint calibration of caps and bonds with consistent families must be properly carried out as a MOO problem. In doing so, we choose an *a posteriori articulation* of preferences, that is, we delay the selection of the solution from the palette of solutions after the weighted-method runs. In response to this articulation of preferences, the decision-maker imposes preferences directly on a set of the potential solution points which depicts the Pareto front. In such a context, weights are typically chosen such that

$$\sum_{i=1}^{Q} \omega_i = 1 \tag{5.41}$$

with $\boldsymbol{\omega} \geq \mathbf{0}$ leading to a convex combination of objectives and this choice can be more helpful than unrestricted weights so as not to repeat any weighting vectors in terms of its relative values [40, Sect. 5.3, pp. 73–74]. According to this, the weighted-sum method was run for every date in sample with the fixed weight vector $\boldsymbol{\omega}$ satisfying (5.41). In doing so, we assume the same ten uniformly spread values

$$\omega_1 = \frac{1}{10} j \qquad j = 1, 2, \ldots, 10 \tag{5.42}$$

and $\omega_2 = 1 - \omega_1$ as the second vector weight component for all trading dates.

We use the following transformation scheme of the objectives, which is often called *scaling* [48]:

$$l_i^{\tau} = \frac{l_i(\boldsymbol{p})}{s_i} \tag{5.43}$$

$$\text{with} \quad \frac{l_1(\boldsymbol{p}_0)}{s_1} \approx \frac{l_2(\boldsymbol{p}_0)}{s_2} \tag{5.44}$$

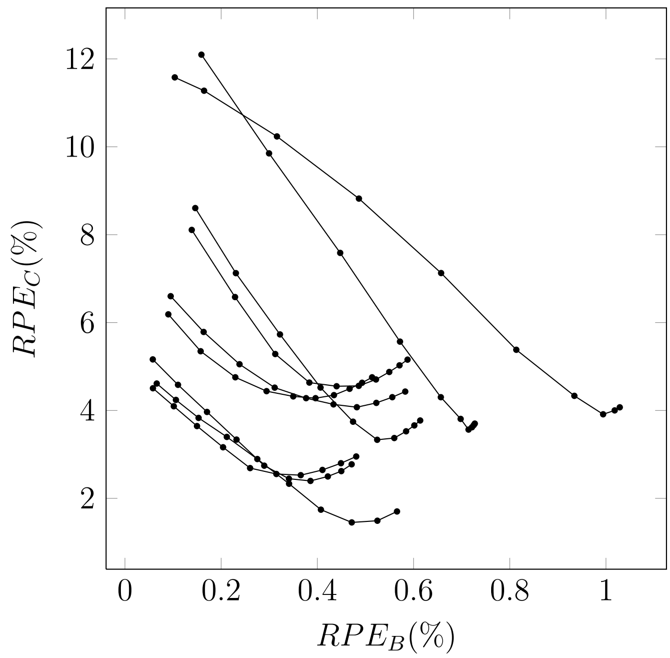
where s_i are scalar coefficients, and $\boldsymbol{p}_0$ is a feasible starting point. This approach ensures the objective functions have similar orders of magnitude. Thus, the way to solve our joint calibration problem is to use the weighted-sum method, which is finally stated as:

$$\min_{\boldsymbol{p}} \left(\omega_1 \frac{l_1(\boldsymbol{p})}{s_1} + (1 - \omega_1) \frac{l_2(\boldsymbol{p})}{s_2} \right) \tag{5.45}$$

with ω_1 discrete values given by (5.42).

Figure 5.2 shows the in-sample fitting results performed by the MIN family for some dates of the sample under analysis. First of all, notice that efficient frontiers with regular shapes appear nicely revealing the intrinsic multi-objective nature of the consistent calibration. Moreover, note that it can be found different topologies for this frontiers depending on the date. All the days the objectives are conflicting, and beyond a certain point of the Pareto front the better we fit the discount bonds the worse we calibrate the caps portfolio. However, note that, moving on to the Pareto curve, we can achieve better results for both components of the vector objective without a trade-off until the above-mentioned Pareto point is reached. In other words, the MOO calibration may provide a *better* set of results for the calibration of caps that would produced by single optimizing the scalar objective $l_1(\boldsymbol{p})$.

Figure 5.2: Some daily calibration results for the minimal consistent family.



The tables on Figure 5.3 show, as a numerical example, two different Pareto sets restricting ourselves to the MIN family, the family with the lowest dimensionality. If we look on both tables, it must be noted that for a fixed trading date the best cap fit results may occur with $\omega_1 \neq 1$, even if the objectives are competing.

In Figure 5.4, we analize more deeply the latter fact this time for both, MIN and ANS, consistent families. We plot the daily distribution of the weight ω_1 which performs the best calibration for caps in sample data. As for the lowest dimensional family, most of the days the weight vector ($\omega_1 = 0.7, \omega_2 = 0.3$) produces the best cap calibration results and there is a non-negligible number of bussiness dates where other weights than $\omega_1 = 1.0$ produce better goals than it. As for the ANS family, we observe an entirely different distribution, but once again, we note that the best cap calibration results may be reached with weights

Figure 5.3: Efficient points in the $RPE_B - RPE_C$ space using the method of convex combinations for two different days in sample.

Day 1

ω_1	ω_2	RPE_B (%)	RPE_C (%)
0.1	0.9	0.1036	11.5795
0.2	0.8	0.1645	11.2761
0.3	0.7	0.3160	10.2359
0.4	0.6	0.4864	8.8214
0.5	0.5	0.6573	7.1267
0.6	0.4	0.8136	5.3813
0.7	0.3	0.9343	4.3329
0.8	**0.2**	**0.9941**	**3.9132**
0.9	**0.1**	**1.0181**	**4.0016**
1.0	**0.0**	**1.0287**	**4.0718**

Day 2

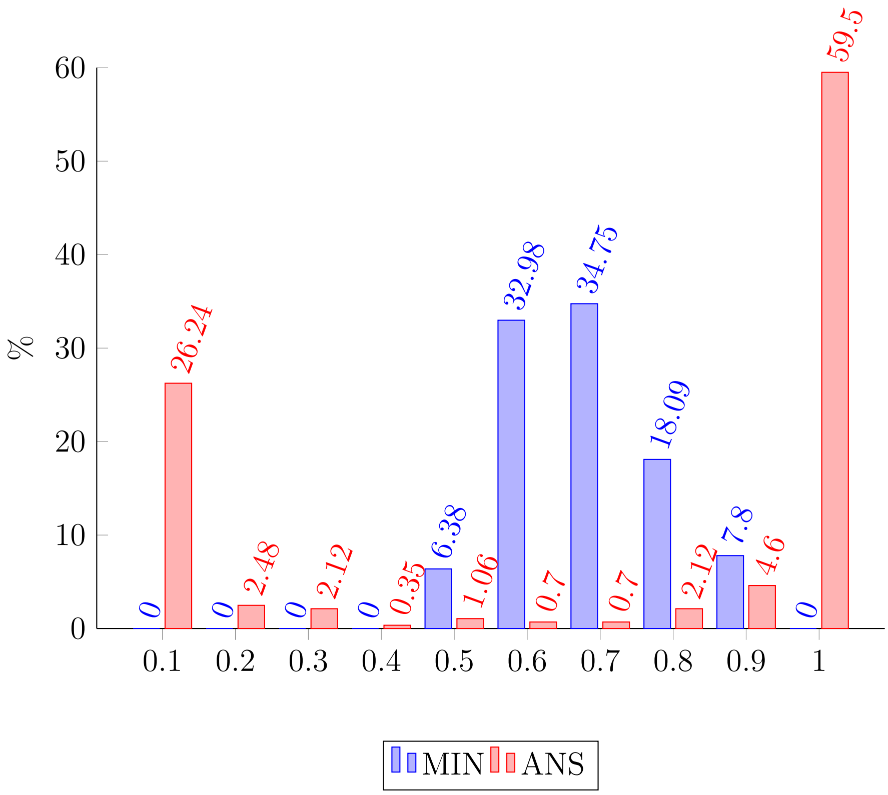
ω_1	ω_2	RPE_B (%)	RPE_C (%)
0.1	0.9	0.0902	6.1874
0.2	0.8	0.1574	5.3478
0.3	0.7	0.2295	4.7551
0.4	0.6	0.2944	4.4395
0.5	0.5	0.3499	4.3203
0.6	**0.4**	**0.3962**	**4.2797**
0.7	**0.3**	**0.4345**	**4.3490**
0.8	**0.2**	**0.4670**	**4.4933**
0.9	**0.1**	**0.4930**	**4.6303**
1.0	**0.0**	**0.5138**	**4.7546**

The partial objectives, ω_1 and ω_2 are strong conflicting, for the Day 1 (top). In contrast, the latter ones are more cooperative for the Day 2 (bottom).

different than $\omega_1 = 1.0$ and even get a large number of them with the smallest weight possible, $\omega_1 = 0.1$.

For the shake of simplicity, from now on we will only consider the calibration results obtained with daily weights choices that produce the best calibration for the caps on every trading date. This *a posteriori* articulation of preferences may be followed by a decision-maker which want to use consistent calibration as a good risk management practice or as an extrapolation tool for marking to

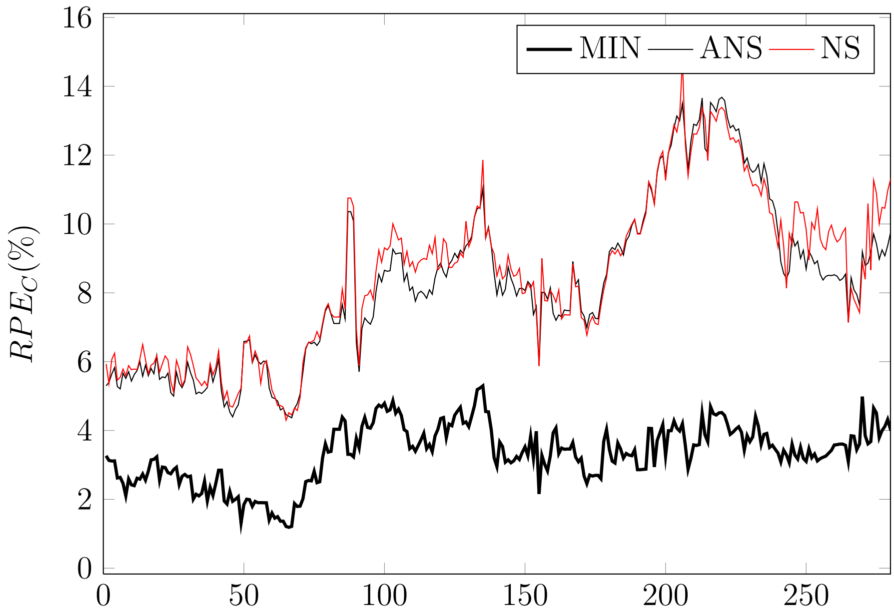
Figure 5.4: Daily empirical distribution of weights with the best RPE_C for both consistent families as produced by the multi-objective calibration.

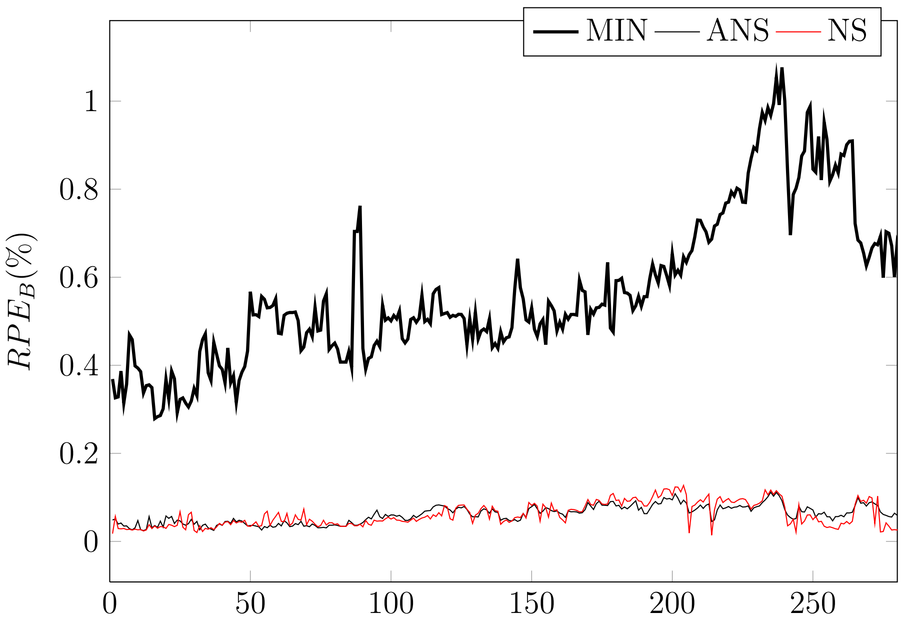


market less liquid interest-rate derivatives. Following this particular articulation of preferences we choose a single Pareto optimum from the set that estimates the complete Pareto curve. In Figure 5.6, we compare summary statistics of the parameter estimates and the in-sample fit measures reported by NS, MIN and ANS families. In addition, Figure 5.5 shows the comparison of in-sample fitting results in time series.

The two consistent families under study report better RPE results when we restrict the analysis to cap data. For RPE on bonds, only the ANS family outperforms NS in the sample. Recall that this fact is acceptable since the MIN family is a forward curve with less number of parameters than the other ones proposed. Moreover, on caps, note that the MIN family appears to give better results than its consistent counterpart, ANS. Now, this behaviour can be explained because the major of dates considered, market data make the objective functions l_1 and l_2 of this family to strongly conflict as seen in Figure 5.4.

Figure 5.5: Time Series Comparison.





5.6 Concluding Remarks

When calibrating the Hull-White model, a TSIR curve choice to fit a few market data observations is needed. In particular it seems to be natural to use families of curves which do not modify their structure under the future evolution of the model, the so-called consistent families.

Figure 5.6: Summary statistics for the calibration results. In-sample descriptive statistics are carried out using the daily Pareto points with the best derivative fit outcomes.

Summary Statistics

	MIN	ANS	NS
σ	0.0145	0.0127	0.0164
a	-0.0096	0.0243	0.0256
$C_v(\sigma)$	0.1113	0.3118	0.14
$C_v(a)$	-3.2659	2.4268	1.8831
RPE_C (%)	3.3361	8.3011	8.5458
RPE_B (%)	0.5688	0.061	0.0609

In this work, we choose three families of curves (two consistent families and the popular Nelson-Siegel family) and we conclude that this choice have an effective impact on the quality of in-sample fitting for US-market data. Moreover, this paper extends the seminal calibration algorithm proposed in Angelini and Herzel [2].

In a consistent approach the parameters of the model are estimated jointly with the esmation of initial discount bond curve. Therefore, from a rigorous point of view, joint calibration of caps and bonds must be viewed as a multi-objective optimization nonlinear problem. Although the main purpose of the algorithm is to minimize the relative differences of cap prices too, note that the bi-objective extension of the consistent calibration presents more general features. Such extension is structured to allow more numerical outcomes and we observe that it allows to better fit results for both, caps and bonds, than the above mentioned. In particular, it is possible to find better cap calibration outcomes with $\omega_1 \neq 1$, and this is definitively different from what worked Angelini and Herzel [2] on Hull-White model, where only the fixed $\omega_1 = 1$ seems to be considered for all consistent families. The empirical findings of this paper show that, in general, consistent calibration on every date must to be carried out by analyzing the entire shape of the Pareto curve.

In this sense, this work confirms and complements the shown by Angelini and Herzel [2, 3] restricted to a Euro data set. We articulate preferences restricting possible outcomes on every date, by choosing the Pareto points which are responsible of better fit results on caps. Then the minimal consistent family gives the best performance in terms of caps pricing errors and becomes a good candidate for the calibration of the Hull-White model. The ANS consistent family performs very close to the Nelson-Siegel family, though it seems to be the best solution for estimating the discount bond function. Now, this could be explained in the context of vector optimization. We show empirically the usual competing behaviour followed by the objectives through the sample considered. Then, the minimal parameterized consistent family relax the performance on the estimation of the discount bond curve function, allowing minor relative pricing errors on caps.

Chapter 6

Multiojective Calibration: More Empirical Evidence

In Chapter 5 we have shown that the consistency hypothesis stated by Björk [10, 11], implies that the discount bond curve has to be determined at the same time as the parameters of the model. Angelini and Herzel [2, 3], originally proposed the use of a optimization program related to the mentioned daily calibrations, which is compatible with this joint estimation. The milestone of this methodology is the use of an objective function based on an error measure for just the portfolio of caps. Then, the theoretical prices for the caps along the minimization of this measure can be calculated at the same time that the discount bond curve is fitted. This is an efficient method because consistent families of discount bond curves have good analytical properties under the Gaussian hypothesis, i.e. deterministic *forward* volatility.

In last chapter, we have also provided an extension of the above strategy which involves a multi-objective framework. As a direct consequence, the objective function above-mentioned, is subtituted by a *scalarized* form of the intrinsic bi-objective problem. Now, the error measure for the discount bonds is evaluated in each iteration and could even dominate the joint optimization problem. To this scope, we construct this *scalarized* form using a convex combination of both the cap and the bond error measures, by means of a set of restricted weights. As a matter of fact, this approach is richer in possible outcomes.

6.1 Introduction

Given the theoretical tools we have developed in the previous chapters, we want to analyze further empirical results which support the use of consistent families and the multi-objective calibration techniques.

To this end, we extend such analysis to a particular humped volatility HJM model, proposed by Ritchen and Chuang [49] and Mercurio and Moraleda [41], independently. We have chosen this model because it is quite popular and analytically treatable. In particular, it provides closed formulas for Europeans interest-rate options. Moreover, it is a one-factor Gaussian model that seems to be more capable than the Hull-White model for reproducing large humps in the implied cap volatility curve.

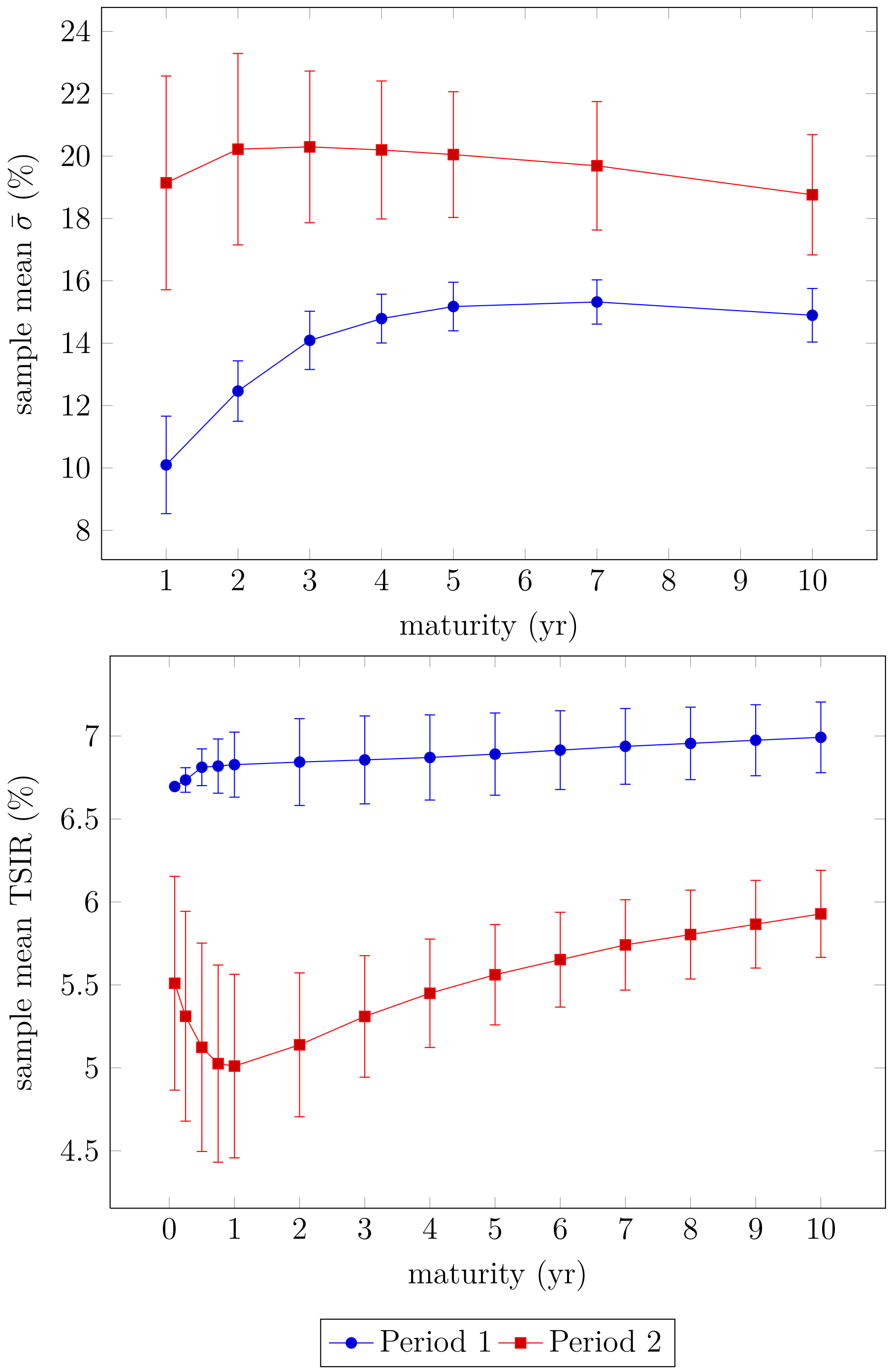
We perform our study by calibrating this model, first by using simulated data and second by focusing on market a data set composed by US discount bonds and at-the-money flat cap volatilities quotes in two different periods, as shown by the Figure 6.1.

With regard to the real market data, the first scenario depicts a market situation where the implied cap volatility curves have a large long-term humped shape and the term structure of interest rates is closer to be flat. On the other hand, the second scenario has periodically resurfaced in the market and may be considered more typical. In this situation the peak of the hump is at about the two year point. Moreover, the TSIR is not monotonic increasing nor flat, being initially-inverted with a local minimum at short-term maturities.

This rest of this chapter is organized as follows[1]. In Section 6.2 we give a brief overview of the model we want to consider and later we discuss how to construct consistent families with such a model. Section 6.3 is devoted to empirical results, first comparing the consistent calibration with the non-consistent approach by means of simulated data, then presenting the results of the fitting of the different methods with market data. In the last section we give some final conclusions and remarks.

[1] This chapter is based on [24, 26].

Figure 6.1: Market TSIR and TSV data in the two different market scenarios.



6.2 Consistent Curves with The Model

Our work is devoted to the one-dimensional Gaussian HJM humped volatility model of the form:

$$df_t(x) = \{\dots\}\,dt + (\alpha + \beta x)\,e^{-ax}\,dW_t. \tag{6.1}$$

Thus, the forward rate volatility function $\sigma(x)$ is deterministic depending only on time to maturity. Note that $\sigma(x)$ is a QE function that admits the matrix representation (5.19). Therefore, $\sigma(x)$ may be rewritten as

$$\begin{aligned} \sigma(x) &= ce^{Ax}b, \text{ where} \\ c &= [\alpha \;\; \beta - a\alpha], \\ A &= \begin{bmatrix} 0 & -a^2 \\ 1 & -2a \end{bmatrix}, \\ b &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \end{aligned} \tag{6.2}$$

For this model, the forward rate decomposition $f_t(x) = \delta_t(x)+q_t(x)$, as seen before in Sect. 5.3, eqs. (5.21)–(5.22), has the corresponding q_t-process dynamics:

$$dZ_t = AZ_t\,dt + b\,dW_t, \quad Z_0 = 0, \tag{6.3}$$

$$q_t(x) = C(x)Z_t, \tag{6.4}$$

with A, b as in (6.2) and $C(x) = ce^{Ax}$. Therefore, the analytical expression of the forward rate curve for the model is given by

$$f_t(x) = f^o(x+t) + \frac{1}{2}\left[S^2(t+x) - S^2(x)\right] + C(x)Z_t, \tag{6.5}$$

being $S(x) = \int_0^x \sigma(u)\,du$. After some algebraic manipulations like using the explicit expansion for the stochastic term $C(x)Z_t$

$$ce^{Ax}\begin{bmatrix} Z_t^1 \\ Z_t^2 \end{bmatrix} = e^{-ax}\,[\alpha \;\; \beta - a\alpha]\begin{bmatrix} 1+ax & -a^2x \\ x & 1-ax \end{bmatrix}\begin{bmatrix} Z_t^1 \\ Z_t^2 \end{bmatrix}$$

$$= e^{-ax}\left(\alpha Z_t^1 - a\alpha Z_t^2 + \beta Z_t^2\right) + xe^{-ax}\left(\beta Z_t^1 - a\beta Z_t^2\right),$$

and expanding the deterministic term $\frac{1}{2}\left[S^2(t+x) - S^2(x)\right]$ which is of the form

$$g_1(t)e^{-2ax} + g_2(t)xe^{-2ax} + g_3(t)x^2e^{-2ax} + h_1(t)e^{-ax} + h_2(t)xe^{-2ax},$$

(6.5) may be written as

$$\begin{aligned} f_t(x) = f^o(x+t) + g_1(t)e^{-2ax} + g_2(t)xe^{-2ax} + g_3(t)x^2e^{-2ax} + \\ \left(h_1(t) + \alpha Z_t^1 - a\alpha Z_t^2 + \beta Z_t^2\right)e^{-ax} + \left(h_2(t) + \beta Z_t^1 - a\beta Z_t^2\right)xe^{-ax}. \end{aligned} \tag{6.6}$$

Note that this formula, as its corresponding counterpart in previous chapter (5.28), is relevant for consistency, because it shows which curves the model produces for a given initial curve $f^o(x)$.

6.2.1 The Minimal Consistent family

Proposition 11 *The family*

$$G_{HMC}(z,x) = (z_1 + z_2x)e^{-ax} + (z_3 + z_4x + z_5x^2)e^{-2ax}, \tag{6.7}$$

is the minimal dimension consistent family with the model characterized by deterministic volatility $\sigma(x) = (\alpha + \beta x)e^{-ax}$.

Proof. From (6.6) we see that a family which is invariant under time translation is consistent with the model if and only if it contains the linear space $\{e^{-ax}, xe^{-ax}, e^{-2ax}, xe^{-2ax}, x^2e^{-2ax}\}$. □

Similar results as discussed along the lines of Sect. 5.2.1 will turn up over and over again, so we list some concluding remarks which the reader may immediately derive.

Lemma 3 *The following hold for the humped volatility Heath-Jarrow-Morton model characterized by $\sigma(x) = (\alpha + \beta x)e^{-ax}$.*

- *The NS*

  $$G_{NS}(z,x) = z_1 + z_2e^{-z_4x} + z_3xe^{-z_4x},$$

  is not consistent with this model.

- *The family*

  $$G_{HMC}(z,x) = (z_1 + z_2x)e^{-ax} + (z_3 + z_4x + z_5x^2)e^{-2ax},$$

it is the lowest dimension family consistent with the model (hereafter HMC).

- *The family*

$$G_{ANS+}(z,x) = z_1 + z_2 e^{-ax} + z_3 x e^{-ax} + (z_4 + z_5 x + z_6 x^2) e^{-2ax},$$

is the simplest adjustment based on restricted NS family that allows model consistency (hereafter ANS+).

6.3 Empirical Results

We compare four different estimations of the initial discount bond curve based on NS, HMC, ANS+ and cubic spline interpolation (hereafter SP).

The US data set consists of 126 daily observations divided in two periods: first period covers from 3/7/2000 to 29/09/2000 (64 trading dates) and the second one starts in 4/1/2001 and finish on 30/3/2001 (62 trading dates).

With regard to the market the data set is composed of US discount bond of fourteen maturities (1, 3, 6 and 9 months and from 1 to 10 years) and of implied volatilities of at-the-money interest rate caps with maturities 1, 2, 3, 4, 5, 7 and 10 years. This two windows of data comes from the same database explained before in Chapter 5 being kindly provided by Thomson Reuters Datastream. On the other hand, the simulated data was obtained from 360 extractions of bond and cap prices with identical maturities as its real-market equivalents as produced by the model under study.

Simulations

We simulate the forward rate curves of the humped volatility model at time t when initiliazed from alternative starting curves $f^o(x)$ using (6.6).

Next, we compute the fourteen prices of the set of discount bonds by integrating the forward curve $f_t(x)$ in

$$P(t,x) = e^{-\int_0^x f_t(u)\,du},$$

and the seven prices of the ATM caps by using equations (5.29), (3.13), (3.18), and by working out the integral (3.19) to obtain the model implied volatility function:

$$\vartheta^2(0, x_{j-1}) = \int_0^{x_{j-1}} \left[\int_{x_{j-1}}^{x_j} (\alpha + \beta(T-t))e^{-a(T-t)}\, dT \right]^2 dt. \tag{6.8}$$

The fixed model parameters, $\boldsymbol{p_0} = [0.002\, 0.007\, 0.35]^T$, have been taken. This particular choice has similar order of magnitude as the empirical estimations for this model reported by Angelini and Herzel [3]. As alternative starting curves, we choose HMC, ANS+ and NS fitted to the zero coupon bond prices shown in Figure 6.2.

Figure 6.2: Discrete data for initial yield-curve estimation.

Maturity, x	0.083	0.25	1	2	3	4
Discount Bond, $P^o(x)$	0.9962	0.9886	0.9538	0.9069	0.8602	0.8142
Maturity, x	5	6	7	8	9	10
Discount Bond, $P^o(x)$	0.7693	0.7260	0.6843	0.6445	0.6066	0.5706

Starting from the initial fitted curves, which may be denoted with $f^o_{HMC}(x)$, $f^o_{ANS+}(x)$ and $f^o_{NS}(x)$, and according to (6.5), the corresponding three different model evolutions are calibrated to HMC, ANS+ and NS, restricting the palette of the possible Pareto-front approximants produced by the *scalarized* program (5.45), to $\omega_1 = 1$. In order to make calibration results more comparable, Monte Carlo simulations are built in from the identical random sequence $(Z^1_t,\ Z^2_t)$ in all three cases.

Following the expression (6.6), it is easy to observe that there are two consistent families, G_{HMC} and G_{ANS+}, for the first simulation E1, just one, G_{ANS+}, for the second simulation E2, and no one for the last simulation E3.

Figure 6.3 shows main consequences of the theory when the model is the *truth* model. Notice that perfect calibration just occurs, although model parameters are fixed *a priori*, when the used family to perform calibrations is consistent with all the future forward curves generated from initial curve $f^o(x)$. This fact explains, for instance, the bad performance for the NS family even on E3 experiment. Indeed, as may be seen in Figure 6.4, an incorrect discount bond choice selection produces parameter instability and imprecision.

Figure 6.3: Summary statistics for calibration results with simulated data.

		HMC	ANS+	NS
E1: $f_0(x) = f^o_{HMC}(x)$	$\varepsilon_r(\alpha)$	0	0	0.23
	$\varepsilon_r(\beta)$	0	0	0.13
	$\varepsilon_r(a)$	0	0	$8.7\ 10^{-2}$
	$C_v(\alpha)$	0	0	0.18
	$C_v(\beta)$	0	0	0.14
	$C_v(a)$	0	0	$9.7\ 10^{-2}$
	σ_{LS}	0	0	$1.9\ 10^{-3}$
E2: $f_0(x) = f^o_{ANS+}(x)$	$\varepsilon_r(\alpha)$	0.25	0	0.28
	$\varepsilon_r(\beta)$	0.16	0	0.16
	$\varepsilon_r(a)$	0.12	0	$9.5\ 10^{-2}$
	$C_v(\alpha)$	$3.8\ 10^{-2}$	0	0.117
	$C_v(\beta)$	$3.9\ 10^{-2}$	0	$9.1\ 10^{-2}$
	$C_v(a)$	$3.2\ 10^{-2}$	0	$4.8\ 10^{-2}$
	σ_{LS}	$2.6\ 10^{-4}$	0	$6.7\ 10^{-4}$
E3: $f_0(x) = f^o_{NS}(x)$	$\varepsilon_r(\alpha)$	0.313	$2.7\ 10^{-4}$	0.18
	$\varepsilon_r(\beta)$	0.20	$2.10\ 10^{-4}$	0.10
	$\varepsilon_r(a)$	0.16	$1.6\ 10^{-5}$	$6.7\ 10^{-2}$
	$C_v(\alpha)$	$2.3\ 10^{-2}$	$1.4\ 10^{-4}$	0.17
	$C_v(\beta)$	$2.6\ 10^{-2}$	$1.0\ 10^{-4}$	0.111
	$C_v(a)$	$2.2\ 10^{-2}$	$8.3\ 10^{-5}$	$6.3\ 10^{-2}$
	σ_{LS}	$3.8\ 10^{-4}$	$3.9\ 10^{-9}$	$3.5\ 10^{-4}$

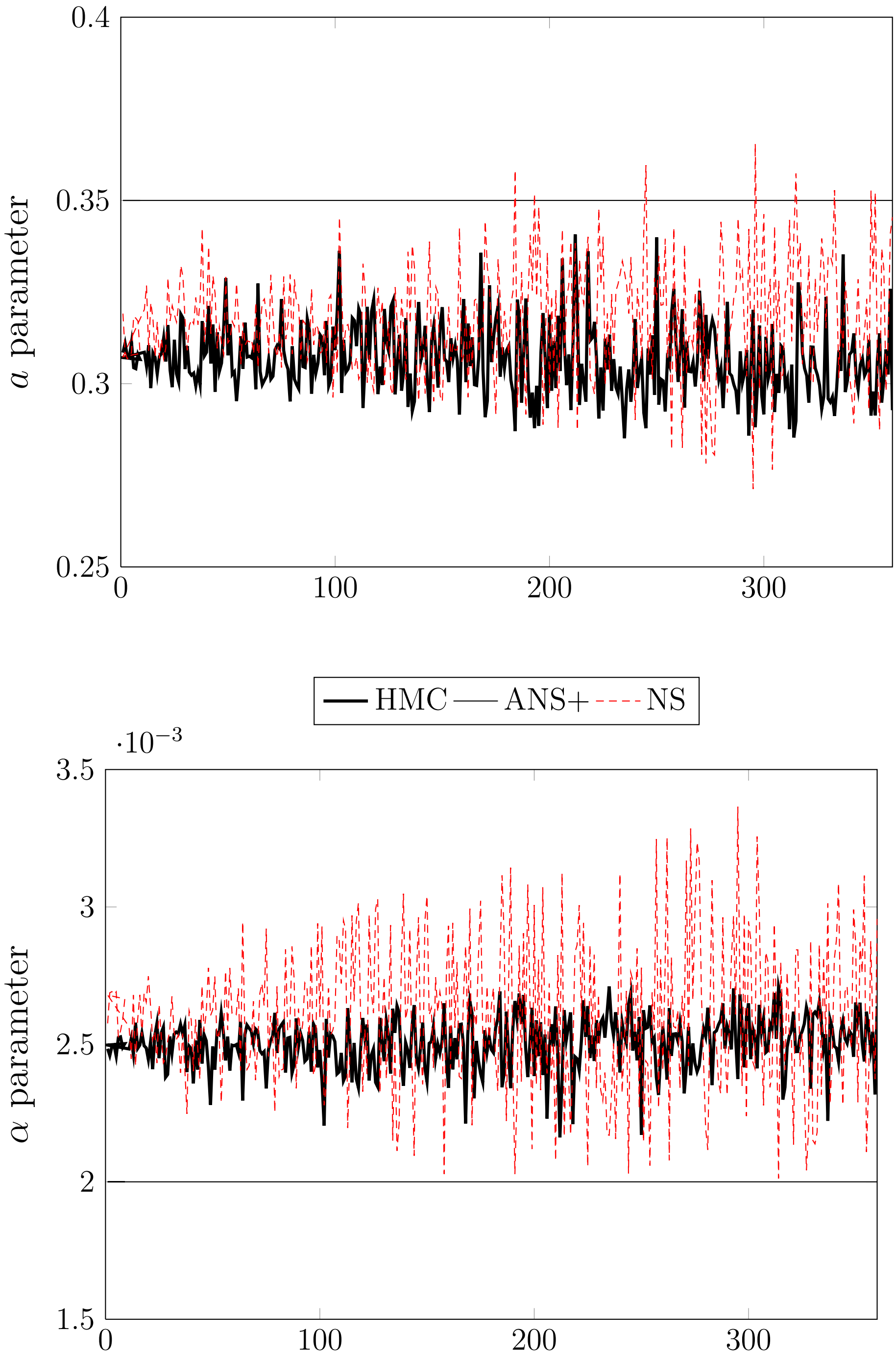
Sample statistics of the calibration on simulated data. Relative errors of the parameters estimates are expressed in absolute value. We set to 0 table entries with value $< 10^3 \cdot$`eps` (variable `eps` $\sim 10^{-16}$ measures MATLAB internal accuracy).

Real Data

The main purpose of this section is to compare the performance of the two different calibration approaches introduce in Chapter 5 along the two different periods of real trading dates described before. Therefore, from now on we will only consider the calibration results obtained with the market data. For simplicity, the consistent calibrations are carried out by means of just the lowest dimension family, the HMC family.

Concerning the real data, calibration with consistent families are carried out by setting the weights palette $(\omega_1,\ \omega_2)$, as defined in (5.42), Sect. 5.5. With regard to the consistent calibration, the *scalarized* MOO program (5.45) has also been used. The table on Figure 6.5 exhibits the sample mean of the daily error

Figure 6.4: Daily estimates of parameters a and α for data simulated from the model with $\alpha = 0.002$ and $a = 0.35$ and starting forward curve $f_0(x) = f^o_{ANS+}(x)$.



The straight line corresponds to daily calibration results belonging ANS+ family, the irregular black line to the MC family and the dashed red one to the NS family.

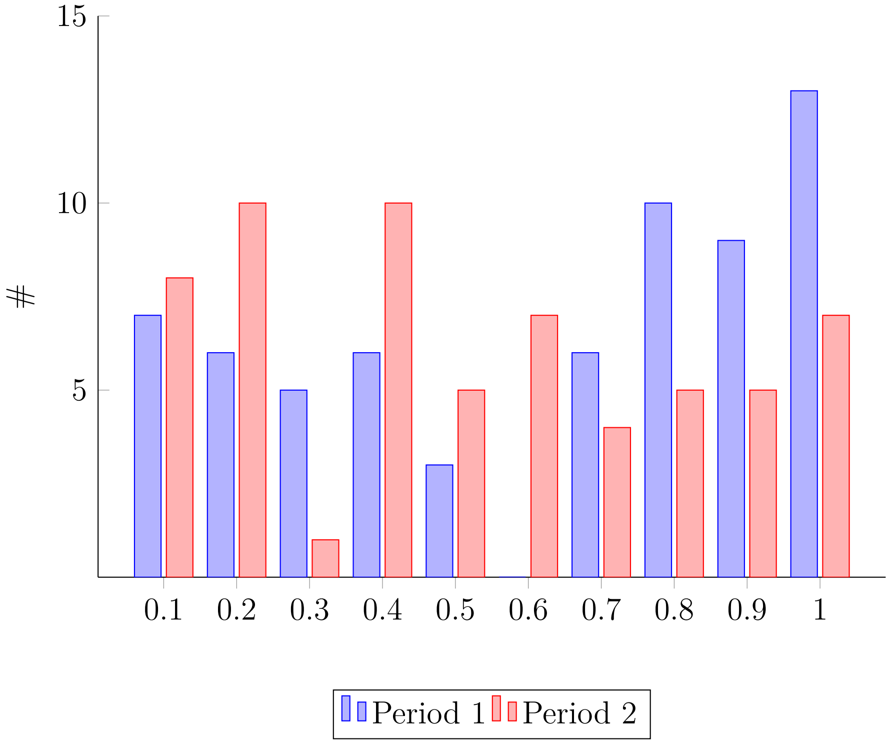
fitting measures, namely RPE_C and RPE_B, and the mean and the coefficient of variation of parameter estimates.

Figure 6.5: Summary statistics for calibration results with US data on both periods.

		HMC	NS	SP
Period 1	α	0.0093	0.0098	0.01
	β	0.0007	0.0013	0.0008
	a	0.0024	0.0961	0.064
	$C_v(\alpha)$	0.11	0.05	0.05
	$C_v(\beta)$	0.33	0.71	1.30
	$C_v(a)$	0.52	0.54	0.92
	$RPE_C(\%)$	2.2	2.7	2.75
	$RPE_B(\%)$	0.019	0.047	
Period 2	α	0.0087	0.0091	0.0085
	β	0.0041	0.0039	0.0052
	a	0.1469	0.176	0.2129
	$C_v(\alpha)$	0.15	0.1	0.11
	$C_v(\beta)$	0.72	0.38	0.35
	$C_v(a)$	0.96	0.3	0.27
	RPE_C	1.49	1.56	1.36
	RPE_B	0.031	0.043	

On the other hand, Figure 6.7 shows in-sample fitting time series. The HMC family under study report good in-sample fitting results as compared with non-consistent approaches. However, when we look in detail at the Period 2, the families NS and HMC and even the cubic spline based interpolants perform slightly similar with regard to caps calibration. In fact, the non-consistent approach based with cubic spline interpolation, marginally outperforms all the rest. In Figure 6.6, we plot the daily distribution of the weight ω_1 which performs the best calibration for caps in both samples of data. As for the Period 1, we observe that the cap contribution of the scalarized objective is more dominant, because of the reason that when articulating *a posteriori* preferences as we made in previous chapter, the weights closer to the WPO ($\omega_1 = 1$) have the capability to reproduce better daily fits of derivatives. As for the Period 2, the individual objectives, $[l_1\ l_2]^T$ appear to be more cooperative relaxing the performance of the fit results of the implied cap volatility curve. This fact, may explain the very similar results provided for all three methods in the second window.

Figure 6.6: Not normalized daily empirical distribution of weights with the best RPE_C for both sample periods as produced by the multi-objective calibration.



6.4 Conclusions

In this chapter, we analyze two new consistent families of curves (the HMC and ANS+ families) for comparing to another non-consistent approaches like cubic spline interpolation as well as the well-known Nelson and Siegel family. In so doing, we have tried to support and extend to another treatable Gaussian model the empirical findings of Chapter 6.

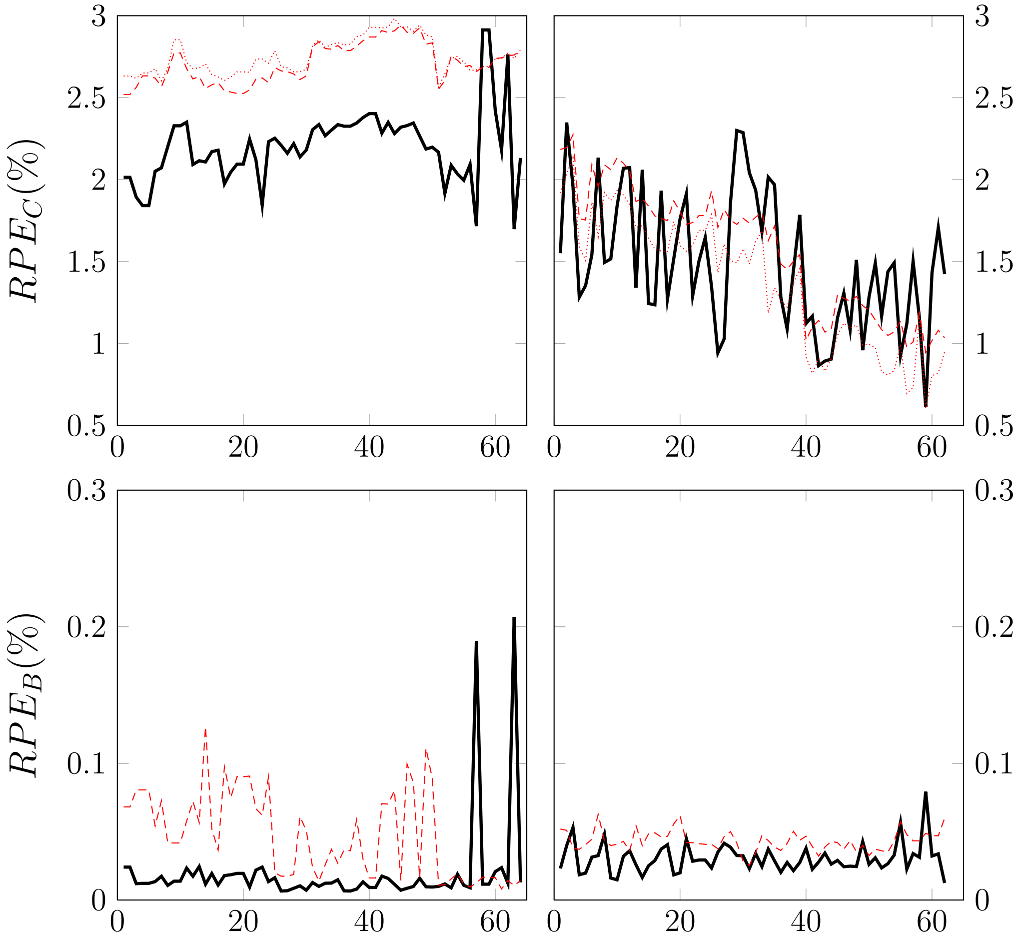
When using simulated data it is very clear that the consistent families for the E1 ans E2 experiments performs much better than the non-consistent ones. Moreover, Nelson-Siegel family does not work even if it is chosen as the starting yield-curve (recall E3 experiment). These empirical facts constitute a nice demonstration of the theory introduced in Chapter 4, in the sense that even on absence of *model risk* only when consistent families are used, perfect calibration may occur.

Translation of these consequences to real data is less clear, due to *model risk* and quality of data but we can infer the following concluding remarks. In this case, the introduction of a sufficiently rich *consistent families* like HMC, which is well motivated theoretically by Björk et al., improves in-sample fitting capabilities on caps on bonds complementing what Angelini and Herzel [2, 3] empirically found with a different set of data.

According to the results reported for the humped volatilty model in this chapter and the Hull-White model in Chapter 5, multi-objective calibration would lead generally to better results in caps calibration as compared to non-extended consistent calibration (originally introduced in [2, 3]) and the more traditional non-consistent methodologies.

Finally, future empirical research on the matter should include multi-factor models for capturing more appropriately the TSIR and TSV observed in the market.

Figure 6.7: In-sample fitting time series for the first period (left) and the second period (right) with real market data.



The thick black line corresponds to the minimal consistent family, the dashed red line to the NS family and the dotted one to cubic spline interpolation.

Chapter 7

Pricing Options with a Consistent HJM Model

7.1 Introduction

One of the main goals of financial mathematics is to determine the prices of derivatives. In this chapter we adapt the sophisticated machinery introduced in literature in past years, in order to study which pricing methods are available for interest rate derivatives in the event that consistent families are used.

Therefore, we will study the behaviour of some efficient numerical implementations of the model introduced in Chapter 5, pricing the most liquid and traded derivatives such as vanilla caps as well as directly related deals such as bond options and binary caps. However, to calculate prices for derivatives when models such as considered in Chapter 6 are used, more expensive computational methods like Monte Carlo simulation have to be used.

The rest of this chapter is arranged as follows. First, we show in Section 7.2 how deterministic methods may be set up in a Hull-White economy. How to implement finite difference methods for the Hull-White model is explained in Section 7.3. Section 7.4 presents and dicusses the computational results reported by the several lattice methods aforementioned. Finally, basic Monte Carlo procedures for valuation are outlined for the model analyzed in Chapter 6.

7.2 Partial Differential Equation of the HW Model

We recall the $\mathbb{Q}$-dynamics of the HW model as introduced in Sect. 5.1:

$$dr(t) = [\Phi(t) - ar(t)]\, dt + \sigma dW(t). \tag{7.1}$$

A method of transforming this stochastic evolution problem into a deterministic one is to use the well-known *Feynman-Kač* formula.

Theorem 5 (Feynman-Kač) *The partial differential equation*

$$V_t + \frac{1}{2}\sigma(x,t)^2 V_{xx} + \mu(x,t)V_x - r(x,t)V = 0$$

with boundary conditon $h(x,T)$ has the solution

$$V(x,t) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r(X,s)ds} h(X,T)\right],$$

where the expectation is taken with respect to the process X defined by

$$dX = \mu(X,t)dt + \sigma(X,t)dW.$$

Proof. A proof can be found in [46]. □

Therefore, under an arbitrage free economy, the value of an interest rate derivative security V, solves the Partial Differential Equation (PDE, for short):

$$\begin{cases} V_t + \frac{1}{2}\sigma^2 V_{rr} + [\Phi(t) - ar]\, V_r - rV &= 0 \\ V(r,T) &= h(r). \end{cases} \tag{7.2}$$

For instance, for the discount bond value we have

$$h(r) = 1,$$

and for the European-style option the corresponding payoff is

$$h(r) = [\phi(P(T,S) - K)]^+$$

at option expiry time T, $T < S$, where binary unit $\phi = +1$ for the call and $\phi = -1$ for the put, as we discussed in Sect. 2.4.

7.2.1 Bypassing Forward Induction

In a recent work, Daglish [21] provided an extension of the Hull and White [35] approach that allows the use of implicit methods by noting that the calculation of Arrow-Debreu prices for interest rate securities is analogous to discretize the Forward Fokker-Planck equation which describes the future evolution of the transition probability densities. This approach substitutes the traditional explicit scheme with the superior implicit Crank-Nicolson method for approximating the aforementioned equation.

In this section we show how the HW model we are considering can be fitted to the initial term structure analytically which is definitively different from both the original algorithm of Hull and White and the above mentioned improvement in which the use of the *forward induction* technique stands [37]. Later, numerical examples confirm that this approach improves computation times and accuracy for derivative pricing whatever the discretization method is used, explicit or implicit.

Consider the following transformation of variables

$$\begin{aligned} x &= r - \Omega(t) \\ \Omega(t) &:= e^{-at}\left(r_0 + \int_0^t e^{au}\Phi(u)du\right). \end{aligned} \tag{7.3}$$

Let us note that $\Omega(t)$ is chosen in such a way that $x(0) = 0$. The price of any derivative in terms of the new variable can be written as $w(x, x)$. We can immediately infer the following relations between V and w

$$\begin{aligned} V(r,t) &\equiv w(x,t) = w(r - \Omega(t), t) \\ V_t &= w_t - (-a\Omega(t) + \Phi(t))\, w_x \\ V_r &= w_x \\ V_{rr} &= w_{xx} \end{aligned} \tag{7.4}$$

where we have used that

$$\frac{d}{dt}\Omega(t) + a\Omega(t) = \Phi(t)$$

Substituing this identities into (7.10) and using $r = x + \Omega(t)$, the PDE reduces to

$$w_t + \frac{1}{2}\sigma^2 w_{xx} - axw_x + (x + \Omega(t))\, w = 0,$$

where the function $\Omega(t)$ may be easily obtained by direct integration of (7.3):

$$\Omega(t) = F(0,t) + \frac{\sigma^2}{2a^2}\left(1 - e^{-at}\right)^2. \tag{7.5}$$

From now we show that with an additional transformation of function we can make the PDE independent of this function $\Omega(t)$. If we define the function $u(x,t)$ as

$$u(x,t) = e^{\int_t^T \Omega(q)dq}\, w(x,t), \tag{7.6}$$

we get the following PDE for u

$$u_t + \frac{1}{2}\sigma^2 u_{xx} - axu_x - xu = 0 \tag{7.7}$$

and we see that this PDE has no longer coefficients dependent on t. By solving (7.7) numerically, approximate values $U(x,t)$ for the exact solution $u(x,t)$ can be calculated. Recall that the price $V(r,t)$ of any interest rate option is recovered from $u(x,t)$ via

$$V(r,t) = e^{-\int_t^T \Omega(q)dq}\, u(r - \Omega(t), t), \tag{7.8}$$

where using the analytic formula for $\Omega(t)$ given in (7.5), the integral of Ω can be calculated as

$$\begin{aligned}\int_t^T \Omega(q)dq = &- \left(\log P(0,T) - \log P(0,t)\right) + \\ &+ \frac{\sigma^2}{2a^3}\left(a(T-t) - 2(e^{-at} - e^{-aT}) + \frac{1}{2}(e^{-2at} - e^{-2aT})\right),\end{aligned} \tag{7.9}$$

which allows to match the initial discount bond curve by means of a consistent family with the model and with no need of *forward induction*.

7.3 Finite-Difference Implementation

Let us consider a general form of the homogeneous one-dimensional parabolic equation given by:

$$V_t + a(x,t)V_{xx} + b(x,t)V_x + c(x,t)V = 0 \tag{7.10}$$

subject to the final condition

$$V(x,T) = g(x). \tag{7.11}$$

We approximate the PDE (7.10) in the bounded domain $[x_{min}, x_{max}] \times [t_{min}, t_{max}]$. Let us define the finite difference operator:

$$\mathcal{L}_h V_n^m := \frac{a_n^m}{h^2}(V_{n+1}^m - 2V_n^m + V_{n-1}^m) + \frac{b_n^m}{2h}(V_{n+1}^m - V_{n-1}^m) + c_n^m V_n^m,$$

where $l = (t_{max} - t_{min})/M$ and $h = (x_{max} - x_{min})/N$. Our main goal is to construct an approximation, $\widehat{V}(x,t)$ of the true solution $V(x,t)$ by using a weighted finite-difference (FD, henceforth) scheme given by

$$\frac{\widehat{V}_n^{m+1} - \widehat{V}_n^m}{l} + (1-\nu)\mathcal{L}_h \widehat{V}_n^{m+1} + \nu \mathcal{L}_h \widehat{V}_n^m = 0. \tag{7.12}$$

As it is well-known, we recall that if:

1. $\nu = 0$ we get the explicit finite-difference scheme,

2. $\nu = \frac{1}{2}$ we get the Crank-Nicolson implicit finite-difference scheme,

3. $\nu = 1$ we get the fully implicit finite-difference scheme.

It may be shown that when $\nu = \frac{1}{2}$ the discretization error is $O(h^2, l^2)$ whereas this error is $O(h^2, l)$ otherwise, see for instance [22, Chap. 8, pp. 139–156] or [51, Chap. 2, pp. 39 and 93]. The implicit methods are also absolutely stable while the explicit method has the avoidable disadvantage of certain stability conditions as we will see later.

Applying $\mathcal{L}_h$ to the FD equation (7.12), we get

$$\begin{aligned} 0 = {} & \widehat{V}_n^{m+1} - \widehat{V}_n^m + \frac{(1-\nu)l a_n^{m+1}}{h^2}(\widehat{V}_{n+1}^{m+1} - 2\widehat{V}_n^{m+1} + \widehat{V}_{n-1}^{m+1}) + \\ & + \frac{\nu l a_n^m}{h^2}(\widehat{V}_{n+1}^m - 2\widehat{V}_n^m + \widehat{V}_{n-1}^m) + \frac{(1-\nu)l b_n^{m+1}}{2h}(\widehat{V}_{n+1}^{m+1} - \widehat{V}_{n-1}^{m+1}) + \\ & + \frac{\nu l b_n^{m+1}}{2h}(\widehat{V}_{n+1}^m - \widehat{V}_{n-1}^m) + (1-\nu) l c_n^{m+1} \widehat{V}_n^{m+1} + \nu l c_n^m \widehat{V}_n^m. \end{aligned} \tag{7.13}$$

Simplifying, we obtain a general form of the weighted scheme discretization

$$\left(-\frac{\nu la_n^m}{h^2}+\frac{\nu lb_n^m}{2h}\right)\widehat{V}_{n-1}^m+\left(1+\frac{2\nu la_n^m}{h^2}-\nu lc_n^m\right)\widehat{V}_n^m+\left(-\frac{\nu la_n^m}{h^2}-\frac{\nu lb_n^m}{2h}\right)\widehat{V}_{n+1}^m=$$
$$=\left(\frac{(1-\nu)la_n^{m+1}}{h^2}-\frac{(1-\nu)lb_n^{m+1}}{2h}\right)\widehat{V}_{n-1}^{m+1}+\left(1-\frac{2(1-\nu)lb_n^{m+1}}{h^2}+(1-\nu)lc_n^{m+1}\right)\widehat{V}_n^{m+1}+$$
$$+\left(\frac{(1-\nu)la_n^{m+1}}{h^2}+\frac{(1-\nu)lb_n^{m+1}}{2h}\right)\widehat{V}_{n+1}^{m+1}, \tag{7.14}$$

Let us introduce the coefficients:

$$\begin{aligned} A_n^m &= \frac{la_n^m}{h^2}-\frac{lb_n^m}{2h} \\ B_n^m &= -\frac{2la_n^m}{h^2}+lc_n^m, \\ C_n^m &= \frac{la_n^m}{h^2}+\frac{lb_n^m}{2h}. \end{aligned} \tag{7.15}$$

Substituting these into (7.14) we finally arrive to the more compact:

$$\begin{aligned} &-\nu A_n^m\widehat{V}_{n-1}^m+(1-\nu B_n^m)\,\widehat{V}_n^m-\nu C_n^m\widehat{V}_{n+1}^m= \\ &=(1-\nu)A_n^{m+1}\widehat{V}_{n-1}^{m+1}+\left(1+(1-\nu)B_n^{m+1}\right)\widehat{V}_n^{m+1}+(1-\nu)C_n^{m+1}\widehat{V}_{n+1}^{m+1}, \end{aligned} \tag{7.16}$$

$$m=0,\ldots,\ M.$$

7.3.1 A Stable Explicit Scheme

Suppose we construct a grid with steps $\Delta x = h$ symmetric along the x-axis where

$$x_{min}=-x_{max}\ , \qquad \left\lfloor\frac{N}{2}\right\rfloor=\left\lfloor\frac{x_{max}}{h}\right\rfloor$$

and steps $\Delta t = l$ along the t-axis. A node (m, n) on the grid is a point where:

$$x_n = nh, \quad n=-\lfloor N/2\rfloor,-\lfloor N/2\rfloor+1,...,0,1,...,\lfloor N/2\rfloor-1,\lfloor N/2\rfloor \tag{7.17}$$
$$t_m = ml, \quad m=0,1,...,M. \tag{7.18}$$

We will next apply the following explicit scheme to the PDE (7.7)

$$\begin{aligned} 0 = U_n^{m+1} - U_n^m + \frac{la_n^{m+1}}{h^2}(U_{n+1}^{m+1} - 2U_n^{m+1} + U_{n-1}^{m+1}) + \\ + \frac{lb_n^{m+1}}{2h}(U_{n+1}^{m+1} - U_{n-1}^{m+1}) + lc_n^m U_n^m. \end{aligned} \tag{7.19}$$

where we have used the notation U_n^m for an approximation to $u_n^m = u(x_n, t_m)$. So we get after solving for U_n^m

$$U_n^m = \frac{1}{1+nhl}\left[\left(\frac{l\sigma^2}{2h^2} + \frac{1}{2}anl\right)U_{n-1}^{m+1} + \left(-\frac{l\sigma^2}{h^2} + 1\right)U_n^{m+1}\left(\frac{l\sigma^2}{2h^2} - \frac{1}{2}anl\right)U_{n+1}^{m+1}\right], \tag{7.20}$$

where we have used the identities

$$\begin{aligned} a_n^{m+1} &:= \frac{1}{2}\sigma^2 \\ b_n^{m+1} &:= -nah \\ c_n^m &:= -nh. \end{aligned} \tag{7.21}$$

By setting $\frac{l\sigma^2}{h^2} = \frac{1}{3}$ the backward recursion (7.20) reduces to

$$U_n^m = \frac{1}{1+nhl}\left(q_d U_{n-1}^{m+1} + q_m U_n^{m+1} + q_u U_{n+1}^{m+1}\right), \tag{7.22}$$

with

$$\begin{cases} q_d := \frac{1}{6} + \frac{1}{2}anl \\ q_m := \frac{2}{3} \\ q_u := \frac{1}{6} - \frac{1}{2}anl\ . \end{cases}$$

Lemma 4 (Stability Condition.) *Let u be the solution of the PDE (7.7) and let U be the solution of (7.22). If $q_i > 0$, and provided that $q_u + q_m + q_d = 1$, then*

$$max\ |u_n^m - U_n^m| \leq AT(l + h^2)$$

for $x_{min} \leq x_n \leq x_{max}$, and $0 \leq t_m \leq T$.

Proof. It is a direct consequence of the results stated in [1, Chap. 2, pp. 44–45] and [51, Chap. 2, pp 45–47] for the canonical heat equation and its explicit discretization. □

For $-\frac{1}{3}\frac{1}{al} < n < \frac{1}{3}\frac{1}{al}$, the numbers $q_i(n)$ are all positive, satisfying stability condition. In order to prevent these quantities from going negative, we cannot use a finite difference grid that is arbitrarily large. At some level $\hat{n} < \frac{1}{3}al$, we want to express $U^m_{\hat{n}}$ in terms of $U^{m+1}_{\hat{n}}$, $U^{m+1}_{\hat{n}-1}$ and $U^{m+1}_{\hat{n}-2}$. By doing so we avoid using $U^{m+1}_{\hat{n}+1}$ and the grid will remain bounded at $\hat{n}$. If we use the following approximations of the *spatial* partial derivatives

$$\begin{aligned} u_{xx} &\approx \frac{U^{m+1}_{\hat{n}} - 2U^{m+1}_{\hat{n}-1} + U^{m+1}_{\hat{n}-2}}{h^2} \\ u_x &\approx \frac{3U^{m+1}_{\hat{n}} - 4U^{m+1}_{\hat{n}-1} + U^{m+1}_{\hat{n}-2}}{2h} \end{aligned} \tag{7.23}$$

we can express $U^m_{\hat{n}}$ as

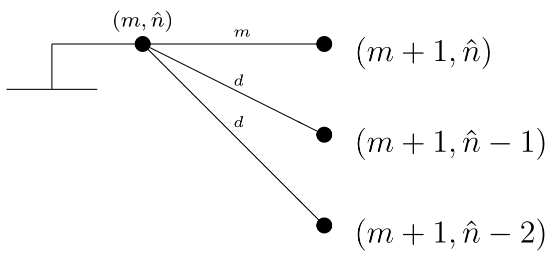
$$U^m_{\hat{n}} = \frac{1}{1+nhl}\left(\hat{q}_{dd}U^{m+1}_{\hat{n}-2} + \hat{q}_d U^{m+1}_{\hat{n}-1} + \hat{q}_m U^{m+1}_{\hat{n}}\right) , \tag{7.24}$$

with

$$\begin{cases} \hat{q}_{dd} &:= \frac{1}{6} - \frac{1}{2}a\hat{n}l \\ \hat{q}_d &:= -\frac{1}{3} + 2a\hat{n}l \\ \hat{q}_m &:= \frac{7}{6} - \frac{3}{2}a\hat{n}l \ . \end{cases}$$

These coefficients are all positive for $\frac{1}{6}\frac{1}{al} < \hat{n} < \frac{1}{3}\frac{1}{al}$.

Figure 7.1: Downward Branching



We can analogously proceed for bounding the grid from below imposing a level $\check{n} > -\frac{1}{3}\frac{1}{al}$. At $\check{n}$ we get

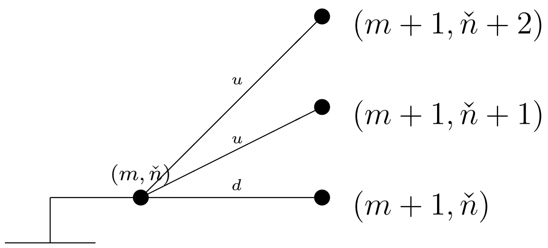
$$U^m_{\check{n}} = \frac{1}{1+nhl}\left(\check{q}_{uu}U^{m+1}_{\check{n}+2} + \check{q}_u U^{m+1}_{\check{n}+1} + \check{q}_m U^{m+1}_{\check{n}}\right) , \tag{7.25}$$

with quantities:

$$\begin{cases} \check{q}_{uu} &:= \frac{1}{6} + \frac{1}{2}a\check{n}l \\ \check{q}_u &:= -\frac{1}{3} - 2a\check{n}l \\ \check{q}_m &:= \frac{7}{6} + \frac{3}{2}a\check{n}l \ . \end{cases}$$

which ara all positive for $-\frac{1}{3}\frac{1}{al} < \check{n} < -\frac{1}{6}\frac{1}{al}$.

Figure 7.2: Upward Branching



Let us note that from the definition (7.6) for exact solutions, we may infer the identical relation for the approximants

$$U_n^m \equiv U(x_n, t_m) = e^{\int_{t_m}^{T} \Omega(q)dq} W_n^m \tag{7.26}$$

which allows us to rewrite the differencing scheme as

$$W_n^m = \frac{e^{-\int_{ml}^{(m+1)l} \Omega(q)dq}}{1+nhl} \left(q_d W_{n-1}^{m+1} + q_m W_n^{m+1} + q_u W_{n+1}^{m+1} \right) .$$

Note that for every m, the integral expression can be calculated analitically by adapting (7.9) to the grid

$$\begin{aligned}\int_{ml}^{(m+1)l} \Omega(q)dq = &- \left(\log P_{m+1}(0) - \log P_m(0) \right) + \\ &+ \frac{\sigma^2}{2a^3} \left(al - 2(1-e^{-al})e^{-aml} + \frac{1}{2}(1-e^{-2al})e^{-2aml} \right), \\ &m = 0, 1, \ldots, M-1.\end{aligned} \tag{7.27}$$

We also note that the quantities

$$\log P_0(0) = 0, \, \log P_1(0), \, \ldots, \, \log P_M(0)$$

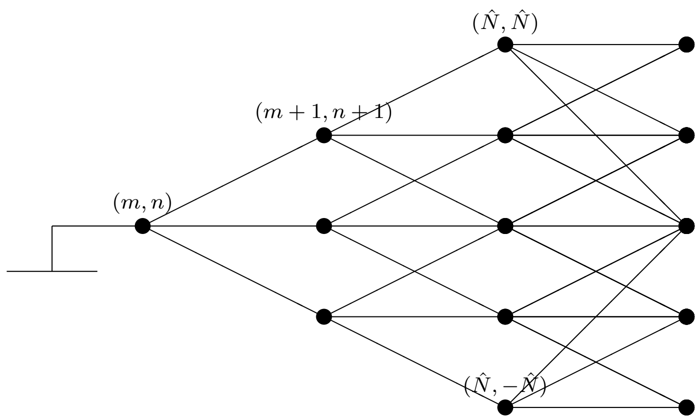
may be computed by means of consistent calibration, deeply analyzed in previous chapters. With the results provided by the joint calibration for the parameters $\hat{\boldsymbol{p}}$ and $\hat{z}(\hat{\boldsymbol{p}})$ from a set of instrument observations $[P_1^o \, \ldots \, P_{M'}^o]$ and $[\, C_1^o \, \ldots \, C_{N'}^o \,]$, we recall that

$$\log P_m(0) = -\int_0^{x_m} G(z(\hat{\boldsymbol{p}}), \hat{\boldsymbol{p}}, q)dq = \sum_{j=1}^{n_p} M_{mj}(\hat{\boldsymbol{p}})\hat{z}_j.$$

It is also important to point out that for every node point (m, n) the approximate price $\widehat{V}_n^m$ of any derivative is equivalent to W_n^m by construction.

The finite difference method outlined below can be implemented as follows. For a derivative with maturity T, and a given number of steps M we may calculate the step-sizes as $l = T/M$ and $h = \sigma\sqrt{3l}$. Then, we set $\hat{N} = \lceil \frac{1}{6al} \rceil$ which is the first integer value on the right of the lowest bound of $\hat{n}$ for which central and downward branching produce positive coefficients in the backward recursions (7.22) and (7.24). Therefore, reasoning by symmetry, making $\hat{n} = -\check{n} = \hat{N}$ we can bound the grid without making this explicit scheme unstable.

Figure 7.3: An example of stable explicit grid for x with $l = 10a = 1$.



For $m = 0, \ldots, \hat{N}$ we may set up a normal branching with coefficients q_u, q_m, q_d.
For $m = \hat{N}+1, \ldots, \ldots, M$ we build a bounded scheme at $n = \hat{N}$ and $n = -\hat{N}$.

We remark that

$$W_n^M = h(x_n, Ml)$$

for $n = -\hat{N}, \ldots, 0, \ldots, \hat{N}$ are the known payoff values which allow to start the backward recursion. Finally, we recall that due to the fact that the grid is bounded additional *spatial* boundary conditions are no needed.

7.3.2 A Crank-Nicolson Scheme

From now, we apply the Crank-Nicolson implicit scheme to the PDE (7.7). In this case, introducing the following ratios

$$\varrho_1 = \frac{l}{h^2},\ \varrho_2 = \frac{l}{h}$$

and particularizing ν to $\frac{1}{2}$ the general weighted scheme (7.16) may be written as

$$-\alpha_n U_{n-1}^m + (1-\beta_n)\,U_n^m - \gamma_n U_{n+1}^m = \alpha_n U_{n-1}^{m+1} + (1+\beta_n)\,U_n^{m+1} + \gamma_n U_{n+1}^{m+1}, \quad (7.28)$$

where

$$\begin{cases} \alpha_n &= \frac{1}{2}A_n^m = \frac{1}{4}\sigma^2\varrho_1 + \frac{1}{4}nah\varrho_2\,, \\ \beta_n &= \frac{1}{2}B_n^m = -\frac{1}{2}\sigma^2\varrho_1 - \frac{1}{2}nhl\,, \\ \gamma_n &= \frac{1}{2}C_n^m = \frac{1}{4}\sigma^2\varrho_1 - \frac{1}{4}nah\varrho_2\,. \end{cases}$$

In this case, we consider again a symmetric *spatial* domain $[-x_{max}, x_{max}]$ where the grid points (m,n) are defined as

$$\begin{array}{lllllll} x_n &=& -x_{max} + nh, & n &=& 0,1,\ldots,N \\ t_m &=& ml, & m &=& 0,1,\ldots,M. \end{array} \quad (7.29)$$

being $h = 2x_{max}/N$ and $l = T/M$. As suggested by Cairns [17] and Daglish [21] we fix x_{max} to $5\frac{\sigma}{2a}$ and impose as well homogeneous Dirichlet boundary conditions

$$u(x_{max}, t_m) = u(x_{min}, t_m) = 0 \quad m = 0,1,\ldots,M. \quad (7.30)$$

Therefore, we recall that

$$U_0^m = u_0^m = 0,\ U_N^m = u_N^m = 0,\ U_n^M = h(x_n, Ml)$$

for $m = 0,1,\ldots,M$ and $n = 0,1,\ldots,N$ are known values.

In order to clarify how to compute the approximate values for the solution at the grid points, taking the above mentioned conditions into account, we may break down the difference equation (7.28) as follows

$$\begin{cases} n=0 & \rightarrow \quad U_0^m = 0 \\ n=1 & \rightarrow \quad (1-\beta_1)U_1^m - \gamma_1 U_2^m = (1+\beta_1)U_1^m + \gamma_1 U_2^m \\ n=2,\ldots,N-2 & \qquad \text{Eqn. (7.28)} \\ n=N-1 & \rightarrow \quad -\alpha_{N-1}U_{N-2}^m + (1-\beta_{N-1})U_{N-1}^m = \alpha_{N-1}U_{N-2}^m + (1+\beta_{N-1})U_{N-1}^m \\ n=N & \rightarrow \quad U_N^m = 0 \end{cases} \tag{7.31}$$

Written in matrix form the above problem we have the linear system:

$$D_1 \boldsymbol{U}^m = D_2 \boldsymbol{U}^{m+1} \tag{7.32}$$

where

$$D_1 = \begin{bmatrix} 1-\beta_1 & -\gamma_1 & & & & \\ -\alpha_2 & 1-\beta_2 & -\gamma_2 & & & \\ & -\alpha_3 & 1-\beta_3 & -\gamma_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\alpha_{N-2} & 1-\beta_{N-2} & -\gamma_{N-2} \\ & & & & -\alpha_{N-1} & 1-\beta_{N-1} \end{bmatrix},$$

$$D_2 = \begin{bmatrix} 1+\beta_1 & \gamma_1 & & & & \\ \alpha_2 & 1+\beta_2 & \gamma_2 & & & \\ & \alpha_3 & 1+\beta_3 & \gamma_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \alpha_{N-2} & 1+\beta_{N-2} & \gamma_{N-2} \\ & & & & \alpha_{N-1} & 1+\beta_{N-1} \end{bmatrix},$$

are nearly tridiagonal and

$$\boldsymbol{U}^m = \left[U_1^m \, U_2^m \, \ldots \, U_{N-1}^m \right]^T.$$

Thus, provided we exactly know the terminal condition

$$\boldsymbol{U}^M = \boldsymbol{u}^M = \boldsymbol{h}(Ml) = [\, h(x_1, Ml) \, \ldots \, h(x_{N-1}, Ml) \,]^T$$

we are able to solve this difference equation obtaining the matrix

$$\boldsymbol{U} = \left[\boldsymbol{U}^0\,\boldsymbol{U}^1\,\ldots\,\boldsymbol{U}^M\right] ,$$

which represents the approximate solution to the matrix

$$\boldsymbol{u} = \left[\boldsymbol{u}^0\,\boldsymbol{u}^1\,\ldots\,\boldsymbol{u}^M\right] .$$

Finally, note that if we apply the relation (7.26) that links the approximants $\boldsymbol{U}^m$ and $\boldsymbol{W}^m$ next we may to construct the following two-level difference equation:

$$\begin{aligned} D_1\boldsymbol{W}^m &= D_2\boldsymbol{Z}^{m+1} \quad m = 0, 1, \ldots, M-1 \\ \boldsymbol{W}^M &= \boldsymbol{w}^M , \end{aligned} \tag{7.33}$$

where $\boldsymbol{Z}^m = e^{-\int_{ml}^{(m+1)l}\Omega(q)dq}\boldsymbol{W}^m$ is known at any time stage on the above backward recursion formula and $\boldsymbol{w}^M = \boldsymbol{h}(Ml)$.

7.4 Numerical Examples

The explicit finite difference (EFD henceforth) and the implicit Crank-Nicolson (CN for short) are different from the trinomial tree approach of Hull and White [35] (HWT hereafter), as we have analyzed in the previous sections.

Due to the fact that with the HWT algorithm *forward induction* is needed, this method is slower than the EFD algorithm even if we use the Arrow-Debreu prices for partially avoiding backward recursion as we suggested in [44, Chap. 2, pp. 23] for vanilla European-style bond options.

Figure 7.4: Discrete data for initial discount bond estimation. All rates are expressed with continuous compounding.

MATURITY, x	0.083	0.25	1	2	3	4
ZERO RATE, $R^o(x)$	3.46%	3.54%	4.02%	4.51%	4.79%	4.98%
MATURITY, x	5	6	7	8	9	10
ZERO RATE, $R^o(x)$	5.13%	5.24%	5.35%	5.44%	5.51%	5.56%

Bond Options

As first examples, we consider the pricing of a two-year and three-year vanilla put options, written on a five-year discount bond. We assume parameters $a = 0.1$ and $\sigma = 0.01$, which are of similar order of magnitude of the ones observed on the markets. In our analysis we also use the discount curve given in the table on Figure 7.4 which is estimated by means of the lowest dimensional consistent family with the model introduced in (5.27) as the MIN family. For finite-difference EFD and CN methods, both bond and option prices are computed on the same grid. The bond price at time $t = S = Ml$ is subject to the well-known condition $P_n^M = 1$. The option price at time $t = T = M'l$, with $M' < M$, is subject to the following condition

$$W_n^{M'} = (K - P_n^{M'})^+. \tag{7.34}$$

We did not use closed-form formulas for the bond price. Therefore we, first of all, compute de bond price starting recursion form $t = S$ up to $t = T$. Then we apply the final condition for the option price. The latter means to replace the bond values with option values computed by (7.34) on the same grid (m, n). Then we compute option prices using the grid up to $t = 0$.

For the trinomial tree approach HWT, we tried to price faster European-style options by using elementary Arrow-Debreu prices $Q_n^{M'}$, evaluated at the options' maturity following [44, Sect. 2.2.4, pp. 22–23]. By means of *forward induction* and restricting the backward recursion to just computing the bond grid values at time stages $m = M', \ldots, M - 1$ we partially avoid the need of a full *forward-then-backward* methodology.

With regard to European-style options, the table on Figure 7.5 and Figure 7.6 confirms that both EFD and CN algorithms are superior than HWT approach in terms of accuracy. We note that CN algorithm converge very fast due that the Crank-Nicolson method is second order accuracy in time. Convergence to within two decimals of the price in basispoints is reached with $l \approx 0.01$. On the other hand, the EFD converges slightly faster than the HWT algorithm, but the difference is not very large.

As concern to time consumption performance, we report in the table on Figure 7.7 the calculation times needed for the three algorithms. We see that the EFD algorithm is by far faster than the HWT algorithm. The CN method is slower than both of them due to its more matricial nature and because we use a LU

Figure 7.5: Prices (in basispoints of the notional) for put options on 5yr discount bond.

		European-style			American-style		
Mat. / Strike	l	EFD	CN	HWT	EFD	CN	HWT
	0.50	0.44	1.07	1.34	18.3	78.6	18.4
	0.40	1.07	1.69	2.11	27.5	78.6	27.6
	0.25	0.75	1.10	1.19	45.3	86.0	45.5
	0.20	0.94	1.08	1.32	53.7	85.5	53.8
$T = 2yr$	0.10	0.91	1.04	1.08	72.4	86.7	72.5
$K = 0.78$	0.05	1.00	1.04	1.09	82.9	89.8	83.0
	0.04	1.00	1.04	1.06	85.0	90.4	85.0
	0.025	1.03	1.04	1.07	87.8	91.2	87.8
	0.02	1.03	1.04	1.06	88.6	91.5	88.6
	0.01	1.04	1.04	1.05	90.2	92.0	90.2
	Exact		1.04				
	0.50	4.14	5.01	5.82	647	779	648
	0.40	2.79	2.89	3.69	680	779	680
	0.25	4.84	4.79	5.65	716	779	717
	0.20	4.12	4.84	4.73	729	779	730
$T = 3yr$	0.10	4.73	4.83	5.04	755	779	755
$K = 0.85$	0.05	4.67	4.82	4.82	767	779	767
	0.04	4.80	4.82	4.92	769	779	769
	0.025	4.83	4.82	4.90	773	779	773
	0.02	4.83	4.82	4.89	774	779	774
	0.01	4.82	4.82	4.85	776	779	776
	Exact		4.82				

Note: We recall that HWT refers to the Hull-White trinomial tree, and EFD and CN to the stable explicit and the Crank-Nicolson implicit algorithms, respectively. For the HWT and EFD method, we assume $h = \sigma\sqrt{3l}$ as in [35]. For the Crank-Nicolson method we set $h = \sigma\sqrt{2l}$ with boundaries at $\pm 5\frac{\sigma}{2a}$ following [17] and [21].

tridiagonal solver to solve the system. However, we remark that for achieving similar accuracies on European-style valuation, the Crank-Nicolson may be even the fastest method with less consumption of time for the involved calculations.

While pricing a vanilla European bond option by finite differences is certainly instructive in order to give an insight of which numerical method may be more efficient, it is not very practical in the real market situations because we are

Figure 7.6: Relative pricing errors of vanilla 2yr put on 5yr discount bond.

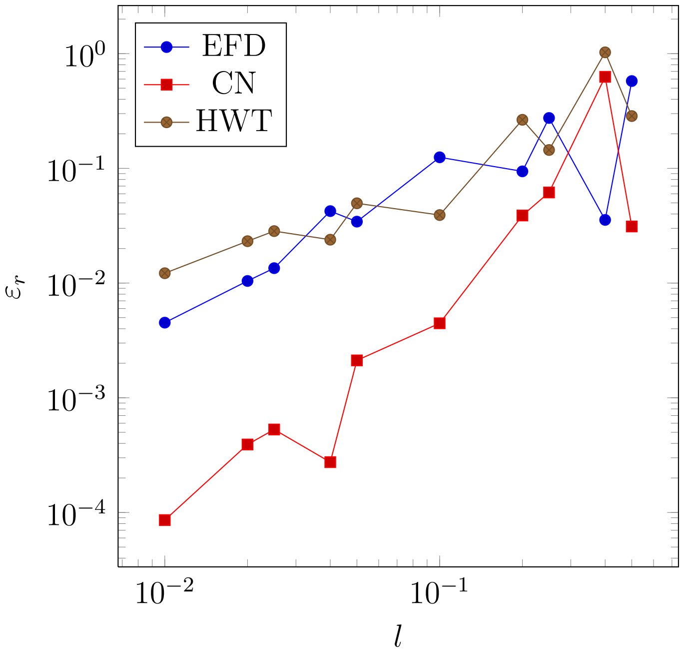


Figure 7.7: Calculation times in milliseconds, running MATLAB on an Intel Core 2 Duo P8600 @ 2.39GHz computer for a $2yr$ option on a $5yr$ discount bond, with $a = 0.1$ and $\sigma = 0.01$.

	American-style		
l	EFD	CN	HWT
0.50	0.97	2.2	1.7
0.40	0.67	2.5	1.8
0.25	0.86	2.9	2.7
0.20	0.98	3.7	3.2
0.10	1.82	17.7	6.6
0.05	4.22	104	13.6
0.04	4.82	401	17.4
0.02	10.8	3367	48.1
0.01	40.2	33366	142

equipped with well-known closed-form solutions for the HW model. Therefore, we may apply these schemes to American options, for which exact formulas are not available. To avoid arbitrage, the option value at each point in the grid (m, n) cannot be less than the intrinsic value (the immediate payoff if the option is exercised). For instance, for a vanilla American-style put on a discount bond, this means

$$w(x,t) \geq (K - P(T,S))^+, \quad t < T.$$

From a strictly practical point of view, taking this condition into account is not very difficult. After computing W_n^m, we should check for the possibility of early exercise, and set

$$W_n^m = \max(W_n^m, K - P_n^m).$$

Therefore, if we want to price American-style options, we need to construct the full grid containing the bond prices, P_n^m, and the separate grid with option prices, W_n^m.

Due to accuracy issues, we might prefer adopting a Crank-Nicolson scheme. We remark that in such a case for each time layer m we have the scheme (7.33) and we may equally compute the chance of early excercise after the calculation of the vector $\boldsymbol{W}^m$

$$\boldsymbol{W}^m = \max(\boldsymbol{W}^m, K - \boldsymbol{P}^m),$$

allowing the implementation of this method with a LU direct solver.

We report in the table on Figure 7.5 the results for all three methods. The results for the CN method are the best in terms of the speed of convergence and compatible with the numerical approximations reported by commercial black boxes such as DerivaGem or FINCAD. As can be seen, the results for the EFD and HWT are slightly similar with less time consumption for the stable explicit method (Figure 7.7).

Interest Rate Caps

Earlier in this work we have shown how the value of a vanilla cap can be expressed as a portfolio of puts on discount bonds.

As is explained in Sect. 3.2, the price of such a cap contract with strike K and resettlement dates $x_0, \ldots, x_{n-1}$ may be determined wih the following representation for the stream of payoffs

$$h_{\gamma_j}(x_{j-1}) = (1 + \tau K)\,(\kappa - P_j(x_{j-1}))^+ \quad j = 1, \ldots, n \tag{7.35}$$

where $\kappa = (1+\tau K)^{-1}$. Therefore, the numerical problem reduces to the valuation of the corresponding portfolio of these European-style bond options which we have discussed in detail above.

A *digital cap* is an instrument that has the same characteristics as a vanilla cap, except that the payoff is a fixed amount paid if the final floating rate is above the strike. Therefore, the payoff of any digital j-caplet which composes it can be represented as

$$h_{\delta_j}(x_j) = \mathbb{1}_{\{L_j(x_{j-1})-K>0\}} \qquad j = 1, \dots, n \tag{7.36}$$

if we take as unitary the notional amount. For instance, typical Chicago Board of Trade binary options on the target US federal funds rate take as notional the amount of \$1000.

By following similar algebraic manipulations as we have used in Sect. 3.2 for the payoff of a vanilla we may numerically compute the price of a digital cap by considering that the sequence of payoffs at times $x_0, \dots, x_{n-1}$,

$$h_{\delta_j}(x_{j-1}) = P_j(x_{j-1})\mathbb{1}_{\{P_j(x_{j-1})-\kappa<0\}} \qquad j = 1, \dots, n, \tag{7.37}$$

is equivalent to (7.36). We remark that now, κ is $(1 + \tau K)^{-1}$ once again.

As next examples, we analyze the pricing of a five-year and ten-year vanilla and digital caps, with semi-annual tenor. We assume the same parameters for the HW model, $a = 0.1$ and $\sigma = 0.01$, as we have used in the case of bond options. Moreover, we employ as well the MIN family to estimate the zero rate curve from the discrete data reported by table on Figure 7.4. In the Appendix C we derive closed-form formulas for the digital caps under the assumptions of Gaussian Heath-Jarrow-Morton models.

From table on the Figure 7.8 and Figure 7.9 we see first that even with the stable explicit procedure described in this work, we already get vanilla cap prices accurate to within one basispoint taking time-steps close to the month ($l \approx 0.1$). Also we see that the vanilla numerical approximations converge much more quickly to the theoretical prices than the corresponding digital approximants, maybe a direct consequence of the severe non-smooth nature of the stream of payoffs which determine the value of these binary options.

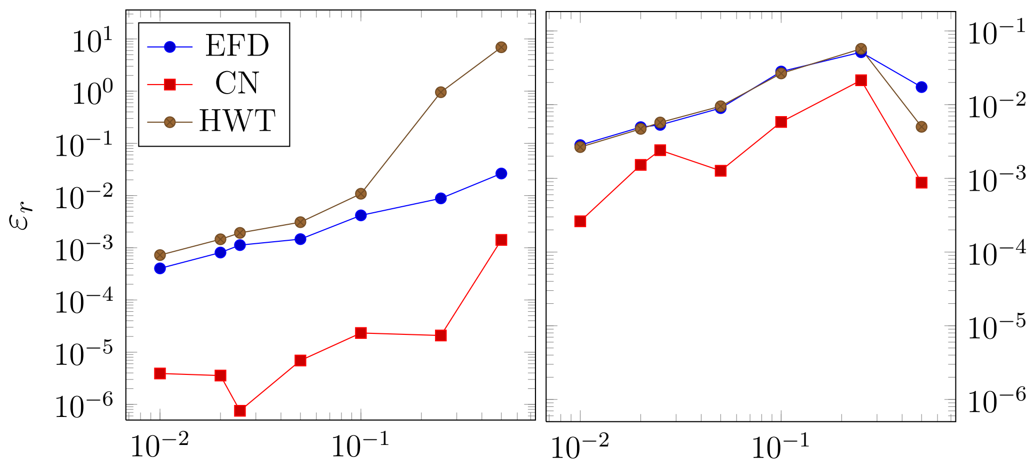
We note that in both cases, digital or plain vanilla, the implicit Crank-Nicolson scheme converges faster than the other procedures considered.

Figure 7.8: Prices of vanilla/digital European-style caps with semi-annual tenor.

		Vanilla			Digital		
Mat. / Strike	l	EFD (%)	CN (%)	HWT (%)	EFD	CN	HWT
	0.50	5.64	5.50	43.6	7.59	7.46	7.49
	0.25	5.55	5.50	10.7	7.07	7.61	7.03
$T = 10yr$	0.1	5.52	5.50	5.56	7.67	7.41	7.65
	0.05	5.51	5.50	5.52	7.39	7.45	7.39
$K = 5.5\%$	0.025	5.51	5.50	5.51	7.42	7.47	7.41
	0.02	5.50	5.50	5.51	7.49	7.47	7.49
	0.01	5.50	5.50	5.50	7.48	7.46	7.48
	Exact	5.50			7.46		
	0.50	3.11	3.16	41.1	4.33	4.47	4.20
	0.25	3.14	3.16	8.34	4.67	4.50	4.62
$T = 5yr$	0.1	3.16	3.16	3.20	4.63	4.51	4.63
	0.05	3.16	3.16	3.17	4.69	4.51	4.68
$K = 5\%$	0.025	3.16	3.16	3.17	4.52	4.53	4.52
	0.02	3.16	3.16	3.17	4.53	4.52	4.53
	0.01	3.16	3.16	3.16	4.51	4.52	4.51
	Exact	3.16			4.52		

The numerical approximations for vanilla caps are expressed in percent of the notional amount whereas the approximants for digital caps are presented in unitary terms. Exact valuation for binary options is worked out in Appendix C.

Figure 7.9: Relative valuation errors of 10yr vanilla/digital caps with semi-annual tenor.



7.5 Monte Carlo Simulation for Consistent HJM Models

As we have discussed above with the finite difference methods, we are typically in front of a derivative pricing problem where we cannot evaluate analytically the fundamental arbitrage-free equation

$$\begin{aligned} V(h,0) &= \mathbb{E}^{\mathbb{Q}}\left[D(0,T)h(T)\right] = \\ &= \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_0^T r(u)\,du\right)h(T)\right]. \end{aligned} \tag{7.38}$$

whether we combine or not it with the Feynman-Kač formula in order to produce the corresponding valuation PDE.

The lattice methods described in the previous sections assumed that there was a short-rate realization for the HJM model under consideration. When the HJM model considered is not associated to a low dimensional markovian system being the implied short-rate process r, one of the state variables, lattice-based computing times increase very significantly and could even be impossible to implement. Then, Monte Carlo methods offer an effective and popular alternative to lattice methods.

7.5.1 Basic Monte Carlo

If we recover the one-factor HJM model considered in Chapter 6

$$dF(t,T) = \{\dots\}\,dt + (\alpha + \beta(T-t))\,e^{-a(T-t)}\,dW(t), \tag{7.39}$$

we notice, in fact, that the spot rate process r is not Markovian since does not belong to the Ritchken and Sankarasubramanian class [50].

Even though the short-rate process is not Markovian, there may yet exist a higher-dimensional Markov process having the short rate as one of its components. At this point, we remark that the volatility function $\sigma(t,T)$ of the model (7.39) can be expressed as a sum of separable into time and maturity dependent factors

$$\begin{aligned} \sigma(t,T) &= e^{at}\left((\alpha+\beta T)e^{-aT}\right) - te^{at}\left(\beta e^{-aT}\right) = \\ &= \sum_{i=1}^{2} \varsigma_i(t)\varrho_i(T). \end{aligned} \tag{7.40}$$

Therefore, as it is shown by [18, Propos. 1, pp. 4] in this case we need $\frac{2}{2}(2+3)$ state variables to determine the forward rate via a suitable markovian system where two of the state variables are stochastic and describe the non-Markovian nature of the short rate process. Thus, assuming we know an analytically treatable relation between these stochastic variables and the spot rate process, we finally conclude that, at best, two dimensions and time lattice-based schemes are needed in order to approximate derivative prices. Consequently a brief analysis of simulation techniques have full sense for this kind of HJM model.

From now, let us take as an example the model the one-factor humped volatility model we have analyzed in previous chapter. We remark that the following discussion may be easily extended for any HJM model. As we shown in (6.6) we know how the forward curve produced by the model evolves. In particular, note that the expression

$$r(t) \equiv f_t(0) = f^o(t) + g_1(t) + h_1(t) + \alpha Z_1(t) + (\beta - a\alpha)\, Z_2(t), \tag{7.41}$$

describes how the spot rate evolves in time. We will assume, without loss of generality, that, as before, we want to evaluate the price at time 0 of a security V with maturity in time T.

Let us consider the following procedure:

1. Discretize the period $[0, T]$ into M intervals of equal length $l = T/M$ and define $t_m = ml$.

2. For S simulations denoted by ω, simulate M i.i.d. standard normal random variables $\xi(t_m) \sim \mathcal{N}(0, 1)$ for $m = 1, \ldots, M$. Note that the vector SDE (6.3) is linear in the narrow sense [39], with explicit solution

$$Z_t = \Phi_t \int_0^t \Phi_s^{-1} b\, dW_s, \tag{7.42}$$

where

$$\Phi_t = e^{At} = e^{-at} \begin{bmatrix} 1 + at & -a^2 t \\ t & 1 - at \end{bmatrix}.$$

Therefore, both $Z_{1,2}$-variables are centered Gaussian variables.

3. Calculate the simulated path ω of $r(t,\cdot)$ as follows; for $m = 1,\ldots,M$ let

$$r(t_m,\omega) = f^o(t_m)+g_1(t_m)+h_1(t_m)+\alpha\sqrt{v_{1,m}}\xi(t_m,\omega)+(\beta - a\alpha)\sqrt{v_{2,m}}\xi(t_m,\omega) \tag{7.43}$$

where deterministic quantities

$$\begin{aligned} v_{1,m} &= \frac{5}{4a}\left(1-\exp(-2at_m)\right) - \left(\frac{3}{2}t_m + \frac{1}{2}at_m^2\right)\exp(-2at_m), \text{ and;} \\ v_{2,m} &= \frac{1}{4a^3}\left(1-\exp(-2at_m)\right) - \left(\frac{1}{2a^2}t_m + \frac{1}{2a}t_m^2\right)\exp(-2at_m), \end{aligned}$$

are derived from *Itô isometry* property of the stochastic integral (7.42).

4. Evaluate $V(T,\omega) \equiv V(T, r(t_m,\omega))$ for simulation ω.

5. Evaluate the random discounted value of the derivative payoff

$$X(\omega) = \exp\left(-\sum_{m=0}^{M-1} r(t_m,\omega)l\right) V(T,\omega). \tag{7.44}$$

6. Conclude with the calculation of the mean value

$$\bar{X} = \frac{1}{S}\sum_{\omega\in\Omega} X(\omega).$$

which is our Monte Carlo estimate of the price.

We point out that, by construction, this simulation procedure naturally makes suitable the use of initial consistent families $f^o(\cdot)$ with the model. As an instructional example, we consider the pricing of a one-year European-style option on a three-year discount bond.

For evaluating the payoff

$$V(T,\omega) \equiv (K - P_T(x,\omega))^+$$

we use the same strategy outlined in Sect. 6.3 under the Musiela parameterization. In this case, we have to pay our attention in just the two-year point, and we directly compute it by simulating forward curves $f_T(x,\omega)$ up to T-time –one-year forward in this example. Then we have to integrate them over time-to-maturity up to x-point in order to determine each simulated realization $P_1(2,\omega)$ of the discount bond composing the payoff.

We assume the model parameters $\alpha = 0.0075$, $\beta = 0.005$ and $a = 0.15$. We employ the HMC family as initial family $f^o(\cdot)$ for the Monte Carlo runs (6.6)–(7.43). Finally, we also remark that this consistent estimation is again carried out from the discrete data reported by table on Figure 7.4.

From table on Figure 7.10 we see that this procedure gives rise to two types of error: simulation error and discretization error. First, the numer of sample paths, S, is finite. In fact, this means that $\bar{X}$ is a random variable. Second, the discretization of the period $[0, T]$ result in one more error: the approximation of the Riemann integral with non-smooth integrand

$$\int_0^T r(u, \omega)\, du,$$

by the sum in (7.44). We see that both errors can be reduced to a limited extent by increasing the number of simulations and reducing the step size l. We must be very careful here. We are showing a particular and basic *implementation* of simulation approach, and could be certainly improved by reworking the simulation based on (6.6)–(7.43) dynamics in a more efficient way or by enhancing the elementary discretization scheme we have used to approximate the integral expression which involves discounting.

Figure 7.10: Prices (in percent of the notional) for a 1yr put option on 3yr discount bond.

	MC Estimator			MC Standard Error		
S / l	0.1	0.05	0.01	0.1	0.05	0.01
100	1.974	2.252	2.478	0.18	0.17	0.17
1000	2.332	2.321	2.315	0.06	0.06	0.06
10000	2.347	2.333	2.346	0.018	0.018	0.018
100000	2.334	2.337	2.330	0.006	0.006	0.006
Exact		2.377				

Appendix A

Forward Rate Models

A.1 The HJM Framework

Proposition 1 (Leibniz Rule for Stochastic Integrals) *Consider for any fixed parameter* $s \in [t_0, T]$*, the Itô process defined by*

$$dg(t,s) = \alpha(t,s)dt + \sum_{j=1}^{q} \beta_j(t,s)dW_j(t) \tag{A.1}$$

with $t \in [t_0, s]$*. Then, the dynamics for the stochastic process* $H(t,T) = \int_t^T g(t,s)ds$ *is*

$$dH(t,T) = \left[\left(\int_t^T \alpha(t,s)ds\right) - g(t,t)\right]dt + \sum_{j=1}^{q}\int_t^T \beta_j(t,s)ds \; dW_j(t). \tag{A.2}$$

Proof. Assume that the $\mathbb{R}^q$-valued stochastic processes

$$\beta(t,T) = [\; \beta_1(t,T) \quad \beta_2(t,T) \quad \dots \quad \beta_q(t,T) \;]$$

$$W(t) = [\; W_1(t) \quad W_2(t) \quad \dots \quad W_q(t) \;]^T$$

are given.

Then for any $t_1 \in [t_0, T]$, the differential (A.1) may be written in integral form:

$$g(t_1,s) = g(t_0,s) + \int_{t_0}^{t_1} \alpha(t,s)dt + \int_{t_0}^{t_1} \beta(t,s)dW(t) \tag{A.3}$$

Thus,

$$\begin{array}{rcl}
H(t_1,T) & = & \int_{t_1}^{T} g(t_1,s)ds \\
 & = & \int_{t_1}^{T} g(t_0,s)ds + \int_{t_1}^{T}\left(\int_{t_0}^{t_1}\alpha(t,s)dt\right)ds + \int_{t_1}^{T}\left(\int_{t_0}^{t_1}\beta(t,s)dW(t)\right)ds \\
* & = & \int_{t_1}^{T} g(t_0,s)ds + \int_{t_0}^{t_1}\left(\int_{t_1}^{T}\alpha(t,s)ds\right)dt + \int_{t_0}^{t_1}\left(\int_{t_1}^{T}\beta(t,s)ds\right)dW(t) \\
 & = & \int_{t_0}^{T} g(t_0,s)ds - \int_{t_0}^{t_1} g(t_0,s)ds + \int_{t_0}^{t_1}\left(\int_{t}^{T}\alpha(t,s)ds\right)dt - \int_{t_0}^{t_1}\left(\int_{t}^{t_1}\alpha(t,s)ds\right)dt + \\
 & & + \int_{t_0}^{t_1}\left(\int_{t}^{T}\beta(t,s)ds\right)dW(t) - \int_{t_0}^{t_1}\left(\int_{t}^{t_1}\beta(t,s)ds\right)dW(t) \\
** & = & H(t_0,T) + \int_{t_0}^{t_1}\left(\int_{t}^{T}\alpha(t,s)ds\right)dt + \int_{t_0}^{t_1}\left(\int_{t}^{T}\beta(t,s)ds\right)dW(t) - \\
 & & - \int_{t_0}^{t_1} g(t_0,s)ds - \int_{t_0}^{t_1}\left(\int_{t_0}^{s}\alpha(t,s)dt\right)ds - \int_{t_0}^{t_1}\left(\int_{t_0}^{s}\beta(t,s)dW(t)\right)ds \\
 & = & H(t_0,T) + \int_{t_0}^{t_1}\left(\int_{t}^{T}\alpha(t,s)ds\right)dt + \int_{t_0}^{t_1}\left(\int_{t}^{T}\beta(t,s)ds\right)dW(t) - \\
 & & - \int_{t_0}^{t_1}\left(g(t_0,s) + \int_{t_0}^{s}\alpha(t,s)dt + \int_{t_0}^{s}\beta(t,s)dW(t)\right)ds \\
*** & = & H(t_0,T) + \int_{t_0}^{t_1}\left(\int_{t}^{T}\alpha(t,s)ds\right)dt + \int_{t_0}^{t_1}\left(\int_{t}^{T}\beta(t,s)ds\right)dW(t) - \int_{t_0}^{t_1} g(s,s)ds \\
 & = & H(t_0,T) + \int_{t_0}^{t_1}\left[\int_{t}^{T}\alpha(t,s)ds - g(t,t)\right]dt + \int_{t_0}^{t_1}\left(\int_{t}^{T}\beta(t,s)ds\right)dW(t)
\end{array}$$

the differential for the process $H(t,T)$ may be deduced. We have used the Fubini Theorem in its classical version and the extended version for Stochastic Integrals –see Ikeda and Watanabe [36] and Heath et al. [31]. For the identity (***), the equation (A.3) has been used. □

A.2 From HJM to Short-Rate Models

Proposition 2 *Suppose that $F(0,T)$, $\alpha(t,T)$ and $\sigma(t,T)$ are differentiable in T with $\int_0^T |\partial_u F(0,u)|\, du < \infty$. Then the short-rate process is an Itô process of the form*

$$dr(t) = \zeta(t)dt + \sigma(t,t)dW(t), \tag{A.4}$$

where

$$\zeta(t) = \alpha(t,t) + \partial_t F(0,t) + \int_0^t \partial_t\alpha(s,t)\, ds + \int_0^t \partial_t\sigma(s,t)dW(s)$$

Proof. Fix a time s with $0 \le s \le t < \infty$, we can rewrite $\alpha(s,t)$ and $\sigma(s,t)$ vector as follows,

$$\begin{aligned}
\alpha(s,t) &= \alpha(s,s) + \int_s^t \frac{\partial\alpha}{\partial z}(s,z)dz \\
\sigma(s,t) &= \sigma(s,s) + \int_s^t \frac{\partial\sigma}{\partial z}(s,z)dz
\end{aligned} \tag{A.5}$$

With $s = 0$, we can express $F(0,t)$ as

$$F(0,t) = r(0) + \int_0^t \frac{\partial F}{\partial z}(0,z)dz. \tag{A.6}$$

Recall now that

$$r(t) = F(t,t) = F(0,t) + \int_0^t \alpha(s,t)\, ds + \int_0^t \sigma(s,t)dW(s). \tag{A.7}$$

Thus, by substituting equations (A.5) and (A.6) into (A.7), we have

$$\begin{aligned} r(t) = r(0) + \int_0^t \frac{\partial F}{\partial z}(0,z)dz + \int_0^t \alpha(z,z)dz + \int_0^t \left(\int_s^t \frac{\partial \alpha}{\partial z}(s,z)dz \right) ds + \\ + \int_0^t \left(\int_s^t \frac{\partial \sigma}{\partial z}(s,z)dz \right) dW(s) + \int_0^t \sigma(z,z)dW(z) \end{aligned} \tag{A.8}$$

Applying the Fubini Theorem in its classical and extended version for Stochastic Integrals, gives

$$\begin{aligned} r(t) = r(0) + \int_0^t \frac{\partial F}{\partial z}(0,z)dz + \int_0^t \alpha(z,z)dz + \int_0^t \left(\int_0^z \frac{\partial \alpha}{\partial z}(s,z)ds \right) dz + \\ + \int_0^t \left(\int_0^z \frac{\partial \sigma}{\partial z}(s,z)dW(s) \right) dz + \int_0^t \sigma(z,z)dW(z), \end{aligned} \tag{A.9}$$

and reordering:

$$\begin{aligned} r(t) &= r(0) + \int_0^t \left[\left(\frac{\partial F}{\partial z}(0,z) + \alpha(z,z) + \int_0^z \frac{\partial \alpha}{\partial z}(s,z)ds + \right.\right. \\ &\left.\left. + \int_0^z \frac{\partial \sigma}{\partial z}(s,z)dW(s) \right) dz + \sigma(z,z)dW(z) \right] = \\ &= r(0) + \int_0^t \left(\zeta(z)dz + \sigma(z,z)dW(z) \right), \end{aligned} \tag{A.10}$$

the differential (A.4) can be finally deduced. □

Appendix B

Geometric Interest Rate Theory

B.1 Setup

Proposition 6 (The Musiela HJM formulation.) *Under the martingale measure $\mathbb{Q}$ the f-dynamics are given by*

$$\begin{cases} df(t,x) &= \left(\frac{\partial f(t,x)}{\partial x} + \widetilde{\sigma}(t,x)\int_0^x \widetilde{\sigma}(t,u)^T\,du\right)dt + \widetilde{\sigma}(t,x)dW(t) \\ f(0,x) &= f^o(0,x). \end{cases} \tag{B.1}$$

where $\widetilde{\sigma}(t,x) := \sigma(t,t+x)$.

Proof.

Stage 1:
Recall the integral representation for the instaneous spot rate given by equation (A.7)

$$r(t) = F(0,t) + \int_0^t \alpha(s,t)ds + \int_0^t \sigma(s,t)dW(s),$$

and (A.8)

$$\begin{aligned} r(t) = r(0) + \int_0^t \partial_z F(0,z)dz + \int_0^t \alpha(z,z)dz + \int_0^t \sigma(z,z)dW(z) + \\ + \int_0^t \left(\int_0^u \partial_u \alpha(z,u)dz + \int_0^u \partial_u \sigma(z,u)dW(z)\right)du \end{aligned} \tag{B.2}$$

On the other hand, we have the following integral representation for the instantaneous forward rate

$$F(t,T) = F(0,T) + \int_0^t \alpha(z,T)dz + \int_0^t \sigma(z,T)dW(z)$$

Let us fix $T = t = u$. By taking partial differentials, we have the following expression:

$$\partial_2 F(u,u) - \partial_u F(0,u) = \int_0^u \partial_u \alpha(z,u)dz + \int_0^u \partial_u \sigma(z,u)dW(z), \tag{B.3}$$

where the operator ∂_2 means partial differentiation with respect to the second argument. By substituting into (B.2), we have

$$r(t) = r(0) + \int_0^t (\alpha(s,s) + \partial_2 F(s,s))\, ds + \int_0^t \sigma(s,s)dW(s), \tag{B.4}$$

where it should be emphasized that s is a "mute" variable of the same type as u in equation (B.3).

Stage 2: Consider us given an arbitrary $x \geq 0$, the parametrization

$$f(t,x) := F(t, t+x)$$

as well as the definitions

$$\begin{aligned} \hat{F}(t,T) &:= F(t,T+x) \\ \hat{\alpha}(z,T) &:= \alpha(z,T+x) \\ \hat{\sigma}(z,T) &:= \sigma(z,T+x) \end{aligned}$$

It should also be noted that $f(t,x) = \hat{F}(t,t) := \hat{r}(t)$. Now, from equation (B.4), an analogous integral equation for the new redefined spot rate, $\hat{r}(t)$, may be easily inferred:

$$\hat{r}(t) = \hat{r}(0) + \int_0^t \left(\hat{\alpha}(s,s) + \partial_2 \hat{F}(s,s)\right) ds + \int_0^t \hat{\sigma}(s,s)dW(s). \tag{B.5}$$

Returning to the old variables we have:

$$\begin{aligned} f(t,x) = f(0,x) + \int_0^t (\alpha(s,s+x) + \partial_x f(s,x))\, ds \quad + \\ + \int_0^t \sigma(s,s+x)dW(s). \end{aligned} \tag{B.6}$$

Let us introduce right now the modified volatility process $\widetilde{\sigma}(t,x) := \sigma(t,t+x)$. Then

$$\begin{aligned}\alpha(s,s+x) = \sigma(s,s+x)\int_s^{s+x}\sigma(t,z)dz = \\ = \widetilde{\sigma}(s,x)\int_0^x \widetilde{\sigma}(t,z)dz := \widetilde{\alpha}(s,x) - \partial_x f(s,x),\end{aligned} \tag{B.7}$$

which combined with (B.6), brings us the formal differential given in (B.1). Similar arguments are followed by Björk in [12] or Musiela in [42]. □

B.2 The Invariance Conditions

Theorem 4 (Consistency Conditions.) *The forward curve manifold $\mathcal{G}$ is f invariant for the forward rate process $f(t,x)$ in $\mathcal{M}$ iff*

$$G_x(z,\cdot) + \sigma(t,\cdot)\int_0^{\cdot}\sigma(t,u)^T\, du + \phi(t,\cdot) \in Im\left[G_z(z,\cdot)\right], \tag{B.8}$$

$$\sigma(t,\cdot) \in Im\left[G_z(z,\cdot)\right]. \tag{B.9}$$

$\forall (t,z) \in \mathbb{R}_+ \times \mathcal{Z}$. *Here, G_z and G_x denote the Jacobian of G w.r.t. to z and x, provided some minimal smoothness of the mapping G.*

Proof. Let us start with the prove of sufficiency. Thus, we assume (B.8)-(B.9). We may select $\gamma : \mathbb{R}_+ \times \mathcal{Z}^d$ and $\psi : \mathbb{R}_+ \times \mathbb{R}^{d\times q}$ satisfying

$$\begin{aligned} G_x(z,\cdot) + \sigma(t,\cdot)\textstyle\int_0^{\cdot}\sigma(t,u)^T\, du + \phi(t,\cdot) &= G_z(z,\cdot)\gamma(z,t), \\ \sigma(t,\cdot) &= G_z(z,\cdot)\psi(z,t). \end{aligned}$$

$\forall\ (t,z) \in \mathbb{R}_+ \times \mathcal{Z}$. Let $y_s(\cdot) \in \mathcal{G}$, thus for some $z^o \in \mathcal{Z}$, $y_s(\cdot) = G(z^o,\cdot)$. Define Z as the solution to (4.10) with initial condition $Z(s) = z^o$ and define the infinite-dimensional process $y_t(x)$ by the relation $y_t(x) = G(Z(t),x)$. Then

$$\begin{aligned} dy_t(x) &= G_z(Z(t),x)\gamma(Z(t),t)dt + G_z(Z(t),x)\psi(Z(t),t)\circ dW(t) \\ &= \Big(G_x(Z(t),x) + \sigma(t,x)\textstyle\int_0^x \sigma(t,u)^T\, du + \phi(t,x)\Big)\, dt + \sigma(t,x)\circ dW(t) \\ &= \Big(\tfrac{\partial}{\partial x}y_t(x) + \sigma(t,x)\textstyle\int_0^x \sigma(t,u)^T\, du + \phi(t,x)\Big) + \sigma(t,x)\circ dW(t). \end{aligned} \tag{B.10}$$

Thus, y solves the Fisk-Stratonovich SDE (4.8) with drift defined as in (4.9).

Now, let us prove necessity. Then, we have to assume f-invariance. If we take the differential

$$dy_t(x) = d(G(Z(t), x) = G_z(Z(t), x)\gamma(Z(t), t)\, dt + G_z(Z(t), x)\psi(Z(t), t) \circ dW(t).$$

Comparing the latter differential with (4.6) and equating the drift and volatility coefficients yields (B.8)-(B.9) for any $t \geq s$, but with $z = Z(t)$. Since time and initial point over the manifold may be chosen arbitrarily, (B.8)-(B.9) are general results. □

Appendix C

Pricing Options with a Consistent HJM Model

Digital Caps

Consider a general HJM model under the risk-neutral measure $\mathbb{Q}$ specified by (2.1). We also assume that

$$\sigma(t,T) = [\, \sigma_1(t,T) \ \dots \ \sigma_q(t,T) \,]$$

are deterministic functions of (t,T), and hence forward rates $F(t,T)$ are Gaussian distributed. Let us consider now a European-style binary option, with expiration date T and exercise price K, on an underlying bond with maturity S (where of course $T < S$). The following general pricing formula may be derived:

Proposition 7 *The price, at $t = 0$ of the security*

$$h(T) = P(T,S)\mathbb{1}_{\{P(T,S)-K\geq 0\}}$$

is given by

$$V(h,0) = P(0,S)N(d_+) \tag{C.1}$$

where

$$\begin{aligned} d_+ &:= \frac{\log\left(\frac{P(0,S)}{KP(0,T)}\right) + \frac{1}{2}\vartheta^2(T,S)}{\vartheta(T,S)} \,, \\ \vartheta^2(T,S) &:= \int_0^T \|\varsigma(u;T,S)\|^2 \, du \,; \end{aligned} \tag{C.2}$$

and,

$$\varsigma(t;T,S) := S(t,S) - S(t,T) = -\int_T^S \sigma(t,s)\,ds. \tag{C.3}$$

Proof. Let us start with the fundamental arbitrage-free equation

$$V = \mathbb{E}\left[D(0,T)P(T,S)\mathbb{1}_{\{P(T,S)\geq K\}}\right] \tag{C.4}$$

where we are taking the expectations with respect the equivalent martingale measure $\mathbb{Q}$ associated to the money-market numeraire $B(\cdot)$.

The Radon-Nikodym derivative that changes S-forward measure $\mathbb{Q}^S$ into the money-market measure $\mathbb{Q}$ will be given by

$$\lambda^S(T) = \frac{d\mathbb{Q}^S}{d\mathbb{Q}} = \frac{P(T,S)/P(0,S)}{B(T)/B(0)} = \frac{D(0,T)P(T,S)}{P(0,S)}.$$

Substituting into (C.4), and combining with the measurability at $t=0$ of $P(0,S)$

$$\begin{aligned} V &= \mathbb{E}\left[P(0,S)\lambda^S(T)\mathbb{1}_{\{P(T,S)\geq K\}}\right] = \\ &= P(0,S)\mathbb{Q}^S\left(P(T,S)\geq K\right). \end{aligned}$$

Now we have the value V for the call option in terms of the forward measure $\mathbb{Q}^S$. For computing the probability $\mathbb{Q}^S$, first, note the following:

$$\mathbb{Q}^S(P(T,S)\geq K) = \mathbb{Q}^S\left(\frac{P(T,T)}{P(T,S)} \leq \frac{1}{K}\right) = \mathbb{Q}^S\left(\log\frac{P(T,T)}{P(T,S)} \leq -\log K\right).$$

We remark that it is enough to introduce the auxiliary processes,

$$W_{T,S}(t) := \frac{P(t,T)}{P(t,S)},$$

and,

$$Z_{T,S}(t) := \log W_{T,S}(t),$$

to conclude that $Z_{T,S}(T)$ distributes like

$$Z_{T,S}(T) \sim \mathcal{N}\left(\log\frac{P(0,T)}{P(0,S)} - \frac{1}{2}\vartheta^2(T,S), \vartheta^2(T,S)\right),$$

and then

$$\mathbb{Q}^S(Z_{T,S}(T) \leq -\log K) = N(d_+).$$

□

Lemma 5 *The price at $t = 0$ of the security*

$$h(T) = P(T,S)\mathbb{1}_{\{P(T,S)-K<0\}}$$

is given by

$$\Pi(h,0) = P(0,S)N(-d_+) \tag{C.5}$$

where the quantities d_+ are completely determined by the identities (2.17) to (2.18).

Proof. First consider the identity option at time $t = 0$

$$\begin{aligned} V + \Pi &= \mathbb{E}\left[D(0,T)P(T,S)\left(\mathbb{1}_{\{P(T,S)-K\geq 0\}} + \mathbb{1}_{\{P(T,S)-K<0\}}\right)\right] \\ &= \mathbb{E}\left[D(0,T)P(T,S)\right], \end{aligned}$$

under the risk-neutral martingale measure. By using the *Change of Numeraire Theorem* in an identical way to that shown by Proposition 7 in this appendix, recall first that the likelihood

$$\lambda^S(T) = \frac{d\mathbb{Q}^S}{d\mathbb{Q}} = \frac{P(T,S)/P(0,S)}{B(T)/B(0)} = \frac{D(0,T)P(T,S)}{P(0,S)},$$

induces the change of the S-forward measure into the risk-neutral measure. Therefore we have:

$$\mathbb{E}\left[D(0,T)P(T,S)\right] = \mathbb{E}\left[\lambda^S(T)P(0,S)\right] = \mathbb{E}^S\left[P(0,S)\right],$$

and then the relation:

$$V + \Pi = P(0,S) \tag{C.6}$$

is inferred. Finally,

$$\Pi = P(0,S) - V = P(0,S)\left(1 - N(d_+)\right) = P(0,S)N(-d_+) \tag{C.7}$$

□

Corollary 4 (Digital Caplet Pricing for Gaussian Forward Rates.) *The price* *at $t = 0$ of the binary j-caplet, δ_j, with payoff:*

$$h_{\delta_j}(x_j) = \mathbb{1}_{\{L_j(x_{j-1})-K\}}$$

is given by

$$\delta_j(h_{\delta_j}, 0) = (1 + \tau_j K) P_j(0) N(-d_+) \tag{C.8}$$

where

$$d_+ := \frac{\log\left(\frac{P_j(0)}{\kappa P_{j-1}(0)}\right) + \frac{1}{2}\vartheta^2(0, x_{j-1})}{\vartheta(0, x_{j-1})}, \tag{C.9}$$

$$\vartheta^2(0, x_{j-1}) := \int_0^{x_{j-1}} \|\varsigma(u; x_{j-1}, x_j)\|^2 \, du ; \tag{C.10}$$

$$\varsigma(t; x_{j-1}, x_j) := -\int_{x_{j-1}}^{x_j} \sigma(t, s) \, ds. \tag{C.11}$$

Bibliography

[1] Ames, W. F. (1977). *Numerical Methods for Partial Differential Equations.* Academic Press.

[2] Angelini, F., & S. Herzel (2002). Consistent Initial Curves for Interest Rate Models, *Journal of Derivatives* 9(4), 8–18.

[3] Angelini, F., & S. Herzel (2005). Consistent Calibration of HJM Models to Cap Implied Volatilities, *Journal of Futures Markets* 25, 1093-1120.

[4] BIS (2005). Zero-coupon yield curves: Technical documentation, *BIS Papers* 25. Bank for International Settlements, Basle.

[5] Björk, T. (2004). *Arbitrage Theory in Continuous Time*, Second Edition, Oxford.

[6] Björk, T. (2003). On the Geometry of Interest Rate Models, en: Carmona, R.A., Cinlar, E., Ekeland, I., Jouini, E., Scheinkman, J.A., Touzi, N., *Paris-Princeton Lectures on Mathematical Finance 2003*, Lecture Notes in Mathematics, 1847, Springer.

[7] Björk, T., & C. Landen (2002). On the construction of finite dimensional realizations for nonlinear forward rate models, *Finance and Stochastics* 6(3), 303-331.

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

[9] Björk, T. (2001). A Geometric View of Interest Rate Theory, en: Jouini, E., Cvitanic, J. & M. Musiela, *Option Pricing, Interest Rates and Risk Management*, Cambridge University Press.

[10] Björk, T., & A. Gombani (1999). Minimal Realizations of interest rate models. *Finance and Stochastics* 3(4), 413-432.

[11] Björk, T., & B.J. Christensen (1999). Interest Rate Dynamics and Consistent Forward Rate Curves, *Mathematical Finance* 9, 323-348.

[12] Björk, T. (1997). Interest Rate Theory, en: Runggaldire, W., *Financial Mathematics*, Lecture Notes in Mathematics, 1656, Springer-Verlag.

[13] Black, F., & P. Karasinski (1991). Bond and Option Pricing when Short Rates are Lognormal, *Financial Analysts Journal* 4, 52-59. 127–155.

[14] Black, F. The Pricing of Commodity Contracts, *Journal of Financial Economics* 3, 167–179.

[15] Brigo, D. & F. Mercurio (2006). *Interest Rate Models: Theory and Practice.* Cambridge University Press.

[16] Buraschi, A., & F. Corielli (2005). Risk Management Implications of Time-Inconsistency: Model Updating and Recalibration of No-Arbitrage Models. *Journal of Banking & Finance* 29, 2883-2907.

[17] Cairns, A. J. G (2004). *Interest Rate Models. An Introduction* Princeton University Press.

[18] Cheyette, O. (1996). Markov Representation of the Heath-Jarrow-Morton Model. *BARRA Inc.*. Working Paper.

[19] Da Prato, G. & J. Zabczyk (1992). *Stochastic Equations in Infinite Dimensions.* Cambridge University Press.

[20] De Rossi, G. (2004). Kalman Filtering of Consistent Forward Rate Curves: a Tool to Estimate and Model Dynamically the Term Structure, *Journal of Empirical Finance* 11, 277–308.

[21] Daglish, T. (2010). Lattice Methods for No-Arbitrage Pricing of Interest Rate Securities, *Journal of Derivatives* 18(2), 7–19.

[22] Dewynne, J., Howison S. & P. Wilmott (1995). *The Mathematics of Financial Derivatives*, Cambridge University Press.

[23] ESCHENAUER, H., KOSKI J. & A. OSYCZKA (1990). *Multicriteria Design Optimization*, Springer.

[24] FALCÓ A., NAVARRO LL. & J. NAVE. (2011). On the Calibration of a Gaussian Heath-Jarrow-Morton model using Consistent Forward Rate Curves. *Quantitative Finance* 11(4), 495–504.

[25] FALCÓ A., NAVARRO LL. & J. NAVE (2009). The Hull-White Model and Multiobjective Calibration with Consistent Curves: Empirical Evidence. *RACSAM Rev. R. Acad. Cien. Serie A. Mat.* 103(2), 235–249.

[26] FALCÓ A., NAVARRO LL., & J. NAVE (2008). A multiobjective approach using consistent rate curves to the calibration of a Gaussian Heath-Jarrow-Morton model. *WP-AD Series* 2008–09, IVIE.

[27] FILIPOVIC D., (2009). *Term-Structure Models*, Springer.

[28] FILIPOVIC, D., (2001). *Consistency Problems for Heath-Jarrow-Morton Interest Rate Models*, Lecture Notes in Mathematics, 1760, Springer-Verlag.

[29] FILIPOVIC D., (1999). A Note on the Nelson and Siegel Family. *Mathematical Finance* 9(4), 349–359.

[30] HARRISON J.M. & D KREPS (1978). Martingales and Arbitrage in Multiperiod Securities Markets. *Journal of Economic Theory* 20, 381–408.

[31] HEATH, D. C., R. JARROW & A. MORTON (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation, *Econometrica* 60(1), 77–105.

[32] HO, T.S.Y, & S.B. LEE (1986). Term Structure Movements and the Pricing of Interest Rate Contingent Claims. *Journal of Financial and Quantitative Analysis* 411011-1029.

[33] HULL, J., & A. WHITE (1990). Pricing Interest Rate Derivatives Securities. *The Review of Financial Studies* 3(4), 573–592.

[34] HULL, J., & A. WHITE (1993). Single-factor interest rate models and the valuation of interest rate derivative securities. *Journal of Financial and Quantitative Analysis* 3(4), 573–592.

[35] Hull, J., & A. White (1994). Numerical Procedures for Implementing Term Structure Models, I: Single-Factor Models. *The Journal of Derivatives* 2, 7–16.

[36] Ikeda, N., & S. Watanabe (1981). *Stochastic Differential Equations and Diffusion Processes*, North-Holland.

[37] Jamshidian, F. (1991). Forward Induction and Construction of Yield Curve Difussion Models. *The Journal of Fixed Income* 1(1), 62–74.

[38] Kim, I. Y., & O.L. de Weck (2005). Adaptive weighted-sum method for bi-objective optimization: Pareto front generation. *Struct. Multidisc. Optim.* 29, 149–158.

[39] Kloeden, P. E., & E. Platen (1999). *Numerical solution of stochastic differential equations*, Springer.

[40] Marler, T., (2005). A study of multi-objective optimization methods for engineering applications, Ph.D. Dissertation, *University of Iowa*, Iowa City, IA.

[41] Mercurio, F., & J.M. Moraleda (2000). An analytically tractable interest rate model with humped volatility. *European Journal of Operational Research* 120, 205–214.

[42] Musiela, M., (1993). Stochastic PDEs and term structure models. Working Paper, J. Intern. Finance, IGR-AFFI, La Baule.

[43] Musiela, M., & M. Rutkowski (2004). *Martingale methods in financial modelling*, Springer.

[44] Navarro, Ll., (2003). Métodos Numéricos para la Valoración de Derivados sobre Tipos de Interés , Master Thesis Dissertation, *University of Valencia*, Valencia.

[45] Nelson, C. R., & A. F. Siegel (1987). Parsimonious Modelling of Yield Curves. *Journal of Business* 60(4), 473–489.

[46] Øksendal, B. (1998). *Stochastic Differential Equations*. Springer

[47] Pareto, V. (1906). Manuale di Economia Politica. *Societa Editrice Libraria.*

[48] Rao, S. S. (1987). Game Theory Approach for Multiobjective Structural Optimization. *Computers and Structures* 25(1), 119–127.

[49] Ritchken, P., & I. Chuang (1999). Interest rate option pricing with volatility humps. *Review of Derivatives Research* 3, 237–262.

[50] Ritchken, P. and L. Sankarasubramanian (1992). On Markovian Representations of the Term Structure, Working Paper.

[51] Smith, G. D. (1985). *Numerical Solution of Partial Differential Equations*, Oxford University Press