The LIBOR Market Model

Nevena Šelić

School of Computational and Applied Mathematics

University of the Witwatersrand

A dissertation submitted for the degree of

Master of Science in

Advanced Mathematics of Finance

May 2006

Declaration

I declare that this is my own, unaided work. It is being submitted for the Degree of Master of Science to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to any other University.

(Signature)

(Date)

Acknowledgements

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed, and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

I would like to thank my supervisor Professor David Taylor for his help, encouragement and support in writing this dissertation. I would also like to thank my friend Jacques du Toit for his comments and suggestions.

Nevena Šelić
May 2006

Table of Contents

Chapter 1

Introduction

The over-the-counter (OTC) interest rate derivative market is large and rapidly developing. In March 2005, the Bank for International Settlements published its "Triennial Central Bank Survey" which examined the derivative market activity in 2004 (http://www.bis.org/publ/rpfx05.htm). The reported total gross market value of OTC derivatives stood at $6.4 trillion at the end of June 2004. The gross market value of interest rate derivatives comprised a massive 71.7% of the total, followed by foreign exchange derivatives (17.5%) and equity derivatives (5%). Further, the daily turnover in interest rate option trading increased from 5.9% (of the total daily turnover in the interest rate derivative market) in April 2001 to 16.7% in April 2004. This growth and success of the interest rate derivative market has resulted in the introduction of exotic interest rate products and the ongoing search for accurate and efficient pricing and hedging techniques for them.

Interest rate caps and (European) swaptions form the largest and the most liquid part of the interest rate option market. These vanilla instruments depend only on the level of the yield curve. The market standard for pricing them is the Black (1976) model. Caps and swaptions are typically used by traders of interest rate derivatives to gamma and vega hedge complex products. Thus an important feature of an interest rate model is not only its ability to recover an arbitrary input yield curve, but also an ability to calibrate to the implied at-the-money cap and swaption volatilities. The LIBOR market model developed out of the market's need to price and hedge exotic interest rate derivatives consistently with the Black (1976) caplet formula. The focus of this dissertation is this popular class of interest rate models.

The fundamental traded assets in an interest rate model are zero-coupon bonds. The evolution of their values, assuming that the underlying movements are continuous, is driven by a finite number of Brownian motions. The traditional approach to modelling the term structure of interest rates is to postulate the evolution of the instantaneous short or forward rates. Contrastingly, in the LIBOR market model, the discrete forward rates are modelled directly. The additional assumption imposed is that the volatility function of the discrete forward rates is a deterministic function of time. In Chapter 2 we provide a brief overview of the history of interest rate modelling which led to the LIBOR market model. The general theory of derivative pricing is presented, followed by a exposition and derivation of the stochastic differential equations governing the forward LIBOR rates.

The LIBOR market model framework only truly becomes a model once the volatility functions of the discrete forward rates are specified. The information provided by the yield curve, the cap and the swaption markets does not imply a unique form for these functions. In Chapter 3, we examine various specifications of the LIBOR market model. Once the model is specified, it is calibrated to the above mentioned market data. An advantage of the LIBOR market model is the ability to calibrate to a large set of liquid market instruments while generating a realistic evolution of the forward rate volatility structure (Piterbarg 2004). We examine some of the practical problems that arise when calibrating the market model and present an example calibration in the UK market.

The necessity, in general, of pricing derivatives in the LIBOR market model using Monte Carlo simulation is explained in Chapter 4. Both the Monte Carlo and quasi-Monte Carlo simulation approaches are presented, together with an examination of the various discretizations of the forward rate stochastic differential equations. The chapter concludes with some numerical results comparing the performance of Monte Carlo estimates with quasi-Monte Carlo estimates and the performance of the discretization approaches.

In the final chapter we discuss numerical techniques based on Monte Carlo simulation for

pricing American derivatives. We present the primal and dual American option pricing problem formulations, followed by an overview of the two main numerical techniques for pricing American options using Monte Carlo simulation. Callable LIBOR exotics is a name given to a class of interest rate derivatives that have early exercise provisions (Bermudan style) to exercise into various underlying interest rate products. A popular approach for valuing these instruments in the LIBOR market model is to estimate the continuation value of the option using parametric regression and, subsequently, to estimate the option value using backward induction. This approach relies on the choice of relevant, i.e. problem specific predictor variables and also on the functional form of the regression function. It is certainly not a "black-box" type of approach.

Instead of choosing the relevant predictor variables, we present the sliced inverse regression technique. Sliced inverse regression is a statistical technique that aims to capture the main features of the data with a few low-dimensional projections. In particular, we use the sliced inverse regression technique to identify the low-dimensional projections of the forward LIBOR rates and then we estimate the continuation value of the option using nonparametric regression techniques. The results for a Bermudan swaption in a two-factor LIBOR market model are compared to those in Andersen (2000).

Chapter 2

LIBOR Market Model Theory

Mathematics possesses not only truth, but beauty - a beauty cold and austere, like that of a sculpture.

– Bertrand Russell

The London Inter-Bank Offered Rates or LIBOR are benchmark short term simple interest rates at which banks can borrow money from other banks in the London interbank market. LIBOR rates are fixed daily by the British Bankers' Association.[1] They are quoted for various maturities and currencies.

	USD	GBP	CAD	EUR	JPY
O/N	3.53688	4.79750	2.80000	2.09500	0.03813
1WK	3.62563	4.66375	2.80000	2.10150	0.03938
2WK	3.69250	4.60500	2.80000	2.10375	0.04000
1MO	3.74000	4.59125	2.80667	2.11300	0.04438
2MO	3.79000	4.59125	2.85500	2.12425	0.05063
3MO	3.85000	4.59000	2.90417	2.13363	0.05750
4MO	3.90000	4.57250	2.92667	2.14713	0.06000
5MO	3.95000	4.56063	2.95667	2.15038	0.06500
6MO	3.99000	4.55000	2.98417	2.15200	0.06813
7MO	4.02000	4.54000	3.00000	2.15950	0.07625
8MO	4.05000	4.53125	3.02000	2.16113	0.07813
9MO	4.07750	4.52625	3.03833	2.16725	0.08188
10MO	4.10163	4.52000	3.05833	2.17438	0.08563
11MO	4.12213	4.51563	3.08000	2.18288	0.08875
12MO	4.14625	4.51375	3.10167	2.19000	0.09375

Table 2.1: British Bankers' Association LIBOR rate quotes on 9 September 2005 (source: *Reuters*)

The LIBOR market model is an approach to modelling the term structure of interest rates based on simple rather than instantaneous rates. It was developed by Miltersen, Sandmann & Sondermann (1997), Brace, Gątarek & Musiela (1997), Musiela & Rutkowski (1997) and Jamshidian (1997). However, this approach was certainly used by practitioners even before the first three references appeared as working papers in 1995. Quoting Rebonato (2002) on the LIBOR market model:

> "...before any of the now-canonical papers appeared, it was simultaneously and independently 'discovered' by analysts and practitioners who, undaunted by the expected occurrence of the log-normal explosion, went ahead and discretized a log-normal forward-rate-based HJM implementation."

We briefly trace the history of modelling the discretely compounded interest rates that led to the development of the LIBOR market model.

[1] The rates are based on an arithmetic average of the offered interbank deposit rates - the deposit rates are ranked and the second and third quartiles are averaged to produce LIBOR rates (http://www.bba.org.uk).

In the early 1990's, an important feature of an interest rate model was not only its ability to recover an arbitrary input yield curve, but also an ability to calibrate to the implied at-the-money cap and (European) swaption volatilities (Rebonato 2004). Caps and swaptions comprise the largest and the most liquid part of the interest rate derivative market. They are typically used by traders of interest rate derivatives to gamma and vega hedge complex products. The LIBOR market model developed out of the need to price and hedge exotic interest rate derivatives consistently with the Black caplet formula.

The traditional approach to modelling the term structure of interest rates was to postulate the evolution of instantaneous short or instantaneous forward rates. A tractable class of models, allowing "Black-like" closed-form formula for caplets, are the Gaussian instantaneous short rate and the Gaussian instantaneous forward rate models. An example of the former is the extended Vasicek model of Hull & White (1990), while an example of the latter is the Gaussian Heath, Jarrow & Morton (1992) (HJM) model developed by Brace & Musiela (1994). The problem with Gaussian models is that they lead to theoretical arbitrage opportunities - interest rates can become negative with positive probability. While the most natural way to exclude negative interest rates is through the lognormal distributional assumption, this too has difficulties. In the lognormal short rate models, such as Black, Derman & Toy (1990) and Black & Karasinski (1991), the expected value of the accumulation factors is infinite over a finite time horizon, i.e. $\mathbb{E}_\tau\left[B(t,T)^{-1}\right]=\infty$ for any $0\leq\tau<t<T$, where $B(t,T)$ is the price at t of a zero-coupon bond maturing at T (Sandmann & Sondermann 1997). The lognormal structure for the instantaneous forward rates leads to rates exploding to positive infinity with positive probability (Heath et al. 1992).

Sandmann & Sondermann (1993, 1994, 1997) noticed that the problems with the lognormal assumption arise as a consequence of modelling the instantaneous rates. The focus was shifted from modelling the instantaneous short rate $r(t)$ to modelling the effective annual rate $r_e(t)$, defined by the formula $e^{r(t)}=1+r_e(t)$. The authors proposed a binomial model for the effective annual rate $r_e(t)$, whose limiting dynamics are geometric Brownian motion with time dependent drift and volatility functions.[2] For this limit model, the instantaneous short rate $r(t)$ follows a combination of normal and lognormal diffusions - it approaches the lognormal diffusion for small values of $r(t)$ and a normal diffusion for large values of $r(t)$. Goldys, Musiela & Sondermann (1994, 2000) extended these results in the HJM framework. The specification of the volatility structure was shifted from the instantaneous forward rates $f(t,T)$ to rates $j(t,T)$ defined by the formula $e^{f(t,T)}=1+j(t,T)$. A deterministic volatility function for the rates $j(t,T)$ was proposed and the authors proved that the resulting model has a unique positive solution, with no dreaded explosion of the instantaneous forward rates. The breakthrough came in Sandmann, Sondermann & Miltersen (1994), when the effective annual forward rates $f_a(t,T,\delta)$ at time t for the interval $[T,T+\delta]$, defined by

$$\big(1+f_a(t,T,\delta)\big)^{\delta}=\exp\left(\int_T^{T+\delta}f(t,u)du\right)$$

were modelled with a deterministic volatility function. It was shown that for $\delta=1$, closed-form solutions for zero-coupon bond options were computable. This exciting observation led to the Miltersen, Sandmann & Sondermann (1994, 1997) papers, where the simple forward rates $f_s(t,T,\delta)$ at time t for the interval $[T,T+\delta]$, defined by

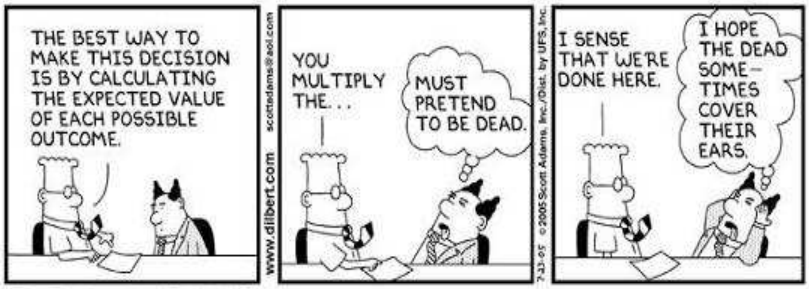
$$1+\delta\, f_s(t,T,\delta)=\exp\left(\int_T^{T+\delta}f(t,u)du\right)$$

were modelled with a deterministic volatility function. A closed-form expression for an option with exercise date T, written on a zero-coupon bond with maturity date $T+\delta$, was obtained. In particular, caplets were priced according to the market standard Black caplet formula. Brace et al. (1997) derived the dynamics of the simple forward rates $f_s(t,T,\delta)$ under the risk-neutral measure and proved the existence and uniqueness of a solution to the resulting stochastic differential equations.

An interest rate model implied by the assumption of a deterministic volatility function for the simple forward rates is known as the LIBOR market model (LMM) or the Brace-Gątarek-Musiela (BGM) model. The construction of this model is presented in Section 2.2. In the following section we review the general theory of derivative pricing.

[2]Suppose that W is a standard Brownian motion and that X satisfies $dX(t)/X(t)=\mu(t)dt+\sigma(t)dW(t)$. We refer to $\mu(t)$ as the *drift function* and $\sigma(t)$ as the *volatility function*. For a more general process X satisfying $dX(t)=\mu(t,\omega)dt+\sigma(t,\omega)dW(t)$, we refer to $\mu(t,\omega)$ as the *drift coefficient* and $\sigma(t,\omega)$ as the *diffusion coefficient*.

2.1 Theory of Derivative Pricing



The history of modelling risky asset prices can be traced back to 1900, when French mathematician Louis Bachelier, under the supervision of Henri Poincaré, proposed arithmetic Brownian motion for the movement of stock prices in his PhD thesis, *"Théorie de la Spéculation"*. To be more precise, 29 March 1900, the date on which Bachelier defended his thesis is considered to be the birth of mathematical finance (Courtault et al. 2000). The problem with Bachelier's construction is that arithmetic Brownian motion is not a plausible model for asset prices because, amongst other things, it allows the asset prices to become negative. Sixty-five years later, Brownian motion was reintroduced in finance by Paul Samuelson in a paper written with Henry P. McKean Jr. (Samuelson 1965). Here he postulated that stock prices follow geometric Brownian motion, which circumvents the problems associated with Bachelier's model.[3]

The breakthrough in derivative pricing came in the seminal papers of Black & Scholes (1973) and Merton (1973). The authors assumed geometric Brownian motion for the dynamics of the stock price and noted that a long position in a stock combined with a specific short position in a European call option on the stock will have a riskless return over an infinitesimally small period of time. To avoid arbitrage, the return must equal the prevailing risk-free rate. This observation led to the Black-Scholes partial differential equation for the derivative price, with explicit solutions for European call and put options. What is truly surprising about their result is the fact that a European option can be replicated by trading in the underlying stock and a riskless asset.

The Black-Scholes partial differential equation is independent of the expected return on the stock. This interesting property led to the discovery of risk-neutral valuation by Cox & Ross (1976). The concept was formalized and extended by Harrison & Kreps (1979) and Harrison & Pliska (1981, 1983). In accordance with J. Michael Harrison and Stanley R. Pliska's seminal work, we now present the theory of derivative pricing in a frictionless market with continuous trading up to some fixed (finite) time horizon T.[4]

2.1.1 Model of the Financial Market

Consider a financial market with a fixed trading horizon $[0, T]$. The uncertainty in the economy is modelled by a filtered complete probability space $\left(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P}\right)$, where the filtration $\mathbb{F} = \{\mathscr{F}_t\}_{0 \le t \le T}$ satisfies the usual hypotheses.[5] Assume that $\mathscr{F}_T = \mathscr{F}$ and that $\mathscr{F}_0$ is trivial, i.e. for $A \in \mathscr{F}_0$, $\mathbb{P}(A) \in \{0, 1\}$. The only role of a probability measure $\mathbb{P}$ is to determine the null sets. As the choice of a measure is arbitrary, the uncertainty in the economy can alternatively be modelled by a family of filtered probability spaces $\left(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P}\right)$, $\mathbb{P} \in \mathscr{P}$, where $\mathscr{P}$ is a class of equivalent probability measures on $\left(\Omega, \mathscr{F}_T\right)$ (Musiela & Rutkowski 2005).[6] The financial interpretation of the assumption of a class $\mathscr{P}$ is that investors agree on which outcomes are possible, but their probability assessments of these outcomes differ.

The financial market consists of n primary securities (traded assets). Denote the price process of the primary securities by $\mathbf{S} = \{\mathbf{S}_t, 0 \le t \le T\}$, where $\mathbf{S}_t = \left(S_t^1, \ldots, S_t^n\right)'$. We model these

[3] Jarrow & Protter (2004) provided a fascinating description of the history of stochastic integration.

[4] To avoid inserting the same reference every few lines, we note that all the mathematical definitions in the following sections are taken directly from Protter (2004).

[5] A filtered complete probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$ is said to satisfy the *usual hypotheses* if $\mathscr{F}_0$ contains all the $\mathbb{P}$-null sets of $\mathscr{F}$ and $\mathscr{F}_t = \bigcap_{u>t} \mathscr{F}_u$, for all $0 \le t < T$.

[6] Consider a measurable space $(\Omega, \mathscr{F})$ with measures $\mathbb{P}$ and $\mathbb{Q}$ defined on this space. $\mathbb{P}$ is *absolutely continuous* with respect to $\mathbb{Q}$, written $\mathbb{P} \ll \mathbb{Q}$, if $\mathbb{P}(A) = 0$ whenever $\mathbb{Q}(A) = 0$ for all $A \in \mathscr{F}$. If $\mathbb{P} << \mathbb{Q}$ and $\mathbb{Q} << \mathbb{P}$, written $\mathbb{P} \sim \mathbb{Q}$, then $\mathbb{P}$ and $\mathbb{Q}$ are called *equivalent* measures.

processes as strictly positive continuous semimartingales. A semimartingale, to be defined below, is the most general stochastic process for which the stochastic integral can be reasonably defined (Bichteler 1981). The semimartingale model is also quite a natural assumption, as it can be shown that, loosely speaking, the semimartingale model for asset prices is implied by the existence of an equivalent martingale measure (Delbaen & Schachermayer 1994*b*, Theorem 7.2). In this general model, both the arrival of random market information and deterministic components, such as the pull-to-par effects of zero-coupon bonds, can be incorporated. We now define these two components mathematically, followed by the definition of a semimartingale.

Definition 2.1.1. *An adapted, càdlàg process $M = \{M_t, 0 \le t \le T\}$ is a **local martingale** if there exists a sequence of stopping times T_m, with $\lim_{m\to\infty} T_m = T$, almost surely, such that the stopped process $\{M_{t\wedge T_m}, 0 \le t \le T\}$ is a uniformly integrable martingale for each m.*[7]

Definition 2.1.2. *An adapted, càdlàg process $A = \{A_t, 0 \le t \le T\}$ is a **finite variation process** if, almost surely, the paths of A have finite variation on each compact interval of $[0, T]$.*

Definition 2.1.3. *An adapted, càdlàg process $X = \{X_t, 0 \le t \le T\}$ is a **semimartingale** if there exist processes M, A with $M_0 = A_0 = 0$ such that*

$$X_t = X_0 + M_t + A_t \tag{2.1}$$

where M is a local martingale and A is a finite variation process.

Decomposition (2.1) is not always unique because there exist finite variation martingales. If X is a continuous semimartingale, then the decomposition is unique, and then both the local martingale and the finite variation process in the decomposition are continuous (Protter 2004, page 130).

In addition to the primary securities in the market, we have a European contingent claim maturing at time T that we want to price and hedge.[8] The payoff of the contingent claim may depend on the entire path of the primary securities up to and including the option maturity T. Thus, the contingent claim is modelled as a nonnegative, $\mathscr{F}_T$-measurable random variable ϑ.

The pricing and hedging of contingent claims is based on the concept of a replicating portfolio. Suppose that we can construct a trading strategy, using the primary securities, which requires no cash inflow or outflow, except at inception, such that the final value of this portfolio matches the value of the contingent claim for all $\omega \in \Omega$ almost surely. Then, by no-arbitrage arguments, the value of the contingent claim must equal the value of the portfolio at inception.

To formulate the contingent claim pricing problem formally, we need to define predictable and locally bounded processes. These technical assumptions are sufficient to ensure that the stochastic integral of a process satisfying these restrictions, with respect to a semimartingale, exists.

Definition 2.1.4. *Let $\mathbb{L}$ denote the space of adapted processes with càglàd (left continuous with right limits) paths. The **predictable σ-algebra** $\mathfrak{P}$ on $[0, T] \times \Omega$ is $\mathfrak{P} = \sigma\{X : X \in \mathbb{L}\}$, the σ-algebra generated by all the processes in $\mathbb{L}$. A process is **predictable** if it is measurable with respect to $\mathfrak{P}$.*

Definition 2.1.5. *A process $X = \{X_t, 0 \le t \le T\}$ is said to be **locally bounded** if there exists a sequence of stopping times T_m, with $\lim_{m\to\infty} T_m = T$, almost surely, such that the stopped process $\{X_{t\wedge T_m}, 0 \le t \le T\}$ is bounded for each m.*

Definition 2.1.6. *An adapted n-dimensional stochastic process $\boldsymbol{\phi} = \{\boldsymbol{\phi}_t, 0 \le t \le T\}$, where $\boldsymbol{\phi}_t = \left(\phi_t^1, \ldots, \phi_t^n\right)'$, is a **trading strategy** if $\boldsymbol{\phi}$ is locally bounded and predictable.*

A trading strategy $\boldsymbol{\phi}_t$ represents the number of assets held at time t, which is revised continuously through time. In a frictionless market, any quantity of assets can be both bought and sold at zero cost. The assumption that the process $\boldsymbol{\phi}$ is predictable has a financial interpretation: one establishes the portfolio just before time t and rebalances after the prices of the primary securities at time t have been observed. The assumption that a predictable process $\boldsymbol{\phi}$ is locally bounded has two implications. Firstly, the stochastic integral of a locally bounded (predictable) process with

[7] A stochastic process $X = \{X_t, 0 \le t \le T\}$ is said to be *adapted* if $X_t \in \mathscr{F}_t$ for each $t \in [0, T]$. It is *càdlàg* if it almost surely has sample paths which are right continuous with left limits. The process is *uniformly integrable* if $\lim_{m\to\infty} \sup_{0\le t\le T} \int_{|X_t|\ge m} |X_t| d\mathbb{P} = 0$.

[8] Risk-neutral valuation of American derivatives was developed by Bensoussan (1984) and Karatzas (1988, 1989). The American option pricing problem will be formulated in Chapter 5.

respect to a local martingale is itself a local martingale, a statement that is not true in general (Protter 2004, page 171). Secondly, the stochastic integral can be defined component-wise, that is the stochastic integral of a trading strategy with respect to a semimartingale is equal to the sum of the stochastic integrals of the relevant vector components (cf. equation (2.3) below) (Musiela & Rutkowski 2005, page 281).

Associated with each trading strategy $\boldsymbol{\phi}$ is the *value process* $V(\boldsymbol{\phi}) = \{V_t(\boldsymbol{\phi}), 0 \leq t \leq T\}$, defined by

$$V_t(\boldsymbol{\phi}) = \boldsymbol{\phi}_t \cdot \mathbf{S}_t = \sum_{i=1}^{n} \phi_t^i \, S_t^i, \tag{2.2}$$

Assuming that the primary securities do not generate any cashflows such as dividends, the *gains process* $G(\boldsymbol{\phi}) = \{G_t(\boldsymbol{\phi}), 0 \leq t \leq T\}$ is defined by

$$G_t(\boldsymbol{\phi}) = \int_0^t \boldsymbol{\phi}_u \cdot d\mathbf{S}_u = \sum_{i=1}^{n} \int_0^t \phi_u^i \, dS_u^i, \tag{2.3}$$

The gains process is adapted and continuous, because it is a stochastic integral with respect to a continuous semimartingale. It represents the total capital gain when trading strategy $\boldsymbol{\phi}$ is followed. A trading strategy that requires no external cashflows is termed a self-financing trading strategy.

Definition 2.1.7. *A **self-financing trading strategy** is a trading strategy $\boldsymbol{\phi}$ whose value process satisfies*

$$V_t(\boldsymbol{\phi}) = V_0(\boldsymbol{\phi}) + G_t(\boldsymbol{\phi}), \quad 0 \leq t \leq T \tag{2.4}$$

Equation (2.4) can be written in a more familiar form as $d\left(\boldsymbol{\phi}_t \cdot \mathbf{S}_t\right) = \boldsymbol{\phi}_t \cdot d\mathbf{S}_t$. Note that the self-financing condition ensures that the value process is continuous.

Let us restate the pricing problem in terms of the introduced notation: we are looking for a self-financing trading strategy $\boldsymbol{\phi}$ such that $V_T(\boldsymbol{\phi}) = \vartheta$, almost surely. If we can find such a $\boldsymbol{\phi}$, then by no-arbitrage arguments, the value of the claim at any time t must be $V_t(\boldsymbol{\phi})$. However, it turns out that one cannot naively allow all self-financing trading strategies. There are two problems that need to be addressed. Firstly, we need to remove doubling strategies that turn "nothing into something" because they represent arbitrage opportunities. A classical example of a doubling strategy is the coin toss game, where if heads comes up, the payout is two times the bet amount. A player bets one unit of currency on the first bet and if he looses he doubles his bet. The player stops at the time of the first win, which is inevitable, even if the coin is not fair. However, one needs to be able to fund arbitrarily large losses until the eventual win. It is possible to construct these strategies in the current framework because trading takes place continuously, and hence infinitely many times in the interval $[0, T]$.[9] Secondly, we need to remove suicide strategies that turn "something into nothing" because they lead to non-unique pricing. If suicide strategies are permitted, one may find two self-financing trading strategies for a claim whose value processes have different initial values.[10] The necessary modifications to a class of self-financing trading strategies depend on the notion of an equivalent martingale measure.

2.1.2 Equivalent Martingale Measures

At the heart of mathematical finance is the assumption that there are no arbitrage opportunities in well-functioning markets.

Definition 2.1.8. *A self-financing trading strategy $\boldsymbol{\phi}$ is called an **arbitrage opportunity** if the value process satisfies the following set of conditions*

$$V_0(\boldsymbol{\phi}) = 0, \quad \mathbb{P}(V_T(\boldsymbol{\phi}) \geq 0) = 1, \quad \mathbb{P}(V_T(\boldsymbol{\phi}) > 0) > 0$$

The "no-arbitrage pricing" approach postulates that there are no arbitrage opportunities in the market. In the Black-Scholes model, the assumption of an arbitrage-free market implies, and is implied by, the existence of a unique equivalent measure such that the stock prices normalized by the money-market account are martingales under this measure. In the general framework one needs to choose an asset, called the numéraire, to normalize the other assets in the market.

[9] For an example of a doubling strategy in the Brownian motion setting, see Duffie (1996, Chapter 6.C).
[10] For an example of a suicide trading strategy, see Harrison & Pliska (1981, Section 6).

Definition 2.1.9. *A **numéraire** is a price process $X = \{X_t, 0 \le t \le T\}$ that is, almost surely, strictly positive for all $t \in [0, T]$.*

All the price processes of the primary securities are strictly positive by assumption. Without loss of generality, choose security S^1 to be the numéraire.[11] The *deflator process* $Y = \{Y_t, 0 \le t \le T\}$ is a strictly positive semimartingale, defined by $Y_t = 1/S_t^1$ through Itô's formula for continuous semimartingales (Karatzas & Shreve 1991, page 149). The *normalized asset price process* is denoted by $\mathbf{Z} = \{\mathbf{Z}_t, 0 \le t \le T\}$, where $\mathbf{Z}_t = \left(1, Z_t^2, \ldots, Z_t^n\right)'$ and $Z_t^i = Y_t\, S_t^i$ for $i = 1, \ldots, n$. The *normalized value process* $V^*(\boldsymbol{\phi})$ and the *normalized gains process* $G^*(\boldsymbol{\phi})$ of a trading strategy $\boldsymbol{\phi}$ are defined by

$$V_t^*(\boldsymbol{\phi}) = Y_t V_t(\boldsymbol{\phi}) = \boldsymbol{\phi}_t \cdot \mathbf{Z}_t = \phi_t^1 + \sum_{i=2}^{n} \phi_t^i Z_t^i, \quad 0 \le t \le T \tag{2.5}$$

$$G_t^*(\boldsymbol{\phi}) = \int_0^t \boldsymbol{\phi}_u \cdot d\mathbf{Z}_u = \sum_{i=1}^{n} \int_0^t \phi_u^i \, dZ_u^i = \sum_{i=2}^{n} \int_0^t \phi_u^i \, dZ_u^i, \quad 0 \le t \le T \tag{2.6}$$

Self-financing trading strategies were defined as trading strategies whose value processes satisfy equation (2.4). To show that self-financing trading strategies remain self-financing after a numéraire change, we need to define the quadratic covariation process.

Definition 2.1.10. *Let X and Y be semimartingales. The **quadratic covariation** of X, Y, also called the (**square**) **bracket process**, is defined by*

$$[X, Y]_t = X_t Y_t - \int_0^t X_{u-} \, dY_u - \int_0^t Y_{u-} \, dX_u \tag{2.7}$$

*where $X_-(Y_-)$ is the left-continuous version of $X(Y)$. The **quadratic variation** of X is $[X, X]$.*

Equation (2.7) is also known as the *integration by parts* formula. To obtain a better understanding of the quadratic covariation process, suppose X and Y are continuous local martingales. Then the process

$$X_t Y_t - [X, Y]_t = \int_0^t X_u \, dY_u + \int_0^t Y_u \, dX_u$$

is a continuous local martingale. Heuristically, $d[X, Y]_t$ is the conditional expectation just before t of $d(XY)_t$ (Back 2001). For a standard Brownian motion W, which is a continuous local martingale, we know that $[W, W]_t = t$ for all $t \ge 0$.

Consider a self-financing trading strategy $\boldsymbol{\phi}$. We now show that self-financing trading strategies remain self-financing after a numéraire change, i.e. that $d\left(\boldsymbol{\phi}_t \cdot \mathbf{Z}_t\right) = \boldsymbol{\phi}_t \cdot d\mathbf{Z}_t$.

$$\begin{aligned} d\big(Y_t V_t(\boldsymbol{\phi})\big) &= Y_t dV_t(\boldsymbol{\phi}) + V_t(\boldsymbol{\phi}) dY_t + d[Y, V(\boldsymbol{\phi})]_t \\ &= Y_t \left(\boldsymbol{\phi}_t \cdot d\mathbf{S}_t\right) + \left(\boldsymbol{\phi}_t \cdot \mathbf{S}_t\right) dY_t + d[Y, \boldsymbol{\phi} \cdot \mathbf{S}]_t \\ &= \boldsymbol{\phi}_t \cdot \left(Y_t d\mathbf{S}_t + \mathbf{S}_t dY_t + d[Y, \mathbf{S}]_t\right) \\ &= \boldsymbol{\phi}_t \cdot d\left(Y_t \mathbf{S}_t\right) \end{aligned}$$

Thus

$$V_t^*(\boldsymbol{\phi}) = V_0^*(\boldsymbol{\phi}) + G_t^*(\boldsymbol{\phi}), \quad 0 \le t \le T \tag{2.8}$$

Substituting equations (2.5) and (2.6) into (2.8), we see that ϕ^1 can be used to form a self-financing trading strategy from an arbitrary trading strategy $\boldsymbol{\phi}$, by setting

$$\phi_t^1 = V_0^*(\boldsymbol{\phi}) + \sum_{i=2}^{n} \int_0^t \phi_u^i \, dZ_u^i - \sum_{i=2}^{n} \phi_t^i \, Z_t^i, \quad 0 \le t \le T$$

Let $\mathscr{P}^*$ be the set (possibly empty) of equivalent martingale measures, defined as a set of equivalent measures such that, for $\mathbb{P}^* \in \mathscr{P}^*$, the normalized asset prices $\mathbf{Z}$ are $\mathbb{P}^*$-martingales. The link between equivalent martingale measures and the absence of arbitrage is known as the *Fundamental Theorem of Asset Pricing*. This theorem states that, for a stochastic process $\mathbf{Z}$, the existence of

[11] In the general equity market setting, security S^1 is assumed to be the money-market account.

an equivalent martingale measure is *essentially* equivalent to the absence of arbitrage opportunities (Delbaen & Schachermayer 1994*b*). It is extremely difficult to prove the exact mathematical conditions that the normalized asset price process $\mathbf{Z}$ needs to satisfy for the absence of arbitrage to imply the existence of an equivalent martingale measure. These restrictions have been established in a series of papers by Delbaen (1992) and Delbaen & Schachermayer (1994*a*,*b*, 1998), for increasingly general classes of processes. We will assume that $\mathscr{P}^*$ is non-empty and examine the reverse implication of the existence of an equivalent martingale measure implying the absence of arbitrage opportunities.

To remove the doubling strategies discussed in the previous subsection, one can require that the value processes be bounded from below. Harrison & Pliska (1981) defined Φ as a class of self-financing trading strategies $\boldsymbol{\phi}$ whose value process satisfies

$$V_t(\boldsymbol{\phi}) \geq 0, \quad 0 \leq t \leq T$$

For $\boldsymbol{\phi} \in \Phi$, the normalized value process $V^*(\boldsymbol{\phi})$ is also nonnegative. This is due to the fact that the deflator is a strictly positive process. Note that ϕ^1 can still be used appropriately to define a self-financing trading strategy, as long as the constructed normalized value process is nonnegative.

In order to eliminate suicide strategies, Harrison & Pliska (1981) fixed a measure $\mathbb{P}^* \in \mathscr{P}^*$ and defined a class of *admissible trading strategies* $\Phi(\mathbb{P}^*)$ as self-financing trading strategies $\boldsymbol{\phi} \in \Phi$ whose normalized value processes $V^*(\boldsymbol{\phi})$ are $\mathbb{P}^*$-martingales. This restriction is sufficient to remove suicide strategies because due to the martingale property, $V_T^*(\boldsymbol{\phi})$ cannot be zero, almost surely, if $V_0^*(\boldsymbol{\phi})$ is positive.

Theorem 2.1.1. *Assume that $\mathscr{P}^*$ is non-empty. Then the model is arbitrage-free.*

Proof. The proof is given in Appendix A. □

The question that we now address is, "How do we construct an equivalent martingale measure $\mathbb{P}^*$ when dealing with semimartingale processes?"

2.1.3 Construction of Equivalent Measures

In this section we examine the construction of a probability measure $\mathbb{Q}$ on $(\Omega, \mathscr{F}_T)$ that is equivalent to the underlying probability measure $\mathbb{P}$. Following Protter (2004), we know that if $\mathbb{Q} \ll \mathbb{P}$, there exists a nonnegative $\mathbb{P}$-integrable random variable ζ_T, called the *Radon-Nikodým derivative*, satisfying $\mathbb{E}^{\mathbb{P}}[\zeta_T] = 1$, such that for all $A \in \mathscr{F}_T$

$$\mathbb{Q}(A) = \int_A \zeta_T \, d\mathbb{P}$$

The Radon-Nikodým derivative ζ_T is denoted by $\frac{d\mathbb{Q}}{d\mathbb{P}}$. The Radon-Nikodým derivative process is the càdlàg version of the following uniformly integrable martingale[12]

$$\zeta_t = \mathbb{E}^{\mathbb{P}}\left[\zeta_T \,|\, \mathscr{F}_t\right], \quad 0 \leq t < T \tag{2.9}$$

If $\mathbb{P} \ll \mathbb{Q}$ as well, then $\frac{d\mathbb{P}}{d\mathbb{Q}} = \left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)^{-1}$. Thus if we can construct the Radon-Nikodým derivative, we can construct an equivalent probability measure.

The construction of an equivalent probability measure when the underlying processes are semimartingales is due to Christopeit & Musiela (1994). Note that in general, even if the underlying processes are continuous semimartingales, the Radon-Nikodým derivative process may be discontinuous because the underlying filtration is not necessarily Brownian (Musiela & Rutkowski 2005, page 295). We now introduce the Doléans-Dade exponential.

Definition 2.1.11. *The **Doléans-Dade exponential** $\mathscr{E}(D)$ is the unique solution of the stochastic differential equation*

$$d\mathscr{E}(D)_t = \mathscr{E}(D)_{t-} dD_t, \qquad \mathscr{E}_0(D) = 1 \tag{2.10}$$

The explicit solution to (2.10) *is given by*

$$\mathscr{E}(D)_t = \exp\left(D_t - \frac{1}{2}[D,D]_t^c\right) \prod_{0 \leq u \leq t} (1 + \Delta D_u) e^{-\Delta D_u} \tag{2.11}$$

[12] One of the consequences of the usual hypotheses is that every martingale has a version that is càdlàg (Protter 2001).

where $\Delta D_t = D_t - D_{t-}$ *and* $[D,D]^c$ *is the path-by-path continuous part of the quadratic variation process* $[D,D]$

$$[D,D]^c_t = [D,D]_t - \sum_{0\le u\le t} (\Delta D_u)^2$$

Theorem 2.1.2. *Consider* $\mathbb{Q} \sim \mathbb{P}$ *and the Radon-Nikodým derivative process* ζ*, as defined. Suppose that there exists a* $\mathbb{P}$*-local martingale* D *with* $D_0 = D_{0-} = 0$ *satisfying*

$$\Delta D_t > -1, \quad 0 \le t \le T \tag{2.12}$$

$$\mathbb{E}^{\mathbb{P}}\left[\mathscr{E}(D)_T\right] = 1 \tag{2.13}$$

where $\Delta D_t = D_t - D_{t-}$ *and* $\mathscr{E}(D)$ *is the Doléans-Dade exponential. Then there exists a one-to-one correspondence between* ζ *and* D*, given by*

$$\zeta_t = \mathscr{E}(D)_t\,, \quad 0 \le t \le T$$

Theorem 2.1.2 states that the Radon-Nikodým derivative process is the Doléans-Dade exponential of a local martingale satisfying certain conditions. Firstly, condition (2.12) is a condition on the jump sizes. From equation (2.11), it is easily seen that this restriction makes the Doléans-Dade exponential strictly positive. Secondly, the Doléans-Dade exponential is a uniformly integrable martingale if and only if condition (2.13) is satisfied.

Girsanov's Theorem, also known as the Girsanov-Meyer Theorem, provides us with the semimartingale decomposition under an equivalent probability measure $\mathbb{Q}$.

Theorem 2.1.3. *Let* X *be a continuous semimartingale under* $\mathbb{P}$ *with the decomposition*

$$X_t = X_0 + M_t + A_t, \quad 0 \le t \le T$$

where M *is a continuous local martingale and* A *is a continuous finite variation process. Let* $\mathbb{Q}$ *be an equivalent measure and let the Radon-Nikodým derivative* $\frac{d\mathbb{Q}}{d\mathbb{P}}$ *be defined by the Doléans-Dade exponential of a local martingale* D *(cf. Theorem 2.1.2). Since* M *has bounded jumps* $(\Delta M_t = 0)$*, the cross-variation process* $\langle M, D\rangle$ *exists (Christopeit & Musiela 1994, Corollary 1).*[13] *Then* X *is a continuous semimartingale under* $\mathbb{Q}$ *with the decomposition*

$$X_t = X_0 + L_t + C_t, \quad 0 \le t \le T \tag{2.14}$$

where L *is a* $\mathbb{Q}$*-local martingale*

$$L_t = M_t - \langle M, D\rangle_t, \quad 0 \le t \le T$$

and C *is a* $\mathbb{Q}$ *finite variation process*

$$C_t = A_t + \langle M, D\rangle_t, \quad 0 \le t \le T$$

In particular, X *is a local martingale under* $\mathbb{Q}$ *if and only if*

$$A_t + \langle M, D\rangle_t = 0, \quad 0 \le t \le T \tag{2.15}$$

The Black-Scholes Model

The two theorems presented above can be made more intuitive in the familiar framework of Black and Scholes. The normalized asset price process Z is the discounted stock price process. Under $\mathbb{P}$, Z satisfies

$$dZ_t = (\mu - r)Z_t dt + \sigma Z_t dW_t \tag{2.16}$$

where μ is the expected return on the asset, σ is the return volatility, r is the risk-free continuously compounded rate of interest and W is a standard Brownian motion. The parameters μ, r and σ are constants. The Martingale Representation Theorem (Karatzas & Shreve 1991, Theorem 4.2) is a beautiful result that demonstrates how one can represent martingales in terms of Brownian motion, a fundamental continuous martingale.

[13] The *cross-variation process* $\langle X, Y\rangle$, also called the *conditional quadratic covariation*, is defined to be the compensator of $[X,Y]$, i.e. a unique predictable finite variation process such that $[X,Y] - \langle X,Y\rangle$ is a local martingale. If X and Y are continuous semimartingales, $[X,Y] = \langle X,Y\rangle$.

Theorem 2.1.4 (Martingale Representation Theorem)**.** *Suppose* $\mathbf{W} = (W^1, \ldots, W^n)'$ *is a* n*-dimensional Brownian motion and let* $\mathbb{F}^W$ *be its completed natural filtration. Then every continuous local martingale* M *for* $\mathbb{F}^W$ *has a representation*

$$M_t = M_0 + \sum_{i=1}^{n} \int_0^t H_u^i dW_u^i$$

where H^i *are measurable adapted processes satisfying* $\mathbb{P}\left[\int_0^t \left(H_u^i\right)^2 du < \infty\right] = 1$ *for all* $0 \leq t \leq T$ *and* $i = 1, \ldots, n$*. Further, if* L *and* N *are two continuous local martingales for* $\mathbb{F}^W$ *with the representations*

$$L_t = L_0 + \sum_{i=1}^{n} \int_0^t H_u^i dW_u^i$$

$$N_t = N_0 + \sum_{i=1}^{n} \int_0^t \widetilde{H}_u^i dW_u^i$$

where H_u^i *and* $\widetilde{H}_u^i$ *satisfy the condition stated above, then*

$$\langle L, N \rangle_t = \sum_{i=1}^{n} \int_0^t H_u^i \widetilde{H}_u^i du \tag{2.17}$$

In the Black-Scholes model, the underlying filtration is the augmented filtration generated by the Brownian motion W. In particular, this implies that the Radon-Nikodým derivative process is a continuous martingale. From the Martingale Representation Theorem

$$\zeta_t = 1 + \int_0^t H_u dW_u$$

where $\zeta_0 = 1$, due to the condition $\mathbb{E}^{\mathbb{P}}[\zeta_T] = 1$. Following Protter (2004), we assume that ζ is "well behaved enough" to define a process $\lambda = \{\lambda_t, 0 \leq t \leq T\}$ through the formula $H_t = \lambda_t \zeta_t$. The dynamics of ζ become

$$\frac{d\zeta_t}{\zeta_t} = \lambda_t dW_t \tag{2.18}$$

The explicit solution of the stochastic differential equation (2.18) is

$$\zeta_t = \exp\left(-\frac{1}{2}\int_0^t \lambda_u^2 du + \int_0^t \lambda_u dW_u\right) \tag{2.19}$$

Theorem 2.1.2 states that the Radon-Nikodým derivative process is the Doléans-Dade exponential of a local martingale D satisfying conditions (2.12) and (2.13). It follows from equation (2.19) that the local martingale D is given by $D_t = \int_0^t \lambda_u dW_u$. This can easily be verified: the jump sizes of the continuous process D are zero, $\Delta D_t = 0$, and the Doléans-Dade exponential of D is

$$\begin{aligned}
\mathscr{E}(D)_t &= \exp\left(\int_0^t \lambda_u dW_u - \frac{1}{2}[D, D]_t\right) \\
&= \exp\left(\int_0^t \lambda_u dW_u - \frac{1}{2}\int_0^t dD_u dD_u\right) \\
&= \exp\left(\int_0^t \lambda_u dW_u - \frac{1}{2}\int_0^t \lambda_u^2 du\right)
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[\mathscr{E}(D)_T\right] &= \exp\left(-\frac{1}{2}\int_0^T \lambda_u^2 du\right) \mathbb{E}^{\mathbb{P}}\left[\exp\left(\int_0^T \lambda_u dW_u\right)\right] \\
&= 1
\end{aligned}$$

The last equality follows from the fact that if a random variable X is normally distributed with mean μ and variance σ^2, written $X \sim N(\mu, \sigma^2)$, its moment generating function is given by

$\mathbb{E}\left[e^{tX}\right] = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. The mean of the normal random variable $\int_0^T \lambda_u dW_u$ is zero, because an Itô integral is a martingale, while the variance of $\int_0^T \lambda_u dW_u$ is $\int_0^T \lambda_u^2 du$, by Itô isometry.

In particular, from Theorem 2.1.2, the local martingale $D_t = \int_0^t \lambda dW_u$ defines an equivalent probability measure for any constant $\lambda \in \mathbb{R}$. What we are really interested in is an equivalent martingale measure. For this we turn to Theorem 2.1.3.

The normalized asset price process Z is an Itô process and hence a continuous semimartingale under $\mathbb{P}$. In particular, using the notation of Theorem 2.1.3, we have $Z = Z_0 + M_t + A_t$ where

$$M_t = \int_0^t \sigma Z_u dW_u \text{ and } A_t = \int_0^t (\mu - r) Z_u du$$

For an equivalent probability measure $\mathbb{Q}$ to be an equivalent martingale measure, Z must be a $\mathbb{Q}$-martingale. Theorem 2.1.3 states that Z is a local martingale under $\mathbb{Q}$ if and only if

$$A_t + \langle M, D \rangle_t = 0$$

Substituting in for A, M and D, and using equation (2.17)

$$\begin{aligned} &\int_0^t (\mu - r) Z_u du + \int_0^t \sigma Z_u \lambda_u dt = 0 \\ \Rightarrow \quad &(\mu - r) Z_t dt + \sigma Z_t \lambda_t dt = 0 \\ \Rightarrow \quad &\lambda_t = -\frac{\mu - r}{\sigma} \end{aligned}$$

Given this choice of λ_t, the $\mathbb{Q}$-local martingale Z has the decomposition $Z = Z_0 + L_t$, where

$$\begin{aligned} L_t &= M_t - \langle M, D \rangle_t \\ &= \int_0^t \sigma Z_u dW_u - \int_0^t \sigma Z_u \left(-\frac{\mu - r}{\sigma} \right) du \\ &= \int_0^t \sigma Z_u dW_u + \int_0^t (\mu - r) Z_u du \\ &= \int_0^t \sigma Z_u \left(dW_u + \frac{\mu - r}{\sigma} du \right) \end{aligned}$$

is a $\mathbb{Q}$-local martingale. Following Protter (2004), let

$$dW_u^{\mathbb{Q}} = dW_u + \frac{\mu - r}{\sigma} du$$

denote this local martingale. By Lévy's Theorem, $W^{\mathbb{Q}}$ is Brownian motion because $\langle W^{\mathbb{Q}}, W^{\mathbb{Q}} \rangle_t = \langle W, W \rangle_t = t$. Under $\mathbb{Q}$, Z satisfies

$$dZ_t = \sigma Z_t dW_t^{\mathbb{Q}}$$

As σ is a constant, hence bounded, Z is a $\mathbb{Q}$-martingale. Thus, as expected, only when λ_t is the negative of the market price of risk does $D_t = \int_0^t \lambda_u dW_u$ define an equivalent martingale measure.

2.1.4 Arbitrage-free Pricing

A contingent claim ϑ is said to be *attainable* if there exists $\boldsymbol{\phi} \in \Phi(\mathbb{P}^*)$ such that $V_T(\boldsymbol{\phi}) = \vartheta$. Assuming that the random variable $Y_T \vartheta$ is $\mathbb{P}^*$-integrable (termed *integrable* claim), the arbitrage-free price of the attainable contingent claim ϑ is given by

$$\begin{aligned} V_t(\boldsymbol{\phi}) &= \frac{1}{Y_t} V_t^*(\boldsymbol{\phi}) \\ &= \frac{1}{Y_t} \mathbb{E}^{\mathbb{P}^*} \left[V_T^*(\boldsymbol{\phi}) \,|\, \mathscr{F}_t \right] \\ &= \frac{1}{Y_t} \mathbb{E}^{\mathbb{P}^*} \left[Y_T \vartheta \,|\, \mathscr{F}_t \right] \end{aligned}$$

for all $0 \leq t \leq T$. This is the fundamental pricing equation: the normalized contingent claim price $Y_t V_t(\boldsymbol{\phi})$ is a martingale under an equivalent martingale measure $\mathbb{P}^*$.

Theorem 2.1.5. *Suppose that the contingent claim ϑ is attainable in $\mathbb{P}_1^* \in \mathscr{P}^*$ and $\mathbb{P}_2^* \in \mathscr{P}^*$. Then the arbitrage-free prices will be equal, i.e.*

$$\mathbb{E}^{\mathbb{P}_1^*}\left[Y_T\vartheta \,|\, \mathscr{F}_t\right] = \mathbb{E}^{\mathbb{P}_2^*}\left[Y_T\vartheta \,|\, \mathscr{F}_t\right], \quad 0 \le t \le T$$

Proof. The proof follows Musiela & Rutkowski (2005) and is given in Appendix B. □

The question of whether every integrable claim is attainable, termed market completeness, depends on the class of admissible trading strategies.

2.1.5 Market Completeness

A market is referred to as complete if every integrable claim is attainable. Thus far, the trading strategies identified turn out to be too restricted if one wants to replicate all integrable claims. In particular, the assumption of locally bounded processes is problematic. Harrison & Pliska (1983) extended the set of trading strategies to a larger class $\mathscr{L}(\mathbf{Z})$, the set of vector-valued, predictable processes that are integrable with respect to a semimartingale $\mathbf{Z}$.[14] The main result of Harrison & Pliska (1981, 1983) follows.

Theorem 2.1.6. *Let $\mathbf{Z}$ denote the normalized asset price process and let $\mathscr{M}$ denote the set of all $\mathbb{P}^*$-martingales. Then the following statements are equivalent:*

1. *The market is complete under $\mathbb{P}^*$.*
2. *Every martingale $M \in \mathscr{M}$ can be represented as*

$$M = M_0 + \int \boldsymbol{\phi} \cdot d\mathbf{Z}, \quad \text{for some } \boldsymbol{\phi} \in \mathscr{L}(\mathbf{Z})$$

3. *$\mathscr{P}^*$ is a singleton, i.e. there is only one equivalent martingale measure $\mathbb{P}^* \in \mathscr{P}^*$.*

According to Protter (2001), few martingales possess the second property in the previous theorem. In fact, the only examples given in the paper are Brownian motion, the Compensated Poisson process and Azéma martingales.

Theorem 2.1.6 shows that prefect hedging of a contingent claim is possible only if the integral representation of the normalized claim price (a martingale) exists. In particular, suppose that a contingent claim ϑ is integrable. Then ϑ is attainable if the normalized claim price

$$M_t = \mathbb{E}^{\mathbb{P}^*}\left[Y_T\vartheta \,|\, \mathscr{F}_t\right], \quad 0 \le t \le T$$

admits the integral representation

$$M_t = M_0 + \sum_{i=2}^{n} \int_0^t \phi_u^i dZ_u^i, \quad 0 \le t \le T$$

In a complete market, the equivalent martingale measure $\mathbb{P}^*$ is unique and every integrable claim is attainable. In an incomplete market, the problem of hedging non-attainable claims becomes one of selecting an equivalent martingale measure that is optimal in some sense. Musiela & Rutkowski (2005) provide a brief introduction to this active area of current research.

2.2 LIBOR Market Model Theory

In this section we construct the LIBOR market model under the assumption that there are no "smile effects" in the interest rate market. This means that the caplet and swaption implied volatility surfaces are assumed to be, for a fixed expiry, a flat function of the strike.[15]

[14] Jarrow & Madan (1994) examined the distinction between vector- and component-wise stochastic integrals in the Brownian motion setting. They note that the earlier paper, Harrison & Pliska (1981), misspecified the class $\mathscr{L}(\mathbf{Z})$. The class implicitly assumed by Harrison & Pliska (1983), this being the set of vector-valued, predictable processes that are integrable with respect to a semimartingale $\mathbf{Z}$, is the correct one for Theorem 2.1.6 to be valid.

[15] Note that there is an implied swaption volatility surface for each individual underlying swap length.

2.2.1 Brace-Gątarek-Musiela Model

Motivated by the lack of an arbitrage-free term structure model consistent with the market practice of pricing caps and floors (which generally comprise the largest part of any interest rate derivative book), Brace et al. (1997) set out to construct such a model in the HJM framework. We provide an overview of the HJM framework, followed by details of the BGM construction of the market model.

Heath-Jarrow-Morton Framework. Heath et al. (1992) developed a framework for modelling the term structure of interest rates based on an exogenous specification of the evolution of the instantaneous forward rates. Interest rate modelling is assumed to take place in continuous time over the interval $[0, T^*]$, on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The probability space is equipped with a filtration $\mathbb{F} = \{\mathscr{F}_t\}_{0 \le t \le T^*}$, the augmented filtration generated by a d-dimensional Brownian motion $\mathbf{W} = \left(W^1, \ldots, W^d\right)'$. Each component of the vector Brownian motion $\mathbf{W}$ represents an independent, exogenous source of uncertainty in the financial market.

Assumption (Heath et al. 1992, C.1)**.** *For fixed $T \in [0, T^*]$ define the instantaneous, continuously compounded forward rate at time t for maturity T, $f(t,T)$, by*

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

where $B(t,T)$ is the price at t of a zero-coupon bond maturing at T. Then, under $\mathbb{P}$, the instantaneous forward rates $f(t,T)$ satisfy

$$df(t,T) = \alpha(t,T,\omega)dt + \sum_{i=1}^{d} \sigma_i(t,T,\omega)dW_t^i$$

where $\omega \in \Omega$ and

1. *the initial forward rate curve, $f(0,\cdot) : [0,T^*] \to \mathbb{R}$ is a Borel-measurable function,*
2. *$\alpha : \{(t,s) : 0 \le t \le s \le T\} \times \Omega \to \mathbb{R}$ and $\sigma_i : \{(t,s) : 0 \le t \le s \le T\} \times \Omega \to \mathbb{R}$ are adapted processes such that, almost surely,*

$$\int_0^T |\alpha(u,T,\omega)|du + \int_0^T |\sigma_i(t,T,\omega)|^2 dt < \infty, \quad \text{for } i = 1, \ldots, d$$

Some extra conditions are needed for the regularity of the zero-coupon bond price processes and the money-market account (Heath et al. 1992, C2 and C3). The money-market account is defined as

$$B^*(t) = \exp\left(\int_0^t r(u)du\right) \tag{2.20}$$

where r is the instantaneous, continuously compounded short rate, $r(t) = \lim_{T \to t} f(t,T)$.

The primary securities are zero-coupon bonds of different maturities. Heath et al. (1992) identified the restrictions that the assumption of an arbitrage-free market imposes on the evolution of the term structure of interest rates. In particular, after constructing a portfolio consisting of a finite number of zero-coupon bonds, the authors derived the necessary and sufficient conditions on the drift of the instantaneous forward rates for the existence of a unique equivalent martingale measure $\mathbb{Q}$, corresponding to the money-market account numéraire. Under $\mathbb{Q}$, the instantaneous forward rates $f(t,T)$ satisfy

$$df(t,T) = \boldsymbol{\sigma}(t,T) \cdot \boldsymbol{\sigma}^*(t,T)dt + \boldsymbol{\sigma}(t,T) \cdot d\mathbf{W}_t^{\mathbb{Q}}, \quad \boldsymbol{\sigma}^*(t,T) = \int_t^T \boldsymbol{\sigma}(t,u)du \tag{2.21}$$

where $\boldsymbol{\sigma}(t,T) = (\sigma_1(t,T,\omega), \ldots, \sigma_d(t,T,\omega))'$ and $\mathbf{W}^{\mathbb{Q}}$ is a d-dimensional $\mathbb{Q}$-Wiener process. Under $\mathbb{Q}$, the zero-coupon bond prices $B(t,T)$ satisfy

$$\frac{dB(t,T)}{B(t,T)} = r(t)dt - \boldsymbol{\sigma}^*(t,T) \cdot d\mathbf{W}_t^{\mathbb{Q}} \tag{2.22}$$

The assumption of no arbitrage implies that the drift coefficient of the instantaneous forward rates is uniquely determined once the diffusion coefficient is specified. This is in great contrast to instantaneous short rate models, where the drift coefficient can be specified independently of the

diffusion coefficient and is in fact used to calibrate the model to the initial yield curve. In the HJM framework, the initial condition for the stochastic differential equation (2.21) is the initial forward rate $f(0,T)$. The calibration to the initial yield curve is automatic because the current yield curve is a function of the initial instantaneous forward rates, $B(0,T) = \exp(-\int_0^T f(0,u)du)$.

Brace-Gatarek-Musiela Construction. Brace et al. (1997) considered a simple interest rate defined over a finite accrual period $\delta > 0$. Denote the simple forward rate at time t for the time interval $[T, T+\delta]$ by $f_s(t,T)$, for $0 \le t \le T \le T^* - \delta$. Using no-arbitrage arguments, the simple forward rates $f_s(t,T)$ are defined in terms of the zero-coupon bond prices

$$1 + \delta f_s(t,T) = \frac{B(t,T)}{B(t,T+\delta)} \tag{2.23}$$

or, equivalently, in terms of the instantaneous forward rates

$$1 + \delta\, f_s(t,T) = \exp\left(\int_T^{T+\delta} f(t,u)du\right) \tag{2.24}$$

Brace et al. (1997) derived the dynamics of the simple forward rates under the equivalent martingale measure $\mathbb{Q}$ using the dynamics of the instantaneous forward rates. However, it is much easier to identify the dynamics of the simple forward rates using relationship (2.23) and equation (2.22).

Lemma 2.2.1. *Suppose that X and Y are two Itô processes satisfying*

$$\frac{dX_t}{X_t} = \mu_t^X dt + \sigma_t^X dW_t, \qquad \frac{dY_t}{Y_t} = \mu_t^Y dt + \sigma_t^Y dW_t$$

where the adapted functions $\mu_t^X, \mu_t^Y, \sigma_t^X$ and σ_t^Y satisfy the conditions needed for the existence and uniqueness of strong solutions of the above stochastic differential equations.[16] *Then*

$$d\left(\frac{X_t}{Y_t}\right) = \left(\frac{X_t}{Y_t}\right)\Big(\big(\mu_t^X - \mu_t^Y\big)dt + \big(\sigma_t^X - \sigma_t^Y\big)\big(-\sigma_t^Y dt + dW_t\big)\Big)$$

The proof is a straightforward application of Itô's formula. Applying this result in equation (2.23), together with equation (2.22), we see that under $\mathbb{Q}$, the simple forward rates satisfy

$$\begin{aligned} df_s(t,T) &= \frac{1}{\delta}\, d\left(\frac{B(t,T)}{B(t,T+\delta)}\right) \\ &= \delta^{-1}\left(\frac{B(t,T)}{B(t,T+\delta)}\right)\big(\boldsymbol{\sigma}^*(t,T+\delta) - \boldsymbol{\sigma}^*(t,T)\big)\cdot\Big(\boldsymbol{\sigma}^*(t,T+\delta)dt + d\mathbf{W}_t^{\mathbb{Q}}\Big) \\ &= \delta^{-1}\big(1+\delta f_s(t,T)\big)\big(\boldsymbol{\sigma}^*(t,T+\delta) - \boldsymbol{\sigma}^*(t,T)\big)\cdot\Big(\boldsymbol{\sigma}^*(t,T+\delta)dt + d\mathbf{W}_t^{\mathbb{Q}}\Big) \end{aligned} \tag{2.25}$$

This is the HJM model for the simple forward rates $f_s(t,T)$. Brace et al. (1997) *postulated* that the simple forward rates have a deterministic volatility function

$$df_s(t,T) = \zeta_t\, dt + f_s(t,T)\,\boldsymbol{\lambda}(t,T)\cdot d\mathbf{W}_t^{\mathbb{Q}} \tag{2.26}$$

with some (to be determined) drift function ζ_t and a deterministic, bounded, piecewise continuous function $\boldsymbol{\lambda} : \mathbb{R}_+^2 \to \mathbb{R}^d$. The volatility function $\boldsymbol{\lambda}(t,T)$ is exogenously specified.

The motivation for this assumption was the Black caplet formula. As we shall show in the following chapter, the assumption of a deterministic volatility function for the simple forward rates is consistent with the Black model, meaning that the resulting model caplet pricing formula agrees with the market standard Black caplet formula.

Equating the diffusion coefficients of equations (2.25) and (2.26), we obtain

$$\boldsymbol{\sigma}^*(t,T+\delta) - \boldsymbol{\sigma}^*(t,T) = \frac{\delta f_s(t,T)}{1+\delta f_s(t,T)}\,\boldsymbol{\lambda}(t,T) \tag{2.27}$$

[16] These conditions are measurability, Lipschitz condition, linear growth and the initial value condition (Kloeden & Platen 1999, Section 4.5).

To determine $\boldsymbol{\sigma}^*(t, T+\delta)$, we can use as a recursion relationship (2.27)

$$\begin{aligned}\boldsymbol{\sigma}^*(t,T+\delta) &= \boldsymbol{\sigma}^*(t,T) + \frac{\delta f_s(t,T)}{1+\delta f_s(t,T)}\boldsymbol{\lambda}(t,T) \\ &= \boldsymbol{\sigma}^*(t,T-\delta) + \frac{\delta f_s(t,T-\delta)}{1+\delta f_s(t,T-\delta)}\boldsymbol{\lambda}(t,T-\delta) + \frac{\delta f_s(t,T)}{1+\delta f_s(t,T)}\boldsymbol{\lambda}(t,T) \\ &= \boldsymbol{\sigma}^*(t,T-k\delta) + \sum_{j=0}^{k}\frac{\delta f_s(t,T-j\delta)}{1+\delta f_s(t,T-j\delta)}\boldsymbol{\lambda}(t,T-j\delta)\end{aligned}$$

where $k = \left\lfloor \frac{T-t}{\delta} \right\rfloor$ and $t \le T-k\delta < t+\delta$.[17] To start the recursion one needs to assign values to $\boldsymbol{\sigma}^*(t, T-k\delta)$, this being the volatility function of zero-coupon bonds with maturities shorter than δ (cf. equation (2.22)). Brace et al. (1997) assumed that $\boldsymbol{\sigma}^*(t,T) = 0$ for $0 \le T-t < \delta$. Then

$$\boldsymbol{\sigma}^*(t,T+\delta) = \sum_{j=0}^{\lfloor \delta^{-1}(T-t)\rfloor}\frac{\delta f_s(t,T-j\delta)}{1+\delta f_s(t,T-j\delta)}\boldsymbol{\lambda}(t,T-j\delta) \tag{2.28}$$

Substituting equations (2.27) and (2.28) in (2.25), we obtain the dynamics of the simple forward rates $f_s(t,T)$ under $\mathbb{Q}$

$$df_s(t,T) = f_s(t,T)\boldsymbol{\lambda}(t,T)\cdot\left(\sum_{j=0}^{\lfloor \delta^{-1}(T-t)\rfloor}\frac{\delta f_s(t,T-j\delta)}{1+\delta f_s(t,T-j\delta)}\boldsymbol{\lambda}(t,T-j\delta)dt + d\mathbf{W}_t^{\mathbb{Q}}\right) \tag{2.29}$$

Brace et al. (1997) proved the existence and uniqueness of a solution to the stochastic differential equation (2.29). Note that in the BGM model, it is the forward rate volatility function $\boldsymbol{\lambda}(t,T)$ that is exogenously specified, not the zero-coupon bond price volatility function $\boldsymbol{\sigma}^*(t,T)$.

The technical problem with this analysis is that in the HJM framework, the zero-coupon bond price volatility function $\boldsymbol{\sigma}^*(t,T)$ has to be sufficiently smooth for the instantaneous forward rates to exist. From the definition of this function in equation (2.21), we see that it must be differentiable with respect to T. This restricts the choice of a deterministic simple forward rate volatility function $\lambda(t,T)$ because of its relationship with the zero-coupon bond price volatility function, equation (2.28). In particular, Brace et al. (1997) parameterized the simple forward rate volatility function as piecewise continuous, which is not a differentiable function. Thus, theoretically, the piecewise continuous specification cannot be analyzed in the HJM framework. As the assumption of the existence of instantaneous forward rates is not convenient, we now examine an alternate approach of modelling the zero-coupon bond price process without any reference to the instantaneous forward rates.

2.2.2 Forward Measures

The existence of a unique equivalent martingale measure $\mathbb{Q}$ associated with the money-market account numéraire implies, and is implied by, an arbitrage-free and complete interest rate system. Under $\mathbb{Q}$, the value of any traded asset normalized by the money-market account is a martingale. Thus the current value of a traded asset is the expected value, under $\mathbb{Q}$, of the discounted terminal asset value. This is the so-called risk-neutral valuation.

When pricing interest rate derivatives, it is frequently convenient to use a zero-coupon bond as a numéraire.[18] The measure associated with the zero-coupon bond maturing at time T is called the *T-forward measure*. Geman, El Karoui & Rochet (1995) showed that the T-forward measure $\mathbb{Q}^T$ is a probability measure on $\left(\Omega, \mathscr{F}_T\right)$, equivalent to $\mathbb{Q}$, defined by the Radon-Nikodým derivative

$$\frac{d\mathbb{Q}^T}{d\mathbb{Q}} = \frac{1}{B(0,T)B^*(T)}$$

where B^* is the money-market account.[19] Under $\mathbb{Q}^T$, any traded asset normalized by the zero-coupon bond maturing at time T is a martingale. We can reformulate the opening statement as

[17] $\lfloor x \rfloor$ is the largest integer less than or equal to x.

[18] For example, when an option has a payoff at time T, it is convenient to use the zero-coupon bond maturing at time T as the numéraire because the value of this numéraire at T is one. This avoids having to derive the joint distribution of the payoff and the numéraire.

[19] In an arbitrage-free and complete market there exists one equivalent martingale measure associated with each numéraire.

follows: the existence of a unique equivalent martingale measure $\mathbb{Q}^T$ associated with the zero-coupon bond maturing at time T as a numéraire implies, and is implied by, an arbitrage-free and complete interest rate system.

A modern interest rate model consists of a numéraire and a set of stochastic differential equations that a family of zero-coupon bonds satisfy. Following Musiela & Rutkowski (1997), we define an interest rate system by imposing the following assumptions.

Assumption 1. *The family of zero-coupon bond prices $B(t,T)$, $0 < t \leq T \leq T^*$, are modelled as strictly positive, continuous semimartingales. A deterministic initial term structure of interest rates $B(0,T)$, $T \in [0,T^*]$, is exogenously specified.*

Assumption 2. *There exists a unique equivalent martingale measure $\mathbb{Q}^{T^*}$ such that for every $T \in [0,T^*)$, the forward process*

$$F(t,T,T^*) = \frac{B(t,T)}{B(t,T^*)}, \quad 0 \leq t \leq T$$

is a strictly positive, continuous martingale under $\mathbb{Q}^{T^*}$.

The first assumption is a fairly general specification of the zero-coupon bond price processes. We will further assume that the underlying filtration is Brownian (cf. Section 2.2.1). The second assumption implies that the interest rate system is arbitrage-free and complete.

By the Martingale Representation Theorem 2.1.4, for every $T \in [0,T^*)$, the forward process $F(t,T,T^*)$ has the following representation under $\mathbb{Q}^{T^*}$

$$dF(t,T,T^*) = F(t,T,T^*)\boldsymbol{\gamma}(t,T,T^*) \cdot d\mathbf{W}_t^{\mathbb{Q}^{T^*}}, \quad 0 \leq t \leq T \tag{2.30}$$

where $\mathbf{W}^{\mathbb{Q}^{T^*}}$ is a d-dimensional $\mathbb{Q}^{T^*}$-Wiener process and $\boldsymbol{\gamma}(t,T,T^*)$ is a $\mathbb{R}^d$-valued, adapted process satisfying $\mathbb{Q}^{T^*}\left[\int_0^T ||\boldsymbol{\gamma}(u,T,T^*)||^2 du < \infty\right] = 1$.

From here one could construct an interest rate model given an exogenous specification of the volatilities of the forward processes $\boldsymbol{\gamma}(t,T,T^*)$. However, we are interested in modelling the simple forward rates $f_s(t,T)$ with an exogenously specified volatility function $\boldsymbol{\lambda}(t,T)$. Define

$$F(t,T,T+\delta) = \frac{F(t,T,T^*)}{F(t,T+\delta,T^*)} = \frac{B(t,T)}{B(t,T+\delta)}, \quad 0 \leq t \leq T$$

Using Lemma 2.2.1 and equation (2.30), under $\mathbb{Q}^{T^*}$, the forward process $F(t,T,T+\delta)$ satisfies

$$\begin{aligned} dF(t,T,T+\delta) &= \frac{F(t,T,T^*)}{F(t,T+\delta,T^*)}\Big(\boldsymbol{\gamma}(t,T,T^*) - \boldsymbol{\gamma}(t,T+\delta,T^*)\Big) \cdot \Big(-\boldsymbol{\gamma}(t,T+\delta,T^*)dt + dW_t^{\mathbb{Q}^{T^*}}\Big) \\ &= F(t,T,T+\delta)\boldsymbol{\gamma}(t,T,T+\delta) \cdot \Big(-\boldsymbol{\gamma}(t,T+\delta,T^*)dt + dW_t^{\mathbb{Q}^{T^*}}\Big) \end{aligned}$$

where $\boldsymbol{\gamma}(t,T,T+\delta) = \boldsymbol{\gamma}(t,T,T^*) - \boldsymbol{\gamma}(t,T+\delta,T^*)$. Define the $T+\delta$-forward measure $\mathbb{Q}^{T+\delta}$ on $\left(\Omega, \mathscr{F}_{T+\delta}\right)$ by the Radon-Nikodým derivative

$$\frac{d\mathbb{Q}^{T+\delta}}{d\mathbb{Q}^{T^*}} = \exp\left(-\frac{1}{2}\int_0^{T+\delta} ||\boldsymbol{\gamma}(t,T+\delta,T^*)||^2 dt + \int_0^{T+\delta} \boldsymbol{\gamma}(t,T+\delta,T^*) \cdot d\mathbf{W}_t^{\mathbb{Q}^{T^*}}\right)$$

By Girsanov's Theorem,

$$\mathbf{W}_t^{\mathbb{Q}^{T+\delta}} = \mathbf{W}_t^{\mathbb{Q}^{T^*}} - \int_0^t \boldsymbol{\gamma}(u,T+\delta,T^*)du, \quad 0 \leq t \leq T+\delta$$

is a d-dimensional $\mathbb{Q}^{T+\delta}$-Wiener process. Under $\mathbb{Q}^{T+\delta}$, the forward process $F(t,T,T+\delta)$ satisfies

$$dF(t,T,T+\delta) = F(t,T,T+\delta)\boldsymbol{\gamma}(t,T,T+\delta) \cdot d\mathbf{W}_t^{\mathbb{Q}^{T+\delta}} \tag{2.31}$$

The simple forward rates $f_s(t,T)$ are defined as (cf. equation (2.23))

$$f_s(t,T) = \frac{1}{\delta}\left(F(t,T,T+\delta) - 1\right)$$

Under $\mathbb{Q}^{T+\delta}$, the simple forward rates $f_s(t,T)$ satisfy

$$\begin{aligned} df_s(t,T) &= \frac{1}{\delta}F(t,T,T+\delta)\boldsymbol{\gamma}(t,T,T+\delta)\cdot d\mathbf{W}_t^{\mathbb{Q}^{T+\delta}} \quad (2.32)\\ &= \frac{1}{\delta}\left(1+\delta\, f_s(t,T)\right)\boldsymbol{\gamma}(t,T,T+\delta)\cdot d\mathbf{W}_t^{\mathbb{Q}^{T+\delta}} \\ &= f_s(t,T)\boldsymbol{\lambda}(t,T)\cdot d\mathbf{W}_t^{\mathbb{Q}^{T+\delta}} \quad (2.33)\end{aligned}$$

where the forward rate volatility function is given by

$$\boldsymbol{\lambda}(t,T) = \frac{1+\delta\, f_s(t,T)}{\delta f_s(t,T)}\boldsymbol{\gamma}(t,T,T+\delta) \tag{2.34}$$

In practice, a finite number of simple forward rates is modelled, not a continuum of forward rates with a fixed compounding period δ. In the following subsection we construct a discrete-tenor LIBOR market model from the interest rate system defined by Assumptions 1 and 2.

2.2.3 Discrete-tenor LIBOR Market Model

The simple forward rates that are usually modelled are the three month rates (e.g. GBP, USD) or the six month rates (e.g. EUR). From now on we use the generic term *forward LIBOR rates* for any family of simple forward rates. The tenor is typically chosen to match the convention in the cap market. The *tenor structure* $\mathfrak{T} = \{T_1, T_2, \ldots, T_{n-1}\}$ is a set of reset times for a family of spanning forward LIBOR rates, with $0 = T_0 < T_1 < T_2 < \ldots < T_{n-1} < T_n = T^*$.[20] For $i = 1, \ldots, n-1$, the forward LIBOR rate at time t for the interval $[T_i, T_{i+1}]$ is denoted by $L_i(t)$ and defined in terms of the forward process $F(t, T_i, T_{i+1})$ as

$$1 + \delta_i L_i(t) = F(t, T_i, T_{i+1}), \quad 0 \le t \le T_i$$

where δ_i is the year-fraction for the interval $[T_i, T_{i+1}]$, using a prespecified day-count convention. Following Musiela & Rutkowski (1997), we construct a discrete-tenor LIBOR market model by backward induction.

Define an interest rate system by imposing Assumptions 1 and 2 discussed in the previous subsection. Under the T_n-forward measure $\mathbb{Q}^{T_n}$ (previously denoted by $\mathbb{Q}^{T^*}$) the forward LIBOR rate $L_{n-1}(t)$ satisfies (cf. equation (2.33) and (2.34))

$$dL_{n-1}(t) = L_{n-1}(t)\boldsymbol{\lambda}(t, T_{n-1})\cdot d\mathbf{W}_t^{\mathbb{Q}^{T_n}}, \quad 0 \le t \le T_{n-1} \tag{2.35}$$

where $\mathbf{W}^{\mathbb{Q}^{T_n}}$ is a d-dimensional $\mathbb{Q}^{T_n}$-Wiener process and

$$\boldsymbol{\lambda}(t, T_{n-1}) = \frac{1+\delta_{n-1}L_{n-1}(t)}{\delta_{n-1}L_{n-1}(t)}\,\boldsymbol{\gamma}(t, T_{n-1}, T_n) \tag{2.36}$$

Assumption 3. *The volatility functions $\lambda(t,T_i)$, $T_i \in \mathfrak{T}$, are exogenously specified, $\mathbb{R}^d$-valued, bounded and* ***deterministic*** *functions of t and T_i.*

The exogenously specified volatility function $\boldsymbol{\lambda}(t, T_{n-1})$ completely determined the dynamics of the forward LIBOR rate $L_{n-1}(t)$. We now construct the family of forward LIBOR rate processes using a backward induction procedure.

Define an equivalent martingale measure $\mathbb{Q}^{T_{n-1}}$ on $(\Omega, \mathscr{F}_{T_{n-1}})$ by the Radon-Nikodým derivative

$$\frac{d\mathbb{Q}^{T_{n-1}}}{d\mathbb{Q}^{T_n}} = \exp\left(-\frac{1}{2}\int_0^{T_{n-1}} ||\boldsymbol{\gamma}(t, T_{n-1}, T_n)||^2 dt + \int_0^{T_{n-1}} \boldsymbol{\gamma}(t, T_{n-1}, T_n)\cdot d\mathbf{W}_t^{\mathbb{Q}^{T_n}}\right)$$

Then, by Girsanov's Theorem

$$W_t^{\mathbb{Q}^{T_{n-1}}} = W_t^{\mathbb{Q}^{T_n}} - \int_0^t \boldsymbol{\gamma}(u, T_{n-1}, T_n)du, \quad 0 \le t \le T_{n-1} \tag{2.37}$$

[20] A family of spanning forward LIBOR rates means that the maturity date of the first forward LIBOR rate is the reset date of the second forward LIBOR rate. The maturity date of the second forward LIBOR rate is the reset date of the third forward LIBOR rate and so on.

is a d-dimensional $\mathbb{Q}^{T_{n-1}}$-Wiener process. Under the $\mathbb{Q}^{T_{n-1}}$, the forward LIBOR rate $L_{n-2}(t)$ satisfies

$$dL_{n-2}(t) = L_{n-2}(t)\boldsymbol{\lambda}(t, T_{n-2}) \cdot d\mathbf{W}_t^{\mathbb{Q}^{T_{n-1}}}, \quad 0 \leq t \leq T_{n-2}$$

Using equations (2.37) and (2.35) it is easy to see that the forward LIBOR rate $L_{n-1}(t)$ satisfies

$$\begin{aligned} dL_{n-1}(t) &= L_{n-1}(t)\boldsymbol{\lambda}(t, T_{n-1}) \cdot d\mathbf{W}_t^{\mathbb{Q}^{T_n}} \\ &= L_{n-1}(t)\boldsymbol{\lambda}(t, T_{n-1}) \cdot \left(\boldsymbol{\gamma}(t, T_{n-1}, T_n)dt + d\mathbf{W}_t^{\mathbb{Q}^{T_{n-1}}}\right) \\ &= L_{n-1}(t)\boldsymbol{\lambda}(t, T_{n-1}) \cdot \left(\frac{\delta_{n-1}L_{n-1}(t)}{1+\delta_{n-1}L_{n-1}(t)}\boldsymbol{\lambda}(t, T_{n-1})dt + d\mathbf{W}_t^{\mathbb{Q}^{T_{n-1}}}\right) \end{aligned}$$

We repeat this backward procedure until we have constructed the entire family of forward LIBOR processes, such that, under $\mathbb{Q}^{T_{i+1}}$, for $i = 1, \ldots, n-1$

$$dL_i(t) = L_i(t)\boldsymbol{\lambda}(t, T_i) \cdot d\mathbf{W}_t^{\mathbb{Q}^{T_{i+1}}}, \quad 0 \leq t \leq T_i \tag{2.38}$$

and for $j = i+1, \ldots, n-1$

$$dL_j(t) = L_j(t)\boldsymbol{\lambda}(t, T_j) \cdot \left(\sum_{k=i+1}^{j} \frac{\delta_k L_k(t)}{1+\delta_k L_k(t)}\boldsymbol{\lambda}(t, T_k)dt + d\mathbf{W}_t^{\mathbb{Q}^{T_{i+1}}}\right), \quad 0 \leq t \leq T_j \tag{2.39}$$

where $\mathbf{W}_t^{\mathbb{Q}^{T_{i+1}}}$ is a d-dimensional $\mathbb{Q}^{T_{i+1}}$-Wiener process. To fully specify the dynamics of the forward LIBOR rates under each measure $\mathbb{Q}^{T_{i+1}}$, it remains to derive the dynamics of $L_j(t)$ for $j = 1, \ldots, i-1$. This can easily be achieved using "forward" induction. Under $\mathbb{Q}^{T_2}$, the forward LIBOR rate $L_1(t)$ satisfies

$$dL_1(t) = L_1(t)\boldsymbol{\lambda}(t, T_1) \cdot d\mathbf{W}_t^{\mathbb{Q}^{T_2}}, \quad 0 \leq t \leq T_1$$

The measure $\mathbb{Q}^{T_2}$ is defined by the Radon-Nikodým derivative

$$\frac{d\mathbb{Q}^{T_2}}{d\mathbb{Q}^{T_3}} = \exp\left(-\frac{1}{2}\int_0^{T_2} ||\boldsymbol{\gamma}(t, T_2, T_3)||^2 dt + \int_0^{T_2} \boldsymbol{\gamma}(t, T_2, T_3) \cdot d\mathbf{W}_t^{\mathbb{Q}^{T_3}}\right)$$

where

$$\boldsymbol{\gamma}(t, T_2, T_3) = \frac{\delta_2 L_2(t)}{1+\delta_2 L_2(t)}\boldsymbol{\lambda}(t, T_2)$$

Since $\mathbb{Q}^{T_2}$ and $\mathbb{Q}^{T_3}$ are equivalent measures, $\frac{d\mathbb{Q}^{T_3}}{d\mathbb{Q}^{T_2}} = \left(\frac{d\mathbb{Q}^{T_2}}{d\mathbb{Q}^{T_3}}\right)^{-1}$ and hence under $\mathbb{Q}^{T_3}$, by Girsanov's Theorem, $L_1(t)$ satisfies

$$dL_1(t) = L_1(t)\boldsymbol{\lambda}(t, T_1) \cdot \left(d\mathbf{W}_t^{\mathbb{Q}^{T_3}} - \frac{\delta_2 L_2(t)}{1+\delta_2 L_2(t)}\boldsymbol{\lambda}(t, T_2)dt\right), \quad 0 \leq t \leq T_1$$

This generalizes under $\mathbb{Q}^{T_{i+1}}$ to

$$dL_j(t) = L_j(t)\boldsymbol{\lambda}(t, T_j) \cdot \left(d\mathbf{W}_t^{\mathbb{Q}^{T_{i+1}}} - \sum_{k=j+1}^{i} \frac{\delta_k L_k(t)}{1+\delta_k L_k(t)}\boldsymbol{\lambda}(t, T_k)dt\right), \quad 0 \leq t \leq T_j \tag{2.40}$$

for all $j = 1, \ldots, i-1$. This concludes the construction of forward LIBOR rates under the forward measures $\mathbb{Q}^{T_{i+1}}$, $i = 1, \ldots, n-1$. Note that for the purpose of pricing interest rate derivatives, one would choose a single measure. However, as mentioned previously, depending on the derivative in question, some choices are more convenient than others.

Spot LIBOR Measure

There is one particular self-financing trading strategy that is analogous to the money-market account. This trading strategy is as follows: at $t = 0$, invest one unit of currency in a zero-coupon bond maturing at T_1. At $t = T_1$, invest the proceeds in a zero-coupon bond maturing

at T_2 and so on. The value at time t of this self-financing trading strategy, termed the *simply compounded money-market account* is

$$B^*(t) = B\left(t, T_{\eta(t)}\right) \prod_{k=0}^{\eta(t)-1} \left(1 + \delta_k L_k(T_k)\right), \quad 0 \le t \le T_n \tag{2.41}$$

where $L_0(T_0)$ is the spot (current) LIBOR rate and η is the index of the next tenor date, a left-continuous function $\eta : [0, T_n) \to \{1, \ldots, n\}$ such that $\eta(t)$ is the unique integer satisfying $T_{\eta(t)-1} \le t < T_{\eta(t)}$ with $\eta(T_n) = n$. For $i = 1, \ldots, n$, the normalized bond prices

$$\begin{aligned} \frac{B(t,T_i)}{B^*(t)} &= \frac{B\left(t, T_{\eta(t)}\right) \prod_{j=\eta(t)}^{i-1} \left(1 + \delta_j L_j(t)\right)^{-1}}{B\left(t, T_{\eta(t)}\right) \prod_{k=0}^{\eta(t)-1} \left(1 + \delta_k L_k(T_k)\right)}, \quad 0 \le t \le T_i \\ &= \left(\prod_{k=0}^{\eta(t)-1} \left(1 + \delta_k L_k(T_k)\right)^{-1} \right) \left(\prod_{j=\eta(t)}^{i-1} \left(1 + \delta_j L_j(t)\right)^{-1} \right) \end{aligned} \tag{2.42}$$

are functions of the forward LIBOR rates only.[21]

The *spot LIBOR measure* $\mathbb{Q}^*$ is a unique measure on $(\Omega, \mathscr{F}_{T_n})$, equivalent to $\mathbb{Q}^{T_n}$, such that the normalized bond prices are martingales under $\mathbb{Q}^*$. By the Martingale Representation Theorem 2.1.4, for all $i = 1, \ldots, n$, there exist a $\mathbb{R}^d$-valued process $\boldsymbol{\nu}_i = \left\{\boldsymbol{\nu}_i(t),\ 0 \le t \le T_i\right\}$ satisfying $\mathbb{Q}^*\left[\int_0^{T_i} ||\boldsymbol{\nu}_i(u)||^2 du < \infty\right] = 1$, such that

$$d\left(\frac{B(t,T_i)}{B^*(t)}\right) = \left(\frac{B(t,T_i)}{B^*(t)}\right) \boldsymbol{\nu}_i(t) \cdot d\mathbf{W}_t^{\mathbb{Q}^*} \tag{2.43}$$

where $\mathbf{W}^{\mathbb{Q}^*}$ is a d-dimensional $\mathbb{Q}^*$-Wiener process. It remains to determine ν. By Itô's formula

$$d\ln\left(\frac{B(t,T_i)}{B^*(t)}\right) = -\frac{1}{2}||\boldsymbol{\nu}_i(t)||^2 dt + \boldsymbol{\nu}_i(t) \cdot d\mathbf{W}_t^{\mathbb{Q}^*} \tag{2.44}$$

From the definition of the normalized bond price process, equation (2.42), we have

$$d\log\left(\frac{B(t,T_i)}{B^*(t)}\right) = -\sum_{j=\eta(t)}^{i-1} d\log\left(1 + \delta_j L_j(t)\right) \tag{2.45}$$

because $\sum_{k=0}^{\eta(t)-1} \log\left(1 + \delta_k L_k(T_k)\right)$ is constant with respect to t. Girsanov's Theorem states that for an equivalent measure change, the diffusion coefficient of the process remains unchanged. From equations (2.38), (2.39) and (2.40), the diffusion coefficient of the forward LIBOR rate $L_j(t)$,$j = 1, \ldots, n-1$, under any measure is $L_j(t)\boldsymbol{\lambda}(t, T_j)$. By Itô's formula, the diffusion coefficient of the process $\log\left(1 + \delta_j L_j(t)\right)$ is

$$\frac{\delta_j}{1 + \delta_j L_j(t)} L_j(t) \boldsymbol{\lambda}(t, T_j)$$

Equating the diffusion coefficients of equations (2.44) and (2.45) yields

$$\boldsymbol{\nu}_i(t) = -\sum_{j=\eta(t)}^{i-1} \frac{\delta_j L_j(t) \boldsymbol{\lambda}(t, T_j)}{1 + \delta_j L_j(t)} \tag{2.46}$$

Now that we have obtained an expression for the processes ν, using Lemma 2.2.1 we can derive the stochastic differential equations that the forward LIBOR rates satisfy under the spot measure

[21] Note that the simply compounded money market account at time t is a function of the realized forward LIBOR rates, while the zero-coupon bond price is a function of the still-alive forward LIBOR rates at time t.

$\mathbb{Q}^*$. For $i = 1, \ldots, n-1$ and $0 \le t \le T_i$ we have

$$\begin{aligned} dL_i(t) &= \frac{1}{\delta_i} d\left(\frac{B(t,T_i)}{B(t,T_{i+1})}\right) \qquad (2.47)\\ &= \delta_i^{-1} d\left(\frac{B(t,T_i)}{B^*(t)} \Big/ \frac{B(t,T_{i+1})}{B^*(t)}\right) \\ &= \delta_i^{-1}\left(\frac{B(t,T_i)}{B(t,T_{i+1})}\right)\left(\boldsymbol{\nu}_i(t) - \boldsymbol{\nu}_{i+1}(t)\right) \cdot \left(-\boldsymbol{\nu}_{i+1}(t)dt + d\mathbf{W}_t^{\mathbb{Q}^*}\right) \\ &= \delta_i^{-1}\big(1+\delta_i L_i(t)\big)\left(\frac{\delta_i L_i(t)\boldsymbol{\lambda}(t,T_i)}{1+\delta_i L_i(t)}\right) \cdot \left(\sum_{j=\eta(t)}^{i} \frac{\delta_j L_j(t)\boldsymbol{\lambda}(t,T_j)}{1+\delta_j L_j(t)} dt + d\mathbf{W}_t^{\mathbb{Q}^*}\right) \\ &= L_i(t)\boldsymbol{\lambda}(t,T_i) \cdot \left(\sum_{j=\eta(t)}^{i} \frac{\delta_j L_j(t)\boldsymbol{\lambda}(t,T_j)}{1+\delta_j L_j(t)} dt + d\mathbf{W}_t^{\mathbb{Q}^*}\right) \qquad (2.48) \end{aligned}$$

for $0 \le t \le T_i$. This completes the construction of the discrete-tenor LIBOR market model. We summarize the LIBOR market model stochastic differential equations under the various measures in the box below.

Under the T_{i+1}-forward measure $\mathbb{Q}^{T_{i+1}}$, $i = 1, \ldots, n-1$

$$\begin{aligned} dL_j(t) &= L_j(t)\boldsymbol{\lambda}(t,T_j) \cdot \left(-\sum_{k=j+1}^{i} \frac{\delta_k L_k(t)}{1+\delta_k L_k(t)} \boldsymbol{\lambda}(t,T_k)dt + d\mathbf{W}_t^{\mathbb{Q}^{T_{i+1}}}\right), \quad j = 1, \ldots, i-1 \\ dL_i(t) &= L_i(t)\boldsymbol{\lambda}(t,T_i) \cdot d\mathbf{W}_t^{\mathbb{Q}^{T_{i+1}}} \\ dL_j(t) &= L_j(t)\boldsymbol{\lambda}(t,T_j) \cdot \left(\sum_{k=i+1}^{j} \frac{\delta_k L_k(t)}{1+\delta_k L_k(t)} \boldsymbol{\lambda}(t,T_k)dt + d\mathbf{W}_t^{\mathbb{Q}^{T_{i+1}}}\right), \quad j = i+1, \ldots, n-1 \end{aligned}$$

where $\mathbf{W}_t^{\mathbb{Q}^{T_{i+1}}}$ is a d-dimensional $\mathbb{Q}^{T_{i+1}}$-Wiener process.

Under the spot measure $\mathbb{Q}^*$

$$dL_i(t) = L_i(t)\boldsymbol{\lambda}(t,T_i) \cdot \left(\sum_{j=\eta(t)}^{i} \frac{\delta_j L_j(t)\boldsymbol{\lambda}(t,T_j)}{1+\delta_j L_j(t)} dt + d\mathbf{W}_t^{\mathbb{Q}^*}\right), \quad i = 1, \ldots, n-1$$

where $\mathbf{W}_t^{\mathbb{Q}^*}$ is a d-dimensional $\mathbb{Q}^*$-Wiener process.

2.2.4 Arbitrage-free Interpolation

The discrete-tenor LIBOR market model framework presented in the previous subsection does not determine a continuum of zero-coupon bond prices. Because of this, there are interest rate instruments that do not fit into the framework. An example is a trigger swap that triggers whenever a certain LIBOR rate passes a reference level, where the trigger can occur at any time (Joshi 2003).

Musiela & Rutkowski (1997) fix a compounding period δ and extend the model to a continuous-tenor case. They apply the assumptions that Brace et al. (1997) imposed in their derivation, namely that all the forward rates with a compounding period of δ are lognormally distributed and that the zero-coupon bonds with maturities less than δ have zero volatilities. Musiela & Rutkowski (1997, page 289) show that if a family of zero-coupon bond prices is constructed using these assumptions, it cannot be guaranteed that zero-coupon bond prices will not exceed unity.

Schlögl (2001) presented an arbitrage-free interpolation of the discrete-tenor model. The value of a zero-coupon bond, for $0 \le t_1 < t_2 \le T_n$, is given by

$$B(t_1,t_2) = B\left(t_1, T_{\eta(t_1)}\right)\left(\prod_{j=\eta(t_1)}^{\eta(t_2)-1} (1+\delta_j L_j(t_1))^{-1}\right)\left(\frac{B(t_1,t_2)}{B\left(t_1,T_{\eta(t_2)}\right)}\right)$$

The forward process $\frac{B(t_1,t_2)}{B\left(t_1,T_{\eta(t_2)}\right)}$ is a martingale under the $T_{\eta(t_2)}$-forward measure. Hence

$$\begin{aligned} B(t_1,t_2) &= B\left(t_1,T_{\eta(t_1)}\right)\left(\prod_{j=\eta(t_1)}^{\eta(t_2)-1}(1+\delta_j L_j(t_1))^{-1}\right)\mathbb{E}^{\mathbb{Q}^{T_{\eta(t_2)}}}\left[\frac{B(t_2,t_2)}{B\left(t_2,T_{\eta(t_2)}\right)}\,\middle|\,\mathscr{F}_{t_1}\right] \\ &= B\left(t_1,T_{\eta(t_1)}\right)\left(\prod_{j=\eta(t_1)}^{\eta(t_2)-1}(1+\delta_j L_j(t_1))^{-1}\right)\mathbb{E}^{\mathbb{Q}^{T_{\eta(t_2)}}}\left[B\left(t_2,T_{\eta(t_2)}\right)^{-1}\,\middle|\,\mathscr{F}_{t_1}\right] \end{aligned}$$

This means that to determine the continuum of zero-coupon bonds from the discrete forward LIBOR rates, we only need to determine a way of calculating the value of short-dated bonds, $B\left(t,T_{\eta(t)}\right)$, $0 \le t \le T_n$.

The interpolation method presented here is based on the assumption that the volatility of the short-dated bonds is zero. This means that the short-dated zero-coupon bonds are interpolated from the spot LIBOR rate. In the simplest case of linear interpolation

$$B\left(t,T_{\eta(t)}\right) = \left(1+\left(T_{\eta(t)}-t\right)L_{\eta(t)-1}\left(T_{\eta(t)-1}\right)\right)^{-1}$$

we obtain monotonically increasing family of zero-coupon bond prices over $[T_{\eta(t)-1}, T_{\eta(t)}]$ satisfying

$$B\left(T_{\eta(t)-1},T_{\eta(t)}\right) = \left(1+\left(T_{\eta(t)}-T_{\eta(t)-1}\right)L_{\eta(t)-1}\left(T_{\eta(t)-1}\right)\right)^{-1}$$

and

$$B\left(T_{\eta(t)},T_{\eta(t)}\right) = 1$$

The linear interpolation scheme produces an explicit solution for the expectation

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}^{T_{\eta(t_2)}}}\left[B\left(t_2,T_{\eta(t_2)}\right)^{-1}\,\middle|\,\mathscr{F}_{t_1}\right] &= 1+\left(T_{\eta(t_2)}-t_2\right)\mathbb{E}^{\mathbb{Q}^{T_{\eta(t_2)}}}\left[L_{\eta(t_2)-1}\left(T_{\eta(t_2)-1}\right)\,\middle|\,\mathscr{F}_{t_1}\right] \\ &= 1+\left(T_{\eta(t_2)}-t_2\right)L_{\eta(t_2)-1}(t_1) \end{aligned}$$

since the forward rate $L_{\eta(t_2)-1}$ is a martingale under $\mathbb{Q}^{T_{\eta(t_2)}}$.

Chapter 3

Calibration of the LIBOR Market Model

It's not the least use to me to foresee the future; I have never known how to avoid it.

– Jean-Jacques Rousseau

The specification of the LIBOR market model relies on the specification of the deterministic volatility functions $\lambda(t, T_i)$, $T_i \in \mathfrak{T}$. These functions determine both the volatility level of the forward rates and the correlation between the rates. The selected covariance structure should match the observable dynamics of the forward rates, such as the number and shape of the underlying principal components (James & Webber 2000). Once the forward rate volatility functions are specified (parameterized), the chosen model is calibrated to the current forward rate curve and to liquid interest rate derivatives, namely caps and swaptions. The calibration to the current forward rate curve is automatic because the current forward rates are initial conditions, and hence inputs, for the LIBOR market model stochastic differential equations. Calibration to the cap and swaption prices is achieved by choosing the forward rate volatility functions such that the model prices of these derivatives match the market prices as closely as possible. The result of a successful calibration is that the pricing and hedging of exotic or less liquid LIBOR derivatives can be achieved consistently with the market data.

The two main advantages of the LIBOR market model that are often stated in the literature are, firstly, the ability of the model to calibrate to a large set of liquid market instruments and secondly, the ability of the model to generate a realistic evolution of the forward rate volatility structure (Piterbarg 2004). However, this is not always seen as a real advantage over modelling the instantaneous forward rates directly. Quoting Andreasen (2004):

> "In my view, the best thing about the 'LMM movement' is that it got us thinking about using closed-form approximations for caps and swaptions in calibration of models."

As we will show in Section 3.3, while the LIBOR market model is consistent with the Black caplet formula, there is no closed-form formula for swaption prices. This is an extremely undesirable model feature because without a closed-form formula, swaption prices will have to be calculated using Monte Carlo simulation at each iteration step in the calibration procedure.[1] This results in an unstable and lengthy calibration process. Fortunately, many approximate closed-form swaption formulas have been developed (Andersen & Andreasen 2000, Daniluk & Gątarek 2005, Hull & White 2000, Jäckel & Rebonato 2003, Kawai 2003). This is an important point because accurate approximations to the prices of liquid interest rate derivatives make implementation of more complex models possible in practice (for example the stochastic volatility separable HJM models of Andersen & Andreasen (2002) and Andreasen (2004, 2005)).

In Section 3.1 we discuss the choice of the number of factors in the LIBOR market model. In Section 3.2 we deal with a related issue of rank reduction of correlation matrices. The set of market data commonly used for calibration is discussed in Section 3.3. Model specifications are described in the final section, together with an example calibration of the LIBOR market model.

[1] The necessity, in general, of pricing derivatives in the LIBOR market model using Monte Carlo simulation will be examined in the following chapter.

3.1 Model Dimension

In the previous chapter we derived the no-arbitrage dynamics for a family of spanning forward LIBOR rates under various measures. In the literature (Brigo & Mercurio 2001, Rebonato 2002), the dynamics of the $n-1$ forward LIBOR rates are frequently formulated in terms of $n-1$ correlated Brownian motions (full-factor formulation). For example, under $\mathbb{Q}^*$, for $i = 1, \ldots, n-1$

$$\frac{dL_i(t)}{L_i(t)} = \sum_{j=\eta(t)}^{i} \frac{\delta_j L_j(t)\sigma(t,T_j)\sigma(t,T_i)\rho_{ij}(t)}{1+\delta_j L_j(t)}dt + \sigma(t,T_i)dW_i(t), \quad 0 \le t \le T_i \tag{3.1}$$

where $\sigma(t,T_i)$ is the instantaneous volatility function of the forward LIBOR rate $L_i(t)$, a $\mathbb{R}$-valued, bounded and deterministic function of t and T_i, $\boldsymbol{\rho}(t) \in \mathbb{R}^{(n-1)\times(n-1)}$ is a time-dependent instantaneous correlation matrix and $\mathbf{W} = (W_1, \ldots, W_{n-1})'$ is a vector of correlated Brownian motions (under $\mathbb{Q}^*$) satisfying

$$dW_i(t)dW_j(t) = \rho_{ij}(t)dt$$

To relate formulation (3.1) to the previous formulation (2.48), we need to express the former in terms of $n-1$ independent Brownian motions. From the properties of multivariate normal distributions, if $\mathbf{X} = (X_1, \ldots, X_m)'$ is $N(\mathbf{0}, \mathbf{I})$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ is given by $\mathbf{Y} = \mathbf{D}\mathbf{X}$ for some $m \times m$ matrix $\mathbf{D}$ of rank $\le m$ then $\mathbf{Y}$ is $N(\mathbf{0}, \mathbf{D}\mathbf{D}')$ (Grimmett & Stirzaker 2001). Thus given a decomposition of the correlation matrix $\boldsymbol{\rho}(t) = \mathbf{Q}(t)\mathbf{Q}(t)'$, where $\mathbf{Q}(t) \in \mathbb{R}^{(n-1)\times(n-1)}$,[2] we can express formulation (3.1) as

$$\frac{dL_i(t)}{L_i(t)} = \sum_{j=\eta(t)}^{i} \frac{\delta_j L_j(t)\sigma(t,T_j)\sigma(t,T_i)\,\mathbf{q}_j(t)\cdot\mathbf{q}_i(t)}{1+\delta_j L_j(t)}dt + \sigma(t,T_i)\,\mathbf{q}_i(t)d\mathbf{Z}(t) \tag{3.2}$$

where $\mathbf{q}_i(t)$ is the i'th row of $\mathbf{Q}(t)$ and $\mathbf{Z} = (Z_1, \ldots, Z_{n-1})'$ is a vector of independent Brownian motions under $\mathbb{Q}^*$. The relationship between the two formulations is now clear:

$$\sigma(t,T_i)\,\mathbf{q}_i(t)' \equiv \boldsymbol{\lambda}(t,T_i) \tag{3.3}$$

if the dimension $d = n-1$.[3] The factor loadings $\boldsymbol{\lambda}(t,T_i)$ can be "decomposed" into two components: the instantaneous volatility of the forward LIBOR rate, $\sigma(t,T_i)$, and a vector $\mathbf{q}_i(t)$ determining the instantaneous correlation between the forward rates.

The forward LIBOR rates that are modelled are typically the three or the six month forward rates. The maturity of an interest rate derivative may be as long as twenty or thirty years. This means that in the full-factor formulation (3.1), the number of driving Brownian motions may become rather large. In the US government bond market, Litterman & Scheinkman (1991) identified three main factors that describe the variation in the returns. They termed these *level*, *steepness* and *curvature* and together, according to their study, they explain at least 96% of the variability. Knez, Litterman & Scheinkman (1994) found that three factors explain, on average, 86% of the variation of short-dated money market instruments and that a fourth factor was needed. Morini & Webber (2004) estimated a historical correlation matrix for the annual EUR forward rates, going out to 20 years. Their results show that 61.6% of the variance is explained by the first factor, 79.1% by the first three and 88.5% by the first six. It seems safe to say that 3 to 6 factors are needed to explain most of the variation in the forward rates. Interest rate models employed in practice, in South Africa, are typically one or two-factor models.

To reduce the full-factor formulation to a formulation with $d << n-1$ independent Brownian motions, d being the required number of factors, we need to find a rank-d correlation matrix that is close to $\boldsymbol{\rho}(t)$ in some sense. The problem is to find a $(n-1)\times d$ matrix $\widetilde{\mathbf{Q}}(t)$ such that $\widetilde{\mathbf{Q}}(t)\widetilde{\mathbf{Q}}(t)'$ is a correlation matrix that is close to $\boldsymbol{\rho}(t)$. Possible approaches will be reviewed in the following section. Then the d-factor dynamics are

$$\frac{dL_i(t)}{L_i(t)} = \sum_{j=\eta(t)}^{i} \frac{\delta_j L_j(t)\sigma(t,T_j)\sigma(t,T_i)\,\widetilde{\mathbf{q}}_j(t)\cdot\widetilde{\mathbf{q}}_i(t)}{1+\delta_j L_j(t)}dt + \sigma(t,T_i)\,\widetilde{\mathbf{q}}_i(t)d\mathbf{X}(t) \tag{3.4}$$

[2] A possible way to decompose a correlation matrix is by using a singular value decomposition or through a Cholesky factorization - see Golub & Van Loan (1996).

[3] Note that in formulation (2.48), the instantaneous correlation between forward rates L_i and L_j is defined as

$$\rho_{ij}(t) = \frac{d\langle L_i, L_j\rangle(t)}{\sqrt{d\langle L_i, L_i\rangle(t)d\langle L_j, L_j\rangle(t)}} = \frac{\boldsymbol{\lambda}(t,T_i)\cdot\boldsymbol{\lambda}(t,T_j)}{||\boldsymbol{\lambda}(t,T_i)||\,||\boldsymbol{\lambda}(t,T_j)||}, \quad 0 \le t \le \min\{T_i, T_j\}$$

This is consistent with equation (3.3).

where $\widetilde{\mathbf{q}}_i(t)$ is the i'th row of $\widetilde{\mathbf{Q}}(t)$ and $\mathbf{X} = (X_1, \ldots, X_d)'$ is a vector of d independent Brownian motions under $\mathbb{Q}^*$. This is equivalent to the general d-factor formulation (2.48)

$$\sigma(t, T_i)\, \widetilde{\mathbf{q}}_i(t)' \equiv \boldsymbol{\lambda}(t, T_i) \tag{3.5}$$

3.2 Rank Reductions

A correlation matrix is a symmetric, positive semidefinite matrix with a unit diagonal.[4] When modelling interest rate derivatives, we are often faced with the problem of finding a low-rank correlation matrix to best approximate some exogenous correlation matrix. For example, it is frequently assumed that the instantaneous correlation matrix $\boldsymbol{\rho}(t)$ is constant as a function of time $\boldsymbol{\rho}(t) = \boldsymbol{\rho}$. The matrix $\boldsymbol{\rho}$ is then estimated either from historical yield curve data or implied from the current market data. If $\boldsymbol{\rho}$ is estimated from historical data, the estimated matrix will, in general, be a full-rank matrix. If $\boldsymbol{\rho}$ is implied from the market data, the rank of the implied matrix will depend on the chosen parametrization. However, as discussed in the previous section, the required interest rate model is typically two dimensional. This corresponds to $\boldsymbol{\rho}$ being a rank-two matrix. In general, the estimated instantaneous correlation matrix will not have the required rank. We are then faced with the problem of finding a rank-two correlation matrix closest to $\boldsymbol{\rho}$ in some sense. This problem is termed rank reduction.

Let us state the problem formally. Suppose that we have an exogenous $m \times m$ correlation matrix $\mathbf{C}$ and that we are interested in finding the closest rank-$r < m$ correlation matrix. To measure distance on the space of matrices, we require a matrix norm. The most frequently used norm is the Frobenius matrix norm

$$\begin{aligned} ||\mathbf{A}||_F^2 &= \text{trace}(\mathbf{A}\mathbf{A}') \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m} |a_{ij}|^2 \end{aligned}$$

From now on, when we refer to the "closest" matrix it should be interpreted as the closest matrix in the Frobenius norm. The closest rank-r correlation matrix $\mathbf{C}_r$ is given by $\mathbf{C}_r = \mathbf{Q}_r\mathbf{Q}_r'$, where $\mathbf{Q}_r$ is a solution to the following problem

$$\begin{aligned} &\min_{\mathbf{Q}\in\mathbb{R}^{m\times r}} ||\mathbf{C} - \mathbf{Q}\mathbf{Q}'||_F^2 \\ &\textbf{subject to } \text{diag}(\mathbf{Q}\mathbf{Q}') = \text{diag}(\mathbf{I}) \end{aligned} \tag{3.6}$$

The solution $\mathbf{Q}_r\mathbf{Q}_r'$ will automatically be positive semidefinite. Thus the constraint need not be incorporated explicitly. We now briefly discuss various methods that have been proposed for finding an approximate solution to this problem, followed by a simple comparison test.

One of the most widely used rank reduction technique in practice is the *normalized principal components* method of Rebonato & Jäckel (1999) and Brigo (2002). This is certainly due, in part, to its conceptual simplicity and rapid computation. Principal components analysis (PCA) is a technique that linearly transforms a group of variables into orthogonal variables (principal components) which then form an orthogonal basis for the data space. PCA is based on the following result from linear algebra, known as the *singular value decomposition* (Golub & Van Loan 1996).

Theorem 3.2.1. *Suppose $\mathbf{S} \in \mathbb{R}^{m\times m}$ is a symmetric positive semidefinite matrix. The singular value decomposition of $\mathbf{S}$ is given by*

$$\mathbf{S} = \mathbf{U}\text{diag}(\lambda_1, \ldots, \lambda_m)\mathbf{U}'$$

a product of a $m \times m$ column-orthonormal matrix $\mathbf{U}$, a $m \times m$ diagonal matrix of eigenvalues $\text{diag}(\lambda_1, \ldots, \lambda_m)$, ordered in descending order, and the transpose of $\mathbf{U}$. Then the Eckart-Young decomposition

$$\mathbf{S}_r = \mathbf{U}\text{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)\mathbf{U}'$$

is the closest rank-r symmetric positive semidefinite matrix. Further,

$$||\mathbf{S} - \mathbf{S}_r||_F^2 = \sum_{i=r+1}^{m} \lambda_i^2$$

[4] This implies that $|a_{ij}| \leq 1$, because a positive semidefinite matrix satisfies $|a_{ij}| \leq \sqrt{a_{ii}a_{jj}}$, $i \neq j$.

The Eckart-Young decomposition of a correlation matrix $\mathbf{C}$ will not necessarily have a unit diagonal. Suppose that $\mathbf{U}\text{diag}(\lambda_1, \ldots, \lambda_m)\mathbf{U}'$ is the singular value decomposition of $\mathbf{C}$ and let $\widetilde{\mathbf{C}}$ be the Eckart-Young decomposition $\widetilde{\mathbf{C}} = \mathbf{U}\text{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)\mathbf{U}'$. To ensure that $\widetilde{\mathbf{C}}$ is a valid correlation matrix, we *normalize* it by pre- and post-multiplying $\widetilde{\mathbf{C}}$ by the scaling matrix $\mathbf{S}$

$$\mathbf{S} = \text{diag}\left(\frac{1}{\sqrt{\widetilde{c}_{11}}}, \ldots\ldots, \frac{1}{\sqrt{\widetilde{c}_{mm}}}\right)$$

where $\widetilde{c}_{ii}$ is the i'th diagonal element of $\widetilde{\mathbf{C}}$. Then $\mathbf{S}\widetilde{\mathbf{C}}\mathbf{S}$ is a rank-r correlation matrix. However, it is exactly in this normalization that we lose the optimality of the solution. That is, $\mathbf{S}\widetilde{\mathbf{C}}\mathbf{S}$ is *not* the closest rank-r correlation matrix.

The second rank reduction technique that we discuss is the *angles parametrization* method of Rebonato & Jäckel (1999), Brigo (2002) and Rapisarda, Brigo & Mercurio (2002). The unit diagonal constraint in the problem formulation (3.6) can be expressed as

$$q_{i1}^2 + q_{i2}^2 + \ldots + q_{ir}^2 = 1, \quad \text{for all } i = 1, \ldots, m$$

As a consequence, Rebonato & Jäckel (1999) noted that, for each i, $(q_{i1}, q_{i2}, \ldots, q_{ir})'$ is a r-tuple defining a r-hypersphere with unit radius. Each element of the r-tuple can be expressed in terms of spherical coordinates as

$$q_{ik} = \begin{cases} \cos(\theta_{ik}) \prod_{j=1}^{k-1} \sin(\theta_{ij}) & 1 \leq k < r \\ \prod_{j=1}^{k-1} \sin(\theta_{ij}) & k = r \end{cases}$$

Rebonato & Jäckel (1999) parameterized each row of the matrix $\mathbf{Q}$ using the above spherical coordinates. The angles $\{\theta_{i1}, \ldots, \theta_{i(r-1)}\}_{i=1}^m$ are optimized until an exogenously specified convergence criterion is reached.[5] In general, the angles parametrization method produces an estimate of the rank-r correlation matrix with smaller error than the normalized principal components method.

Another rank reduction technique is the extension of the *alternating projections* method of Higham (2002). Higham (2002) applied the alternating projections theory to find the nearest rank-m correlation matrix to an exogenously given approximate correlation matrix in $\mathbb{R}^{m \times m}$. An approximate correlation matrix arises in, for example, stress-testing situations in which some part of the existing correlation matrix is altered, leading to a matrix that is not positive semidefinite. This theory has been extended to the problem of rank reduction by Weigel (2004) and Morini & Webber (2004). Before we describe this approach we must stress that the extension proposed by latter authors is no longer guaranteed to converge. However, numerical results on historical forward rate correlation matrices have indicated that the convergence of the algorithm is fast and that accurate results are obtained.

We first state the underlying theory. Consider a Hilbert space $\mathscr{H}$ and a subset $\mathscr{K}$ of $\mathscr{H}$. The *projection* of x onto $\mathscr{K}$ is a point in $\mathscr{K}$ closest to $x \in \mathscr{H}$, i.e. a map $P_{\mathscr{K}} : \mathscr{H} \to \mathscr{K}$, defined by $P_{\mathscr{K}}(x) = \{y \in \mathscr{K} : y = \text{argmin}_{u \in \mathscr{K}} ||u - x||\}$. von Neumann (1950) proved that for $x \in \mathscr{H}$ and $\mathscr{S}$ and $\mathscr{U}$ closed subspaces of $\mathscr{H}$, iteratively projecting onto subspaces

$$x \leftarrow P_{\mathscr{U}}\left(P_{\mathscr{S}}(x)\right)$$

converges to a point in the intersection of $\mathscr{U}$ and $\mathscr{S}$ that is closest to the starting point x. Higham (2002) formulated the problem of finding a rank-m correlation matrix on the Hilbert space $\mathscr{H}$ of symmetric $m \times m$ matrices, where the constraint set $\mathscr{S}$ is the set of positive semidefinite matrices $\mathscr{S} = \left\{\mathbf{A} \in \mathscr{H} : \mathbf{A} \succeq 0\right\}$ while the constraint set $\mathscr{U}$ is the set of matrices with the diagonal elements equal to one $\mathscr{U} = \left\{\mathbf{A} \in \mathscr{H} : a_{ii} = 1, i = 1, \ldots, m\right\}$. However, the constraint sets $\mathscr{S}$ and $\mathscr{U}$ are closed convex sets and not subspaces (see Grubišić (2002) for a proof of this statement and a thorough examination of this algorithm). For the alternating projections to converge to an optimal point, Higham (2002) showed that Dykstra's algorithm (Dykstra 1983) can be applied. This modifies the iteration so that every time the projection onto a convex set is applied, the changes made by the projection at the last iteration are added back. Higham (2002) derived the projections of $\mathbf{A} \in \mathscr{H}$ onto $\mathscr{S}$ and onto $\mathscr{U}$ in the (weighted) Frobenius norm. In particular, in the Frobenius norm, the projection of $\mathbf{A} \in \mathscr{H}$ onto $\mathscr{S}$ is given by

$$P_{\mathscr{S}}(\mathbf{A}) = \mathbf{U}\,\text{diag}\left(\lambda_1 1_{\{\lambda_1 \geq 0\}}, \ldots, \lambda_m 1_{\{\lambda_m \geq 0\}}\right)\,\mathbf{U}'$$

[5] Starting values for the angles can be backed out of the normalized principal components estimate $\mathbf{S}\widetilde{\mathbf{C}}\mathbf{S}$ of the closest rank-r correlation matrix.

where 1_B is the characteristic function of a set B and $\mathbf{U}\text{diag}(\lambda_1, \ldots, \lambda_m)\mathbf{U}'$ is the singular value decomposition of $\mathbf{A}$. The projection of $\mathbf{A} \in \mathscr{H}$ onto $\mathscr{U}$ is the matrix $\mathbf{A}$ with the diagonal elements replaced by ones.

Finding the closest rank-r correlation matrix requires projection onto the constraint set $\mathscr{K}_r$ of rank-r matrices, as well as projections onto $\mathscr{S}$ and $\mathscr{U}$. From Theorem 3.2.1, we know that in the Frobenius norm, projection of $\mathbf{A} \in \mathscr{H}$ onto $\mathscr{K}_r$ is given by

$$P_{\mathscr{K}_r}(\mathbf{A}) = \mathbf{U}\,\text{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)\,\mathbf{U}'$$

where $\mathbf{U}\text{diag}(\lambda_1, \ldots, \lambda_m)\mathbf{U}'$ is the singular value decomposition of $\mathbf{A}$. However, the constraint set $\mathscr{K}_r$ is only closed, not convex (Grubišić 2002). This means that the existing alternating projections theory does not apply, and the algorithm is not guaranteed to converge. Weigel (2004) experimented with iteratively projecting forward rate correlation matrices onto $\mathscr{S}$, $\mathscr{K}_r$ and $\mathscr{U}$. The results indicate that the convergence is fast. Morini & Webber (2004, Algorithm 2) presented a similar algorithm, the *eigenvalue zeroing by iteration* (EZI) method. The approach is to iteratively project the exogenous correlation matrix onto $\mathscr{K}_r$ and $\mathscr{U}$. Once the convergence criterion is reached, for example the changes in the projections onto $\mathscr{K}_r$ are small, the normalized principal components method is applied to the output.

The fourth rank reduction algorithm that we describe is the *Lagrange multipliers* method of Zhang & Wu (2003) and Wu (2002). The authors reformulated the rank reduction problem (3.6) as a minimization problem under constraints and approached it using Lagrange multipliers. In particular, the problem was cast as a minimization-maximization problem

$$\min_{\mathbf{d}} \max_{\mathbf{X} \in \mathscr{K}_r} \mathscr{L}(\mathbf{X}, \mathbf{d}) \tag{3.7}$$

where $\mathbf{d} = (d_1, \ldots, d_m)'$ is a vector of multipliers and the Lagrange function $\mathscr{L}(\mathbf{X}, \mathbf{d})$ is given by

$$\begin{aligned}
\mathscr{L}(\mathbf{X}, \mathbf{d}) &= -||\mathbf{C} - \mathbf{X}||_F^2 - 2\mathbf{d}'\text{diag}(\mathbf{C} - \mathbf{X}) \\
&= -\left\{ \sum_{i,j}^{m} (c_{ij} - x_{ij})^2 + 2\sum_{i=1}^{m} d_i(c_{ii} - x_{ii}) \right\} \\
&= -\left\{ \sum_{i=1}^{m} (c_{ii} - x_{ii} + d_i)^2 - \sum_{i=1}^{m} d_i^2 + \sum_{i,j:\ i \neq j}^{m} (c_{ij} - x_{ij})^2 \right\} \\
&= -||\mathbf{C} + \mathbf{D} - \mathbf{X}||_F^2 + ||\mathbf{d}||^2
\end{aligned}$$

where $\mathbf{D} = \text{diag}(d_1, \ldots, d_m)$. Problem (3.7) was solved in two steps. Firstly, for a fixed value of the Lagrange multiplier $\mathbf{d}$, the inner maximization problem involves minimizing $||\mathbf{C} + \mathbf{D} - \mathbf{X}||_F^2$. From Theorem 3.2.1, we know that the minimum is achieved when $\mathbf{X}$ is the Eckart-Young decomposition of $\mathbf{C} + \mathbf{D}$, $P_{\mathscr{K}_r}(\mathbf{C} + \mathbf{D})$. Zhang & Wu (2003, Theorem 4.1) prove that the local minimum of the inner maximization problem is a global minimum if the r'th and $(r+1)$'th eigenvalues of $\mathbf{C} + \mathbf{D}$ are not equal. This means that the method lacks guaranteed convergence, which is a main criticism of the approach. Secondly, the outer minimization over the Lagrange multipliers $\mathbf{d}$ is solved using gradient-based descending methods, such as the method of steepest descent.

The final rank reduction algorithm that we describe is the *iterative majorization* method of Pietersz & Groenen (2004). Suppose that we need to minimize an objective function f. The idea presented in Pietersz & Groenen (2004) is to construct a function, the majorization function, that is equal to the objective function at a certain point and greater than or equal to the objective function at all the other points in the domain. The majorization function is constructed so that it is easier to minimize than the objective function. Given a starting point, the minimum of the majorization function produces the next point in the iteration.

Problem (3.6) requires minimization of the objective function $f(\mathbf{Q}) = ||\mathbf{C} - \mathbf{Q}\mathbf{Q}'||_F^2$, subject to the unit diagonal constraint. Pietersz & Groenen (2004) derive the majorization function for each row of the matrix $\mathbf{Q}$ (for this and a more general objective function). The algorithm is, starting with the first row, to find the minimum of the majorization function, replace the row by the calculated minimum and loop over all the rows until convergence is achieved. The authors also provided MATLAB code containing the implementation of their algorithm. The algorithm is globally convergent and has a sub-linear local rate of convergence.

3.2.1 Numerical Comparison

To compare the various rank reduction methods, we estimated a historical forward rate correlation matrix for quarterly log forward rates going out to ten years. The data used is the Bank of England's daily commercial bank liability spot curve data from 9 September 2004 to 9 September 2005, obtained from http://www.bankofengland.co.uk/statistics/yieldcurve/index.htm. The quarterly spot rates were available out to five years and thereafter semi-annual rates were interpolated to complete the dataset.

The first principal component explained 80.8% of the variation in the term structure while the second and the third components explained 12.4% and 3% respectively. The historical correlation matrix and its first three principal components are displayed in Figure 3.1.

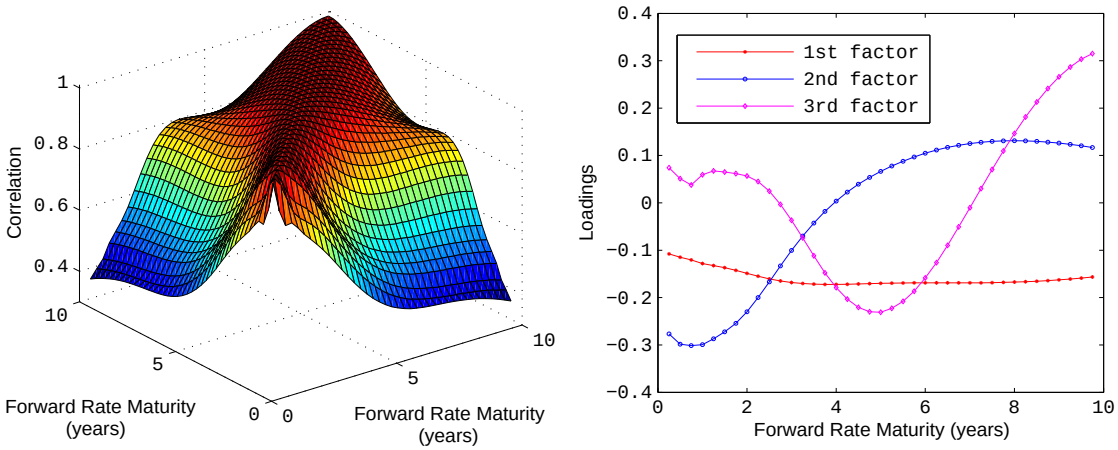


Figure 3.1: GBP historical forward rate correlation matrix (quarterly forward rates) and its first three principal components.

Rank-two, rank-three and rank-four approximations of the historical correlation matrix were examined. The rank reduction methods that were applied were the normalized principal components method (NPCA), the angles parametrization method (Angles), the eigenvalue zeroing by iteration method (EZI), Lagrange multipliers method (Lagrange) and the iterative majorization method (Major). The errors of the various methods are summarized in Table 3.1 below.

	NPCA	Angles	EZI	Major	Lagrange
Rank-two	8.1971	6.3053	6.3145	6.3053	6.3053
Rank-three	2.9529	1.7729	1.8727	1.7729	1.7729
Rank-four	0.7126	0.3614	0.3647	0.3614	0.3614

Table 3.1: Frobenius norm rank reduction errors.

The normalized principal components method performs poorly. The angles parametrization, the iterative majorization and the Lagrange multipliers methods produced the same errors and virtually identical correlation matrices. However, the computational time of the latter two methods was significantly less than the angles parametrization method. The eigenvalue zeroing by iteration method was very fast and produced good results. Due to the lack of guaranteed convergence of the eigenvalue zeroing by iteration and Lagrange multipliers method, the majorization algorithm seems to be the best performing approach.

Figure 3.2 compares the historical correlations with the reduced rank matrix correlations. This graph clearly demonstrates the major drawback of lower-dimensional models on the forward rate dynamics: *too large correlation between neighboring forward rates.* The final figure, Figure 3.3, plots the rank-three correlation matrix obtained when the iterative majorization method is employed to reduce the rank of the historical correlation matrix.

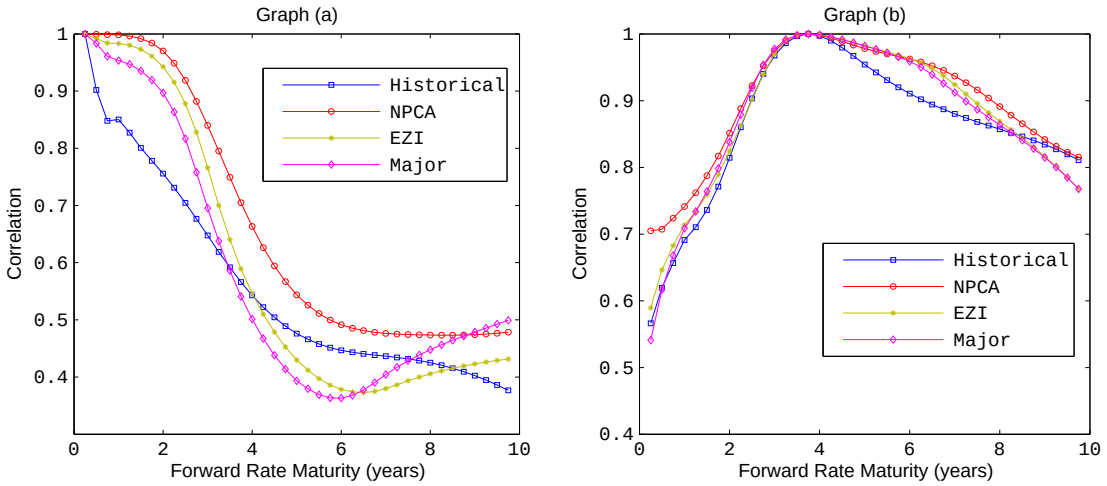


Figure 3.2: A comparison of historical correlations and rank-three correlations. Graph (a) displays correlations between the forward rate maturing in three months time and the rest of the forward rates. Graph (b) displays correlations between the forward rate maturing in four years time and the rest of the forward rates.

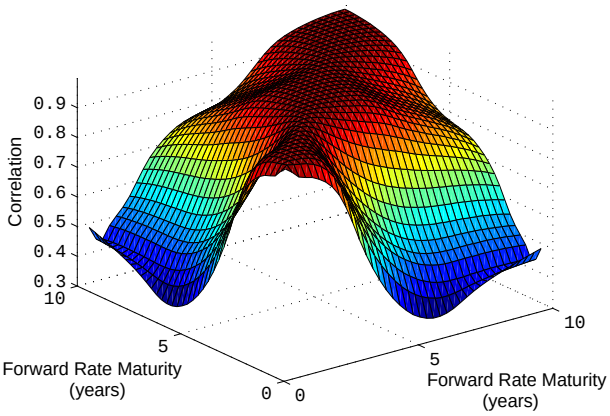


Figure 3.3: Rank-three forward rate correlation matrix, obtained using the iterative majorization algorithm.

3.3 Calibration Instruments

The minimum requirement of an interest rate model is the exact recovery of an arbitrary exogenous yield curve. In the LIBOR market model, the initial forward rates are model inputs.

Interest rate caps and swaptions comprise the largest and the most liquid part of the interest rate derivative market. As they are "similar" products to exotic LIBOR derivatives, caps and swaptions are routinely used in the gamma and vega hedging of these products. Consequently, interest rate models are calibrated to cap and swaption volatilities. This ensures the recovery of these second-order hedging instruments.

3.3.1 Yield Curve

There is no unique way of obtaining a yield curve, even if the input instruments are exogenously specified. One of the main requirements of a desirable bootstrap method is that it correctly prices the instruments used in the construction of the curve. Another vital criterion is that the forward rate curve is positive. Based on these and a few other requirements concerning the stability and localness of the interpolation method, Hagan & West (2005) examined various interpolation algorithms and proposed monotone convex spline interpolation. This method is attractive since, in particular, it overcomes the problem of negative forward rates which appears in the popular

cubic spline interpolation. GBP yield curve inputs and the bootstrapped forward rates, using the monotone convex spline interpolation method, are presented below. The bootstrap was performed using PriceWorX, financial software developed by RiskWorX.

3MO LIBOR	4.590%	5Y swap	4.428%
3X6 FRA	4.455%	6Y swap	4.443%
6X9 FRA	4.320%	7Y swap	4.453%
9X12 FRA	4.285%	8Y swap	4.510%
2Y swap	4.455%	9Y swap	4.463%
3Y swap	4.415%	10Y swap	4.463%
4Y swap	4.417%		

Table 3.2: GBP market rates on 9 September 2005 (source: *Reuters*).

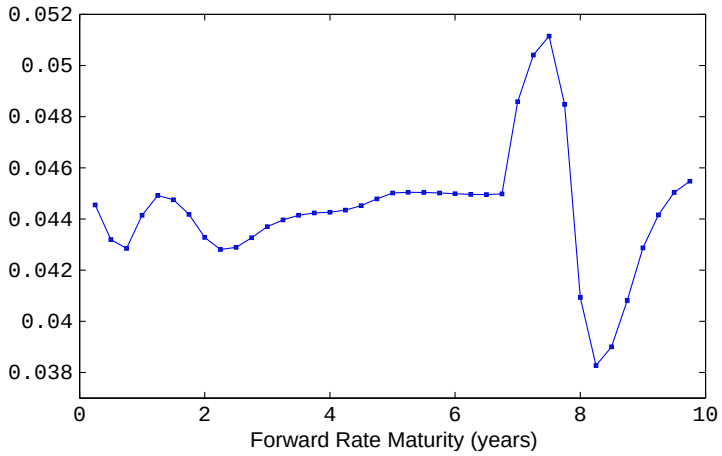


Figure 3.4: GBP bootstrapped quarterly forward rates on 9 September 2005.

GBP swaps have a tenor of six months, while GBP caps have a tenor of three months. The strike of the at-the-money cap is the swap rate for a forward starting swap, starting in three months, with a tenor of three months. A quick check of the bootstrap method is to compare the market quoted strikes with the forward swap rates implied by the bootstrap. Table 3.3 shows that, with the exception of the 2-year cap strike, the bootstrapped results are very good.

	1 y	2 y	3 y	4 y	5 y	6 y	7 y	8 y	9 y	10 y
M	4.35%	4.41%	4.37%	4.38%	4.39%	4.41%	4.42%	4.48%	4.43%	4.43%
B	4.35%	4.35%	4.37%	4.37%	4.39%	4.40%	4.42%	4.43%	4.43%	4.44%

Table 3.3: Comparison of market cap strikes (M) with the forward swap rate implied by the bootstrapped forward rates (B).

3.3.2 Cap Volatilities

Caplets are call options on forward LIBOR rates. These instruments have explicit solutions in the LIBOR market model. Consider a caplet written on the forward LIBOR rate resetting at time T_i, with strike K. From equation (2.38), we know that under the T_{i+1}-forward measure, the forward rate L_i is lognormally distributed

$$L_i(t) = L_i(0) \exp\left\{-\frac{1}{2}\int_0^t ||\boldsymbol{\lambda}(u,T_i)||^2 du + \int_0^t \boldsymbol{\lambda}(u,T_i)\cdot d\mathbf{W}_u^{\mathbb{Q}^{T_{i+1}}}\right\}, \quad 0 \le t \le T_i \tag{3.8}$$

If the value of the caplet at time t is $\vartheta(t)$, then, using risk-neutral pricing,

$$\begin{aligned}\frac{\vartheta(t)}{B(t,T_{i+1})} &= \mathbb{E}^{\mathbb{Q}^{T_{i+1}}}\left[\frac{\vartheta(T_{i+1})}{B(T_{i+1},T_{i+1})}\,\middle|\,\mathscr{F}_t\right] \\ &= \delta_i\,\mathbb{E}^{\mathbb{Q}^{T_{i+1}}}\left[\max\{L_i(T_i)-K,0\}\,|\,\mathscr{F}_t\right]\end{aligned}$$

It is well known that if a random variable V is lognormally distributed with the constant variance of $\log V$ being σ^2, then

$$\mathbb{E}\left[\max\{V-K,0\}\right] = \mathbb{E}[V]\Phi\left(\frac{\log\left(\mathbb{E}[V]/K\right)+\frac{1}{2}\sigma^2}{\sigma}\right) - K\Phi\left(\frac{\log\left(\mathbb{E}[V]/K\right)-\frac{1}{2}\sigma^2}{\sigma}\right) \tag{3.9}$$

where Φ is the cumulative distribution function of a $N(0,1)$ random variable. Now

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}^{T_{i+1}}}\left[L_i(T_i)\right] &= L_i(0)\left(e^{-\frac{1}{2}\int_0^{T_i}||\boldsymbol{\lambda}(u,T_i)||^2du}\right)\mathbb{E}^{\mathbb{Q}^{T_{i+1}}}\left[e^{\int_0^{T_i}\boldsymbol{\lambda}(u,T_i)\cdot d\mathbf{W}_u^{\mathbb{Q}^{T_{i+1}}}}\right] \\ &= L_i(0) \qquad (3.10)\\ \mathbb{E}^{\mathbb{Q}^{T_{i+1}}}\left[\log L_i(T_i)\right] &= \log L_i(0) - \frac{1}{2}\int_0^{T_i}||\boldsymbol{\lambda}(u,T_i)||^2du \qquad (3.11)\\ \mathbb{V}\mathrm{ar}^{\mathbb{Q}^{T_{i+1}}}\left[\log L_i(T_i)\right] &= \mathbb{E}^{\mathbb{Q}^{T_{i+1}}}\left[\left(\log L_i(T_i) - \mathbb{E}^{\mathbb{Q}^{T_{i+1}}}\left[\log L_i(T_i)\right]\right)^2\right] \\ &= \int_0^{T_i}||\boldsymbol{\lambda}(u,T_i)||^2du \qquad (3.12)\end{aligned}$$

where equation (3.10) follows from the fact that if a random variable X is distributed $N(\mu,\sigma^2)$, its moment generating function is given by $\mathbb{E}\left[e^{tX}\right] = e^{\mu t+\frac{1}{2}\sigma^2t^2}$. The mean of the normal random variable $\int_0^{T_i}\boldsymbol{\lambda}(u,T_i)\cdot d\mathbf{W}_u^{\mathbb{Q}^{T_{i+1}}}$ is zero, because an Itô integral is a martingale. The variance of $\int_0^{T_i}\boldsymbol{\lambda}(u,T_i)\cdot d\mathbf{W}_u^{\mathbb{Q}^{T_{i+1}}}$ is $\int_0^{T_i}||\boldsymbol{\lambda}(u,T_i)||^2du$, which follows from Itô isometry. The latter facts explain equations (3.11) and (3.12) as well.

Using equations (3.9), (3.10) and (3.12), the value of a caplet at $t=0$ is given by

$$\vartheta(0) = \delta_iB(0,T_{i+1})\left[L_i(0)\Phi(d_1) - K\Phi(d_2)\right] \tag{3.13}$$

where

$$d_1 = \frac{\log(L_i(0)/K)+\frac{1}{2}\int_0^{T_i}||\boldsymbol{\lambda}(u,T_i)||^2du}{\sqrt{\int_0^{T_i}||\boldsymbol{\lambda}(u,T_i)||^2du}}, \qquad d_2 = \frac{\log(L_i(0)/K)-\frac{1}{2}\int_0^{T_i}||\boldsymbol{\lambda}(u,T_i)||^2du}{\sqrt{\int_0^{T_i}||\boldsymbol{\lambda}(u,T_i)||^2du}}$$

If we define $\sigma_{Black}(T_i)$ as the market implied caplet volatility for the forward LIBOR rate $L_i(t)$, we require the following relation to hold in order to calibrate the LIBOR market model to the market caplet volatilities

$$\int_0^{T_i}||\boldsymbol{\lambda}(u,T_i)||^2du = \sigma_{Black}^2(T_i)T_i \tag{3.14}$$

As mentioned in Section 3.1, the factor loadings are often "decomposed" into the instantaneous volatility component and a vector determining the instantaneous correlations (cf. equations (3.3) and (3.5)). In this case, the caplet calibrating equation (3.14) becomes

$$\int_0^{T_i}\sigma(u,T_i)^2du = \sigma_{Black}^2(T_i)T_i \tag{3.15}$$

because $||\mathbf{q}_i(t)||^2 = ||\widetilde{\mathbf{q}}_i(t)||^2 = 1$.

The discrete set of equations (3.15) for $i=1,\ldots,n-1$ cannot determine the instantaneous volatility function uniquely. Additional assumptions need to be imposed. It is important to note that the choice of the instantaneous volatility function, or equivalently the norms of the factor loadings, affects the correlation of the forward rates at future points in time. One can lower the correlation, without increasing the number of factors, by judiciously redistributing the instantaneous volatility over time. In fact, authors such as Brigo, Mercurio & Morini (2005) stated

that "two factors are usually enough, provided one chooses a flexible volatility structure". A straightforward clarification of this point was given by Sidenius (2000):

Consider a single-factor model over the next six months, and 12×15 and 15×18 forward rates. Suppose that it is a two-period model with two, independent, random shocks driving the interest rates: one in the first three months and one in the last three months. Assume that the 12×15 forward rate has a very large volatility for the first three months and a very small volatility for the following three months. Suppose that the volatility of the 15×18 forward rate behaves exactly the opposite. Then the value of the 12×15 forward rate will be determined mainly by the random shock in the first three months and the value of the 15×18 forward rate mainly by the random shock in the last three months. As these shocks are independent, in six months time, the two rates should have low correlation. While simplistic, this example demonstrates that one can obtain low correlation between the forward rates even when they are perfectly instantaneously correlated.

Different choices for the volatility structure are discussed in Section 3.4. We now examine possible approaches to determining caplet volatilities $\sigma_{Black}(T_i)$ from the market quoted cap volatilities.

Stripping Caplet Volatilities

The market quotes flat cap volatilities from which we need to extract volatilities of the underlying caplets. To fix ideas, let $\vartheta(\sigma, K, T_i)$ be the price of a cap with maturity date T_i, volatility σ and strike K. Let $\vartheta_j(\sigma, K, T_j)$ be the price of the j'th caplet in this cap. By definition, the cap price is a linear sum of the caplet prices

$$\vartheta(\sigma, K, T_i) = \sum_{j=1}^{i-1} \vartheta_j(\sigma, K, T_j) \tag{3.16}$$

We are looking for the so-called forward-forward volatilities σ_j such that

$$\vartheta(\sigma, K, T_i) = \sum_{j=1}^{i-1} \vartheta_j(\sigma_j, K, T_j) \tag{3.17}$$

Hull & White (2000) described the following approach for calculating caplet volatilities: Given flat, at-the-money cap volatilities, one can interpolate between these volatilities to obtain an estimate of the flat cap volatilities for all maturities.[6] Caplet prices for each forward LIBOR rate $L_i(t)$ can be obtained by calculating the value of T_i and T_{i+1}-year caps, with strike equal to the current forward rate $L_i(0)$, and subtracting one from the other. The caplet volatilities are then backed out from the caplet prices by inverting Black's caplet formula (equation (3.13)) using, for example, the Newton-Raphson algorithm.

Alexander (2004) proposed an alternative approach for obtaining individual caplet volatilities given flat cap volatilities for all maturities. A first order Taylor expansion of $\vartheta_j(\sigma_j, K, T_j)$ around σ yields

$$\vartheta_j(\sigma_j, K, T_j) = \vartheta_j(\sigma, K, T_i) + (\sigma_j - \sigma)\upsilon_j(\sigma)$$

where $\upsilon_j(\sigma) = \frac{\partial \vartheta_j}{\partial \sigma_j}(\sigma)$ is the vega of the j'th caplet, evaluated at $\sigma_j = \sigma$. This approximation combined with equations (3.16) and (3.17) yields

$$\sum_{j=1}^{i-1} \vartheta_j(\sigma, K, T_j) = \sum_{j=1}^{i-1} \Big(\vartheta_j(\sigma, K, T_j) + (\sigma_j - \sigma)\, \upsilon_j(\sigma) \Big)$$

Therefore

$$\sigma = \frac{\sum_{j=1}^{i-1} \sigma_j \upsilon_j(\sigma)}{\sum_{j=1}^{i-1} \upsilon_j(\sigma)} \tag{3.18}$$

To first order, the flat cap volatility is approximately equal to the vega-weighted sum of caplet volatilities. When flat cap volatilities are known for all maturities, equation (3.18) defines a recurrence relation for backing out caplet volatilities.

[6] Note that even though each cap has a different strike, interpolating across maturities is not a problem because we are assuming no "smile effects" - the estimated volatility applies to all caps of that maturity irrespective of the strike.

James & Webber (2000) suggested that a parametric form be imposed on the caplet volatilities. The parameters are fitted such that the at-the-money cap prices are recovered as closely as possible. An example of this parametric form is the Nelson & Siegel (1987) family, more commonly known as the "Rebonato form"

$$\sigma_{Black}(T_i) = \beta_0 + (\beta_1 + \beta_2 T_i)e^{-\kappa T_i}$$

or the Svensson (1994) extension of the Nelson & Siegel (1987) family

$$\sigma_{Black}(T_i) = \beta_0 + (\beta_1 + \beta_2 T_i)e^{-\kappa_1 T_i} + \beta_3 T_i e^{-\kappa_2 T_i} \tag{3.19}$$

We now examine stripped caplet volatilities produced by these various approaches, using the GBP at-the-money cap volatility data displayed in Figure 3.5.

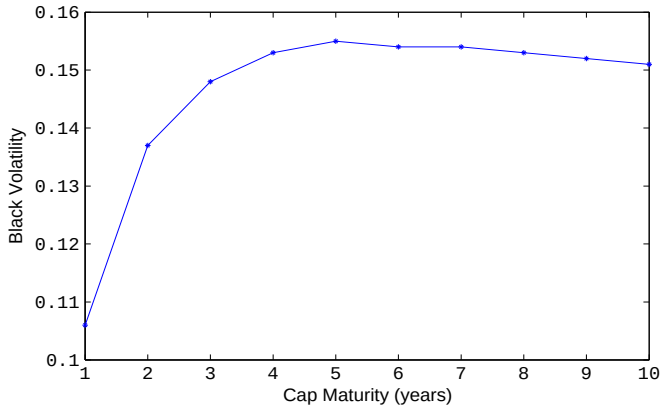


Figure 3.5: GBP at-the-money cap volatilities on 9 September 2005 (source: *Reuters*).

We interpolate the flat cap volatilities (across maturity) using two common methods: linear interpolation and cubic splines. As an alternative to interpolation, we fit a flexible parametric form to the cap volatilities. The chosen parametrization is the Svensson functional form (3.19), which has six free parameters. Because we have ten input cap volatilities, this parametric form has enough free parameters to be almost seen as an "interpolation".

Once the flat cap volatilities have been obtained for all maturities, caplet volatilities can be obtained using the two methods described earlier, which we now term *cap iteration* and *vega iteration*. The former approach obtains caplet prices by subtracting one cap price from the other. The latter approach uses equation (3.18) to back out individual caplet volatilities from flat cap volatilities.

Figure 3.6 (a) displays the results for the cap iteration approach, using linear and cubic splines to interpolate flat cap volatilities. The stripped caplet volatilities are very sensitive to the interpolation algorithm and the resulting oscillations are a result of the interpolation method rather than the data. Identical results are obtained when the vega iteration approach is used. Even though the interpolation approach recovers the input flat cap volatilities exactly, it produces undesirable caplet volatility curves.

Figure 3.6 (b) displays the caplet volatilities obtained using cap iteration and vega iteration methods when the Svensson functional form (3.19) is fitted to the flat cap volatilities. The stripped caplet volatilities are almost identical, but the vega iteration approach seems slightly more stable. The curves produced display the "hump shape" that is typically observed worldwide, with the hump occurring at around two years.

To judge the accuracy of the two approaches presented in Figure 3.6 (b), we compared flat cap volatilities implied by the stripped caplet volatilities with the market quoted volatilities. From Table 3.4 we see that both methods produce good results, but that the vega iteration approach does perform better.

The last approach that we considered was to impose a parametric form onto the caplet volatilities and optimize the parameters to recover at-the-money cap prices as closely as possible. Since the implied cap volatility fits were inferior to those obtained when a parametric form was fitted to cap volatilities, we do not present those results.

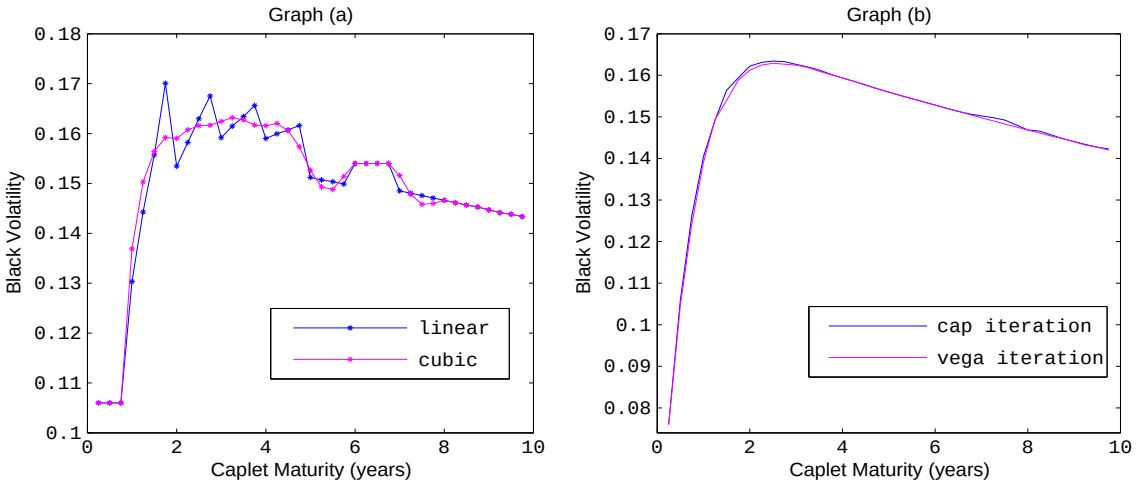


Figure 3.6: Graph (a) displays the stripped caplet volatilities obtained using the cap iteration approach combined with two different interpolations of the flat cap volatilities. Graph (b) displays the stripped caplet volatilities obtained using cap iteration and vega iteration approaches, when the Svensson functional form (3.19) is fitted to the flat cap volatilities.

Maturity	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y
Market	10.6%	13.7%	14.8%	15.3%	15.5%	15.4%	15.4%	15.3%	15.2%	15.1%
CI	10.7%	13.8%	14.9%	15.4%	15.5%	15.5%	15.4%	15.3%	15.2%	15.1%
VI	10.6%	13.7%	14.8%	15.3%	15.4%	15.4%	15.4%	15.3%	15.2%	15.1%

Table 3.4: Comparison of market quoted cap volatilities (Market) with the cap volatilities implied by the cap iteration (CI) and vega iteration (VI) methods, in conjunction with fitting the Svensson functional form (3.19) to the flat cap volatilities.

3.3.3 Swaption Volatilities

Swaptions are options on interest rate swaps. These instruments are frequently used by debt issuers to preserve flexibility throughout their financing cycles, by providing optionality to enter into swaps at some date in the future or to cancel existing swaps (Longstaff et al. 2001). These instruments are sufficiently liquid and standardized to be regarded as "vanilla".

Consider a *payer swaption* maturing at T_p, written on a forward swap, with coupon κ, also starting at T_p with maturity date T_q. This is referred to as a $T_p \times (T_q - T_p)$ payer swaption. If the swap is entered into, the holder will pay a fixed rate κ on reset dates $T_{p+2}, T_{p+4}, \ldots, T_q$ and receive the floating LIBOR rate corresponding to that tenor.[7] The fair swap rate at time t, $S_{p,q}(t)$, is given by

$$S_{p,q}(t) = \frac{B(t,T_p) - B(t,T_q)}{\sum_{j=1}^{(q-p)/2} \tau_j B(t,T_{p+2j})} \tag{3.20}$$

where τ_j is the year fraction for the period $[T_{p+2(j-1)}, T_{p+2j}]$. By no-arbitrage arguments, the value of the payer swaption at expiry T_p, $\mathrm{PS}(T_p)$, is

$$\mathrm{PS}(T_p) = \max\left\{S_{p,q}(T_p) - \kappa, 0\right\} \sum_{j=1}^{(q-p)/2} \tau_j B(T_p, T_{p+2j})$$

Denote by $\mathbb{Q}^{B_{p,q}}$ the measure associated with $B_{p,q}(t) = \sum_{j=1}^{(q-p)/2} \tau_j B(t,T_{p+2j})$ as the numéraire. The process $B_{p,q}(t)$ is a valid numéraire because it is a portfolio of zero-coupon bonds, and hence its value is always positive. The measure $\mathbb{Q}^{B_{p,q}}$ is known as the *forward swap measure*.

Since $B(t,T_p) - B(t,T_q)$ is a portfolio of traded assets, and the value of any traded asset normalized by the numéraire is a martingale in the appropriate measure, the swap rate is a martingale

[7] In practice, caps and swaps have a different tenor. As mentioned in the previous chapter, the LIBOR market model tenor structure $\mathfrak{T} = \{T_1, T_2, \ldots, T_{n-1}\}$ is chosen to match the convention in the cap market.

under $\mathbb{Q}^{B_{p,q}}$. Using risk-neutral valuation, the value of the payer swaption at $t < T_p$ is

$$\frac{\mathrm{PS}(t)}{B_{p,q}(t)} = \mathbb{E}^{\mathbb{Q}^{B_{p,q}}}\left[\max\left\{S_{p,q}(T_p) - \kappa, 0\right\} \middle| \mathscr{F}_t\right] \tag{3.21}$$

The market convention is to price swaptions using the Black model, i.e. by assuming that the swap rate $S_{p,q}(T_p)$ is lognormally distributed under the forward swap measure with the variance of $\log S_{p,q}(T_p)$ being $(\sigma_{p,q}^{Black})^2 T_p$. The quantity $\sigma_{p,q}^{Black}$ is the market quoted Black swaption volatility for a $T_p \times (T_q - T_p)$ swaption. Under this assumption, and using equations (3.9) and (3.21), the price of the swaption at $t = 0$ is

$$\mathrm{PS}(0) = B_{p,q}(0)\left[S_{p,q}(0)\Phi\left(d_1\right) - \kappa\Phi\left(d_2\right)\right]$$

where

$$d_1 = \frac{\ln(S_{p,q}(0)/\kappa) + \frac{1}{2}(\sigma_{p,q}^{Black})^2\, T_p}{\sigma_{p,q}^{Black}\sqrt{T_p}}, \quad d_2 = d_1 - \sigma_{p,q}^{Black}\sqrt{T_p}$$

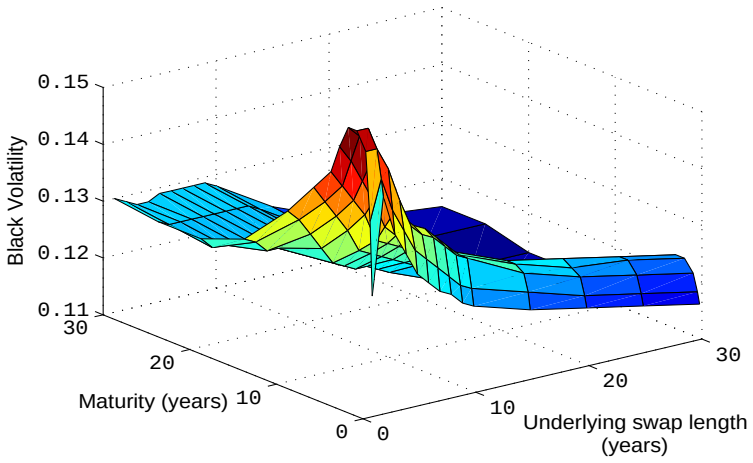


Figure 3.7: GBP swaption volatilities on 9 September 2005 (source: *Reuters*).

We now examine approximate swaption pricing in the LIBOR market model. The fair swap rate, equation (3.20), can be expressed in terms of forward LIBOR rates as

$$S_{p,q}(t) = \frac{1 - \prod_{j=p}^{q-1} \frac{1}{1+\delta_j L_j(t)}}{\sum_{j=1}^{(q-p)/2} \tau_j \prod_{k=p}^{p+2j-1} \frac{1}{1+\delta_k L_k(t)}} \tag{3.22}$$

Using Itô's formula, the fair swap rate satisfies

$$dS_{p,q}(t) = \sum_{i=p}^{q-1} \frac{\partial S_{p,q}}{\partial L_i}(t)\, dL_i(t) + \frac{1}{2}\sum_{i=p}^{q-1}\sum_{j=p}^{q-1} \frac{\partial^2 S_{p,q}}{\partial L_i \partial L_j}(t)\, dL_i(t) dL_j(t)$$

The (conditional) quadratic variation of the swap rate is given by

$$\begin{aligned} d\langle S_{p,q}, S_{p,q}\rangle(t) &= \sum_{i=p}^{q-1}\sum_{j=p}^{q-1} \frac{\partial S_{p,q}}{\partial L_i}(t)\frac{\partial S_{p,q}}{\partial L_j}(t) dL_i(t) dL_j(t) \\ &= \sum_{i=p}^{q-1}\sum_{j=p}^{q-1} \frac{\partial S_{p,q}}{\partial L_i}(t)\frac{\partial S_{p,q}}{\partial L_j}(t) L_i(t) L_j(t) \boldsymbol{\lambda}(t, T_i)\cdot\boldsymbol{\lambda}(t, T_j) dt \end{aligned}$$

The partial derivatives are easily obtained by differentiating equation (3.22). For $i = p, \ldots, q-1$

$$\frac{\partial S_{p,q}}{\partial L_i}(t) = \frac{\delta_i \psi_i(t)}{1 + \delta_i L_i(t)}$$

where

$$\psi_i(t) = \frac{\prod_{j=p}^{q-1} \frac{1}{1+\delta_j L_j(t)}}{\sum_{j=1}^{(q-p)/2} \tau_j \prod_{k=p}^{p+2j-1} \frac{1}{1+\delta_k L_k(t)}} + \frac{\left(1 - \prod_{j=p}^{q-1} \frac{1}{1+\delta_j L_j(t)}\right) \left(\sum_{j=x}^{(q-p)/2} \tau_j \prod_{k=p}^{p+2j-1} \frac{1}{1+\delta_k L_k(t)}\right)}{\left(\sum_{j=1}^{(q-p)/2} \tau_j \prod_{k=p}^{p+2j-1} \frac{1}{1+\delta_k L_k(t)}\right)^2}$$

and $x = \left\lceil \frac{i+1-p}{2} \right\rceil$ is the ceiling of $\frac{i+1-p}{2}$, the smallest integer greater than or equal to $\frac{i+1-p}{2}$.

Define the volatility of the swap rate $\sigma_{p,q}(t)$ by the formula

$$d\langle S_{p,q} dS_{p,q} \rangle(t) = \sigma_{p,q}(t)^2 S_{p,q}(t)^2 dt$$

Then

$$\sigma_{p,q}(t)^2 = \frac{1}{S_{p,q}(t)^2} \left(\sum_{i=p}^{q-1} \sum_{j=p}^{q-1} \frac{\delta_i \psi_i(t)}{1+\delta_i L_i(t)} \frac{\delta_j \psi_j(t)}{1+\delta_j L_j(t)} L_i(t) L_j(t) \boldsymbol{\lambda}(t, T_i) \cdot \boldsymbol{\lambda}(t, T_j) \right) \tag{3.23}$$

The volatility of the swap rate is a function of the forward LIBOR rates and is consequently stochastic. In particular, the forward LIBOR rates and swap rates cannot *both* be lognormal under their respective measures.

The fact that the swap rate has a stochastic volatility precludes closed-form solutions to equation (3.21). Hull & White (2000) and Jäckel & Rebonato (2003) **assume lognormality** by setting the stochastic quantities to their $t = 0$ values i.e. $L_i(t) \approx L_i(0)$, for $i = p, \ldots, q-1$ and $S_{p,q}(t) \approx S_{p,q}(0)$. A model that assumes that the forward LIBOR rates and swap rates are both lognormal under their respective measures is not arbitrage-free. Rebonato (1999*a*) examined the pricing discrepancies when swap rates are assumed lognormal and reached the following conclusion: "...even if and when detectable, the pricing discrepancy is certainly too small to be arbitraged away."

The swaption calibration requirement is equivalent to equation (3.14)

$$\begin{aligned} \left(\sigma_{p,q}^{Black}\right)^2 T_p &= \int_0^{T_p} ||\sigma_{p,q}(t)||^2 dt \\ &= \frac{1}{S_{p,q}(0)^2} \sum_{i=p}^{q-1} \sum_{j=p}^{q-1} \frac{\delta_i \psi_i(0)}{1+\delta_i L_i(0)} \frac{\delta_j \psi_j(0)}{1+\delta_j L_j(0)} L_i(0) L_j(0) \int_0^{T_p} \boldsymbol{\lambda}(t, T_i) \cdot \boldsymbol{\lambda}(t, T_j) dt \end{aligned} \tag{3.24}$$

We conclude this section with an important note on the difference between caps and swaptions. A cap is a portfolio of options on forward rates while a swaption is an option on a portfolio of forward rates. The value of the latter instrument depends on both the instantaneous volatility *and* instantaneous correlation of the forward rates. This should be obvious from equation (3.23).

3.4 Model Choice

The general form of the LIBOR market model is merely a framework. It becomes a model once the forward rate factor loadings $\lambda(t, T_i)$, $T_i \in \mathfrak{T}$ are specified. The information provided by cap and swaption volatilities does not imply a unique form for these loadings. Furthermore, a good calibration to current market prices does not guarantee a good interest rate model. The performance of the model is really determined by how close the model specified evolution of the instantaneous volatility and correlation functions is to the actual evolution. A great advantage of the LIBOR market model is that, due to deterministic forward rate volatility functions, the future evolution of key model structures is deterministic. For example, one can calculate future caplet volatilities once the model is calibrated. These so-called "model diagnostics" are important when it comes to assessing the quality of the model.

Various versions of the piecewise constant factor loadings specification have been proposed by Pedersen (1998), Sidenius (2000) and Hull & White (2000), amongst others. The approach that we have examined is based on the decomposition of factor loadings into two separate components, the instantaneous volatility of the forward rate and a vector determining the instantaneous correlation between the forward rates.

The instantaneous volatility of the forward rate maturing at T_i, $\sigma(t, T_i)$, is by assumption a square-integrable deterministic function of t and T_i. A stationary specification of the instantaneous volatility function, $\sigma(t, T_i) = f(T_i - t)$, is considered to be very desirable. This specification implies that the forward rates with the same time to maturity have the same volatility and that the future caplet volatilities are the same as the current ones. Rebonato (2002, Chapter 6) provides empirical evidence demonstrating that structural features of the implied at-the-money caplet volatility curve remain more or less unchanged through time. Furthermore, De Jong et al. (2004) noted that all the standard interest rate models, for example the affine models of Duffie & Kan (1996), imply that the volatility and correlation parameters are time-homogenous.

It is interesting to note that a stationary instantaneous volatility function can only exist if $\sigma^2_{Black}(T)T$ is an increasing function of T. To prove this, recall that the relationship between the Black caplet volatility $\sigma_{Black}(T_i)$ and the instantaneous volatility function $\sigma(t, T_i)$ is given by $\sigma^2_{Black}(T_i)T_i = \int_0^{T_i} \sigma(t, T_i)^2 dt$ (equation (3.15)). Then, if $\sigma(t, T_i) = f(T_i - t)$, we obtain

$$\begin{aligned} \sigma^2_{Black}(T_{i+1})T_{i+1} - \sigma^2_{Black}(T_i)T_i &= \int_0^{T_{i+1}} f(T_{i+1} - t)^2 dt - \int_0^{T_i} f(T_i - t)^2 dt \\ &= \int_0^{T_{i+1}-T_i} f(T_{i+1} - t)^2 dt \end{aligned}$$

The right-hand side of the above equation is strictly positive, being the integral of a strictly positive quantity. If the left-hand side is not strictly positive, i.e. if $\sigma^2_{Black}(T)T$ is not an increasing function of T, one will end with imaginary instantaneous volatilities. This means that a stationary instantaneous volatility function can only exist if $\sigma^2_{Black}(T)T$ is an increasing function of T.

As stationarity of the volatility structure is important in interest rate modelling, a particularly popular form for the instantaneous volatility function is

$$\sigma(t, T_i) = h(T_i) f(T_i - t) \tag{3.25}$$

This specification splits the instantaneous volatility into two components: a component that depends on the residual maturity of the forward rate $f(T_i - t)$ and a forward rate specific component $h(T_i)$. The inclusion of the forward rate specific component allows all possible market implied caplet volatilities to be recovered, not just ones for which $\sigma^2_{Black}(T)T$ is an increasing function of T. In particular, a perfect fit to the caplet market is always achieved if we set

$$h(T_i)^2 = \frac{\sigma^2_{Black}(T_i)T_i}{\int_0^{T_i} f(T_i - t)^2 dt} \tag{3.26}$$

The function $f(T_i - t)$ is fitted to market caplet volatilities and possibly to market swaption volatilities as well. Then the forward rate specific factor is used to ensure a perfect fit to the caplet market through equation (3.26). However, for formulation (3.25) to be more or less stationary, the forward rate specific factors need to be close to one. If the fitted $h(T_i)$ are significantly varying across maturities T_i, the resulting model would imply that the forward rates with the same time to maturity will have very different responsiveness to the Brownian motion shocks. Rebonato (2002) advises that if this situation arises, specification (3.25) should not be used.

Brigo et al. (2005) considered the following flexible parameterizations of the instantaneous volatility function

$$\sigma(t, T_i) = \zeta_i \psi_{i-\eta(t)} \tag{3.27}$$

$$\sigma(t, T_i) = \zeta_i \left[(a(T_i - t) + d)e^{-b(T_i - t)} + c \right] \tag{3.28}$$

where ζ_i is the forward rate specific constant, $\psi_{i-\eta(t)}$ is the constant volatility applicable when a forward rate has $i - \eta(t)$ full periods between the next reset date and its maturity date and a, b, c, d are simply constants.

The second component in the decomposition of the factor loadings is a vector which determines the instantaneous correlation between the forward rates. From recent research, it would seem that the number of factors used in the model has an extremely important role in accurately pricing exotic options, especially spread options and reset caps. Sidenius (2000) provides numerical examples demonstrating large relative differences in prices when employing a one-factor model, as opposed to a three-factor model. This is due to the restrictions on the dynamics that the lower-dimensional

models typically impose: too large correlation between neighboring forward rates, and too small correlation between forward rates that are further apart. This effect has been mentioned in Section 3.2.1 and can also be found in Rebonato (1999*b*), amongst others.

Rebonato (1999*b*) calibrated the LIBOR market model by extracting correlation information from historical data. Given an exogenously specified forward rate correlation matrix $\boldsymbol{\rho}$, a rank-d correlation matrix $\widehat{\boldsymbol{\rho}}$ can be obtained by any one of the methods described in Section 3.2. Principal component analysis is then used to decompose $\widehat{\boldsymbol{\rho}}$ into $\widehat{\boldsymbol{\rho}} = \mathbf{Q}\mathbf{Q}'$, where $\mathbf{Q} \in \mathbb{R}^{(n-1)\times d}$. The factor loadings are then given by

$$\boldsymbol{\lambda}(t, T_i) = \sigma(t, T_i)(q_{i1}, \ldots, q_{id})'$$

where the function $\sigma(t, T_i)$ has already been calibrated.

Alexander (2004) demonstrated that the historical forward rate correlations, particularly for short dated forward rates, are very volatile. An alternative to using historical data is to imply the instantaneous correlations from liquid market instruments. In the ideal case, one would like to calibrate the model to correlation sensitive derivatives, such as spread options. However, they lack the liquidity to do so and swaption volatilities have to be used. Of course, there are criticisms against fitting correlations to swaption volatilities. The swaption data may be stale because it is unlikely that the entire swaption matrix is updated uniformly. The second criticism is that there has been evidence (Choy et al. 2004, Rebonato 2002) that swaption values are not very sensitive to instantaneous correlations.

If the instantaneous correlations are extracted from the swaptions matrix, a fully nonparametric approach requires estimation of $n(n-1)/2$ parameters, which is certainly going to lead to overfitting and instability. Before we discuss the parametric approach, let us state the additional properties of the forward rate instantaneous correlation matrix.

The forward LIBOR rate correlation matrix should satisfy certain economic constraints. Forward rate correlations should decrease as the maturities between the rates increase. Equally spaced (in terms of the maturities) forward rates should become more correlated as their maturities increase. Mathematically, for a forward LIBOR rate correlation matrix $\boldsymbol{\rho} \in \mathbb{R}^{(n-1)\times(n-1)}$, we require

1. $\rho_{ij} \geq \rho_{i,j+1}$ for $j \geq i$
2. $\rho_{i,i+k}$ is an increasing function of i, for each k

Schoenmakers & Coffey (2003) constructed parsimonious parametric structures for the instantaneous correlation matrix. Given a sequence of positive, strictly increasing real numbers, $\{b_1, \ldots, b_{n-1}\}$, satisfying $b_1 = 1$ and b_i/b_{i+1} strictly increasing in i, the instantaneous correlations are given by

$$\rho_{ij} = \frac{\min(b_i, b_j)}{\max(b_i, b_j)} \tag{3.29}$$

This formulation satisfies the economic constraints mentioned above. Schoenmakers & Coffey (2003) then show that the sequence b_i can be represented, in the general form, as

$$b_i = \exp\left(\sum_{l=2}^{n-1} \min(l-1, i-1)\Delta_l\right)$$

for any non-negative sequence of real numbers $\{\Delta_2, \ldots, \Delta_{n-1}\}$. Substituting into equation (3.29) yields

$$\rho_{ij} = \exp\left(-\sum_{l=i+1}^{n-1} \min(l-i, j-i)\Delta_l\right), \quad i < j$$

There are $n-2$ parameters $\{\Delta_2, \ldots, \Delta_{n-1}\}$ that need to be estimated. Schoenmakers & Coffey (2003) proposed a few examples of parsimonious correlation structures. Based on parameter stability and practical experiments, the following two-parameter structure for the instantaneous correlation matrix is suggested

$$\rho_{ij} = \exp\left\{-\frac{|i-j|}{n-2}\left(-\log(\theta) + \eta\frac{i^2+j^2+ij-3\big((n-1)i+(n-1)j-i-j\big)+2(n-1)^2-n-3}{(n-3)(n-4)}\right)\right\} \tag{3.30}$$

where θ and η and are the two free parameters satisfying $\eta > 0$ and $0 < \eta < -\log(\theta)$.

3.4.1 Market Calibration

We now present an example LIBOR market model calibration. The data consists of GBP market closing prices on 9 September 2005, obtained from Reuters. Tables and graphs of the data used can be found throughout this chapter. The bootstrapped forward rate curve is presented in Figure 3.4. Market implied caplet volatilities, stripped by fitting the Svensson functional form to cap volatilities and applying the vega iteration algorithm, are presented in Figure 3.6. Market swaption volatilities are displayed in Figure 3.7. We will only calibrate to a subset of these swaptions, presented in Table 3.5 below, with the sum of the option maturity and swap length not exceeding ten years. In total, we are fitting 39 caplet and 56 swaption volatilities.

	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y
3M Opt	13.1%	14.8%	14.1%	13.7%	13.3%	13.1%	12.9%	12.8%	12.6%
6M Opt	14.4%	15.1%	14.5%	14.1%	13.7%	13.5%	13.3%	13.1%	12.9%
1Y Opt	15.6%	15.6%	15.2%	14.6%	14.1%	13.9%	13.8%	13.6%	13.4%
2Y Opt	15.9%	15.9%	15.3%	14.8%	14.2%	14.0%	13.9%	13.7%	
3Y Opt	15.9%	15.7%	15.2%	14.7%	14.2%	14.0%	13.8%		
4Y Opt	15.5%	15.2%	14.8%	14.4%	13.9%	13.7%			
5Y Opt	15.0%	14.7%	14.4%	14.1%	13.6%				
7Y Opt	14.6%	14.3%	14.0%						

Table 3.5: GBP market swaption volatilities used for calibrating the LIBOR market model. Columns indicate the length of the underlying swap, while rows indicate swaption maturity.

We fit the instantaneous volatility specification (3.28) to the caplet and swaption volatilities, while simultaneously fitting the parameterized instantaneous correlation matrix, equation (3.30), to swaption volatilities.[8] We have chosen to present the fitting results for a three-factor LIBOR market model, because we found that while a three-factor formulation does not result in significantly smaller fitting errors, it does produce a slightly better evolutions of caplet volatilities and a more satisfactory instantaneous correlation matrix. Note that the two-parameter correlation matrix formulation suggested by Schoenmakers & Coffey (2003), equation (3.30), results in a full-rank correlation matrix. During the optimization it will always be reduced to a rank-three correlation matrix. The chosen rank reduction method is the majorization algorithm.

The parametrization (3.28) is realized by first fitting the stationary part of the instantaneous volatility function, $f(T_i - t) = (a(T_i - t) + d)e^{-b(T_i - t)} + c$, to the swaption volatilities. The parameters of the instantaneous correlation matrix (3.30) are fitted simultaneously. The parameters are fitted such that the squared errors between the market and the model swaption volatilities are minimized. To calculate the model swaption volatilities, we use the lognormal approximation (3.24). A perfect fit to the caplet volatilities is then enforced using the forward rate specific factors ζ_i. The swaption fitting errors are presented in Table 3.6. In Figure 3.8 (a), the fitted forward rate specific factors ζ_i are displayed. Current model caplet volatilities, together with future model caplet volatilities are displayed in Figure 3.8 (b). The rank-three instantaneous correlation matrix implied by the swaption volatilities is displayed in Figure 3.9.

With the exception of the swaptions whose underlying swap length is one year, the fitting seems satisfactory. When expressed as percentage errors, they seem to be in accordance with the errors presented in Brigo et al. (2005, Table 3), which were deemed to be "actually small". However, in our calibration the instantaneous correlation matrix also seems satisfactory.

The graphs in Figure 3.8 indicate that the forward rate specific factors needed to ensure a perfect fit to the caplet market are not all close to one. As a result, the evolution of the caplet volatilities is no longer strictly stationary. However, from Figure 3.8(b) we see that this departure from stationarity is minimal. We found that if we require a more stationary structure, we have to sacrifice the quality of the swaption fitting.

[8] Experiments with unconstrained optimization of parametrization (3.27) did not produce good results. Even when constraints were imposed, the evolution of the caplet volatilities was rather noisy.

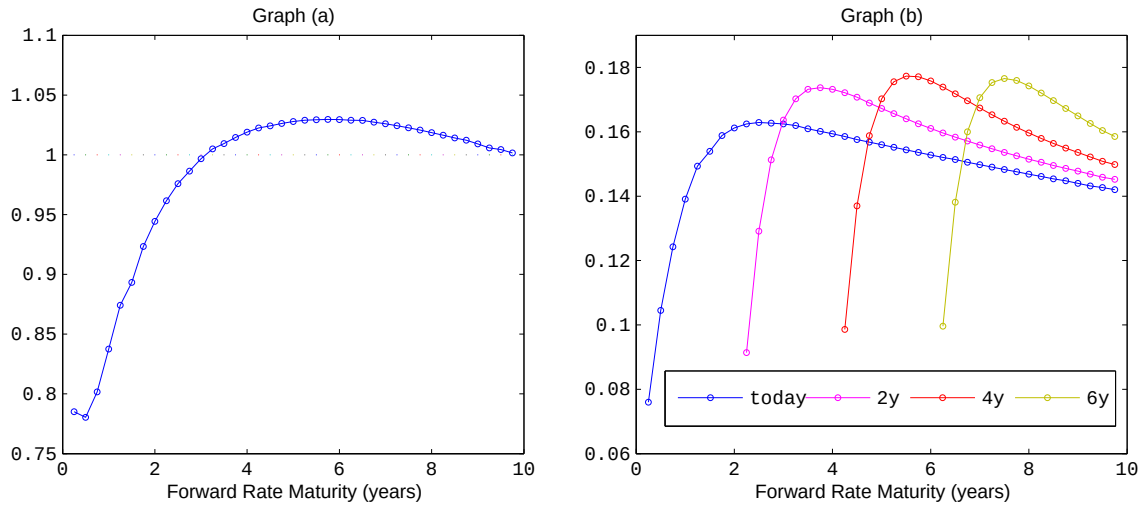


Figure 3.8: Graph (a) plots the forward rate specific factors obtained when parametrization (3.28) is fitted to swaption volatilities together with the instantaneous correlation matrix parameterized by (3.30). Graph (b) plots the evolution of caplet volatilities over time.

	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y
3M Opt	0.2%	0.5%	0.0%	0.0%	-0.2%	-0.2%	-0.2%	-0.1%	-0.3%
6M Opt	0.1%	0.1%	0.0%	0.0%	-0.1%	0.0%	0.0%	-0.1%	-0.1%
1Y Opt	-0.3%	-0.1%	0.2%	0.1%	0.0%	0.1%	0.3%	0.2%	0.1%
2Y Opt	-0.7%	0.0%	0.1%	0.2%	-0.1%	0.1%	0.2%	0.2%	
3Y Opt	-0.5%	0.1%	0.2%	0.2%	0.1%	0.1%	0.1%		
4Y Opt	-0.5%	-0.1%	0.1%	0.2%	-0.1%	-0.1%			
5Y Opt	-0.6%	-0.3%	0.0%	-0.1%	-0.3%				

Table 3.6: Absolute errors in swaption calibration, defined by (Market vol - Model vol).

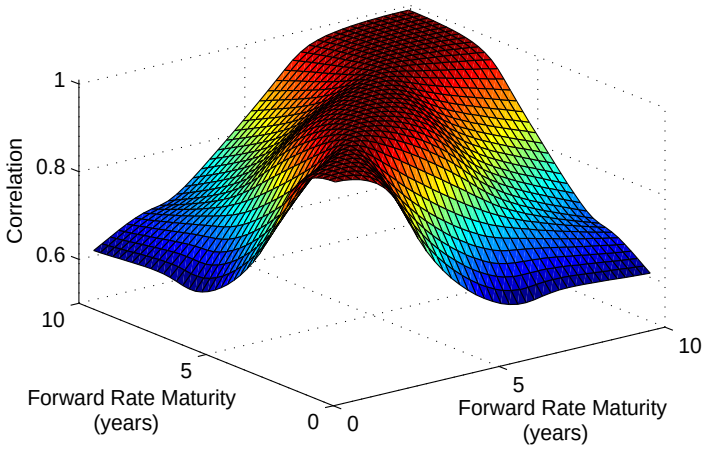


Figure 3.9: Schoenmakers & Coffey (2003) two-parameter instantaneous correlation matrix, fitted to swaption volatilities. The rank-three matrix displayed here was obtained using the majorization algorithm.

Chapter 4

Monte Carlo Basics

Computers are incredibly fast, accurate, and stupid. Human beings are incredibly slow, inaccurate, and brilliant. Together they are powerful beyond imagination.

– Albert Einstein

Systematic development of Monte Carlo methods took place around 1944. These methods stem from the work on the atomic bomb during the second world war, where they were applied to simulation of probabilistic problems regarding the neutron diffusion in fissile material (Hammersley & Handscomb 1964). Monte Carlo simulation was introduced in derivative pricing by Boyle (1979) and is an indispensable valuation tool in the LIBOR market model. While lattice methods are the standard implementation tool for short rate models, the full state-dependence of the drift term of LIBOR rates results in a non-recombining tree (Jäckel 2000), which in general is not a computationally feasible approach. If the instantaneous volatility structure is a separable function of time and maturity of the forward rate, the model can be approximated as a low-dimensional Markovian model and PDE methods applied (Pietersz et al. 2004). However, Piterbarg (2004) argues that the main advantage of the LIBOR market models, namely the ability to calibrate to a large set of instruments whilst maintaining a realistic evolution of the volatility structure, is "severely hampered" when one restricts the volatility function to be separable. Thus, in general, derivative prices in the LIBOR market model are computed using Monte Carlo simulation. This chapter discusses Monte Carlo and quasi-Monte Carlo methods and examines the discretization of the LIBOR market model's stochastic differential equations.

4.1 Monte Carlo Methods

Monte Carlo (pseudo-Monte Carlo) methods and quasi-Monte Carlo methods are designed for the problem of evaluating integrals of the form

$$I(f) = \int_{[0,1)^d} f(\mathbf{x}) d\mathbf{x} \tag{4.1}$$

where f is a square-integrable function[1] and $\mathbf{x} = (x_1, \ldots, x_d)'$. These methods approximate the value of the integral by

$$\widehat{I}(f) = \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}_i) \tag{4.2}$$

where the function f is averaged over the point set $P_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in [0,1)^d$. The difference between Monte Carlo and quasi-Monte Carlo methods lies in the construction of the point set P_n.

Monte Carlo methods are based on the observation that approximating the integral (4.1) is equivalent to approximating the expectation of a function of a multivariate uniform random variable. Suppose $\mathbf{X}$ is a d-variate uniform random variable. Then $I(f) = \mathbb{E}[f(\mathbf{X})]$ and an unbiased

[1] In the Brownian motion setting, the function f transforms uniform random variables to normal random variables, constructs the process for the underlying state variables and calculates the normalized payoff of the derivative.

estimate of the expectation is given by (4.2), where P_n is a set of realizations of independent identically distributed uniform random variables over $[0,1)^d$. The uniform random numbers are generated using deterministic algorithms, the most well known being the linear and nonlinear congruential generators. A classical reference for analysis of these algorithms is Niederreiter (1992).

Two celebrated theorems in probability theory are used to analyze the Monte Carlo estimate of the integral: the law of large numbers and the central limit theorem. By the strong law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n} f(\mathbf{X}_i) \to \mathbb{E}[f(\mathbf{X})] \quad \text{almost surely, as } n \to \infty$$

when $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are independent, identically distributed uniform random variables. By the Central Limit Theorem

$$\left(\frac{1}{n}\sum_{i=1}^{n} f(\mathbf{X}_i) - \mathbb{E}[f(\mathbf{X})]\right) \to \frac{\sigma}{\sqrt{n}}\,\mathbf{z} \quad \text{in distribution, as } n \to \infty$$

where $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$ and $\sigma = \mathbb{V}\mathrm{ar}[f(\mathbf{X})]$ (Grimmett & Stirzaker 2001). The latter theorem shows that the standard error of the estimator tends to zero as $n^{-1/2}$. When $d = 1$, the convergence rate of $\mathscr{O}(n^{-1/2})$ is not particularly appealing, considering that the convergence rate for a simple trapezoidal rule is $\mathscr{O}(n^{-2})$. However, the Monte Carlo convergence rate is independent of the dimension of the problem and becomes competitive when the dimension is greater than or equal to three or four.

Various approaches can be applied to reduce the variability of the Monte Carlo estimate, many of which are reviewed by Boyle, Broadie & Glasserman (1997). The two variance reduction techniques that are frequently used in the LIBOR market model are antithetic variates and control variates. Briefly, antithetic variates are based on the observation that if $\mathbf{x}_i$ is uniformly distributed, so is $\mathbf{x}_i^* = \mathbf{1} - \mathbf{x}_i$. The estimator of the integral using antithetic variates is

$$\widehat{I}(f)^{\mathrm{av}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{f(\mathbf{x}_i) + f(\mathbf{x}_i^*)}{2}\right) \tag{4.3}$$

Constructing antithetic pairs $(\mathbf{x}_i, \mathbf{x}_i^*)$ leads to a more uniformly distributed sample over $[0,1)^d$ than simply doubling the sample size. Boyle et al. (1997) showed that antithetic variates lead to a guaranteed reduction in the variance of the estimator if the payoff function of the derivative is monotone in the underlying state variables.

Control variates rely on the estimate of an integral of another function, say $\widehat{I}(g)$, whose value $I(g)$ is known, to control the estimate $\widehat{I}(f)$. The estimator of the integral using control variates is

$$\widehat{I}(f)^{\mathrm{cv}} = \widehat{I}(f) + \Big(I(g) - \widehat{I}(g)\Big) \tag{4.4}$$

For the control variates method to reduce the variance of the estimate we require

$$\mathbb{V}\mathrm{ar}\left(\widehat{I}(g)\right) < 2\,\mathbb{C}\mathrm{ov}\left(\widehat{I}(g), \widehat{I}(f)\right)$$

In a financial setting, a good control variate is one whose payoff is highly correlated with the payoff of the option under valuation. When pricing American options, the control variate that is frequently chosen is the corresponding European option. In general, a hedge portfolio should be a good control variate. Boyle et al. (1997) further discuss the estimator $\widehat{I}(f) + \beta\left(I(g) - \widehat{I}(g)\right)$ and the choice of β for which this estimator is guaranteed not to have a larger variance than $\widehat{I}(f)$.

The cause of the slow $\mathscr{O}(n^{-1/2})$ convergence is the fact that random and pseudo-random numbers tend to cluster and not fill the space uniformly. One possible way of speeding up the convergence rate is to use low-discrepancy sequences, sequences constructed to avoid gaps and clustering.

4.2 Low-Discrepancy Sequences

The idea behind low-discrepancy sequences is to choose integration points that are well dispersed throughout the integration region. Discrepancy is a notion from number theory that is used to analyze deterministic sequences. It is a measure of how close the empirical distribution of the sequence is to the uniform distribution on $[0,1)^d$. The most researched measure is the star discrepancy.

Definition 4.2.1. *The **star discrepancy** D_n^* of a sequence $\mathbf{x}_1, \ldots, \mathbf{x}_n \in [0,1)^d$ is given by*

$$D_n^* = \sup_{\mathbf{u} \in [0,1)^d} \left| \frac{\sum_{i=1}^n 1_{\prod_{j=1}^d [0,u_j)}(\mathbf{x}_i)}{n} - \prod_{i=1}^d u_i \right|$$

where 1_A is the characteristic function of a set A. The first term is the proportion of points that fall in the hypercube $[0, u_1) \times \cdots \times [0, u_d)$, while the second term is the volume of this hypercube. The interpretation of the star discrepancy is that if a sequence is uniformly distributed in $[0,1)^d$, then the star discrepancy of the sequence will be zero in the limit as $n \to \infty$.

Niederreiter (1992) stated that it is widely believed, although only proved for $d = 1$, that the star discrepancy of the first n points of *any* sequence of numbers in $[0,1)^d$ satisfies

$$D_n^* \geq B(d) \frac{(\log n)^d}{n}$$

where $B(d)$ is a constant that depends only on the dimension d. Low-discrepancy sequences are sequences with a small star discrepancy, namely $\mathscr{O}\left(\frac{(\log n)^d}{n}\right)$, $n \geq 2$. Low-discrepancy sequences are also known as quasi-random, which is misleading as they do not attempt to be random at all.

The Koksma-Hlawka inequality provides deterministic (worst) error bounds for quasi-Monte Carlo approximations. This inequality shows that the absolute value of the integration error is less than or equal to a product of two functions, one being the star discrepancy and the other a function depending on the smoothness of the integrand

$$|\widehat{I}(f) - I(f)| \leq D_n^* V(f)$$

where $V(f)$ is the total variation of the function in the sense of Hardy and Krause (see Niederreiter (1992) for a technical specification of $V(f)$). Thus, effectiveness of quasi-Monte Carlo methods depends on *both* the chosen low-discrepancy sequence and on the integrand. Caflisch, Morokoff & Owen (1997) found that, in practice, only a minimal amount of smoothness of the integrand is necessary for effectiveness of quasi-Monte Carlo. However, they warn that the effectiveness of these methods diminishes for discontinuous integrands.

The presence of the $(\log n)^d$ factor in the convergence rate is a cause for concern when dealing with high-dimensional integration problems. The problems that arise in a financial settings are often high dimensional. For example, in single-factor interest rate models, the number of steps taken to evolve the underlying state variables is the dimension of the problem. In multi-factor models, the number of steps taken multiplied by the number of factors is the dimension of the problem. Brately, Fox & Niederreiter (1992) examined the effectiveness of two low-discrepancy sequences, Niederreiter and Sobol', for integration problems in 12 dimensions or more. Quoting Brately et al. (1992):

> "In high-dimensional problems (say > 12), quasi-Monte Carlo seems to offer no practical advantage over pseudo-Monte Carlo because the discrepancy bound for the former is far larger than $N^{1/2}$ for $N = 2^{30}$, say."

A problem that arises with high-dimensional low-discrepancy sequences is that not all of their two-dimensional projections are uniform. This is a necessary (but not sufficient) condition for uniformity of the sequence (Caflisch et al. 1997). In particular, lower dimensions have better uniformity properties as illustrated in Figure 4.1.

Paskov & Traub (1995) examined the pricing of collateralized mortgage obligations, involving integration in up to 360 dimensions. They found that Halton and Sobol' sequences outperform standard Monte Carlo even in such high dimensions, with the latter low-discrepancy sequence outperforming the former. This result caused quite a stir and studies have tried to provide answers to this outperformance by quasi-Monte Carlo methods in high-dimension. Boyle et al. (1997) attributed this contrasting result to the fact that the integrands in finance are better behaved than the ones examined in Brately et al. (1992), which were highly periodic. A second explanation is the effective dimension hypothesis which will be discussed in Section 4.3.

The performance of Halton sequence is not regarded as competitive with Faure and Sobol' sequences, both in low and high dimensions. There is no conclusive answer on which of the latter two sequences has better performance, but it seems that the financial community prefers Sobol' sequence. In the tests that we performed, which compared the convergence of quasi-Monte Carlo

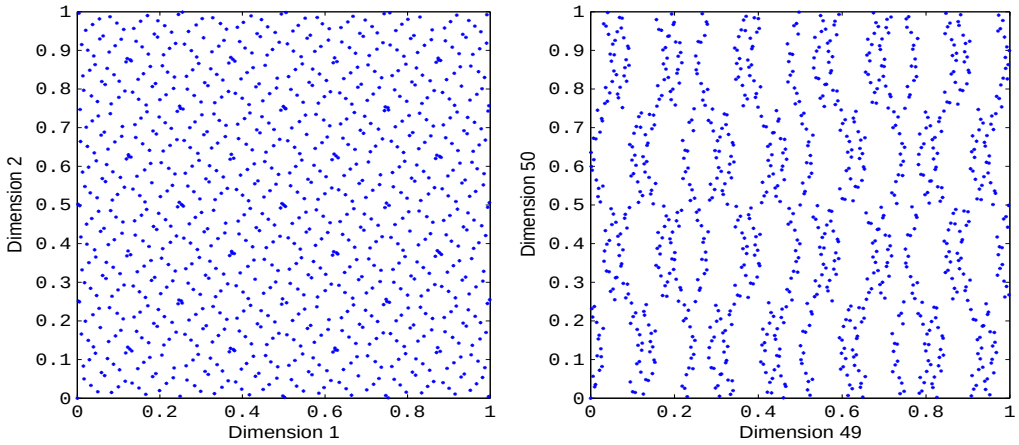


Figure 4.1: Projections of 1000 Sobol' points in 50 dimensions.

cap prices to their closed form solutions, both the Faure and the generalised Faure sequences performed worse than the Sobol' sequence.[2] For this reason, we only discuss the latter sequence.

4.2.1 Sobol' Sequences

Sobol' sequences belong to a class of digital nets. We will not attempt to present the theoretical underpinnings of this method and only present the algorithm following Brately & Fox (1988).

The generation of a Sobol' sequence begins with a choice of a primitive polynomial over $\mathbb{Z}_2$.

Definition 4.2.2. *A **primitive polynomial of degree** d **in** $\mathbb{Z}_2$ is*

$$p(x) = x^d + a_1 x^{d-1} + \ldots + a_{d-1}x + 1$$

where the coefficients are in the field $\mathbb{Z}_2 = \{0, 1\}$ and $p(x)|x^{2^d-1} + 1$.

The coefficients of the primitive polynomial $\{a_1, \ldots, a_{d-1}\}$ are used to generate a sequence $\{m_1, m_2, \ldots\}$. Each element of the sequence is an odd integer satisfying $0 < m_i < 2^i$. The first d elements of the sequence, $\{m_1, \ldots, m_d\}$, need to be initialized. The performance of a Sobol' sequence seems to be *very dependent* on the choice of these initialization numbers. For $i \geq d + 1$, the sequence is generated using the following recursion relation

$$m_i = 2a_1 m_{i-1} \oplus 2^2 a_2 m_{i-2} \oplus \cdots \oplus 2^{d-1} a_{d-1} m_{i-d+1} \oplus 2^d m_{i-d} \oplus m_{i-d}$$

where $\oplus$ is the bitwise exclusive-or (XOR) operation, applied to the binary representations of the operands. The bitwise XOR is defined as $0 \oplus 0 = 1 \oplus 1 = 0$ and $0 \oplus 1 = 1 \oplus 0 = 1$. For example, $3 \oplus 7 = 011_2 \oplus 111_2 = 100_2 = 4$. Once the sequence $\{m_1, m_2, \ldots\}$ is generated, the *direction numbers* $\{v_1, v_2, \ldots\}$ are defined by

$$v_i = \frac{m_i}{2^i}$$

Then the Sobol' sequence $\{x_1, x_2, \ldots\}$ is generated by

$$x_n = b_1 v_1 \oplus b_2 v_2 \oplus \ldots$$

where $(\ldots b_2 b_1)_2$ is the binary representation of n. We will generate the sequence using the method of Antonov and Saleev suggested in Brately & Fox (1988). Utilizing Gray codes,[3] they showed that a Sobol' sequence can be generated using the following recursion relation

$$\begin{aligned} x_0 &= 0 \\ x_{n+1} &= x_n \oplus v_{c(n)} \end{aligned}$$

[2] L'Ecuyer & Lemieux (2002) provided a nice exposition of the Faure and the generalised Faure sequence algorithms.

[3] Some interesting history of Gray codes and their definition can be found in Press et al. (2002).

where $c(n)$ is the position of the rightmost zero-bit in the binary representation of n (e.g. $c(3) = c(11_2) = 3$). This method is considerably faster than the former - Galanti & Jung (1997) reported a 20% increase in speed.

Brately & Fox (1988) discussed additional uniformity properties that Sobol' provided in order to suitably choose the initial values $\{m_1, \ldots, m_d\}$, known as Property A and Property A'. The authors provide implementation for integrals in up to 40 dimensions. In the financial setting, and in the LIBOR market model in particular, we frequently need to evaluate integrals in an even higher dimension than 40.

To generate a Sobol' sequence in more than one dimension, one needs a different primitive polynomial for each dimension. Joe & Kuo (2003) advise that the chosen primitive polynomial should be of as low a degree as possible and that the initial values $\{m_1, \ldots, m_d\}$ should be chosen differently for any two primitive polynomials of the same degree. Joe & Kuo (2003) provided initial values to evaluate integrals up to 1111 dimensions, using primitive polynomials of degree 13 or less that satisfy Property A. We will use these initial values and primitive polynomials to generate multidimensional Sobol' sequences.

The final comment that we make about Sobol' sequences concerns the chosen number of runs. Brately & Fox (1988) stated that the Sobol' sequence has theoretical uniformity properties whenever the sample size is 2^k, $k \geq \max\{2d, \tau_d + d - 1\}$, where d is the dimension and τ_d is a constant defined by Sobol'. Due to the high-dimensionality of the problems encountered in finance, even a sample size of 2^{2d} is *far* too large. In experiments with multidimensional Sobol' sequences, we found that, ignoring the zeroth draw,[4] a sample size of $2^m - 1$, $m = 2, 3, \ldots$, is desirable. The reason is that the mean of the sampled normal random numbers, generated using the inverse cumulative normal function, is zero (for each dimension) if the sample size is $2^m - 1$. The fact that the first moments of the sample and the distribution are matched, results in a more regular sample. This is exactly the intuition behind antithetic variates.

4.3 Effective Dimension

The Paskov & Traub (1995) paper sparked a lot of interest because it showed that Sobol' sequences can and do outperform Monte Carlo methods even in up to 360 dimensional integration problems. To explain the empirically observed fact that quasi-Monte Carlo methods can accurately estimate high-dimensional integrals for practical sample sizes, Paskov (1997) introduced the notion of problem's effective dimension. The basic argument is that the integrands encountered in finance have a lower effective dimension than the actual dimension of the problem. However, while it is an interesting and useful concept, recent research by Tezuka (2005) has shown that low effective dimension is not even a necessary condition for quasi-Monte Carlo to outperform Monte Carlo.

Consider the integral $I(f)$ in equation (4.1). When viewed as a function of independent random variables, $f(\mathbf{x})$ can be decomposed into "main effects", "interactions" and "higher-order interactions" (Efron & Stein 1981). The *functional ANOVA decomposition of* f is a decomposition of f into orthogonal functions

$$f(\mathbf{x}) = \sum_{u \subseteq \{1,\ldots,d\}} f_u(\mathbf{x})$$

where u is a subset of $\{1, \ldots, d\}$ and $f_u(\mathbf{x})$ depends only on $\{x_i, i \in u\}$. For the empty set $\emptyset$, $f_\emptyset(\mathbf{x}) = \mu$, a constant. The random variables $f_u(\mathbf{x})$ are constructed to have a mean of zero and be mutually uncorrelated, i.e. $\int_{[0,1)^d} f_u(\mathbf{x}) d\mathbf{x} = 0$, $u \neq \emptyset$ and $\int_{[0,1)^d} f_u(\mathbf{x}) f_k(\mathbf{x}) d\mathbf{x} = 0$, $u \neq k$. A recurrence relationship for defining these functions is given by Owen (1998). The variance of f is

$$\sigma_f^2 = \int_{[0,1)^d} \big(f(\mathbf{x}) - I(f)\big)^2 d\mathbf{x} = \sum_u \int_{[0,1)^d} \big(f_u(\mathbf{x}) - I(f)\big)^2 d\mathbf{x} = \sum_u \sigma_{f_u}^2$$

which is reminiscent of the ANOVA in statistics, where the decomposition of the total sum of squares is examined.

Caflisch et al. (1997) define *effective dimension in the truncation sense* as the smallest integer d_T such that

$$\sum_{u \subseteq \{1,\ldots,d_T\}} \sigma_{f_u}^2 \geq 0.99 \sigma_f^2$$

[4] The zeroth draw is a vector of zeros. It is removed because the uniform random numbers are used to simulate a normal distributions using the inverse cumulative normal distribution function. This function is zero at minus infinity.

The choice of the 99'th percentile is arbitrary. The integer d_T is interpreted as an indicator of which variables are needed. In particular, if the first d_T random variables, $X_1, \ldots, X_{d_T}$, are uniformly distributed, approximation (4.2) should be good. In financial applications, frequently all the variables are important. The *effective dimension in the superposition sense* is the smallest integer d_S such that

$$\sum_{0<|u|\leq d_s} \sigma_{f_u}^2 \geq 0.99\sigma_f^2$$

where $|u|$ is the cardinality of u. The integer d_S indicates the size of the subsets that are significant in the ANOVA decomposition - the function f is well approximated with a sum of 1 to d_S-dimensional functions. If the d_S-dimensional projections of $X_1, \ldots, X_d$ have low discrepancy, the estimate of the integral should be accurate (Caflisch et al. 1997).

To further reduce the effective dimension, Brownian motion sample paths can be constructed using a Brownian bridge. The Brownian bridge algorithm constructs paths using an initial point of the process and a point at a terminal time T. The construction places most of the paths' variance in the first few steps. This implies that the variance will be determined by the first few dimensions of a low-discrepancy sequence, which are well distributed.[5]

4.3.1 Brownian Bridge Construction

Brownian bridge is a technique for constructing Brownian motion paths. Let W_t be a Brownian motion at time $t \geq 0$. The Brownian bridge technique relies on the generation of W_{t_j} given W_{t_i} and W_{t_k}, $t_i < t_j < t_k$. From the definition of conditional distributions, the conditional density f of $W = W_{t_j}$ given W_{t_i} and W_{t_k} is

$$f(W \,|\, W_{t_i} = w_{t_i}, W_{t_k} = w_{t_k}) = \frac{\frac{1}{\sqrt{(2\pi)^3|\Sigma_3|}} \exp\left\{ -\frac{1}{2} \left(w_{t_i}\ w\ w_{t_k}\right) \Sigma_3^{-1} \left(w_{t_i}\ w\ w_{t_k}\right)' \right\}}{\frac{1}{\sqrt{(2\pi)^2|\Sigma_2|}} \exp\left\{ -\frac{1}{2} \left(w_{t_i}\ w_{t_k}\right) \Sigma_2^{-1} \left(w_{t_i}\ w_{t_k}\right)' \right\}}$$

where

$$\Sigma_3 = \begin{pmatrix} t_i & t_i & t_i \\ t_i & t_j & t_j \\ t_i & t_j & t_k \end{pmatrix}, \qquad \Sigma_2 = \begin{pmatrix} t_i & t_i \\ t_i & t_k \end{pmatrix}$$

After considerable simplification, we can rewrite the conditional expectation as

$$f(W \,|\, W_{t_i} = w_{t_i}, W_{t_k} = w_{t_k}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \left(w - \mu\right)^2 \right\} \tag{4.5}$$

where

$$\sigma^2 = \frac{(t_k - t_j)(t_j - t_i)}{t_k - t_i}, \quad \mu = \left(\frac{t_k - t_j}{t_k - t_i}\right) w_{t_i} + \left(\frac{t_j - t_i}{t_k - t_i}\right) w_{t_k}$$

For simulation purposes, the generation of W_{t_j}, given realizations $W_{t_i} = w_{t_i}$ and $W_{t_k} = w_{t_k}$, $t_i < t_j < t_k$, is given by the linear interpolation

$$W_{t_j} = \left(\frac{t_k - t_j}{t_k - t_i}\right) w_{t_i} + \left(\frac{t_j - t_i}{t_k - t_i}\right) w_{t_k} + \sqrt{\left(\frac{(t_k - t_j)(t_j - t_i)}{t_k - t_i}\right)} Z \tag{4.6}$$

where Z is a $N(0,1)$ random variable. It is easy to see that the interpolation (4.6) preserves the conditional distribution (4.5).

Consider simulating a Brownian motion path over $[0, T]$, with n steps $W_{t_1}, \ldots, W_{t_n}$. The Brownian bridge algorithm first generates the endpoint $W_{t_n} = \sqrt{T}Z$, where Z is a $N(0,1)$ random variable, and then "fills in" the path, conditional on this realization. The next sample point is $W_{t_{\lfloor n/2 \rfloor}}$, generated using $w_{t_0} = 0$ and the realized w_{t_n}. The sequence is generated in the following order, using the closest two points to the one that is being generated

$$W_{t_n}, W_{t_{\lfloor n/2 \rfloor}}, W_{t_{\lfloor n/4 \rfloor}}, W_{t_{\lfloor 3n/4 \rfloor}}, W_{t_{\lfloor n/8 \rfloor}}, \ldots$$

[5]Note that this is not a universally more desirable construction. Papageorgiou (2002) demonstrated a case of a digital option where the Brownian bridge construction performed worse than the standard construction of Brownian motion.

For example, if $n = 10$, the sequence will be generated in the following order

$$W_{t_{10}}, W_{t_5}, W_{t_2}, W_{t_7}, W_{t_1}, W_{t_3}, W_{t_6}, W_{t_8}, W_{t_4}, W_{t_9}$$

To understand how this approach to generating Brownian motion reduces the effective dimension, recall that the standard way of generating Brownian motion is to generate W_{t_j} given the realization $W_{t_i} = w_{t_i}$, $t_i < t_j$, as

$$W_{t_j} = w_{t_i} + \sqrt{(t_j - t_i)}Z \tag{4.7}$$

where Z is a $N(0,1)$ random variable. Comparing the variance of (4.6) with (4.7), we see that the variance of the latter approach is

$$(t_j - t_i) \geq \frac{(t_k - t_j)(t_j - t_i)}{t_k - t_i}$$

Furthermore, most of the variance of the Brownian bridge path is contained in the first few steps.

4.4 Discretizations of the LIBOR Market Model

Forward LIBOR rates can be simulated under any of the forward measures or the spot measure. If one of the forward measures is chosen, the derivative payoff will be normalized by the appropriate zero-coupon bond. If the forward rates are simulated under the spot measure, the discretely compounded money-market account will be used. When the simulated forward rates get very large, the values of the zero-coupon bonds will become very small. This will result in the normalized payoff values under the forward measure becoming very large. Under the spot measure, the value of the money-market account is at least one. For this reason, Glasserman & Zhao (2000) found that simulating forward rates under the spot measure leads to a smaller variance of the caplet prices. To avoid the possible inflation of the variances of the estimators, we will always simulate under the spot measure.

Recall that under the spot measure $\mathbb{Q}^*$, the forward LIBOR rates satisfy equation (2.48)

$$dL_i(t) = L_i(t)\boldsymbol{\lambda}(t,T_i) \cdot \left(\sum_{j=\eta(t)}^{i} \frac{\delta_j L_j(t)\boldsymbol{\lambda}(t,T_j)}{1+\delta_j L_j(t)} + d\mathbf{W}_t^{\mathbb{Q}^*} \right), \quad 0 \leq t \leq T_i$$

for $i = 1, \ldots, n-1$ and $\eta(t)$ a left-continuous function, $T_{\eta(t)-1} \leq t < T_{\eta(t)}$. The simplest simulation scheme is the Euler scheme. The standard approach is to apply the Euler scheme to the log of the forward rates, as this ensures that the forward rates remain positive. The Euler approximation has a strong order of convergence of 0.5 (Kloeden & Platen 1999). In mathematical finance, we are interested in calculating the derivative price as an expectation of the discounted terminal payoff. This does not require pathwise approximation of the Itô processes but simply an approximation of the terminal distribution of the underlying variables. The notion of weak convergence is more relevant in this case. The Euler scheme has weak convergence of order 1.

Fixing a grid of times $0 = t_0 < t_1 < \ldots < t_m$, most commonly taken to be the reset dates, the Euler scheme, applied to $\log L_i(t)$, is given by

$$\begin{aligned} \widetilde{L}_i(t_{k+1}) &= \widetilde{L}_i(t_k) \exp\left\{ \sum_{j=\eta(t_k)}^{i} \frac{\delta_j \widetilde{L}_j(t_k)\boldsymbol{\lambda}(t_k,T_i)\cdot\boldsymbol{\lambda}(t_k,T_j)}{1+\delta_j \widetilde{L}_j(t_k)}(t_{k+1}-t_k) \right. \\ &\quad \left. - \frac{1}{2}||\boldsymbol{\lambda}(t_k,T_i)||^2(t_{k+1}-t_k) + \sqrt{t_{k+1}-t_k}\boldsymbol{\lambda}(t_k,T_i)\cdot\mathbf{Z}_{k+1} \right\} \end{aligned} \tag{4.8}$$

for $i = 0, 1, \ldots, m-1$, where $\mathbf{Z}_1, \ldots, \mathbf{Z}_m$ are independent d-dimensional standard multivariate normal random variables.

The assumption of the Euler scheme (4.8) is that over the interval $[t_k, t_{k+1}]$, the values of $L_j(t)$ and $\boldsymbol{\lambda}(t,T_j)$, $j = \eta(t_k), \ldots, i$, are approximately equal to their values at the beginning of the interval, $L_j(t_k)$ and $\boldsymbol{\lambda}(t_k,T_j)$. It is easy to relax the latter assumption of $\boldsymbol{\lambda}(t,T_j) \approx \boldsymbol{\lambda}(t_k,T_j)$, $t \in [t_k, t_{k+1}]$ (see Hunter et al. (2001)). The former assumption of $L_j(t) \approx L_j(t_k)$, $t \in [t_k, t_{k+1}]$, is the main source of error. The industry standard for improving the discretization error is not to reduce the step size, but to use the methods of Glasserman & Zhao (2000) and Hunter, Jäckel & Joshi (2001) which we now discuss.

4.4.1 Arbitrage-free Discretization

Glasserman & Zhao (2000) regarded a discretization to be arbitrage-free if the discrete normalized bond-prices are martingales. The discrete normalized bond-price processes are not necessarily martingales for the discretization (4.8). Glasserman & Zhao (2000) proposed simulating suitable martingales directly, and recovering forward LIBOR rates from these martingales.

Consider simulating over a grid of times $0 = t_0 < t_1 < \ldots < t_m$. We define the discrete normalized bond-price process, $\{D_i(t_1), \ldots, D_i(t_m)\}$, for each $i = 1, \ldots, n$, as

$$D_i(t_k) = \begin{cases} \frac{B(t_k,T_i)}{B^*(t_k)} & 0 \le t_k < T_i \\ \frac{1}{B^*(T_i)} & t_k \ge T_i \end{cases} \tag{4.9}$$

where B^* is the simply compounded money-market account defined by equation (2.41). A simple approach to keep the discretization arbitrage-free is to simulate the normalized bond-price process itself. Glasserman & Zhao (2000) found that this approach results in approximately the same bias as the Euler scheme application. Instead, they suggested simulating the following transformation

$$V_i(t) = \frac{D_i(t) - D_{i+1}(t)}{B(0,T_1)}$$

To find the dynamics of the process $V_i(t)$ under $\mathbb{Q}^*$, recalling equations (2.43) and (2.46) we have

$$\frac{dD_i(t)}{D_i(t)} = -\sum_{j=\eta(t)}^{i-1} \frac{\delta_j L_j(t) \boldsymbol{\lambda}(t,T_j)}{1+\delta_j L_j(t)} \cdot d\mathbf{W}_t^{\mathbb{Q}^*} \tag{4.10}$$

The normalized bond-price process $D_1(t)$ is a constant $D_1(t) = B(0,T_1)$, $t \in [0,T_n]$, because the value of the simply compounded money market account at $t = 0$ is one. For $i = 2, \ldots, n$

$$\begin{aligned} D_i(t) &= D_{i-1}(t) - B(0,T_1)V_{i-1}(t) \\ &= D_{i-2}(t) - B(0,T_1)V_{i-2}(t) - B(0,T_1)V_{i-1}(t) \\ &= D_1(t) - B(0,T_1)\left(V_1(t) + \cdots + V_{i-1}(t)\right) \\ &= B(0,T_1)(1 - V_1(t) - \cdots - V_{i-1}(t)) \end{aligned} \tag{4.11}$$

Extend the definition of the forward LIBOR rates over the whole interval $[0,T_n]$ by defining the forward LIBOR rate after reset date to equal the reset rate i.e. $L_i(t) = L_i(T_i)$ for $t \in [T_i,T_n]$. Then

$$\delta_i L_i(t) = \frac{D_i(t) - D_{i+1}(t)}{D_{i+1}} = \frac{V_i(t)}{1 - V_1(t) - \cdots - V_i(t)} \tag{4.12}$$

Substituting equations (4.12) and (4.11) into (4.10), we obtain the dynamics of D in terms of V

$$dD_i(t) = B(0,T_1)(1 - V_1(t) - \cdots - V_{i-1}(t)) \left(-\sum_{j=\eta(t)}^{i-1} \frac{V_j(t)\boldsymbol{\lambda}(t,T_j)}{1 - V_1(t) - \cdots - V_{j-1}(t)} \right) \cdot d\mathbf{W}_t^{\mathbb{Q}^*}$$

where $1 - V_1(t) - \cdots - V_{j-1}(t) = 1$ for $j = 1$. The dynamics of V, for $i = 1, \ldots, n-1$, are

$$dV_i(t) = V_i(t) \left(\frac{1 - V_1(t) - \cdots - V_i(t)}{1 - V_1(t) - \cdots - V_{i-1}(t)} \boldsymbol{\lambda}(t,T_i) - \sum_{j=\eta(t)}^{i-1} \frac{V_j(t)\boldsymbol{\lambda}(t,T_j)}{1 - V_1(t) - \cdots - V_{j-1}(t)} \right) \cdot d\mathbf{W}_t^{\mathbb{Q}^*}$$

remembering the convention that $1 - V_1(t) - \cdots - V_{j-1}(t) = 1$ for $j = 1$ and that the empty sum is zero. If the forward rates are non-negative, the zero-coupon bond prices will be non-increasing functions of time-to-maturity, i.e. $0 \le D_{i+1}(t) \le D_i(t)$. From equation (4.11), we have

$$\frac{1 - V_1(t) - \cdots - V_i(t)}{1 - V_1(t) - \cdots - V_{i-1}(t)} = \frac{D_{i+1}(t)}{D_i(t)}, \quad \frac{V_j(t)}{1 - V_1(t) - \cdots - V_{j-1}(t)} = \frac{D_j(t) - D_{j+1}(t)}{D_j(t)}$$

Both of these quantities should not exceed 1 or be negative. To improve computational efficiency in discrete time, Glasserman & Zhao (2000) suggested that these quantities are replaced by

$$\gamma\left(\frac{1 - V_1(t) - \cdots - V_i(t)}{1 - V_1(t) - \cdots - V_{i-1}(t)}\right) \quad \text{and} \quad \gamma\left(\frac{V_j(t)}{1 - V_1(t) - \cdots - V_{j-1}(t)}\right)$$

respectively, where $\gamma(x) = \min\{1, x^+\}$. This does not effect the continuous time limit. The process V is simulated by applying an Euler discretization to $\log V_i$ and the LIBOR rates are backed out using equation (4.12).

4.4.2 Predictor-Corrector Method

Hunter, Jäckel & Joshi (2001) proposed a very simple and effective algorithm for correcting the state-dependent drift: evolve the rates according to (4.8) to obtain the predicted values $\widetilde{L}_i(t_{k+1})$ and recalculate the drift using these evolved LIBOR rates,

$$\mu(t_k)^{\mathrm{c}} = \sum_{j=\eta(t_k)}^{i} \frac{\delta_j \widetilde{L}_j(t_{k+1}) \boldsymbol{\lambda}(t_k, T_i) \cdot \boldsymbol{\lambda}(t_k, T_j)}{1 + \delta_j \widetilde{L}_j(t_{k+1})} (t_{k+1} - t_k)$$

To "correct" the drift of the predicted values, average the initial and calculated drifts

$$\mu(t_k)^{\mathrm{pc}} = \frac{1}{2} \left\{ \mu(t_k)^{\mathrm{c}} + \sum_{j=\eta(t_k)}^{i} \frac{\delta_j \widetilde{L}_j(t_k) \boldsymbol{\lambda}(t_k, T_i) \cdot \boldsymbol{\lambda}(t_k, T_j)}{1 + \delta_j \widetilde{L}_j(t_k)} (t_{k+1} - t_k) \right\}$$

Using the same random numbers, re-evolve the forward rates using the average drift $\mu(t_k)^{\mathrm{pc}}$

$$\widetilde{L}_i(t_{k+1}) = \widetilde{L}_i(t_k) \exp \left\{ \mu(t_k)^{\mathrm{pc}} - \frac{1}{2} ||\boldsymbol{\lambda}(t_k, T_i)||^2 (t_{k+1} - t_k) + \sqrt{t_{k+1} - t_k} \boldsymbol{\lambda}(t_k, T_i) \cdot Z_{k+1} \right\}$$

4.5 Numerical Results

Consider a three-factor LIBOR market model calibrated to the GBP market on 9 September 2005 (see Section 3.4.1). Cap prices are chosen as benchmark instruments because an explicit price for these instruments is given by the Black formula. The instruments that we examine are six year at-the-money, in-the-money and out-of-the-money caps. The grid of times that we simulate over are the reset dates of the forward rates (23 timesteps). Hence the dimension of the problem is $23 \times 3 = 69$ (23 timesteps and 3 factors).

In the first test we simulate forward rates under the spot measure using the standard discretization (4.8). The normal random variates are generated using pseudo- and quasi-random numbers. Figure 4.2 shows that for the six year at-the-money cap, Sobol' sequences do not seem to outperform antithetic Monte Carlo. However, convergence to a stable value is achieved for a relatively small number of sample paths. The same results were obtained when Sobol' sequences were used in conjunction with the Brownian bridge construction of Brownian motion.

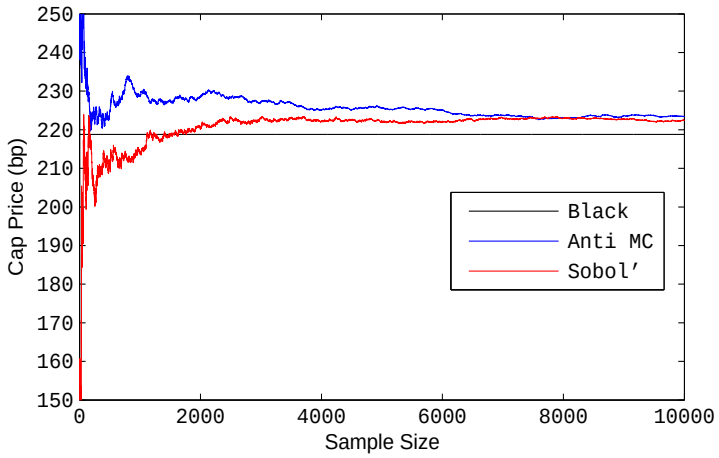


Figure 4.2: Comparison of Monte Carlo and quasi-Monte Carlo cap prices for a six year at-the-money cap.

We now fix the sample size at 5,000 antithetic samples for pseudo-Monte Carlo and at $2^{12} - 1 = 4095$ samples for quasi-Monte Carlo. Define the error in cap price as

$$\text{Error in price} = \text{Black price - Monte Carlo price}$$

Table 4.1 shows the cap price errors for a six year at-the-money, in-the-money and out-of-the-money cap. The strike for in-the-money and out-of-the-money caps is the at-the-money strike minus 50 basis points and at-the-money strike plus 50 basis points respectively. The predictor-corrector approach of Hunter et al. (2001) did not seem to make any visible difference to the standard simulation approach. The Glasserman & Zhao (2000) approach resulted in a slight discretization error reduction. Using the standard discretization (4.8) combined with Sobol' sequences resulted in the smallest cap price errors.

	in-the-money	at-the-money	out-of-the-money
HJJ	-2.7	-6.1	-4.7
GZ	-3.5	-5.5	-3.7
Sobol'	-0.7	-3.5	-2.3

Table 4.1: Pricing errors for a six-year cap, in basis points, when the predictor-corrector approach of Hunter et al. (2001) (HJJ), the arbitrage-free discretization approach of Glasserman & Zhao (2000) (GZ) and the standard discretization combined with Sobol' quasi-random numbers (Sobol') are used to evolve the forward rates.

Chapter 5

American Options

When its a question of money, everybody is of the same religion.

– Voltaire

American options grant the holder the additional right to exercise the option at any time up to the expiry date. The holder of the option should choose their exercise strategy in such a way as to optimize the expected discounted payoff of the option. At every point in time, the state space of the underlying variable(s) can be divided into two regions: the *continuation region*, where it is optimal to continue holding the option and the *exercise region*, where it is optimal to exercise the option. The exercise region is closed while the continuation region is open. The boundary between these two regions is termed the *early exercise boundary* or the *optimal stopping boundary*. American option pricing problems are interesting and complicated because the early exercise boundary must be determined as part of the solution.

The first analysis of this problem appeared in an appendix to Samuelson (1965), where Henry P. McKean Jr. transformed the American option pricing problem into a free boundary problem. The American option value and the early exercise boundary can be formulated as a solution to a system of partial differential equations. This is known as a free boundary problem. McKean analyzed the problem under the real-world probability measure, due to the fact that the paper was written before the seminal works of Black & Scholes (1973) and Merton (1973), and before the discovery of risk-neutral valuation. His analysis was extended by van Moerbeke (1976), who characterized the price of an American option as a solution to a free boundary problem. A rigorous treatment of the American option pricing problem using arbitrage arguments was given by Bensoussan (1984) and Karatzas (1988, 1989). Myneni (1992) provided an excellent summary of these results. The American option pricing problem is formulated in Section 5.1.

Explicit solutions for American option prices do not exist, except for special cases such as the infinite time horizon problems, where the early exercise boundary degenerates to a constant. As a result, numerical procedures have been developed to approximate values of American contingent claims. For the purpose of pricing American options in the LIBOR market model, the only numerical techniques of interest to us are those which are based on Monte Carlo simulation.

Numerical algorithms for pricing American options are typically based on dynamic programming, applied backward-recursively from the final payoff. At expiry, the option value is equal to its payoff. At every other exercise date, the option value is the maximum of the immediate exercise value and the continuation value. This is the dynamic programming principle, explained in more detail in Section 5.1. In contrast to this, Monte Carlo entails simulation of the underlying asset(s) path(s) forward in time. Until recently, "conventional wisdom" advocated that this mismatch in principle approach meant that pricing American derivatives using Monte Carlo was impossible.

In the mid 1990's several authors challenged this conclusion. The first numerical technique for pricing American options using Monte Carlo was proposed by Tilley (1993), who combined dynamic programming with a partitioning algorithm. The so-called bundling algorithm partitions the asset price state space by bundling similarly valued asset prices. The continuation value for each bundle is estimated to be the discounted average of the option prices at the next exercise date. In an extension, Barraquand & Martineau (1995) proposed an algorithm based on partitioning the payoff space, the advantage being that the payoff space is one-dimensional irrespective of the dimension

of the underlying state space. Boyle et al. (1997) provided a summary and comparison of these two algorithms. While these papers were encouraging first steps, a paradigm shift only occurred when the papers by Carriere (1996), Tsitsiklis & Roy (2001), Longstaff & Schwartz (2001), Ju (1998), Andersen (2000) and García (2003) were written. The latter three authors approached the problem of approximating the American option value by approximating the early exercise boundary using a parametric form. The parameters of the early exercise boundary are estimated by maximizing the option value over the relevant parameter space. This approach is examined in Section 5.2. On other hand, the first three authors approached the problem by approximating the continuation value of the option using a parametric form. Carriere (1996) approximated the continuation value using nonparametric regression techniques while Tsitsiklis & Roy (2001) and Longstaff & Schwartz (2001) used parametric regression. These regression-based algorithms are analyzed in Section 5.3.

Our opinion is that the ideas in the paper by Carriere (1996) have not been explored. Nonparametric regression techniques are regarded as being very slow when compared to parametric regression techniques. In Section 5.4 we examine penalized regression splines, an approach which is only slightly slower than parametric regression, yet allows us to utilize the advantages of nonparametric regression.

The accuracy of the regression-based Monte Carlo methods depends on the choice of predictor variables. As an alternative to choosing appropriate predictor variables, which is not a trivial exercise, we propose the sliced inverse regression technique in Section 5.5. The aim of sliced inverse regression is to capture the main features of the data with a few low-dimensional projections. We use this technique to identify the low-dimensional projections of the forward LIBOR rates and subsequently estimate the continuation value of the option using nonparametric regression. In the final section we compare the Bermudan swaption pricing results obtained using this approach to those published in Andersen (2000).

5.1 Primal and Dual Problem Formulations

Consider an American option maturing at time T, written on some underlying Markov (vector) process. We first fix some preliminary notation. Let f_t and ζ_t be the intrinsic value and the value of the numéraire at time t, respectively. The measure associated with the numéraire ζ is denoted by $\mathbb{Q}^\zeta$. The normalized process f_t/ζ_t is known as the discounted reward process.

Following Myneni (1992), from the general theory of optimal stopping, the *Snell envelope* J of the discounted reward process is the smallest supermartingale which dominates the discounted reward process. It is given by[1]

$$J_t = \operatorname*{ess\,sup}_{\tau \in \mathscr{S}_{t,T}} \mathbb{E}^{\mathbb{Q}^\zeta} \left[\frac{f_\tau}{\zeta_\tau} \,\middle|\, \mathscr{F}_t \right], \qquad 0 \le t \le T \tag{5.1}$$

where $\mathscr{S}_{u,v}$ is a collection of stopping times τ of the filtration $\{\mathscr{F}_t\}$ with values in $[u, v]$, $\{\mathscr{F}_t\}$ being the augmented filtration generated by the underlying process. The optimal stopping time, i.e. the solution to problem (5.1), is given by

$$\tau_t^* = \inf \left\{ t \le u \le T \,\middle|\, J_u = \frac{f_u}{\zeta_u} \right\}, \qquad 0 \le t \le T$$

It can be shown that the stopped process $\{J_{u \wedge \tau_t^*}, t \le u \le T\}$ is a martingale.

Define a process V as $V_t = \zeta_t J_t$. Karatzas (1988), and later Myneni (1992), proved that V perfectly hedges the American option, meaning that $V_t \ge f_t$ for $t \in [0, T)$ and $V_T = f_T$, almost surely. Consequently, the value of the American option cannot exceed V_0. More importantly, these authors also proved that, in an arbitrage-free market, V_0 cannot exceed the value of the American option. The value of the American option is thus equal to V_0. The *primal formulation* of the American option pricing problem is

$$\begin{aligned} \frac{V_0}{\zeta_0} &= \sup_{\tau \in \mathscr{S}_{0,T}} \mathbb{E}^{\mathbb{Q}^\zeta} \left[\frac{f_\tau}{\zeta_\tau} \right] \\ &= \mathbb{E}^{\mathbb{Q}^\zeta} \left[\frac{f_{\tau^*}}{\zeta_{\tau^*}} \right] \end{aligned} \tag{5.2}$$

[1] Given an arbitrary collection of measurable functions, $F = \{f_i\}_{i \in I}$, the function $G(\omega) = \operatorname{ess\,sup} F$ is a measurable function which satisfies (i) $G(\omega) \ge f_i(\omega)$ for all $i \in I$ and $\omega \in \Omega$ and (ii) for any function satisfying $\varphi(\omega) \ge f_i(\omega)$ for all $i \in I$ and $\omega \in \Omega$, $\varphi(\omega) \ge G(\omega)$. The difference between a supremum and an essential supremum of a collection of measurable functions is that an essential supremum is defined to be a measurable function.

where the optimal stopping time is given by

$$\tau^* = \inf\left\{0 \le t \le T \,|\, V_t = f_t\right\} \tag{5.3}$$

The primal formulation suggests that one can estimate the price of a derivative by approximating the optimal stopping rule τ^*. The resulting estimate, for a finite sample size, will be biased low, meaning that the expected value of the estimator is less than or equal to the true option value. The reason for this bias is that any stopping rule cannot be better than the optimal stopping rule.

To implement Monte Carlo based numerical methods, one needs to replace the continuous exercise feature of American options with a finite set of exercise times. Options that can be exercised at a discrete set of times are known as Bermudan options. By discretizing the continuous exercise feature of American options, we are approximating the American option price with a Bermudan option price.

A solution of the discrete optimal stopping problem can be obtained using the *dynamic programming principle*. Consider a Bermudan derivative with m exercise times, $\mathfrak{T} = \{t_1, \ldots, t_m\}$, where $0 = t_0 \le t_1 < \ldots < t_m = T$. The primal formulation of the Bermudan option pricing problem is given by equation (5.2) where now $\mathscr{S}_{0,T}$ is the set of stopping times taking values in $\mathfrak{T}$ and $\tau^* = \inf\{t_i \in \mathfrak{T} \,|\, V_{t_i} = f_{t_i}\}$. Lamberton & Lapeyre (1996) provided a detailed examination of the Snell envelope in discrete time. In particular, they showed that the Bermudan price process $V = \{V_{t_i}, i = 0, \ldots, m\}$ is given by the following dynamic programming formulation

$$\begin{aligned} V_{t_m} &= f_{t_m} \\ V_{t_i} &= \max\left\{f_{t_i}, \mathbb{E}^{\mathbb{Q}^\zeta}\left[\frac{\zeta_{t_i}}{\zeta_{t_{i+1}}} V_{t_{i+1}} \,\middle|\, \mathscr{F}_{t_i}\right]\right\}, \quad \text{for } i = 1, \ldots, m-1 \\ V_{t_0} &= \frac{\zeta_{t_0}}{\zeta_{t_1}} V_{t_1} \end{aligned} \tag{5.4}$$

The conditional expectation in recursion (5.4) is known as the *continuation value* or the *holding value*. The dynamic programming formulation states that at every exercise time, the option value is equal to the maximum of immediate exercise value and the continuation value. In the following sections we examine the possible approaches for estimating the continuation value.

As mentioned previously, the American option price estimates for a finite sample size are generally biased low. One requires a method to assess the accuracy of the lower bound approximation. Rogers (2002) constructed a so-called dual formulation of the American option pricing problem. The dual formulation was introduced independently and simultaneously by Haugh & Kogan (2004), in an earlier working paper version of this reference. Let Ψ be the space of $\mathbb{Q}^\zeta$-martingales M satisfying $\sup_{0\le t\le T} |M_t| \in \mathscr{L}^1$, that are null at 0. Then, for $M \in \Psi$

$$\begin{aligned} \frac{V_0}{\zeta_0} &= \sup_{\tau \in \mathscr{S}_{0,T}} \mathbb{E}^{\mathbb{Q}^\zeta}\left[\frac{f_\tau}{\zeta_\tau} - M_\tau + M_\tau\right] \\ &= \sup_{\tau \in \mathscr{S}_{0,T}} \mathbb{E}^{\mathbb{Q}^\zeta}\left[\frac{f_\tau}{\zeta_\tau} - M_\tau\right] + M_0 \\ &\le \mathbb{E}^{\mathbb{Q}^\zeta}\left[\sup_{0\le t\le T}\left(\frac{f_t}{\zeta_t} - M_t\right)\right] \end{aligned}$$

where the second equality is the Optional Sampling Theorem and the last inequality is Jensen's inequality. Since $M \in \Psi$ was chosen arbitrarily

$$\frac{V_0}{\zeta_0} \le \inf_{M\in\Psi}\left\{\mathbb{E}^{\mathbb{Q}^\zeta}\left[\sup_{0\le t\le T}\left(\frac{f_t}{\zeta_t} - M_t\right)\right]\right\}$$

This is an infinite-dimensional minimization problem over a space of martingales. What is quite remarkable is that the relationship holds with equality for a special martingale. The normalized American option price process V/ζ is a Snell envelope, which has a Doob-Meyer decomposition[2]

$$\frac{V_t}{\zeta_t} = \frac{V_0}{\zeta_0} + N_t - A_t \tag{5.5}$$

[2]The Snell envelope is a càdlàg supermartingale with the relevant integrability properties to admit a Doob-Meyer decomposition. It is regular and of class D - see Karatzas & Shreve (1991) and Myneni (1992) for details.

where N is a square-integrable martingale and A is a continuous non-decreasing process, both processes being null at 0. Now

$$\begin{aligned}\inf_{M\in\Psi}\left\{\mathbb{E}^{\mathbb{Q}^\zeta}\left[\sup_{0\le t\le T}\left(\frac{f_t}{\zeta_t}-M_t\right)\right]\right\} &\le \mathbb{E}^{\mathbb{Q}^\zeta}\left[\sup_{0\le t\le T}\left(\frac{f_t}{\zeta_t}-N_t\right)\right]\\ &\le \mathbb{E}^{\mathbb{Q}^\zeta}\left[\sup_{0\le t\le T}\left(\frac{V_t}{\zeta_t}-N_t\right)\right]\\ &= \mathbb{E}^{\mathbb{Q}^\zeta}\left[\sup_{0\le t\le T}\left(\frac{V_0}{\zeta_0}-A_t\right)\right]\\ &= \frac{V_0}{\zeta_0}\end{aligned}$$

so that when the martingale part of the normalized American option price process is used, the infimum is attained. The *dual formulation* of the American option pricing problem is

$$\frac{V_0}{\zeta_0}=\inf_{M\in\Psi}\left\{\mathbb{E}^{\mathbb{Q}^\zeta}\left[\sup_{0\le t\le T}\left(\frac{f_t}{\zeta_t}-M_t\right)\right]\right\}$$

where the infimum is attained by the martingale part of the Doob-Meyer decomposition of the normalized American option price process.

The dual formulation suggests that an upper bound for the option price can be computed from an arbitrary martingale $M\in\Psi$. However, to produce a tight bound we need to find a "good" martingale. Rogers (2002) suggested that as an approximation to the optimal martingale, one should use the martingale part of the corresponding European option. A natural way to define the martingale part M of a European option price process V^E is to set

$$M_t=\frac{V_t^E}{\zeta_t}-\frac{V_0^E}{\zeta_0}$$

However the main difficulty of calculating an upper bound for the option price remains choosing a suitable martingale. Haugh & Kogan (2004) and Andersen & Broadie (2004) suggested an alternative approach to calculating an upper bound for the option price. The method uses the calculated lower bound to extract the martingale component of the approximated normalized American option price process. These methods are computationally very intensive. Jensen & Svenstrup (2005) discussed approaches for enhancing the numerical efficiency of these algorithms.

5.2 Parametric Early Exercise Boundary

It has been observed (Ju 1998) that the value of American equity options is not very sensitive to the exact position of the early exercise boundary. This suggests that one can accurately price American options by approximating the boundary using a parametric function, such as piecewise linear or piecewise exponential. The parameters of the early exercise boundary are estimated by maximizing the option value over the relevant parameter space. García (2003) proved that the value of an American option with a parametric early exercise boundary converges to an optimal value within the parametric class. Again, as any stopping rule cannot be better than the optimal stopping rule, this approach underestimates the option value.

Following García (2003), we now state the problem formally. Consider an underlying Markov vector process $\mathbf{X}=\{\mathbf{X}_t\in\mathbb{R}^n, 0\le t\le T\}$. Approximate the optimal stopping time τ^*, defined in equation (5.3), by

$$\tau^*(\boldsymbol{\theta})=\inf\left\{0\le t\le T\,|\,\mathbf{X}_t\in\mathfrak{E}_t(\boldsymbol{\theta})\right\}$$

where $\mathfrak{E}_t(\boldsymbol{\theta})$ is the closed subset of $\mathbb{R}^n$ where early exercise is optimal at time t (the exercise region), parameterized by $\boldsymbol{\theta}\in\Theta\subset\mathbb{R}^k$, for some constant k. An estimate of the optimal option value within the parametric class provides an estimate of the American option price.

The optimal option value within the parametric class is given by

$$V_0^p=\zeta_0\sup_{\boldsymbol{\theta}\in\Theta}\left\{\mathbb{E}^{\mathbb{Q}^\zeta}\left[\frac{f_{\tau^*(\boldsymbol{\theta})}}{\zeta_{\tau^*(\boldsymbol{\theta})}}\right]\right\}$$

The algorithm for estimating V_0^p is straightforward. We simulate a set of m sample paths of the underlying vector process. Given an initial guess of the parameter values $\boldsymbol{\theta}^{\text{guess}}$, we calculate the

American option value using the simulated paths and the early exercise boundary at parameter $\boldsymbol{\theta}^{\text{guess}}$. This option value is given by

$$\zeta_0 \left(\frac{1}{m} \sum_{i=1}^{m} \frac{f_{\tau^*(\boldsymbol{\theta}^{\text{guess}})}(\omega_i)}{\zeta_{\tau^*(\boldsymbol{\theta}^{\text{guess}})}(\omega_i)} \right) \tag{5.6}$$

where $f_{\tau^*(\boldsymbol{\theta}^{\text{guess}})}(\omega_i)$ and $\zeta_{\tau^*(\boldsymbol{\theta}^{\text{guess}})}(\omega_i)$ are the intrinsic value and the value of the numéraire respectively, at the exercise time of the i'th sample path. The parameter values $\boldsymbol{\theta}^{\text{guess}}$ are then varied using a multidimensional optimization algorithm until the maximum option value is found. García (2003) suggested the simplex method for the optimization, details of which can be found in Press et al. (2002). Fortunately, the optimization problem frequently decomposes into a series of subproblems, which allows one to avoid the suggested high-dimensional "brute-force" approach. We shall see an example of this in Subsection 5.2.1.

Once the optimal parameter values $\boldsymbol{\theta}^{\text{opt}}$, and hence the early exercise boundary is estimated, another set of sample paths, independent of the first set, is simulated. An approximate American option value is determined using the second set of sample paths and the estimated early exercise boundary. The reason that we do not price the option using the same set of sample paths that were used to estimate the early exercise boundary is due to mixing of bias. To understand this, suppose that we approximate the American option value using the same set of m sample paths that were used to the calculate the optimal parameter values $\boldsymbol{\theta}^{\text{opt}}$. Then the estimator of the American option value is

$$\zeta_0 \max_{\boldsymbol{\theta} \in \Theta} \left(\frac{1}{m} \sum_{i=1}^{m} \frac{f_{\tau^*(\boldsymbol{\theta})}(\omega_i)}{\zeta_{\tau^*(\boldsymbol{\theta})}(\omega_i)} \right) = \zeta_0 \left(\frac{1}{m} \sum_{i=1}^{m} \frac{f_{\tau^*(\boldsymbol{\theta}^{\text{opt}})}(\omega_i)}{\zeta_{\tau^*(\boldsymbol{\theta}^{\text{opt}})}(\omega_i)} \right) \tag{5.7}$$

Jensen's inequality states that for a convex function u and a random variable X with a finite mean, $\mathbb{E}[u(X)] \geq u(\mathbb{E}[X])$. Hence the expected value of estimator (5.7) is greater than or equal to optimal value within the parametric class V_0^p, because the maximum function is convex. However, as V_0^p is less than or equal to the American option value, the overall bias of estimator (5.7) is unknown. To ensure that the estimator is biased low, we use a second set of sample paths to approximate the American option value.

When the underlying process is multidimensional, it becomes even more difficult to determine the structure of the early exercise region. In those situations, the current approach is fruitful only if we have a financial understanding of what determines the early exercise decision. In the LIBOR market model, the number of state variables is equal to the number of forward rates being modelled. The dimension of the problem can sometimes be reduced by recognizing that the exercise decision depends mainly on a one-dimensional function of the state vector rather than the state vector itself. We now examine the application of the parametric early exercise boundary approach to pricing Bermudan swaptions in the LIBOR market model.

5.2.1 Application: Bermudan Swaption Pricing

Bermudan swaptions are one of the most liquid exotic interest rate derivatives. It is important to have an accurate and efficient algorithm to price these instruments in the LIBOR market model.

A payer (receiver) Bermudan swaption provides the holder with the option to enter into a payer (receiver) interest rate swap. Two dates characterize a Bermudan swaption: the lockout date T_{loc}, the date before which one cannot exercise the option and the maturity date of the underlying swap, T_{end}. The Bermudan swaption can be exercised on the reset dates of the underlying swap. By exercising a payer Bermudan swaption at time T_{ex}, where $T_{loc} \leq T_{ex} < T_{end}$, the holder enters into a payer swap maturing at time T_{end}. This is called the T_{end} no-call T_{loc} ($T_{end}\text{NC}\,T_{loc}$) payer Bermudan swaption.

Let us fix some notation. Consider a European payer swaption maturing at T_p, written on a forward swap, with coupon κ, also starting at T_p, with maturity date T_q. Denote the value of this option by $\text{PS}_{p,q}(t)$, $0 \leq t \leq T_p$. This payer swaption gives the owner the right to enter into an interest rate swap to pay the fixed rate κ at times $T_{p+1}, \ldots, T_q$, and receive the floating LIBOR rate corresponding to that tenor. Let the value of the corresponding $T_q\text{NC}T_p$ Bermudan swaption be $\text{BS}_{p,q}(t)$. All the dates $T_p, \ldots, T_{q-1}$ coincide with the reset dates of the forward LIBOR rates.

At each exercise date, the decision to exercise the Bermudan swaption will depend on the values of the forward rates underlying the interest rate swap. However, Andersen (2000) noted that the exercise decision relies mainly on the values of the still-alive European swaptions. In particular,

by no-arbitrage arguments, the value of a Bermudan swaption can never be less than the most expensive European swaption component. Andersen (2000) assumed that the early exercise rule is a function of the still-alive European swaptions and a single, time-dependent parameter. Thus, at each exercise date T_i, $T_i \in \{T_p, T_{p+1}, \ldots, T_{q-1}\}$, the early exercise boundary is parameterized by a single parameter θ_{T_i}. In particular, Andersen (2000) proposed three early exercise rules. The simplest early exercise rule is to exercise the Bermudan swaption at time T_i if the intrinsic value is above some level θ_{T_i}

$$f_{T_i} > \theta_{T_i} \tag{5.8}$$

where the intrinsic value f_{T_i} is the value at time T_i of a European swaption maturing at T_i, $\mathrm{PS}_{i,q}(T_i)$. A more complex (and more realistic) exercise rule is to exercise the Bermudan swaption at time T_i if the intrinsic value is above some level θ_{T_i} and the intrinsic value is larger than the most expensive still-alive European swaption

$$f_{T_i} > \theta_{T_i} \quad \text{and} \quad f_{T_i} > \max_{j=i+1,\ldots,q-1} \left\{\mathrm{PS}_{j,q}(T_i)\right\} \tag{5.9}$$

The final proposed exercise rule is to exercise the Bermudan swaption at time T_i if the intrinsic value is above some level θ_{T_i} plus the value of the most expensive still-alive European swaption

$$f_{T_i} > \theta_{T_i} + \max_{j=i+1,\ldots,q-1} \left\{\mathrm{PS}_{j,q}(T_i)\right\} \tag{5.10}$$

The three early exercise rules (5.8), (5.9) and (5.10) reduce the problem of estimating a high-dimensional early exercise boundary to a series of recursive low-dimensional optimization problems. The vector of parameters $\boldsymbol{\theta} = (\theta_{T_p}, \theta_{T_{p+1}}, \ldots, \theta_{T_{q-1}})'$ is found using the following recursive procedure. We simulate a set of forward rate sample paths. At the final exercise date T_{q-1}, the Bermudan swaption will be exercised if the consequential FRA is in-the-money. This means that we set $\theta_{T_{q-1}} = 0$ (in any of the three formulations stated above). The value of $\theta_{T_{q-2}}$ is then calculated to maximize the value of the Bermudan swaption at time T_{q-2}.[3] This backward procedure is repeated until θ_{T_p} is found.

Interestingly, Andersen (2000) found that, when pricing using strategy (5.8), the dependence of the Bermudan swaption prices on the exact location of the early exercise boundary is quite weak. As a result, the early exercise boundary can be estimated using only a small number of simulations. Once the early exercise boundary is estimated, the Bermudan swaption price can be estimated using a new set of simulations. This ensures that the estimate is biased low. The number of simulations used to value the Bermudan swaption is typically much larger than the number of simulations used to determine the early exercise boundary.

Andersen (2000) performed various tests using additional time-dependent parameters. The tests indicate that one parameter specification produces accurate results and that strategy (5.10) outperforms the other two strategies. However, both strategy (5.9) and strategy (5.10) are computationally burdensome because they require approximation of all the still-alive European swaption prices. Jensen & Svenstrup (2005) noted that the most valuable swaption is almost always the first to mature, because its underlying swap is the longest. They proposed that the maximum of the still-alive European swaptions, $\max_{j=i+1,\ldots,q-1}\{\mathrm{PS}_{j,q}(T_i)\}$, in equations (5.9) and (5.10) be approximated by the value of the first European swaption $\mathrm{PS}_{i+1,q}(T_i)$. Finally we note that an alternative parametrization of the early exercise rule has been proposed by Jäckel (2002) that does not require approximation of European swaption prices at all. Instead, the exercise rule is parameterized as a function of the difference between the first forward rate and the forward swap rate starting at the next reset date. The financial intuition behind this choice is that the short rate and the long swap rate are good proxies for the slope and the level of the yield curve, respectively.

To conclude, the Bermudan swaption pricing problem has a rich structure which allows the use of the parametric early exercise boundary approach. In the following section we will review an alternative valuation approach: The least-squares Monte Carlo proposed by Longstaff & Schwartz (2001). When pricing Bermudan swaptions, Jäckel (2002) stated that the approach of Andersen (2000) was superior to the approach of Longstaff & Schwartz (2001). The main contribution of Longstaff & Schwartz (2001) is the ease of use of least-squares Monte Carlo to price path-dependent products when there are multiple driving factors, and when stochastic variables follow general stochastic processes.

[3] The value of the Bermudan swaption at T_{q-2} is estimated as the average of the sample Bermudan swaption values, these being the intrinsic (if exercise occurs) or the present value of the terminal payoff (if no exercise occurs). The exercise rule at T_{q-2} is determined by $\theta_{T_{q-2}}$, a parameter that we are calculating. Andersen (2000) suggested the golden section algorithm be used for the optimizations.

5.3 Parametric Continuation Value

The dynamic programming formulation for the Bermudan option price, equation (5.4), relies on the estimation of the continuation value. This conditional expectation can be interpreted as the best prediction of the discounted Bermudan option value *one time-step ahead* given the current market conditions. Regression estimation is a statistical technique for predicting the value of one (response) variable given the values of other (predictor) variables. Carriere (1996), Tsitsiklis & Roy (2001) and Longstaff & Schwartz (2001) proposed the use of regression techniques for estimation of continuation values in the Monte Carlo framework.

5.3.1 Parametric Regression

Regression estimation is one of the most commonly used statistical techniques. It is believed that it was first thought of by Galileo Galilei in 1632, when he constructed a procedure which can be interpreted as fitting a linear relationship to contaminated observed data (Györfi et al. 2002). Regression estimation is a set of techniques that are used to predict the value of the response variable $Y \in \mathbb{R}$ given the value of a random vector $\mathbf{X} \in \mathbb{R}^n$. The problem is to find a measurable function $f : \mathbb{R}^n \to \mathbb{R}$, such that, typically, the squared error $\mathbb{E}\left[(f(\mathbf{X}) - Y)^2\right]$ is minimized. The minimum is achieved when $f(\mathbf{x}) = \mathbb{E}(Y \,|\, \mathbf{X} = \mathbf{x})$ (Grimmett & Stirzaker 2001, page 346). The conditional mean is denoted by $m(\mathbf{x}) = \mathbb{E}(Y \,|\, \mathbf{X} = \mathbf{x})$ and is termed the *regression function*. The aim of regression estimation is to use the data to estimate the regression function.

The classical approach to estimating a regression function is to assume that the structure of the regression function is known and takes a certain parametric form. This is known as *parametric regression*. The most frequently encountered example of parametric regression is *linear regression*

$$m(\mathbf{x}) = \beta_0 + \sum_{i=1}^{n} \beta_i \, x_i \tag{5.11}$$

where $\mathbf{x} = (x_1, \ldots, x_n)'$ and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_n)'$ is a vector of regression coefficients. In many cases, the relationship between the response and the predictor variables is not linear. A popular approach for capturing nonlinearities in the data is to generalize the above model to

$$m(\mathbf{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i \, \psi_i(\mathbf{x}) \tag{5.12}$$

where $\psi_i : \mathbb{R}^n \to \mathbb{R}$ is a transformation of the inputs. Typical examples of the functions ψ are linear, polynomial and trigonometric expansions. Formulation (5.12) is flexible and, given that the structure of the data is known, parametric regression can capture many nonlinearities in the data. The regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ are estimated using the principle of least squares

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{k+1}}{\operatorname{argmin}} \left\{ \frac{1}{m} \sum_{j=1}^{m} \left(y_j - \beta_0 - \sum_{i=1}^{k} \beta_i \, \psi_i(\mathbf{x}_j) \right)^2 \right\} \tag{5.13}$$

where $\{(\mathbf{x}_j, y_j), j = 1, \ldots, m\}$ is the observed dataset. The solution to problem (5.13) is the least squares estimator

$$\widehat{\boldsymbol{\beta}} = \left(\psi(\mathbf{X})'\psi(\mathbf{X})\right)^{-1} \psi(\mathbf{X})'\mathbf{y} \tag{5.14}$$

where $\psi(\mathbf{X}) = (\mathbf{1} \; \psi_1(\mathbf{X}) \; \psi_2(\mathbf{X}) \; \ldots \; \psi_k(\mathbf{X}))$, $\psi_1(\mathbf{X}) = (\psi_1(\mathbf{x}_1), \ldots, \psi_1(\mathbf{x}_m))'$ and $\mathbf{y} = (y_1, \ldots, y_m)'$. Problems encountered in mathematical finance typically require only a small number of functions ψ. As such, the sample size does not have to be very large for a reliable estimate $\widehat{\boldsymbol{\beta}}$.

The main drawback of parametric regression is that the best estimate of the regression function is restricted to the closest function with the specified parametric form.

5.3.2 Least Squares Monte Carlo

A method that has become very popular with practitioners for pricing Bermudan instruments in the LIBOR market model is the least squares Monte Carlo (LSMC) method proposed by Longstaff & Schwartz (2001). The LSMC approach uses parametric regression techniques to estimate the continuation value.

We first present the mathematical underpinnings of the LSMC approach. Recall the dynamic programming formulation (5.4). Suppose that the conditional expectation

$$C_{t_i} = \mathbb{E}^{\mathbb{Q}^\zeta}\left[\frac{\zeta_{t_i}}{\zeta_{t_{i+1}}} V_{t_{i+1}} \,\middle|\, \mathscr{F}_{t_i}\right]$$

for $i = 1, \ldots, m-1$ belongs to a (Hilbert) space of square integrable functions. The Hilbert space representation states that such a function can be represented as a linear combination of basis vectors for the space (Royden 1968, page 212)

$$C_{t_i} = \sum_{v=0}^{\infty} \beta_v \psi_v \left(\mathbf{X}_{t_i}\right)$$

where $\mathbf{X} = \{\mathbf{X}_{t_j} \in \mathbb{R}^n, j = 0, \ldots, m\}$ is the underlying Markov chain and $\{\psi_v\}_{v=0}^{\infty}$ forms a basis. Longstaff & Schwartz (2001) approximated the continuation value using the first M terms

$$C_{t_i} \approx \sum_{v=0}^{M-1} \beta_v \psi_v \left(\mathbf{X}_{t_i}\right)$$

The choice of basis functions ψ can be different at each time-step t_i. We suppress this dependence for notational simplicity. The suggested basis functions are polynomials such as Laguerre, Hermite or Legendre, although in practice simple monomials are frequently used. When pricing an American put option, numerical studies conducted by Moreno & Navas (2003) resulted in many different polynomials yielding almost identical prices.

We now refer to the LSMC algorithm. Consider a Bermudan derivative with m exercise times, $\mathfrak{T} = \{t_1, \ldots, t_m\}$, where $0 = t_0 \leq t_1 < \ldots < t_m = T$. We simulate a set of independent paths of the underlying process. Using the dynamic programming algorithm, at the terminal time we set the estimated option value to equal to the intrinsic value (for each path)

$$\widehat{V}_{t_m} = f_{t_m}$$

We now apply backward induction. At time t_{m-1}, we choose l_{m-1} basis functions $\psi_1, \ldots, \psi_{l_{m-1}}$. Regression coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{l_{m-1}})'$ are estimated using the principle of least squares, equation (5.14). Longstaff & Schwartz (2001) suggested that only the paths for which the option is in-the-money at time t_{m-1} be used in the regression. This allows for better estimation of the function in the in-the-money region, where exercise matters. A natural estimate of the option value at time t_{m-1}, for each sample path, is the estimator

$$\begin{aligned} \widehat{V}_{t_{m-1}} &= \max\left\{f_{t_{m-1}}, \widehat{C}_{t_{m-1}}\right\} \\ &= \max\left\{f_{t_{m-1}}, \sum_{v=1}^{l_{m-1}} \widehat{\beta}_v \psi_v \left(\mathbf{X}_{t_{m-1}}\right)\right\} \end{aligned} \tag{5.15}$$

This is the approach suggested by Tsitsiklis & Roy (2001). Longstaff & Schwartz (2001) proposed using the realized option value instead of the estimated conditional expectation

$$\widehat{V}_{t_{m-1}} = \begin{cases} f_{t_{m-1}} & \text{if } f_{t_{m-1}} \geq \displaystyle\sum_{v=1}^{l_{m-1}} \widehat{\beta}_v \psi_v \left(\mathbf{X}_{t_{m-1}}\right) \\ \dfrac{\zeta_{t_{m-1}}}{\zeta_{t_m}} V_{t_m} & \text{if } f_{t_{m-1}} < \displaystyle\sum_{v=1}^{l_{m-1}} \widehat{\beta}_v \psi_v \left(\mathbf{X}_{t_{m-1}}\right) \end{cases}$$

Either way, the procedure is repeated recursively until the first exercise date is reached. To ensure that the estimator is biased low, once all the regression coefficients are generated, a new set of sample paths can be simulated to approximate the value of the Bermudan option.

The LSMC algorithm involves two approximations. Firstly, the continuation value is approximated by projections onto a finite number of functions. Secondly, least-squares regression using simulated sample paths is used to estimate these projections. Clément, Lamberton & Protter (2002) and Stentoft (2004) proved the convergence of the Longstaff & Schwartz (2001) approach.

Choice of Basis Functions

The choice of basis functions when pricing Bermudan derivatives in the LIBOR market model is crucial. Longstaff & Schwartz (2001) considered the pricing of Bermudan swaptions in a class of models called string models with deterministic zero-coupon bond price volatility functions. This is in contrast to the LIBOR market model, where discrete forward rate volatility functions are deterministic. However, there is often equivalence between these models and the LIBOR market model, as shown by Kerkhof & Pelsser (2002).

At each exercise date of the Bermudan swaption, Longstaff & Schwartz (2001) used the following set of basis functions: 1 (for the intercept), the still-alive zero-coupon bonds, the underlying swap rate and the underlying swap rate squared and cubed. The problem with this choice of basis functions is that they are highly correlated, leading to possible multi-collinearity problems. Numerical inaccuracies could arise due to almost singular matrices, and the estimation of regression parameters (5.14) becomes more difficult. In practice, for Bermudan swaptions, the net present value of the underlying swap is typically taken to be the single explanatory variable.

For general Bermudan derivative pricing, Piterbarg (2004) stated that *almost universally*, the two most useful variables are the ones representing the level and the slope of an interest rate curve. The proxy for the former is the "core" swap rate, defined by Piterbarg (2004) as the rate that fixes on the exercise date of the option and matures when the whole deal matures. The proxy for the slope of the interest rate curve is taken to be the six-month or one-year LIBOR rate that fixes on the exercise date.

To conclude, parametric regression-based Monte Carlo methods are a flexible approach to pricing Bermudan options using Monte Carlo simulation. The choice of basis functions, which is not a trivial exercise, determines the accuracy of this approach.

5.4 Nonparametric Continuation Value

The LSMC approach presented in the previous section relies on both the appropriate choice of predictor variables and the appropriate choice of a parametric relationship between the response variable and the predictor variables. The number of exercise times for a Bermudan option is typically too large for one to examine the quality of the regressions at each time-step. When a parametric structure of the regression function is not known, flexibility can be gained using nonparametric regression function estimation. The estimation of continuation value using nonparametric regression was proposed by Carriere (1996). Following Györfi et al. (2002), we present the "four paradigms" of nonparametric regression. In particular, we focus on the penalized regression splines approach, which is both flexible and, more importantly, computationally fast.

5.4.1 Paradigms of Nonparametric Regression

In contrast to parametric regression techniques, nonparametric regression does not assume a global functional form of the relationship between the response variable and the predictor variables. Instead it "allows" the data to determine the regression function.

Suppose that we are given a dataset $\{(\mathbf{x}_i, y_i), i = 1, \ldots, m\}$. A simple and intuitive approach for estimating the conditional expectation of Y, given a point $\mathbf{X} = \mathbf{x}$ is to average the observed y_i's whose corresponding $\mathbf{x}_i$'s are close to $\mathbf{x}$ in some metric. This is known as *local averaging*. Local averaging methods were first introduced in the 1960's by Loftsgaarden & Quesenberry (1965). They proposed the *k-nearest neighbors (k-NN) estimator*

$$\widehat{m}(\mathbf{x}) = \sum_{i=1}^{m} w_i(\mathbf{x})\, y_i$$

where the weights $w_i(\mathbf{x})$ are either 1, if $\mathbf{x}_i$ is one of the k points closest to $\mathbf{x}$ in the Euclidean metric, or 0, otherwise.[4] The smoothness of the resulting regression curve is controlled by the parameter k. This estimator led to the *partitioning estimator*, where the state space of $\mathbf{X}$, i.e. $\mathbb{R}^n$,

[4] An important question that arises in nonparametric regression is whether the estimators are universally consistent, meaning that the estimates are valid for all distributions and as the sample size increases the estimator converges to the conditional expectation. Precise definitions of consistent estimators, as well as the consistency proofs, are provided in Györfi et al. (2002).

is partitioned into p sets $\mathscr{P} = \{P_1, P_2, \ldots, P_p\}$ and the regression function is estimated by

$$\widehat{m}(\mathbf{x}) = \frac{\sum_{i=1}^{m} 1_{\{\mathbf{x}_i \in P_j\}}\, y_i}{\sum_{i=1}^{m} 1_{\{\mathbf{x}_i \in P_j\}}} \quad \text{for } \mathbf{x} \in P_j \tag{5.16}$$

where 1_A is the characteristic function of a set A and $\widehat{m}(\mathbf{x}) = 0$ if the denominator is zero.

An extension of the partitioning estimate is the *Nadaraya-Watson kernel regression estimator* (Nadaraya 1964, Watson 1964)

$$\widehat{m}(\mathbf{x}) = \frac{\sum_{i=1}^{m} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right) y_i}{\sum_{i=1}^{m} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)}$$

where $\widehat{m}(\mathbf{x}) = 0$, if the denominator is zero, and $K : \mathbb{R}^n \to \mathbb{R}_+$ is a kernel function, used to control the influence of points close to and far away from $\mathbf{x}$. Typically the kernel function is chosen to have either compact support or to be a quickly decaying function. An example of the kernel function is the Epanechnikov kernel $K(x) = \max\{1 - ||x||^2, 0\}$. The constant $h > 0$ is called the bandwidth, a number which controls the size of the neighborhood over which the averaging is taken. It can be shown (Györfi et al. 2002) that this kernel estimator is locally fitting a constant to the data

$$\widehat{m}(\mathbf{x}) = \underset{\beta \in \mathbb{R}}{\operatorname{argmin}} \left\{ \frac{1}{m} \sum_{i=1}^{m} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right) (y_i - \beta)^2 \right\}$$

While various kernel functions K have been proposed, Fan (1992) demonstrated that the Nadaraya-Watson kernel regression estimator suffers from severe (asymptotic) bias. A generalization of local averaging that eliminates this problem is *local modelling*. The idea behind local modelling is that instead of locally fitting a constant, we fit a more general function. From a Taylor expansion, we know that any smooth function can be represented by a low-degree polynomial in a small neighborhood. The most widely used algorithm for fitting a low-degree polynomial locally is the *LOcally WEighted Scatter plot Smoothing* (LOWESS) proposed by Cleveland (1979) and its extension LOESS by Cleveland & Devlin (1988). Analysis of these procedures is given by Fan & Gijbels (1996). We will not examine local polynomial regression as it is computationally a relatively slow procedure.

The third paradigm of nonparametric regression, *global modelling*, is based on the fact that the regression function $m(\mathbf{x})$ satisfies

$$\mathbb{E}\left[(m(\mathbf{X}) - Y)^2\right] = \inf_{f:\mathbb{R}^n \to \mathbb{R}} \mathbb{E}\left[(f(\mathbf{X}) - Y)^2\right]$$

The infimum is taken over all measurable functions $f : \mathbb{R}^n \to \mathbb{R}$. An estimate of the $\mathscr{L}^2$ error of a function f is

$$\widehat{\mathbb{E}}\left[(f(\mathbf{X}) - Y)^2\right] = \frac{1}{m} \sum_{i=1}^{m} |f(\mathbf{x}_i) - y_i|^2$$

The idea behind global modelling is to search for a function that minimizes the square error. The regression function is estimated as

$$\widehat{m}(\cdot) = \underset{f \in \mathfrak{F}}{\operatorname{argmin}} \left\{ \frac{1}{m} \sum_{i=1}^{m} |f(\mathbf{x}_i) - y_i|^2 \right\}$$

where $\mathfrak{F}$ is a class of functions, typically a set of piecewise polynomials with respect to some partition $\mathscr{P} = \{P_1, P_2, \ldots, P_p\}$ of $\mathbb{R}^n$.[5] However, the selection procedures for the partitions can be cumbersome. The penalized regression splines method that we present in the following subsection is a global modelling technique which circumvents this problem of partition selection.

Finally we mention the *penalized modelling* approach, an alternative to global modelling that does not restrict the class of functions. Instead, the idea is to penalize for overparametrization or the "roughness" of the function. These approaches were explored in Green & Silverman (1994).

[5] We cannot minimize over all measurable functions, because this leads to a function that interpolates the data. One needs to restrict the choice of functions. This choice may depends on the data and on the sample size.

5.4.2 Penalized Regression Splines

Penalized regression splines were introduced by Ruppert & Carroll (1997) and made computationally efficient by Ruppert (2002). In this section we present the methodology for the univariate case. The extension to the multivariate case is the tensor product penalized regression splines (Ruppert & Carroll 1997). We discuss the problems associated with multivariate nonparametric regression techniques in the following section.

Consider the problem of estimating the regression function $m(x) = \mathbb{E}(Y \mid X = x)$ where X is univariate. The regression function can be estimated using a *regression spline of degree p*

$$m(x) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p + \sum_{k=1}^{K} \beta_{p+k}(x - \kappa_k)_+^p \tag{5.17}$$

where $p \geq 1$ is an integer, the function $(x)_+^p = (\max\{x, 0\})^p$, $\kappa_1 < \ldots < \kappa_K$ are fixed knots and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p+K})'$ is a vector of regression coefficients. This produces functions that are piecewise polynomials of degree p between the knots and joined up by continuity of $p-1$ at the knots. The number and location of the knots controls the degree of smoothness of the regression function. The knots are typically placed at locations where the curvature of the function changes.

Ruppert & Carroll (1997) suggested that instead of using automatic knot selection procedures, which are very time-consuming, one keeps a fixed number of knots but limits their influence by shrinking the relevant regression coefficients. This approach renders the exact location of the knots not as crucial. The authors recommend setting $5 \leq K \leq 40$ and letting the knots be the equally spaced sample quantiles i.e. κ_k is the $k/(K+1)$'th sample quantile of the predictor variable.

Given a data-set $\{(x_i, y_i), i = 1, \ldots, m\}$, let $\mathbf{y} = (y_1, \ldots, y_m)'$ and let $\mathbf{X}$ be the design matrix for the regression spline

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & \ldots & x_1^p & (x_1 - \kappa_1)_+^p & \ldots & (x_1 - \kappa_K)_+^p \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_m & \ldots & x_m^p & (x_m - \kappa_1)_+^p & \ldots & (x_m - \kappa_K)_+^p \end{pmatrix}$$

The least-squares estimate of the regression coefficients is

$$\widehat{\boldsymbol{\beta}}(\alpha) = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1+K}}{\operatorname{argmin}} \left\{ ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2 + \alpha \sum_{k=1}^{K} \rho(\beta_{p+k}) \right\} \tag{5.18}$$

where ρ is a nonnegative function α is a smoothing parameter. The larger α is, the smaller the estimate of the regression coefficients $\{\beta_{p+k}\}_{k=1,\ldots,K}$. This means that the regression spline approaches a global pth degree polynomial fit as $\alpha \to \infty$. When the penalty function $\rho(x) = x^2$, the least-squares estimate of the regression coefficients is

$$\widehat{\boldsymbol{\beta}}(\alpha) = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1+K}}{\operatorname{argmin}} \left\{ ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2 + \alpha \boldsymbol{\beta}' \mathbf{D} \boldsymbol{\beta} \right\} \tag{5.19}$$

where $\mathbf{D}$ is a diagonal matrix with the first $(1+p)$ diagonal elements equal to 0, and remaining diagonal elements equal to 1. The explicit solution of equation (5.19) is given by

$$\widehat{\boldsymbol{\beta}}(\alpha) = \left(\mathbf{X}'\mathbf{X} + \alpha \mathbf{D}\right)^{-1} \mathbf{X}'\mathbf{y} \tag{5.20}$$

The quality of the regression function estimate relies on an appropriate selection of the smoothing parameter α. Ruppert & Carroll (1997) stated that if the smoothing parameter is selected appropriately and if $p \geq 2$, different values of K produce similar regression function estimates. Thus the choice of α is much more crucial than either the choice of the degree p of the polynomial basis or the number and location of the knots.

The appropriate choice of α is one that leads to a "good fit". From standard regression theory, we know that the quality of the regression is judged by the residual sum of squares

$$\mathrm{RSS}(\alpha) = ||\mathbf{y} - \widehat{\mathbf{y}}(\alpha)||^2$$

where $\widehat{\mathbf{y}}(\alpha) = \mathbf{X}\widehat{\boldsymbol{\beta}}(\alpha)$. This criterion cannot be used directly to choose α, because direct minimization of the residual sum of squares favours functions that interpolate the data. Ruppert &

Carroll (1997) suggested the use of the generalised cross-validation (GCV) statistic. When applied to bandwidth selection in kernel regression, GCV leads to highly variable choices of smoothing parameters and has a tendency to undersmooth (Hurvich, Simonoff & Tsai 1998). We will use the corrected Akaike's information criterion (AIC_C) proposed by Hurvich et al. (1998). Let

$$\mathbf{H}(\alpha) = \mathbf{X}\left(\mathbf{X}'\mathbf{X} + \alpha\mathbf{D}\right)^{-1}\mathbf{X}'$$

be the hat matrix, also known as the smoother matrix. The former name stems from the fact that the hat matrix puts a hat on $\mathbf{y}$, $\widehat{\mathbf{y}} = \mathbf{H}(\alpha)\mathbf{y}$. The AIC_C is given by

$$\text{AIC}_\text{C} = 2\log\left(||\mathbf{y} - \mathbf{H}(\alpha)\mathbf{y}||\right) + \frac{2\left\{\text{tr}\left(\mathbf{H}(\alpha)\right)+1\right\}}{m - \text{tr}\left(\mathbf{H}(\alpha)\right) - 2} \tag{5.21}$$

The value of α is chosen to minimize the corrected Akaike's information criterion. Following Ruppert (2002), we choose the default grid of 100 equally spaced values between -10 and 12 of $\log_{10}(\alpha)$. If the corrected Akaike's information criterion is minimized at the endpoints, then the grid is expanded at that end.

Demmler-Reinsch Algorithm

We choose the smoothing parameter α by computing the corrected Akaike's information criterion for a range of α values and choosing the one that minimizes this criterion. This requires manipulation of the matrices $\mathbf{H}$ in equation (5.21), a procedure which is in general both costly and numerically unstable. To eliminate both of these problems, Ruppert (2002) proposed a variation of the Demmler-Reinsch algorithm.

Consider $\widehat{\boldsymbol{\beta}}(\alpha)$ as defined in equation (5.20). If $\mathbf{X}'\mathbf{X}$ is symmetric and positive definite,[6] then it admits the Cholesky factorization

$$\mathbf{X}'\mathbf{X} = \mathbf{B}^{-\mathbf{1}}\mathbf{B}^{-\mathbf{T}}$$

where $\mathbf{B}^{-1}$ is a unique lower triangular $m \times m$ matrix, with positive diagonal elements (Golub & Van Loan 1996). Construct a symmetric matrix $\mathbf{B}\mathbf{D}\mathbf{B}^T$ and its singular value decomposition

$$\mathbf{B}\mathbf{D}\mathbf{B}^T = \mathbf{U}\mathbf{C}\mathbf{U}'$$

where $\mathbf{U}$ is orthogonal and $\mathbf{C}$ is a diagonal matrix. This implies that $\mathbf{D} = \mathbf{B}^{-\mathbf{1}}\mathbf{U}\mathbf{C}\mathbf{U}'\mathbf{B}^{-\mathbf{T}}$, because the inverse of the transpose is the transpose of the inverse. Hence

$$\begin{aligned}
\mathbf{H}(\alpha) &= \mathbf{X}\left(\mathbf{X}'\mathbf{X} + \alpha\mathbf{D}\right)^{-1}\mathbf{X}' \\
&= \mathbf{X}\left(\mathbf{B}^{-\mathbf{1}}\mathbf{B}^{-T} + \alpha\mathbf{B}^{-\mathbf{1}}\mathbf{U}\mathbf{C}\mathbf{U}'\mathbf{B}^{-\mathbf{T}}\right)^{-1}\mathbf{X}' \\
&= \mathbf{X}\left(\mathbf{B}^{-\mathbf{1}}\mathbf{U}\left(\mathbf{I} + \alpha\mathbf{C}\right)\mathbf{U}'\mathbf{B}^{-\mathbf{T}}\right)^{-1}\mathbf{X}' \\
&= \mathbf{X}\left(\left(\mathbf{U}'\mathbf{B}^{-\mathbf{T}}\right)^{-1}\left(\mathbf{I} + \alpha\mathbf{C}\right)^{-1}\left(\mathbf{B}^{-\mathbf{1}}\mathbf{U}\right)^{-1}\right)\mathbf{X}' \\
&= \mathbf{X}\left(\mathbf{B}'\mathbf{U}\right)\left(\mathbf{I} + \alpha\mathbf{C}\right)^{-1}\mathbf{U}'\mathbf{B}\mathbf{X}' \\
&= \mathbf{Z}\left(\mathbf{I} + \alpha\mathbf{C}\right)^{-1}\mathbf{Z}'
\end{aligned}$$

where $\mathbf{Z} = \mathbf{X}\left(\mathbf{B}'\mathbf{U}\right)$. Finally, since $\mathbf{Z}'\mathbf{Z} = \mathbf{U}'\mathbf{B}\mathbf{X}'\mathbf{X}\left(\mathbf{B}'\mathbf{U}\right) = \mathbf{I}$

$$\begin{aligned}
\text{trace}(\mathbf{H}(\alpha)) &= \text{trace}\left(\mathbf{Z}\left(\mathbf{I} + \alpha\mathbf{C}\right)^{-1}\mathbf{Z}'\right) \\
&= \text{trace}\left(\left(\mathbf{I} + \alpha\mathbf{C}\right)^{-1}\mathbf{Z}'\mathbf{Z}\right) \\
&= \text{trace}\left(\left(\mathbf{I} + \alpha\mathbf{C}\right)^{-1}\right) \\
&= \sum_{i=1}^{m}\left(1 + \alpha c_{ii}\right)^{-1}
\end{aligned}$$

This algorithm requires the calculation of $\mathbf{Z}$ and $\mathbf{C}$ only once.

[6] Ruppert (2002) noted that when using power basis the matrix $\mathbf{X}'\mathbf{X}$ may not be numerically positive definite. He advises the addition of $10^{-10}\mathbf{D}$ to the matrix $\mathbf{X}'\mathbf{X}$ to increase stability, if $\mathbf{X}'\mathbf{X}$ is ill-conditioned.

5.5 Dimension Reduction

Generalizations of the univariate nonparametric regression techniques to multivariate regression problems break down when the dimension of the predictor variables is more than two. The problem that arises when there are several variables is known as the *curse of dimensionality*, a term first used by Bellman (1961). This refers to the fact that the sample size required to estimate a function of several variables accurately can grow exponentially in the number of variables. The reason that the estimators of functions in high dimensions have large variance is the fact that high dimensional spaces are empty. A local neighborhood in high dimensions is not in fact "local". To illustrate this phenomenon, we restate the following example from Hastie et al. (2001): Consider uniformly distributed inputs in a d-dimensional hypercube. Suppose that we wanted to capture $r\%$ of the observations. We would expect $r\%$ of the observations to lie in a hypercube containing $r\%$ of the unit volume. The expected edge length of the hypercube will be $(r/100)^{1/d}$. Hence in 10 dimensions, to find 1% of the data, one needs to cover $0.01^{0.1} = 63\%$ of the range of each input variable. This leads to problems when trying to estimate multivariate densities.

To deal with the curse of dimensionality, we need to impose additional assumptions on the regression functions to reduce the dimensionality of the problem. The number of possible approaches is extensive (Hastie et al. 2001). For example, one of the first simplifying assumptions was to extend the linear multiple regression model (5.11) by removing the linearity assumption, while retaining the assumption that the regression function is an additive function of its components

$$m(\mathbf{x}) = \sum_{i=1}^{n} m_i(x_i)$$

where $\mathbf{x} = (x_1, \ldots, x_n)'$ and $m_1, \ldots, m_n$ are univariate functions, which are obtained using some nonparametric regression technique. This is the class of *additive models*, described in detail in Hastie & Tibshirani (1990).

In this section we present the sliced inverse regression technique introduced by Li (1991) and Duan & Li (1991). Sliced inverse regression is a tool for reducing the dimensionality of the data. The idea behind this approach is that the "interesting" features of high-dimensional data are retrievable from low-dimensional projections. The sliced inverse regression technique has been successfully applied in the marketing environment, where it outperformed other data reduction techniques (Naik et al. 2000).

5.5.1 Sliced Inverse Regression

Sliced inverse regression (SIR) is a technique that aims to capture the main features of the data with a few low-dimensional projections. In the ideal situation, the projection of a n-dimensional predictor variable $\mathbf{X}$ onto a L-dimensional subspace, where L is smaller than n, captures all the information about the response variable Y, i.e. the regression function is given by

$$m(\mathbf{x}) = f\left(\boldsymbol{\gamma}_1'\mathbf{x}, \ldots, \boldsymbol{\gamma}_L'\mathbf{x}\right) \tag{5.22}$$

where $\mathbf{x} = (x_1, \ldots, x_n)'$, f is an arbitrary unknown function $f : \mathbb{R}^L \to \mathbb{R}$ and $\{\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_L\}$ are L unknown n-dimensional column vectors. Li (1991) termed the linear space generated by vectors $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_L$ the *effective dimension reduction (e.d.r) space*. The main contribution of the paper is the use of sliced inverse regression to estimate the vectors $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_L$.

The name of the approach stems from the fact that the method is based on the *inverse regression*

$$\xi(y) = \mathbb{E}\left(\mathbf{x} \,|\, y\right) \tag{5.23}$$

This function is simple to estimate, because y is a scalar - each coordinate of $\mathbf{x}$ is regressed against y. The resulting function $\xi(y)$ is a n-dimensional function. However, Li (1991) proved that under the model assumption (5.22), if the distribution of $\mathbf{x}$ is elliptically symmetrical,[7] then the centered inverse regression curve

$$\mathbb{E}\left(\mathbf{x} \,|\, y\right) - \mathbb{E}\big(\mathbb{E}\left(\mathbf{x} \,|\, y\right)\big) = \mathbb{E}\left(\mathbf{x} \,|\, y\right) - \mathbb{E}\left(\mathbf{x}\right)$$

[7] The elliptically symmetrical condition can be replaced by a weaker condition that the distribution of $\mathbf{x}$ is such that for all $\mathbf{b} \in \mathbb{R}^n$, $\mathbb{E}(\mathbf{b}'\mathbf{x} \,|\, \boldsymbol{\gamma}_1'\mathbf{x}, \ldots, \boldsymbol{\gamma}_L'\mathbf{x}) = c_0 + c_1(\boldsymbol{\gamma}_1'\mathbf{x}) + \ldots + c_L(\boldsymbol{\gamma}_L'\mathbf{x})$, for some constants $c_0, c_1, \ldots, c_L$. Elliptically symmetrical distributions, such as the normal distribution, satisfy this condition.

is contained in the L-dimensional subspace spanned by $\boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\gamma}_L$, where $\boldsymbol{\Sigma}_{\mathbf{x}}$ is the covariance matrix of $\mathbf{x}$. Let $\boldsymbol{\eta}_i = \boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\gamma}_i$, $i = 1, \ldots, L$. Then Li (1991) showed that the covariance matrix

$$\boldsymbol{\Sigma}_{\eta} = \mathrm{Cov}\big(\mathbb{E}\left(\mathbf{x} \,|\, y\right) - \mathbb{E}(\mathbf{x})\big)$$

is degenerate in any direction orthogonal to the $\boldsymbol{\eta}_i$'s. Hence $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_L$ are the eigenvectors associated with the L nonzero eigenvalues of the covariance matrix $\boldsymbol{\Sigma}_{\eta}$.[8] The vectors in the e.d.r space are given by the transformation $\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\eta}_1, \ldots, \boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\eta}_L$.

We now discuss the implementation of the SIR algorithm. Suppose that we have a dataset $\{(\mathbf{x}_i, y_i), i = 1, \ldots, m\}$. To estimate the inverse regression function $\xi(y)$, the simple partitioning estimator (5.16) is used. In particular, we partition the sample range of y into H intervals $P_1, \ldots, P_H$. Li (1991) suggested that the range of y should be partitioned such that the number of observations in each interval $P_1, \ldots, P_H$ are approximately equal. However, the number of intervals is not crucial to the performance of the SIR algorithm. We follow Duan & Li (1991), who recommended that ten intervals are "probably good enough for most purposes". The partitioning estimate of the inverse regression function is given by

$$\widehat{\xi}(y) = \widehat{\xi}_j = \frac{\sum_{i=1}^{m} 1_{\{y_i \in P_j\}}\, \mathbf{x}_i}{\sum_{i=1}^{m} 1_{\{y_i \in P_j\}}} \quad \text{for } y \in P_j$$

where 1_A is the characteristic function of a set A. Duan & Li (1991) proved that the optimal estimate of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ is given by the $n \times n$ weighted covariance matrix

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}} = \sum_{j=1}^{H} \widehat{p}_j \left(\widehat{\xi}_j - \boldsymbol{\mu}_{\mathbf{x}}\right)\left(\widehat{\xi}_j - \boldsymbol{\mu}_{\mathbf{x}}\right)'$$

where $\widehat{p}_j$ is the sample proportion of y_i's in the interval P_j and $\boldsymbol{\mu}_{\mathbf{x}}$ is the sample mean of $\mathbf{x}$. Then the vectors $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_L$ are the eigenvectors of the matrix $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}}$ associated with the L largest eigenvalues. Li (1991) and Ferre (1998) proposed various tests for determining the number of components L that one should retain. Estimates of the vectors in the e.d.r space are given by the transformations $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}\boldsymbol{\eta}_1, \ldots, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}\boldsymbol{\eta}_L$, where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}$ is the inverse of the sample covariance matrix of $\mathbf{x}$.

The only restriction of this algorithm is that it fails if the function f is symmetric about the mean of $\mathbf{x}$. An example is if $y = x_1^2$ and x_1 is symmetric about its mean, where x_1 is the first coordinate variable in $\mathbf{x}$ (Li 1992). However, Li (1992) stated that the chi-square test (Li 1991) for determining the number of factors to retain will conservatively show that no interesting directions were found. In this case one needs to consider alternative approaches such as the ones presented in Li (1992).

5.6 Numerical Results

We now price Bermudan swaptions in a two-factor LIBOR market model presented in Andersen (2000). The initial quarterly forward rates ($\delta_i = 0.25$ for all i) are flat at 10% and the factor loadings are given by

$$\lambda(t, T_i) = \begin{bmatrix} 0.15 \\ 0.15 - \sqrt{0.009(T_i - t)} \end{bmatrix}$$

A range of Bermudan swaptions with different strikes and maturities is considered. The results are compared with those in Andersen (2000), where the parametric early exercise boundary approach was used.[9]

We price Bermudan swaptions using two approaches. The first approach is the least-squares Monte Carlo method of Longstaff & Schwartz (2001). The chosen predictor variable is the net present value (NPV) of the underlying swap. It has been observed that the value of Bermudan swaptions increases almost linearly with the value of the underlying swap (Pietersz et al. 2004). Thus we will assume the simple linear regression model

$$m(x) = \beta_0 + \beta_1 x$$

[8] Under the model assumption (5.22), the covariance matrix $\boldsymbol{\Sigma}_{\eta}$ has L nonzero eigenvalues.

[9] The Bermudan swaption values displayed in Table 5.1 are results obtained using the third exercise rule (5.10).

where x is the NPV of the underlying swap. The second approach relies on the sliced inverse regression technique to project the yield curve (vector of forward rates) at each point in time onto a one-dimensional subspace. Penalized regression splines are then fitted to estimate the continuation value of the option.

The results are presented in Table 5.1. The Bermudan swaption values are calculated using 10,000 antithetic Monte Carlo simulations to estimate the early exercise boundary and another 10,000 antithetic Monte Carlo simulations to value the option. The prices of the corresponding European swaptions are also presented. With the exception of the long dated 11NC1 Bermudan payer swaption, the sliced inverse regression technique combined with penalized regression splines produces results that are in line with the parametric early exercise boundary and the parametric continuation value approaches.

			European Swaption		Bermudan Swaption		
T_{loc}	T_{end}	Strike	Andersen	MC	Andersen	LSMC	SIRPSMC
0.25	1.25	8%	183.6	183.6	184.0	183.8	183.9
0.25	1.25	10%	31.6	30.6	43.2	42.4	42.5
0.25	1.25	12%	0.6	0.5	5.6	5.1	5.1
1	3	8%	332.2	332.2	339.4	340.4	338.8
1	3	10%	101.1	100.3	125.7	126.4	125.6
1	3	12%	16.7	16.2	36.6	37.1	37.3
1	6	8%	719.8	721.4	751.6	749.7	758.34
1	6	10%	211.3	213.7	319.4	320.04	318.81
1	6	12%	31.6	32.8	129.2	128.3	129.66
1	11	8%	1163.7	1165.9	1253.7	1272.4	1292.4
1	11	10%	352.5	356.7	633.2	643.9	644.9
1	11	12%	56.8	58.1	337.0	338.4	341.5
3	6	8%	429.8	431.8	445.2	445.4	440.1
3	6	10%	199.9	202.2	227.5	226.3	223.3
3	6	12%	79.5	80.3	107.6	106.8	106.7

Table 5.1: Bermudan swaption price comparison between the parametric early exercise boundary approach of Andersen (2000) (Andersen), the parametric continuation value approach of Longstaff & Schwartz (2001) (LSMC) and the proposed nonparametric continuation value approach, combined with sliced inverse regression (SIRPSMC). All the prices are in basis points.

Appendix A

Proof of Theorem 2.1.1

As $\mathscr{P}^*$ is nonempty by assumption, fix $\mathbb{P}^* \in \mathscr{P}^*$. Consider $\boldsymbol{\phi} \in \Phi(\mathbb{P}^*)$ with $V_0(\boldsymbol{\phi}) = 0$. Since $V_0^*(\boldsymbol{\phi}) = Y_0 V_0(\boldsymbol{\phi})$, we have $V_0^*(\boldsymbol{\phi}) = 0$. The nonnegativity of the normalized value process $V^*(\boldsymbol{\phi})$ yields $\mathbb{P}^*(V_T^*(\boldsymbol{\phi}) \geq 0) = 1$. From the martingale property of $V^*(\boldsymbol{\phi})$, $\mathbb{E}^{\mathbb{P}^*}[V_T^*(\boldsymbol{\phi})] = 0$ and hence $\mathbb{P}^*(V_T^*(\boldsymbol{\phi}) > 0) = 0$. As $\mathbb{P}^* \sim \mathbb{P}$, we obtain

$$\mathbb{P}(V_T(\boldsymbol{\phi}) \geq 0) = 1 \quad \text{and} \quad \mathbb{P}(V_T(\boldsymbol{\phi}) > 0) = 0$$

Thus there are no arbitrage opportunities among the trading strategies $\boldsymbol{\phi} \in \Phi(\mathbb{P}^*)$.

Appendix B

Proof of Theorem 2.1.5

Consider two trading strategies, $\boldsymbol{\phi} \in \Phi(\mathbb{P}_1^*)$ and $\boldsymbol{\psi} \in \Phi(\mathbb{P}_2^*)$ such that $V_T(\boldsymbol{\phi}) = \vartheta$ and $V_T(\boldsymbol{\psi}) = \vartheta$. From equation (2.8)

$$V_T^*(\boldsymbol{\phi}) = V_t^*(\boldsymbol{\phi}) + G_T^*(\boldsymbol{\phi}) - G_t^*(\boldsymbol{\phi}), \quad 0 \leq t \leq T \tag{B.1}$$

Under $\mathbb{P}_2^*$, $G^*(\boldsymbol{\phi})$ is a local martingale because it is a stochastic integral of $\boldsymbol{\phi}$ with respect to a $\mathbb{P}_2^*$-local martingale $\mathbf{Z}$. Further, $G^*(\boldsymbol{\phi})$ is nonnegative under $\mathbb{P}_2^*$ because $V^*(\boldsymbol{\phi})$, hence $G^*(\boldsymbol{\phi})$, is nonnegative under $\mathbb{P}_1^*$ and $\mathbb{P}_1^*$ and $\mathbb{P}_2^*$ are equivalent probability measures. We now show that a nonnegative local martingale is a supermartingale. For $0 \leq s \leq t \leq T$ we have

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}_2^*}\left[G_t^*(\boldsymbol{\phi}) \,|\, \mathscr{F}_s\right] &= \mathbb{E}^{\mathbb{P}_2^*}\left[\lim_{n\to\infty} G_{t\wedge T_n}^*(\boldsymbol{\phi}) \,|\, \mathscr{F}_s\right] \\
&= \mathbb{E}^{\mathbb{P}_2^*}\left[\liminf_{n\to\infty} G_{t\wedge T_n}^*(\boldsymbol{\phi}) \,|\, \mathscr{F}_s\right] &\text{(B.2)} \\
&\leq \liminf_{n\to\infty} \mathbb{E}^{\mathbb{P}_2^*}\left[G_{t\wedge T_n}^*(\boldsymbol{\phi}) \,|\, \mathscr{F}_s\right] &\text{(B.3)} \\
&= \liminf_{n\to\infty} G_{s\wedge T_n}^*(\boldsymbol{\phi}) \\
&= G_s^*(\boldsymbol{\phi})
\end{aligned}$$

Equation (B.2) follows from the property that for a convergent sequence lim and lim inf coincide and equation (B.3) follows from Fatou's Lemma. Hence under $\mathbb{P}_2^*$, $G^*(\boldsymbol{\phi})$ is a supermatingale. From the assumption $V_T(\boldsymbol{\phi}) = \vartheta = V_T(\boldsymbol{\psi})$ and equation (B.1) we obtain

$$\begin{aligned}
V_t^*(\boldsymbol{\psi}) &= \mathbb{E}^{\mathbb{P}_2^*}\left[V_T^*(\boldsymbol{\psi}) \,|\, \mathscr{F}_t\right] \\
&= \mathbb{E}^{\mathbb{P}_2^*}\left[V_T^*(\boldsymbol{\phi}) \,|\, \mathscr{F}_t\right] \\
&= V_t^*(\boldsymbol{\phi}) + \mathbb{E}^{\mathbb{P}_2^*}\left[G_T^*(\boldsymbol{\phi}) - G_t^*(\boldsymbol{\phi}) \,|\, \mathscr{F}_t\right] \\
&\leq V_t^*(\boldsymbol{\phi})
\end{aligned}$$

Similarly $V_t^*(\boldsymbol{\phi}) \leq V_t^*(\boldsymbol{\psi})$. Hence $V_t^*(\boldsymbol{\phi}) = V_t^*(\boldsymbol{\psi})$ for all $0 \leq t \leq T$.

Bibliography

Alexander, C. (2004), 'Common Correlation and Calibrating the Lognormal Forward Rate Model', *WILMOTT Magazine* pp. 68–77.

Andersen, L. (2000), 'A simple approach to the pricing of Bermudan swaptions in the multifactor LIBOR market model', *Journal of Computational Finance* **3**(2), 5–32.

Andersen, L. & Andreasen, J. (2000), 'Volatility skews and extensions of the Libor market model', *Applied Mathematical Finance* **7**, 1–32.

Andersen, L. & Andreasen, J. (2002), 'Volatile volatilities', *Risk* **15**(12), 163–168.

Andersen, L. & Broadie, M. (2004), 'Primal-Dual Simulation Algorithm for Pricing Multidimensional American Options', *Management Science* **50**(9), 1222–1234.

Andreasen, J. (2004), Interest rate modeling: From the basic to the advanced. The Inaugural Fixed Income Conference: Prague, Czech Republic.

Andreasen, J. (2005), 'Back to the future', *Risk* pp. 104–109.

Back, K. (2001), 'Asset pricing for general processes', *Journal of Mathematical Economics* **20**(4), 371–395.

Barraquand, J. & Martineau, D. (1995), 'Numerical Valuation of High Dimensional Multivariate American Securities', *Journal of Financial and Quantitative Analysis* **30**(3), 383–405.

Bellman, R. E. (1961), *Adaptive Control Processes*, Princeton University Press.

Bensoussan, A. (1984), 'On the theory of option pricing', *Acta Applicandae Mathematicae* **2**(2), 139–158.

Bichteler, K. (1981), 'Stochastic integration and L^p-theory of semimartingales', *The Annals of Probability* **9**(1), 49–89.

Black, F. (1976), 'The pricing of commodity contracts', *Journal of Financial Economics* **3**, 167–179.

Black, F., Derman, E. & Toy, W. (1990), 'A one-factor model of interest rates and its application to treasury bond options', *Financial Analysts Journal* **46**(1), 33–39.

Black, F. & Karasinski, P. (1991), 'Bond and option pricing when short rates are lognormal', *Financial Analysts Journal* **47**(4), 52–59.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *The Journal of Political Economy* **81**(3), 637–654.

Boyle, P. (1979), 'Options: A Monte Carlo approach', *Journal of Financial Economics* **4**, 323–338.

Boyle, P., Broadie, M. & Glasserman, P. (1997), 'Monte Carlo methods for security pricing', *Journal of Economic Dynamics and Control* **21**(8-9), 1267–1321.

Brace, A., Gątarek, D. & Musiela, M. (1997), 'The market model of interest rate dynamics', *Mathematical Finance* **7**(2), 127–155.

Brace, A. & Musiela, M. (1994), 'A multifactor Gauss Markov implementation of Heath, Jarrow and Morton', *Mathematical Finance* **4**(3), 259–283.

Brately, P. & Fox, B. L. (1988), 'Algorithm 659 implementing Sobol's quasirandom sequence generator', *ACM Transactions on Mathematical Software* **14**(1), 88–100.

Brately, P., Fox, B. L. & Niederreiter, H. (1992), 'Implementation and tests of low-discrepancy sequences', *ACM Transactions on Modelling and Computer Simulation* **2**(3), 195–213.

Brigo, D. (2002), A note on correlation and rank reduction. Banca IMI Working Paper, http://www.damianobrigo.it/.

Brigo, D. & Mercurio, F. (2001), *Interest Rate Models: Theory and Practice*, Springer Finance, Springer-Verlag.

Brigo, D., Mercurio, F. & Morini, M. (2005), 'The LIBOR model dynamics: Approximations, calibration and diagnostics', *European Journal of Operational Research* **163**(1), 30–51.

Caflisch, R. E., Morokoff, W. & Owen, A. (1997), 'Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension', *Journal of Computational Finance* **1**(1), 88–100.

Carriere, J. F. (1996), 'Valuation of early-exercise price of options using simulations and nonparametric regression', *Insurance: Mathematics and Economics* **19**(1), 19–30.

Choy, B., Dun, T. & Schlögl, E. (2004), 'Correlating market models', *Risk* **17**(9), 124–129.

Christopeit, N. & Musiela, M. (1994), 'On the existence and characterization of arbitrage-free measures in contingent claim valuation', *Stochastic Analysis and Applications* **12**(1), 41–63.

Clément, E., Lamberton, D. & Protter, P. (2002), 'An analysis of a least squares regression method for American option pricing', *Finance and Stochastics* **6**(4), 449–471.

Cleveland, W. S. (1979), 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association* **74**(368), 829–836.

Cleveland, W. S. & Devlin, S. J. (1988), 'Locally weighted regression: an approach to regression analysis by local fitting', *Journal of the American Statistical Association* **83**, 597–610.

Courtault, J., Kabanov, Y., Bru, B., Crépel, P., Lebon, I. & Le Marchand, A. (2000), 'Louis Bachelier on the centenary of *Théorie de la Spéculation*', *Mathematical Finance* **10**(3), 341–351.

Cox, J. C. & Ross, S. (1976), 'The valuation of options for alternative stochastic processes', *Journal of Financial Economics* **3**, 145–166.

Daniluk, A. & Gątarek, D. (2005), 'A fully lognormal libor market model', *Risk* pp. 115–118.

De Jong, F., Driessen, J. & Pelsser, A. A. J. (2004), 'On the information in the interest rate term structure and option prices', *Review of Derivatives Research* **7**.

Delbaen, F. (1992), 'Representing martingale measures when asset prices are continuous', *Mathematical Finance* **2**(2), 107–130.

Delbaen, F. & Schachermayer, W. (1994*a*), 'Arbitrage and free lunch with bounded risk for unbounded continuous processes', *Mathematical Finance* **4**(4), 343–348.

Delbaen, F. & Schachermayer, W. (1994*b*), 'A general version of the fundamental theorem of asset pricing', *Mathematische Annalen* **300**, 463–520.

Delbaen, F. & Schachermayer, W. (1998), 'A fundamental theorem of asset pricing for unbounded stochastic processes', *Mathematische Annalen* **312**(2), 215–250.

Duan, N. & Li, K.-C. (1991), 'Slicing regression: A link-free regression method', *The Annals of Statistics* **19**(2), 505–530.

Duffie, D. (1996), *Dynamic Asset Pricing Theory*, second edn, Princeton University Press.

Duffie, D. & Kan, R. (1996), 'A yield-factor model of the interest rates', *Mathematical Finance* **64**, 379–406.

Dykstra, R. L. (1983), 'An algorithm for restricted least squares regression', *Journal of the American Statistical Association* **78**, 837–842.

Efron, B. & Stein, C. (1981), 'The jackknife estimate of variance', *The Annals of Statistics* **9**(3), 586–596.

Fan, J. (1992), 'Design-adaptive nonparametric regression', *Journal of the American Statistical Association* **87**(420), 998–1004.

Fan, J. & Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Monographs on Statistics and Applied Probability 66, Chapman & Hall.

Ferre, L. (1998), 'Determining the dimension in sliced inverse regression and related methods', *Journal of the American Statistical Association* **93**(441), 132–140.

Galanti, S. & Jung, A. (1997), 'Low-discrepancy sequences: Monte Carlo simulation of option prices', *The Journal of Derivatives* **5**(1), 63–83.

García, D. (2003), 'Convergence and biases of Monte Carlo estimates of American option prices using a parametric exercise rule', *Journal of Economic Dynamics & Control* **27**, 1855–1879.

Geman, H., El Karoui, N. & Rochet, J.-C. (1995), 'Changes of numéraire, changes of probability measure and option pricing', *Journal of Applied Probability* **32**(2), 443–458.

Glasserman, P. & Zhao, X. (2000), 'Arbitrage-free discretization of lognormal forward Libor and swap rate models', *Finance and Stochastics* **4**(1), 35–68.

Goldys, B., Musiela, M. & Sondermann, D. (1994), Lognormality of rates and term structure modles. University of Bonn Discussion Paper B–394, http://ideas.repec.org/p/bon/bonsfb/394.html.

Goldys, B., Musiela, M. & Sondermann, D. (2000), 'Lognormality of rates and term structure models', *Stochastic Analysis and Applications* **18**(3), 375–396.

Golub, G. H. & Van Loan, C. F. (1996), *Matrix Computations*, third edn, The Johns Hopkins Univerity Press.

Green, P. J. & Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman & Hall.

Grimmett, G. & Stirzaker, D. (2001), *Probability and Random Processes*, third edn, Oxford Univerity Press.

Grubišić, I. (2002), Interest rate theory: The BGM model, Master's thesis, Leiden University. http://www.math.uu.nl/people/grubisic/publications.html.

Györfi, L., Kohler, M., Krzyżak, A. & Walk, H. (2002), *A Distribution-Free Theory of Nonparametric Regression*, Springer Series in Statistics, Springer-Verlag.

Hagan, P. & West, G. (2005), Interpolation methods for curve construction. To appear in Applied Mathematical Finance, available at http://www.finmod.co.za/research.html.

Hammersley, J. M. & Handscomb, D. (1964), *Monte Carlo Methods*, Methuen's Monographs on Applied Probability and Statistics, Methuen & Co.

Harrison, J. M. & Kreps, D. M. (1979), 'Martingales and arbitrage in multiperiod securities markets', *Journal of Economic Theory* **20**, 381–408.

Harrison, J. M. & Pliska, S. R. (1981), 'Martingales and stochastic integrals in the theory of continuous trading', *Stochastic Processes and their Applications* **11**(3), 215–260.

Harrison, J. M. & Pliska, S. R. (1983), 'A stochastic calculus model of continuous trading: Complete markets', *Stochastic Processes and their Applications* **15**(3), 313–316.

Hastie, T. J. & Tibshirani, R. J. (1990), *Generalized Additive Models*, Monographs on Statistics and Applied Probability 43, Chapman & Hall.

Hastie, T. J., Tibshirani, R. J. & Friedman, J. H. (2001), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer Series in Statistics, Springer-Verlag.

Haugh, M. B. & Kogan, L. (2004), 'Pricing American options: A duality approach', *Operations Research* **52**(2), 258–270.

Heath, D., Jarrow, R. & Morton, A. (1992), 'Bond pricing and the term structure of interest rates: A new methodology', *Econometrica* **60**(1), 77–105.

Higham, N. J. (2002), 'Computing the nearest correlation matrix - a problem from finance', *IMA Journal of Numerical Analysis* **22**, 329–343.

Hull, J. C. & White, A. (1990), 'Pricing interest rate derivative securities', *Review of Financial Studies* **3**.

Hull, J. C. & White, A. (2000), 'Forward rate volatilities, swap rate volatilities, and implementation of the LIBOR market model', *Journal of Fixed Income* **10**(2), 46–62.

Hunter, C., Jäckel, P. & Joshi, M. (2001), 'Getting the drift', *Risk* **14**(7), 81–84.

Hurvich, C. M., Simonoff, J. S. & Tsai, C.-L. (1998), 'Smoothing parameter selection in nonparametric regression using an improved akaike information criterion', *Journal of the Royal Statistical Society: Series B* **60**, 271–293.

Jäckel, P. (2000), Non-recombining trees for the pricing of interest rate derivatives in the BGM/J framework. Quantitative Research Centre, The Royal Bank of Scotland, http://www.quarchome.org/.

Jäckel, P., ed. (2002), *Monte Carlo Methods in Finance*, Chichester: John Wiley & Sons, Ltd.

Jäckel, P. & Rebonato, R. (2003), 'The link between caplet and swaption volatilities in a Brace-Gatarek-Musiela/Jamshidian framework: Approximate solutions and empirical evidence', *Journal of Computational Finance* **6**(4), 41–59.

James, J. & Webber, N. (2000), *Interest Rate Modelling*, Wiley Series in Financial Engineering, John Wiley & Sons.

Jamshidian, F. (1997), 'Libor and swap market models and measures', *Finance and Stochastics* **1**(4), 293–330.

Jarrow, R. A. & Madan, D. B. (1994), 'A characterization of complete security markets on a Brownian filtration', *Mathematical Finance* **1**(3), 31–43.

Jarrow, R. & Protter, P. (2004), 'A short history of stochastic integration and mathematical finance the early years, 1880 – 1970', *IMS Lecture Notes* **45**, 1–17.

Jensen, M. S. & Svenstrup, M. (2005), 'Efficient control variates and strategies for Bermudan swaptions in a LIBOR market model', *The Journal of Derivatives* **12**(4), 20–33.

Joe, S. & Kuo, F. Y. (2003), 'Remark on Algorithm 659: Implementing Sobol's quasirandom sequence genrator', *ACM Transactions on Mathematical Software* **29**(1), 49–57.

Joshi, M. S. (2003), *The Concepts and Practice of Mathematical Finance*, Mathematics, Finance and Risk, Cambridge University Press.

Ju, N. (1998), 'Pricing an American option by approximating its early exercise boundary as a multipiece exponential function', *The Review of Financial Studies* **11**(3), 627–646.

Karatzas, I. (1988), 'On the pricing of American options', *Applied Mathematics and Optimization* **17**(1), 37–60.

Karatzas, I. (1989), 'Optimization problems in the theory of continuous trading', *SIAM Journal on Control and Optimization* **27**(6), 1221–1259.

Karatzas, I. & Shreve, S. E. (1991), *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics 113, second edn, Springer-Verlag.

Kawai, A. (2003), 'A new approximate swaption formula in the LIBOR market model: an asymptotic expansion approach', *Applied Mathematical Finance* **10**, 49–74.

Kerkhof, J. & Pelsser, A. (2002), 'Observational equivalence of discrete string models and market models', *The Journal of Derivatives* **10**(1), 55–61.

Kloeden, P. E. & Platen, E. (1999), *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics 23, Springer-Verlag.

Knez, P. J., Litterman, R. & Scheinkman, J. (1994), 'Exploration into factors explaining money market returns', *Journal of Finance* **49**(5), 1861–1882.

Lamberton, D. & Lapeyre, B. (1996), *Introduction to Stochastic Calculus Applied to Finance*, first edn, Chapman & Hall.

L'Ecuyer, P. & Lemieux, C. (2002), Recent advances in randomized quasi-monte carlo methods, *in* M. Dror, P. L'Ecuyer & F. Szidarovszki, eds, 'Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications', Kluwer Academic Publishers, pp. 419–474.

Li, K.-C. (1991), 'Sliced Inverse Regression for Dimension Reduction', *Journal of the American Statitical Association* **86**, 316–327.

Li, K.-C. (1992), 'On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's Lemma', *Journal of the American Statitical Association* **87**(420), 1025–1039.

Litterman, R. & Scheinkman, J. (1991), 'Common factors affecting bond returns', *Journal of Fixed Income* **1**(1), 54–61.

Loftsgaarden, D. O. & Quesenberry, C. P. (1965), 'A nonparametric estimate of a multivariate density function', *The Annals of Mathematical Statistics* **36**(3), 1049–1051.

Longstaff, F. A., Santa-Clara, P. & Schwartz, E. S. (2001), 'Throwing away a billion dollars: The cost of suboptimal exercise strategies in the swaptions market', *Journal of Financial Economics* **62**(1), 39–66.

Longstaff, F. A. & Schwartz, E. S. (2001), 'Valuing american options by simulation: A simple least-squares approach', *The Review of Financial Studies* **14**(1), 113–147.

Merton, R. R. C. (1973), 'The rational theory of option pricing', *Bell Journal of Economic and Management Science* **4**, 141–183.

Miltersen, K. R., Sandmann, K. & Sondermann, D. (1994), Closed form solutions for term structure derivatives with log-normal interest rates. University of Bonn Discussion Paper B–308, http://ideas.repec.org/p/bon/bonsfb/308.html.

Miltersen, K. R., Sandmann, K. & Sondermann, D. (1997), 'Closed form solutions for term structure derivatives with log-normal interest rates', *Journal of Finance* **52**(1), 409–430.

Moreno, M. & Navas, J. F. (2003), 'On the Robustness of Least-Squares Monte Carlo (LSM) for Pricing American Derivatives', *Review of Derivatives Research* **6**(2), 107–128.

Morini, M. & Webber, N. (2004), An EZI method to reduce the rank of a correlation matrix. Warwick Finance WF04-210 http://www2.warwick.ac.uk/fac/soc/wbs/research/wfri/wpaperseries/.

Musiela, M. & Rutkowski, M. (1997), 'Continuous-time term structure models: Forward measure approach', *Finance and Stochastics* **1**(4), 261–291.

Musiela, M. & Rutkowski, M. (2005), *Martingale Methods in Financial Modelling*, Application of Mathematics: Stochastic Modelling and Applied Probability, second edn, Springer-Verlag.

Myneni, R. (1992), 'The pricing of the American option', *The Annals of Applied Probability* **2**(1), 1–23.

Nadaraya, E. A. (1964), 'On estimating regression', *Theory of Applied Probability* **9**, 141–142.

Naik, P. A., Hagerty, M. R. & Tsai, C.-L. (2000), 'A new dimension reduction approach for data-rich marketing environments', *Journal of Marketing Research* **37**(1), 88–101.

Nelson, C. R. & Siegel, A. F. (1987), 'Parsimonious modelling of yield curves', *Journal of Business* **60**(4), 473–489.

Niederreiter, H. (1992), *Random Number Generation and Quasi-Monte Carlo Methods*, Volume 63 of SIAM CBMS–NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics.

Owen, A. B. (1998), 'Latin supercube sampling for very high-dimensional simulations', *ACM Transactions of Modeling and Computer Simulation* **8**(1), 71–102.

Papageorgiou, A. (2002), 'The Brownian bridge does not offer a consistent advantage in quasi-Monte Carlo integration', *Journal of Complexity* **18**(1), 171–186.

Paskov, S. H. (1997), New methodologies for valuing derivatives, *in* M. A. H. Dempster & S. Pliska, eds, 'Mathematic of Derivative Securities', Cambridge University Press, pp. 545–582.

Paskov, S. H. & Traub, J. F. (1995), 'Faster valuation of financial derivatives', *Journal of Portfolio Management* **22**(1), 113–120.

Pedersen, M. B. (1998), Calibrating Libor market models. SimCorp Financial Research Department Working Paper.

Pietersz, R. & Groenen, P. J. F. (2004), 'Rank reduction of correlation matrices by majorization', *Quantitative Finance* **4**(6), 649–662.

Pietersz, R., Pelsser, A. & van Regenmortel, M. (2004), 'Fast drift approximated pricing in the BGM model', *Journal of Computational Finance* **8**(1), 93–124.

Piterbarg, V. V. (2004), 'Pricing and hedging callable Libor exotics in forward Libor models', *Journal of Computational Finance* **8**(2), 65–117.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2002), *Numerical Recipes in C++: The Art of Scientific Computing*, second edn, Cambridge University Press.

Protter, P. E. (2001), 'A partial introction to financial asset pricing theory', *Stochastic Processes and their Applications* **91**(2), 169–203.

Protter, P. E. (2004), *Stochastic Integration and Differential Equations*, Application of Mathematics: Stochastic Modelling and Applied Probability, second edn, Springer-Verlag.

Rapisarda, F., Brigo, D. & Mercurio, F. (2002), Parameterizing correlations: a geometric interpretation. Product and Business Development Group, Banca IMI Working Paper http://www.fabiomercurio.it/papers.html.

Rebonato, R. (1999*a*), 'On the pricing implications of the joint lognormal assumption for the swaption and cap markets', *Journal of Computational Finance* **2**(3), 57–76.

Rebonato, R. (1999*b*), 'On the simulataneous calibration of multifactor lognormal interest rate models to Black volatilities and to the correlation matrix', *Journal of Computational Finance* **2**(4), 5–27.

Rebonato, R. (2002), *Modern Pricing of Interest Rate Derivatives: the LIBOR Market Model and Beyond*, Princeton University Press.

Rebonato, R. (2004), 'Interest-rate term-structure pricing models: A review', *Proceedings of The Royal Society: Mathematical, Physical and Engineering Sciences* **460**(2043), 667–728.

Rebonato, R. & Jäckel, P. (1999), A methodology for creating a valid correlation matrix for risk management and option pricing purposes. Quantitative Research Centre, The Royal Bank of Scotland, http://www.quarchome.org/.

Rogers, L. C. G. (2002), 'Monte Carlo valuation of American options', *Mathematical Finance* **12**(3), 271–286.

Royden, H. L. (1968), *Real Analysis*, second edn, The Macmillan Company, New York.

Ruppert, D. (2002), 'Selecting the number of knots for penalized splines', *American Statistical Association* **11**(4), 735–757.

Ruppert, D. & Carroll, R. (1997), Penalized regression splines. Cornell University Working Paper.

Samuelson, P. A. (1965), 'Rational theory of warrant pricing', *Industrial Management Review* **6**(2), 13–39.

Sandmann, K. & Sondermann, D. (1993), 'A term structure model and the pricing of interest rate derivatives', *The Review of Futures Markets* **12**(2), 392–423.

Sandmann, K. & Sondermann, D. (1994), On stability of lognormal interest rate models and the pricing of Eurodollar futures. University of Bonn Discussion Paper B–263, http://ideas.repec.org/p/bon/bonsfb/263.html.

Sandmann, K. & Sondermann, D. (1997), 'A note on the stability of lognormal interest rate models and the pricing of Eurodollar futures', *Mathematical Finance* **7**(2), 119–125.

Sandmann, K., Sondermann, D. & Miltersen, K. R. (1994), Closed form term structure derivatives in a Heath-Jarrow-Morton model with log-normal annually compounded interest rates. University of Bonn Discussion Paper B–285, http://ideas.repec.org/p/bon/bonsfb/285.html.

Schlögl, E. (2001), Arbitrage-free interpolation in models of market observable interest rates. University of Technology Working Paper, http://ideas.repec.org/p/uts/rpaper/71.html.

Schoenmakers, J. & Coffey, B. (2003), 'Systematic generation of parametric correlation structures for the LIBOR market model', *International Journal of Theoretical and Applied Finance* **6**(5), 507–519.

Sidenius, J. (2000), 'LIBOR market models in practice', *Journal of Computational Finance* **3**(3), 5–26.

Stentoft, L. (2004), 'Convergence of the least squares Monte Carlo approach to American option valuation', *Management Science* **50**(9), 1193–1203.

Svensson, L. E. O. (1994), Estimating and interpreting forward interest rates: Sweden 1992–94. NBER Working Paper No. 4871, http://www.nber.org/papers/W4871.

Tezuka, S. (2005), 'On the necessity of low-effective dimensions', *Journal of Complexity* **21**(5), 710–721.

Tilley, J. (1993), 'Valuing American options in a path simulation model', *Transactions of the Society of Actuaries* **45**, 83–104.

Tsitsiklis, J. N. & Roy, B. V. (2001), 'Regression methods for pricing complex American-style options', *IEEE Transactions on Neural Networks* **12**(4), 694–703.

van Moerbeke, P. L. J. (1976), 'On optimal stopping and free boundary problem', *Archive for Rational Mechanics and Analysis* **60**(2), 101–148.

von Neumann, J. (1950), *Functional Operators: The Geometry of Orthogonal Spaces*, Vol. second, Princeton University Press.

Watson, G. S. (1964), 'Smooth regression analysis', *Sankhyā, Series A* **26**, 359–372.

Weigel, P. (2004), 'A note on optimal calibration of the LIBOR market model to the correlations', *The Journal of Derivatives* **12**(2), 43–50.

Wu, L. (2002), 'Fast at-the-money calibration of the Libor market model using Lagrange multipliers', *Journal of Computational Finance* **6**(2), 39–78.

Zhang, Z. & Wu, L. (2003), 'Optimal low-rank approximation to a correlation matrix', *Linear Algebra and its Applications* **364**, 161–187.