A Numerical Study of One-factor Interest Rate Models

by

Zhong Ge

A thesis submitted in conformity with the requirements
for the degree of Master of Science
Graduate Department of Computer Science
University of Toronto

Abstract

A Numerical Study of One-factor Interest Rate Models

Zhong Ge

Master of Science

Graduate Department of Computer Science

University of Toronto

1998

This thesis is concerned with a numerical study of one-factor interest rate models. First the basics of option pricing theory are outlined together with a simple derivation of the Heath-Jarrow-Morton (HJM) models. Then we discuss the term-structure-consistent models and their relation with the HJM models. We show that adaptive tree methods, namely, tree methods that include adaptive branching rules to incorporate the mean-reverting properties of the interest rates, are very stable and efficient. We apply the tree methods to five models: the Black-Karasinski model, the Cox-Ingersoll-Ross (CIR) model, the cubic-variance model, the Ho-Lee model and the Hull-White model. Using the market data from the US market near the end of June, 1997, we test and compare the five models. Our results show that all the models except that of Ho-Lee produce similar prices for at-the-money or in-the-money options, but give very different prices for deep-out-of-the-money options. In particular, the more sensitive the volatility of a model is to the level of the interest rates, the higher prices the model produces for deep-out-of-the-money put options. Finally we look at the issue of negative interest rates in the Hull-White model and our results suggest that it is insignificant for today's US market.

Acknowledgements

It is my pleasure to thank Professor Ken Jackson for his encouragement and support, for making comments on the earlier drafts of the thesis and for the weekly meetings in which the plan for the thesis was discussed. I want to thank Professor Christina C. Christara and Dr. Mark Reimers for taking pains in reading the thesis and for the comments and suggestions. I also want to thank Yidong Liu for providing some of the data that was used in the thesis and for numerous conversations, Dr. Izzy Nelken for mentioning the floors at negative rates, Dr. Wulin Suo for a discussion about the negative interest rates in the Hull-White model and Dr. Charles Zhu who explained some tricky points about yield curves. I also want to thank all those with whom I had conversations about interest rate models.

Finally, I want to thank my wife, Fen, for her support and for cheering me up constantly.

Contents

Chapter 1

Introduction

Interest rate options are options whose payoffs are dependent in some way on interest rates. While the Black-Scholes model has been widely used for equity options, it does not work well for a long-term bond option, for example, because of the pull-to-bar phenomenon of the underlying bond price (i.e. the price of a bond at its maturity must equal its face value). For plain vanilla options, it is common to use Black's formula (see Appendix A), which makes the assumption that a underlying variable at the maturity is lognormally distributed but makes no assumption about the movement of the interest rates. For more complicated options, however, one needs a model that describes the possible movement of the interest rates, namely, a yield curve model. So far many yield curve models have been proposed; however, there is little consensus about which one is more suitable for pricing interest rate options. Some of the yield curve models that will be studied in this thesis are

- the Black-Karasinski model

$$d[log(r)] = (\theta - a \cdot log(r)) \cdot dt + \sigma \cdot dZ$$

- the Cox-Ingersoll-Ross model

$$dr = (a - \theta \cdot r) \cdot dt + \sigma \cdot \sqrt{r} \cdot dZ$$

- the cubic variance model

$$dr = (a - \theta \cdot r) \cdot dt + \sigma \cdot r^{3/2} \cdot dZ$$

- the Ho-Lee model

$$dr = \theta \cdot dt + \sigma \cdot dZ$$

- the Hull-White model

$$dr = (\theta - a \cdot r) \cdot dt + \sigma \cdot dZ$$

where r is the instantaneous rate, Z is a Brownian motion, a, σ and θ are parameters independent of r (see Chapter 3 for more details). At present, the market practice is that different institutions use different models to price the same products, and this situation will likely continue in the near future (see Section 3.3.5 for a more detailed discussion). It is our view that this constitutes a model risk,[1]which can be alleviated by a comparison of different models. However, many of the proposed interest rate models do not have closed-form solutions, thus one has to resort to numerical methods.

Unlike many problems in engineering, interest rate models have no boundary conditions. Also unlike the Black-Scholes model, interest rates models are time-dependent (at least in the term-structure consistent approach). Furthermore, one of the inputs to the models, the initial yield curves, can be a non-smooth function, which affects the accuracy of the numerical solutions. These special features make interest rates models more difficult to implement than the Black-Scholes model.

This thesis consists of five chapters besides this introduction. Chapters 2 and 3 are of survey nature, the rest chapters are based on our own research.

In Chapter 2, we review the basics of option pricing theory. In particular, we discuss the following key ingredients:

- market price of risk;
- change of measure and risk-neutral evaluation;

[1]Model risk refers to the danger in the use of models. It was propelled into common parlance when NatWest Markets announced losses of £90 million on mispriced interest rate options in March 1997.

- change of numeraire;
- the Black-Scholes model;
- the Heath-Jarrow-Morton (HJM) models.

This chapter is to provide a theoretic foundation for later chapters. Together with Chapter 3, it justifies the term-structure consistent approach (see Chapter 3) that is used throughout the thesis.

In Chapter 3, we discuss term structure consistent models. In particular, we

- compare the equilibrium approach and the term-structure consistent approach;
- introduce the five interest rate models mentioned in the beginning of this chapter;
- discuss the relation between the term-structure consistent models and the HJM models.

Chapter 4 is mainly concerned with numerical methods for interest rate models. Since interest rate models have no boundary conditions, it is our view that tree methods are more convenient. However, tree methods can be unstable and, as a result, the numerical solutions need not converge to the true solutions. This problem is further complicated by the fact that the interest rate models are both space (interest rate) and time dependent.

In order to improve the stability of the tree methods, we use adaptive tree methods, namely tree methods with adaptive branching rules. Both theoretic analysis and extensive numerical experiments show that adaptive tree methods are very stable. Moreover, as adaptive tree methods incorporate the mean-reverting property [2] and use fewer nodes, they are more efficient. Furthermore, since adaptive tree methods can be applied to a wide range of interest rate models, we are able to implement an object-oriented framework for interest rate models, which makes the addition/change of a model easy.

The aim of Chapter 5 is twofold. We first survey different approaches to calibration, then test and compare the five models using market data (caplet prices and yield curves

[2]This refers to the property of the interest rates that if they are high, they will tend to become lower, and if they are low, they will tend to become higher.

in the U.S. market near the end of June, 1997). Though the testing is limited by the sample size, we hope that it will lead to suggestions for further study.

Some authors suggest that interest rate models, once calibrated, produce similar prices (see, e.g. [Lochoff]). It is our view that they only address the issue partially, since they consider only a small number of models. We revisit the topic and expand the scope in three ways:

1. We look at a broader range of interest rate models.

2. We use the volatilities implied from the U.S. market.

3. We consider a broader range of products including deep-out-of-the-money options.[3]

Our major findings are as follows:

1. All the models, except that of Ho-Lee, produce similar prices for the at-the-money options or in-the-money options. This is consistent with the conclusion of [Hull-White(1993)], but significantly different from the conclusion of [Uhrig-Walter] that uses the historical volatility approach in the calibration.

2. The Ho-Lee model generates a price for the at-the-money option that is much higher than those generated by the other models. This is somewhat surprising as it is considered easy to price at-the-money options.

3. The Black-Karasinski model and the CIR model produce similar prices for both in-the-money and out-of-the-money options. This somehow confirms the conclusion of Canabarro [Canabarro], who compares the CIR model and the Black-Derman-Toy model (which is similar to the Black-Karasinski model), and Lochoff [Lochoff].

4. However, even for slightly out-of-the-money options, both the Hull-White model and the cubic-variance model produce prices that are quite different from thoses generated by the Black-Karasinski and the CIR models. This is interesting as the

[3] Roughly speaking, an out-of-the-money option would lead to a negative cash flow if it were exercised immediately, an at-the-money option would lead to a zero cash flow if it were exercised immediately, and an in-the-money option would lead to a positive cash flow if it were exercised immediately.

cubic-variance model is one of the best models to describe the historic movement of the U.S. interest rates (see [Chan]).

5. Except the Ho-Lee model, for deep-out-of-the-money put options, the more sensitive the volatility of a model is to the level of the interest rate, the higher the prices the model produces.

The Hull-White model has often been criticized for producing negative interest rates (see, e.g. [Rogers]). In Chapter 6, we look at this issue in two ways. We first compute the probability that the interest rates in the model will become negative using the implied volatilities obtained in Chapter 5. Then we modify the Hull-White model by truncating the negative interest rates to zero and compare it with the original model. In both cases we find that the negative interest rates in the Hull-White model are insignificant for today's U. S. market.

Future work

In the calibration of interest rate models we use an optimization package that implements a version of the Broyden-Fletcher-Goldfarb-Shanno algorithm. It turns out that this package is not very suitable for our problem. It would be interesting to improve the optimization procedure in the future.

A challenging problem is to develop efficient and stable numerical methods for two-factor interest rate models. Our experience with one-factor models suggests the following might be useful:

- Tree methods are more convenient, especially for two-factor models, since there are no boundary conditions.

- Adaptive tree methods can greatly increase both the efficiency and the stability of the numerical methods. Our preliminary test shows that the efficiency will increase by 6-9 times with the use of adaptive tree methods.

- An object-oriented approach will make the addition/change of a model easy.

- Empirically comparing multiple interest rate models will alleviate the model risk.

Chapter 2

Basic Option Pricing Theory

2.1 Introduction

In this section, we will review the basic option pricing theory. In particular, we will discuss its following ingredients:

- market price of risk
- change of measure (Girsanov's theorem)
- change of numeraire

Though other ingredients are as important, the above are sufficient for our purpose of understanding, developing and implementing stochastic models for pricing interest rate options.

Notation

- S : the price of a security;
- t : the time variable;
- r : the instantaneous risk-free interest rate;
- Z : the Brownian motion representing the uncertainty affecting the security;

2.1.1 Brownian motion

The risk of an investment is the uncertainty associated with its future value. It often is modeled by a *Brownian motion*, a one-parameter family of random variables $Z(t)$ with the following properties:

1. The initial value $Z(0) = 0$.

2. For $t_1 < t_2 < \cdots < t_n$, the increments

$$Z(t_2) - Z(t_1), Z(t_3) - Z(t_2), \cdots, Z(t_n) - Z(t_{n-1}),$$

   are independent and normally distributed.

3. The random variable $Z(t)$ is normalized so that $E(Z(t)) = 0$, which is often simply written as

$$E(dZ) = 0 \tag{2.1}$$

   where E is the expectation operator.

4. The variance of the increment $Z(t_2) - Z(t_1)$, for $t_2 > t_1$, is

$$Var((Z(t_2) - Z(t_1)) = t_2 - t_1$$

   which is often written as $E((dZ)^2) = dt$.

2.1.2 The law of one price

The pricing of derivatives is derived from the non-arbitrage condition, which in turn is based on the instant law of one price. In this subsection we will explain the law of one price and its instant version.

In a deterministic economy (meaning that the price of any security is predetermined) without transaction costs, two identical securities will sell at the same price, no matter how they are obtained. Otherwise, one can buy the security at the lower price and sell it at the higher price, making a riskless profit.

Moreover, in a deterministic economy the price of any security (if it ever exists) will grow at the risk-free rate. That is, let $S(t)$ be the price of a security at time t, $r(t_1, t_2)$ the risk-free interest rate from time t_1 to t_2, then

$$\frac{S(t_2) - S(t_1)}{S(t_1)} = r(t_1, t_2).$$

It is easy to see why this equality holds. If

$$\frac{S(t_2) - S(t_1)}{S(t_1)} > r(t_1, t_2),$$

then every one would prefer to hold the security, knowing that it would earn a higher return than the risk-free deposit. On the other hand, if

$$\frac{S(t_2) - S(t_1)}{S(t_1)} < r(t_1, t_2),$$

then every one would prefer to hold the risk-free deposit, knowing that it would earn a higher return than the security.

Instant law of one price

For the same reason as above, if the economy is not deterministic, but at some time t_1, all the risk factors affecting the economy disappear, then the instantaneous return on any security at t_1 is $r(t_1, t_1)$. Otherwise, if we assume

$$\frac{S(t_1 + dt) - S(t_1)}{S(t_1)} > r(t_1, t_1),$$

then every one would prefer to hold the security, knowing that it would earn a higher return than the risk-free deposit. On the other hand, if

$$\frac{S(t_1 + dt) - S(t_1)}{S(t_1)} < r(t_1, t_1),$$

then every one would prefer to hold the risk-free deposit, knowing that it would earn a higher return than the security.

By the same token, if at time t_1, all the risk factors affecting a portfolio disappear instantaneously, then the instantaneous return of the portfolio will be equal to the risk-free interest rate. This fact plays an important role in the derivation of the Black-Scholes formula for options.

2.2 Market price of risk

Notation

- μ: the expected instantaneous return on the security:
- σ^2 : the variance of the instantaneous return of the security;

We assume that in the economy each security is affected by only one risk factor represented by a Brownian motion Z

$$\frac{dS}{S} = \mu \cdot dt + \sigma \cdot dZ \tag{2.2}$$

Define the *market price of risk* to be the excess return per volatility

$$\lambda = \frac{\mu - r}{\sigma}. \tag{2.3}$$

The key property of the market price of risk is that two securities driven by the same risk factor Z must have the same market price of risk, as the following result shows.[1]

Lemma 2.1 *Suppose two security prices, S_1, S_2 are driven by the same Brownian motion*

$$\frac{dS_1}{S_1} = \mu_1 \cdot dt + \sigma_1 \cdot dZ, \tag{2.4}$$

$$\frac{dS_2}{S_2} = \mu_2 \cdot dt + \sigma_2 \cdot dZ, \tag{2.5}$$

then their market prices of risk must be the same

$$\frac{\mu_1 - r}{\sigma_1} = \frac{\mu_2 - r}{\sigma_2}. \tag{2.6}$$

We prove this result for the sake of completeness. Before proving the lemma, we discuss the basic idea in the proof. Since S_1 and S_2 share the same source of uncertainty, we can build a portfolio from S_1 and S_2 that eliminates the risk temporarily. Then the

[1]This result can be found in [Hull]. It justifies the estimation of the market price of risk from the initial bond prices in Chapter 4.

portfolio must grow at the risk-free rate temporarily by the instant law of one price. This results in an equation from which (2.6) follows.

Proof of Lemma 2.1. Suppose at time t_0, the prices of S_1 and S_2 are S_{10}, S_{20} respectively. Construct a portfolio

$$P = a \cdot S_1 + b \cdot S_2.$$

where a, b are two constants to be determined later. The instantaneous change of the portfolio at time t_0 is

$$\begin{aligned} dP &= a \cdot dS_1 + b \cdot dS_2 \\ &= a \cdot (\mu_1 \cdot S_{10} \cdot dt + S_{10} \cdot \sigma_1 \cdot dZ) + b \cdot (\mu_2 \cdot S_{20} \cdot dt + S_{20} \cdot \sigma_2 \cdot dZ) \\ &= (a \cdot \mu_1 \cdot S_{10} + b \cdot \mu_2 \cdot S_{20}) \cdot dt + (a \cdot S_{10} \cdot \sigma_1 + b \cdot S_{20} \cdot \sigma_2) \cdot dZ \end{aligned}$$

The stochastic term above is

$$(a \cdot S_{10} \cdot \sigma_1 + b \cdot S_{20} \cdot \sigma_2) \cdot dZ. \tag{2.7}$$

If, at time t_0, (2.7) is zero, then, by the instant law of one price, the expected instantaneous return of the portfolio should be equal to the risk-free interest rate, i.e.

$$\begin{aligned} a \cdot \mu_1 \cdot S_{10} + b \cdot \mu_2 \cdot S_{20} &= r \cdot P(t_0) \\ &= r \cdot (a \cdot S_{10} + b \cdot S_{20}) \end{aligned} \tag{2.8}$$

Now we choose constants a, b such that (2.7) is zero, i.e.

$$a \cdot S_{10} \cdot \sigma_1 + b \cdot S_{20} \cdot \sigma_2 = 0 \tag{2.9}$$

In particular, we take

$$\begin{aligned} b &= \quad S_{10} \cdot \sigma_1 \\ a &= -S_{20} \cdot \sigma_2 \end{aligned} \tag{2.10}$$

Since (2.9) implies (2.8) (as it implies that (2.7) is zero, which in turn implies (2.8)), inserting (2.10) into (2.8), we obtain

$$(\mu_1 - r) \cdot S_{10} \cdot S_{20} \cdot \sigma_2 = (\mu_2 - r) \cdot S_{20} \cdot S_{10} \cdot \sigma_1$$

or

$$\frac{\mu_2 - r}{\sigma_2} = \frac{\mu_1 - r}{\sigma_1}$$

This proves the lemma.

As a result of Lemma 2.1, the market price of risk is independent of the price of any security but depends on the risk represented by dZ and possibly other variables that are not the price of a security (i.e. an interest rate or the time variable).

2.2.1 Application: the Black-Scholes equation

Assume the price of a stock follows

$$\frac{dS}{S} = \mu \cdot dt + \sigma \cdot dZ.$$

Suppose f is an option on S, $f = f(S,t)$. By Ito's lemma (see, for example, [Hull]), f satisfies

$$\frac{df}{f} = \mu_f \cdot dt + \sigma_f \cdot dZ$$

where

$$\mu_f = \frac{1}{f} \cdot \left(\frac{\partial f}{\partial t} + S \cdot \mu \cdot \frac{\partial f}{\partial S} + \frac{S^2 \cdot \sigma^2}{2} \cdot \frac{\partial^2 f}{\partial S^2}\right) \tag{2.11}$$

$$\sigma_f = \frac{S \cdot \sigma}{f} \cdot \frac{\partial f}{\partial S} \tag{2.12}$$

Since S and f share the same source of uncertainty, by Lemma 2.1,

$$\frac{\mu - r}{\sigma} = \frac{\mu_f - r}{\sigma_f}. \tag{2.13}$$

From (2.11), (2.12) and (2.13), we obtain the celebrated Black-Scholes equation

$$\frac{\partial f}{\partial t} + r \cdot S \cdot \frac{\partial f}{\partial S} + \frac{S^2 \cdot \sigma^2}{2} \cdot \frac{\partial^2 f}{\partial S^2} = r \cdot f$$

2.3 Change of measure

In the last subsection, we have seen that the market price of risk is the same for all the securities driven by the same risk factor. In this subsection we will discuss a technique, called *change of measure*, that allows us to eliminate the market price of risk.

First we explain this technique heuristicly. On the surface it is just a substitution of the random variable $Z(t)$ by a new random variable $Z_1(t)$. From the definition of the market price of risk (2.3), we have $\mu = \lambda \cdot \sigma + r$. Thus we can rewrite (2.2) as

$$\frac{1}{S} \cdot dS = r \cdot dt + \sigma \cdot d(Z(t) + \int_0^t \lambda \cdot dt). \tag{2.14}$$

Define the new random variable $Z_1(t)$ by

$$Z_1(t) = Z(t) + \int_0^t \lambda \cdot dt, \tag{2.15}$$

then (2.14) can be written as

$$\frac{1}{S} \cdot dS = r \cdot dt + \sigma \cdot dZ_1. \tag{2.16}$$

The process in (2.16) is a simple one: its drift is equal to the risk-free interest rate, and hence the market price of risk is zero:

$$\lambda = \frac{r - r}{\sigma} = 0.$$

However, Z_1 is *not* a Brownian motion, since

$$E(dZ_1) = \lambda \neq 0$$

which violates the necessary condition (2.1) for a Brownian motion.

Fortunately, the celebrated theorem of Girsanov deals with this problem. It says that *after a change of probability measure, Z_1 becomes a Brownian motion.*

Lemma 2.2 (Girsanov's Theorem) *Introduce*

$$W(t) = exp(-\int_0^t \lambda(u) \cdot dZ(u) - \frac{1}{2} \cdot \int_0^t \lambda(u)^2 \cdot du).$$

Use $W(t)$ as a Radon-Nikodym derivative to define a new probability measure, i.e.,

$$\begin{aligned} P_1(A) &= \int_A W(t) \cdot dP, && \textit{for all } A \in \mathcal{F}(t) \\ E_1[Y] &= E[W(t) \cdot Y] && \textit{for all random variables } Y \end{aligned}$$

where $\mathcal{F}(t)$ is an accompanying filtration generated by the Brownian motion $Z(t)$. Then, under the new measure P_1, the process Z_1 introduced in (2.15) is a Brownian motion.

Often a change of measure is called a change of worlds. In particular, if we change to a measure for which the market price of risk $\lambda = 0$, then the new world is called the *risk-neutral world.*

Of course, we need not change to the risk-neutral world. In fact, we may change to any convenient world. For example, we may introduce a new random variable Z_2

$$Z_2(t) = Z(t) + \int_0^t \delta \cdot dt$$

where δ is a function of time. Again Z_2 is not a Brownian motion; however, after a change of measure, Z_2 becomes one. In the old (real) world, the process is

$$dS = (r + \lambda_{old} \cdot \sigma) \cdot dt + \sigma \cdot dZ$$

where λ_{old} denotes the market price of risk in the real world. In the new world, the process for the security can be written as

$$dS = (r + \lambda_{old} \cdot \sigma - \delta \cdot \sigma) \cdot dt + \sigma \cdot dZ_2$$

and the market price of risk is

$$\begin{aligned} \lambda_{new} &= \frac{(r + \lambda_{old} \cdot \sigma - \delta \cdot \sigma) - r}{\sigma} \\ &= \lambda_{old} - \delta \end{aligned}$$

Thus, as in the old (real) world, the market price of risk in the new world is the same for all the securities driven by the same risk factor.

To summarize, the market price of risk is the same for all the securities driven by the same risk factor. However, once the measure is changed, the market price of risk is changed too. In particular, in the risk-neutral world, the market price of risk is zero.

2.3.1 Application: the HJM model of interest rates

We first introduce some more notation.

Notation

- $P(t,T)$: the price of a *discount bond* at time t maturing at T (i.e. it pays off \$1 at time T);
- $\mathcal{F}(t,T)$: the *instantaneous forward rate,*
$$\mathcal{F}(t,T) = -\frac{\partial P(t,T)}{P\cdot\partial T} = -\frac{\partial[lnP(t,T)]}{\partial T};$$
- $r(t) = \mathcal{F}(t,t)$: the risk-less instantaneous rate;
- Θ: the set of market variables affecting the bond price (e.g. interest rates).
- $\mu = \mu(t,T,\Theta)$: the expected instantaneous return on the bond $P(t,T)$. We will simply write $\mu(t,T,\Theta)$ as μ if there is no confusion.
- $\sigma = \sigma(t,T,\Theta)$ the variance of the instantaneous return. We will simply write $\sigma(t,T,\Theta)$ as σ if there is no confusion.

In the following, we will write (t,T,Θ) simply as (t,T).

Since $P(T,T) = 1$,
$$\sigma(T,T) = 0 \tag{2.17}$$
which says that, as the maturity of the discount bond is approached, the bond volatility will decrease to zero.

For $t < T$, we assume $P(t,T)$ satisfies
$$\frac{dP(t,T)}{P(t,T)} = \mu\cdot dt + \sigma\cdot dZ \tag{2.18}$$

Using Ito's lemma, we obtain from (2.18)
$$d[ln(P(t,T)] = (\mu - \sigma^2/2)\cdot dt + \sigma\cdot dZ.$$

Taking the derivative of the above equation with respect to T, we obtain [2]
$$-d\{\mathcal{F}(t,T)\} \quad = d\{\frac{\partial}{\partial T}[ln(P)]\}$$

[2] Here we assume that we can exchange the order of the operators $\partial/\partial T$ and d

$$
\begin{aligned}
&= \frac{\partial}{\partial T}\{d[ln(P)]\} \\
&= \frac{\partial}{\partial T}\{\mu - \sigma^2/2\} \cdot dt + \frac{\partial \sigma}{\partial T} \cdot dZ \\
&= (\frac{\partial \mu}{\partial T} - \sigma \cdot \frac{\partial \sigma}{\partial T}) \cdot dt + \frac{\partial \sigma}{\partial T} \cdot dZ
\end{aligned}
$$

By the definition of the market price of risk,

$$\frac{\mu - r}{\sigma} = \lambda,$$

and the fact that λ is independent of T (i.e. it is independent of the maturity of a bond), we obtain

$$-d[\mathcal{F}(t,T)] = (\lambda \cdot \frac{\partial \sigma}{\partial T} - \sigma \cdot \frac{\partial \sigma}{\partial T}) \cdot dt + \frac{\partial \sigma}{\partial T} \cdot dZ \tag{2.19}$$

For simplicity, we write $-\partial\sigma(t,T)/\partial T$ as $b(t,T)$. From (2.17),

$$
\begin{aligned}
\sigma(t,T) &= -\int_t^T b(t,\tau) \cdot d\tau + \sigma(t,t) \\
&= -\int_t^T b(t,\tau) \cdot d\tau
\end{aligned}
$$

Thus (2.19) can be written as

$$d[\mathcal{F}(t,T)] = b(t,T) \cdot (\lambda(t) + \int_t^T b(t,\tau) \cdot d\tau) \cdot dt + b(t,T) \cdot dZ$$

This is the celebrated *Heath-Jarrow-Morton (HJM) equation*. In general b is a function of t, T, Θ.

In the risk-neutral world, $\lambda = 0$. Thus the HJM equation simplifies to

$$d[\mathcal{F}(t,T)] = b(t,T) \cdot (\int_t^T b(t,\tau) \cdot d\tau) \cdot dt + b(t,T) \cdot dZ_1$$

2.4 Change of numeraire

Change of numeraire is the change of the units in the valuation of assets/securities. For example, instead of saying that an apple is worth 2 dollars, we can say it is worth 1 pear. In this case, we choose the price of a pear as the numeraire and denominate the values of other assets by the numeraire. Sometimes it is more convenient to choose the price of an asset (i.e. a discount bond) as the numeraire.

Suppose two security prices f and g follow the stochastic processes

$$\frac{df}{f} = \mu_f \cdot dt + \sigma_f \cdot dZ \tag{2.20}$$

$$\frac{dg}{g} = \mu_g \cdot dt + \sigma_g \cdot dZ \tag{2.21}$$

respectively. Choose g as the numeraire and the relative price of f denominated by g is f/g. We will derive the stochastic process followed by f/g, following [Hull-White(1997)].

From (2.20), (2.21) and Ito's lemma, we have

$$d[ln(f)] = (\mu_f - \frac{\sigma_f^2}{2}) \cdot dt + \sigma_f \cdot dZ$$

$$d[ln(g)] = (\mu_g - \frac{\sigma_g^2}{2}) \cdot dt + \sigma_g \cdot dZ.$$

Subtracting the second equation from the first one, we obtain

$$d[ln(f/g)] = (\mu_f - \mu_g - \frac{\sigma_f^2 - \sigma_g^2}{2}) \cdot dt + (\sigma_f - \sigma_g) \cdot dZ.$$

From the market price of risk,

$$\frac{\mu_f - r}{\sigma_f} = \frac{\mu_g - r}{\sigma_g} = \lambda$$

we have

$$d[ln(\frac{f}{g})] = (\lambda - \frac{\sigma_f + \sigma_g}{2}) \cdot (\sigma_f - \sigma_g) \cdot dt + (\sigma_f - \sigma_g) \cdot dZ.$$

Applying Ito's lemma again, we obtain

$$\frac{d[f/g]}{f/g} = (\lambda - \sigma_g) \cdot (\sigma_f - \sigma_g) \cdot dt + (\sigma_f - \sigma_g) \cdot dZ. \tag{2.22}$$

After changing to a world in which the market price of risk satisfies

$$\lambda = \sigma_g \tag{2.23}$$

then the drift term in (2.22) disappears,

$$\frac{d[f/g]}{f/g} = (\sigma_f - \sigma_g) \cdot dZ_g$$

that is, f/g is a martingale ([Baxter-Rennie]). Here Z_g is the Brownian motion in the new world in which $\lambda = \sigma_g$.

We call the world in which the relation (2.23) holds as the *forward risk-neutral world with respect to g*. In this world, by the property of a martingale,

$$\frac{f(0)}{g(0)} = E_g\left(\frac{f(T)}{g(T)}\right) \tag{2.24}$$

where E_g is the expectation in the forward risk-neutral world with respect to g.

Chapter 3

One-Factor Interest Rate Models

In this section we will introduce several one-factor interest rate models that are often used.

Notation.

- Z : a standard Brownian motion driving the interest rate and the bond prices;
- r: the instantaneous interest rate;
- $P(t, T)$: the price at time t of a discount bond that matures at T;
- λ: the market price of risk

3.1 Introduction

3.1.1 The equilibrium approach versus the term structure consistent approach

The *equilibrium approach* starts from a description of the underlying economy. It makes assumptions about the stochastic evolution of state variables in the economy and about the preferences of a representative investor. Equilibrium conditions are used to specify the interest rate and the prices of all derivatives.

An alternative approach is based on no-arbitrage considerations. This approach makes assumptions about the stochastic evolution of interest rates, and derives the prices of all derivatives by imposing the condition that there be no arbitrage opportunities in the economy. This approach is used in Chapter 2.

In spite of their apparent difference, Dybvig and Ross [Dybvig-Ross] conclude that the two approaches are equivalent. For example, no-arbitrage models can always be embedded in an equilibrium model, although the utility function may be state dependent.

In the traditional treatment of both approaches, the interest rate models are time-independent, i.e. all the parameters in the models are constant.

In contrast, the *term-structure consistent approach* [1] allows a parameter in the models to be a function of time to fit the models to the initial term structure. Thus in this approach the current bond prices are used as the input to the models, while in the the traditional equilibrium approach they are the output of the models.

The main advantages of the term structure consistent approach are:

1. The models can reproduce the prices of discount bonds correctly. This is desirable since most interest rate options are a portfolio of options on discount bonds, and discount bonds are often used as the hedging instruments.

2. In this approach, the one-factor models are special cases of the HJM models (see Section 3.2.3).

3. The market price of risk can be estimated from the initial bond prices (see Section 3.2.2 and also [Uhrig-Walter]).

In this chapter we will mainly discuss term structure consistent models.

3.2 General one-factor models

3.2.1 Some often-used models

Many different one-factor interest rate models have been proposed, all of which can be written in the form

$$dr = b \cdot dt + c \cdot dZ \tag{3.1}$$

[1] In [Hull] this approach is called the non-arbitrage approach, which should not be confused with the terminology used here.

where $b = b(t, r)$, $c = c(t, r)$ are functions of time t, interest rate r and possibly some other parameters.

Some well-known examples of one-factor interest rate models are listed in Table 3.1.

Table 3.1. A list of often-used one-factor models

Name of Model	Evolution Equation	Parameters
Black-Karasinski	$dln(r) = (\theta(t) - a \cdot ln(r))dt + \sigma \cdot dZ$	$b = r \cdot \theta(t) - a \cdot ln(r) \cdot r$ $+r \cdot \sigma^2/2$, $c = \sigma \cdot r$
CIR	$dr = (a - \theta(t) \cdot r)dt + \sigma \cdot \sqrt{r} \cdot dZ$	$b = a - \theta(t) \cdot r$, $c = \sigma\sqrt{r}$
Cubic variance	$dr = (a - \theta(t) \cdot r)dt + \sigma \cdot r^{3/2} \cdot dZ$	$b = a - \theta(t) \cdot r$, $c = \sigma r^{3/2}$
Ho-Lee	$dr = \theta(t) \cdot dt + \sigma \cdot dZ$	$b = \theta(t)$, $c = \sigma$
Hull-White	$dr = (\theta(t) - a \cdot r)dt + \sigma \cdot dZ$	$b = \theta(t) - a \cdot r$, $c = \sigma$

One of the main differences among these models is their assumptions on the volatility of the interest rates. The Ho-Lee and Hull-White models assume the volatility is independent of r, and are often called "normal models." The CIR model specifies that the volatility is proportional to $\sqrt{r}$, and is often called a "square root model." The BK model postulates the volatility is proportional to r, and is often called a "lognormal model." Thus the volatility in the BK model is the most sensitive to the level of r, while those in the Ho-Lee and Hull-White models are the least.

Figures 3.1 - 3.4 show some random sample paths generated by the various models. They show that in a model whose volatility is sensitive to the interest rate r, r will likely stay low, but once r becomes high, it can become much higher.

3.2.2 Equation for pricing derivatives

We shall derive an equation for interest rate derivatives from the interest rate process (3.1).

Let $f = f(r, t)$ be the price of an interest rate derivative. Applying Ito's lemma to

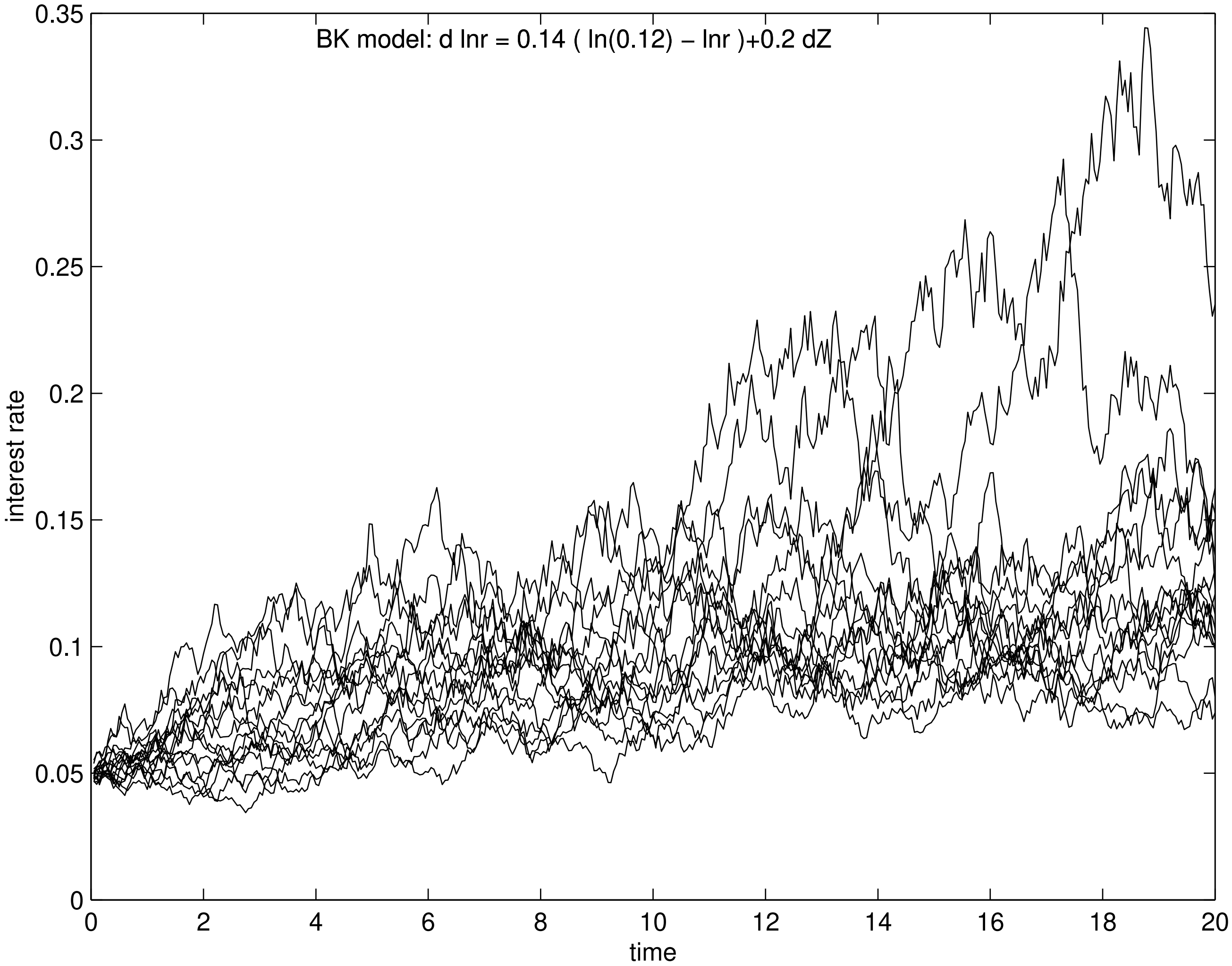


Figure 3.1: Some random sample paths generated by the Black-Karasinski model

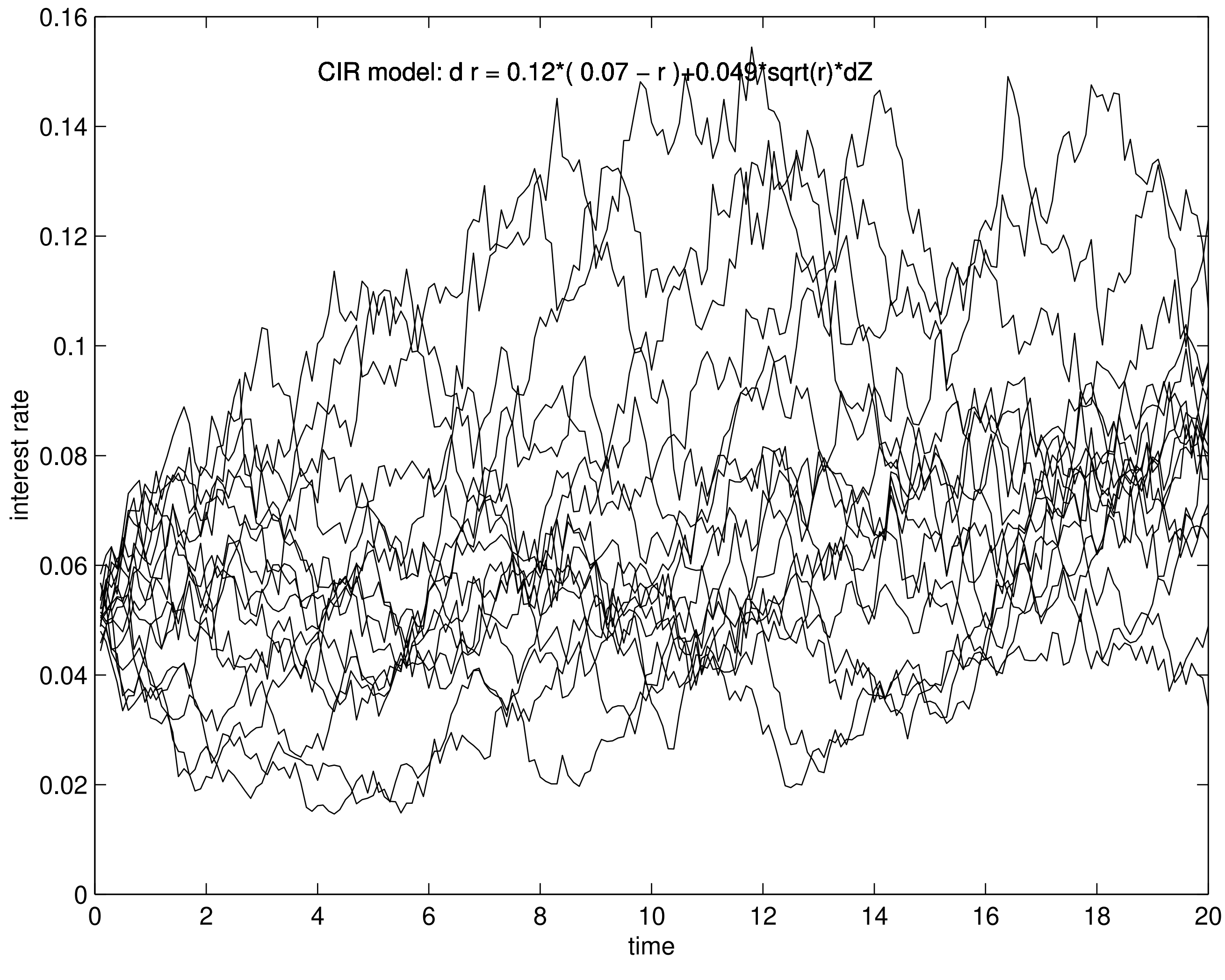


Figure 3.2: Some random sample paths generated by the CIR model

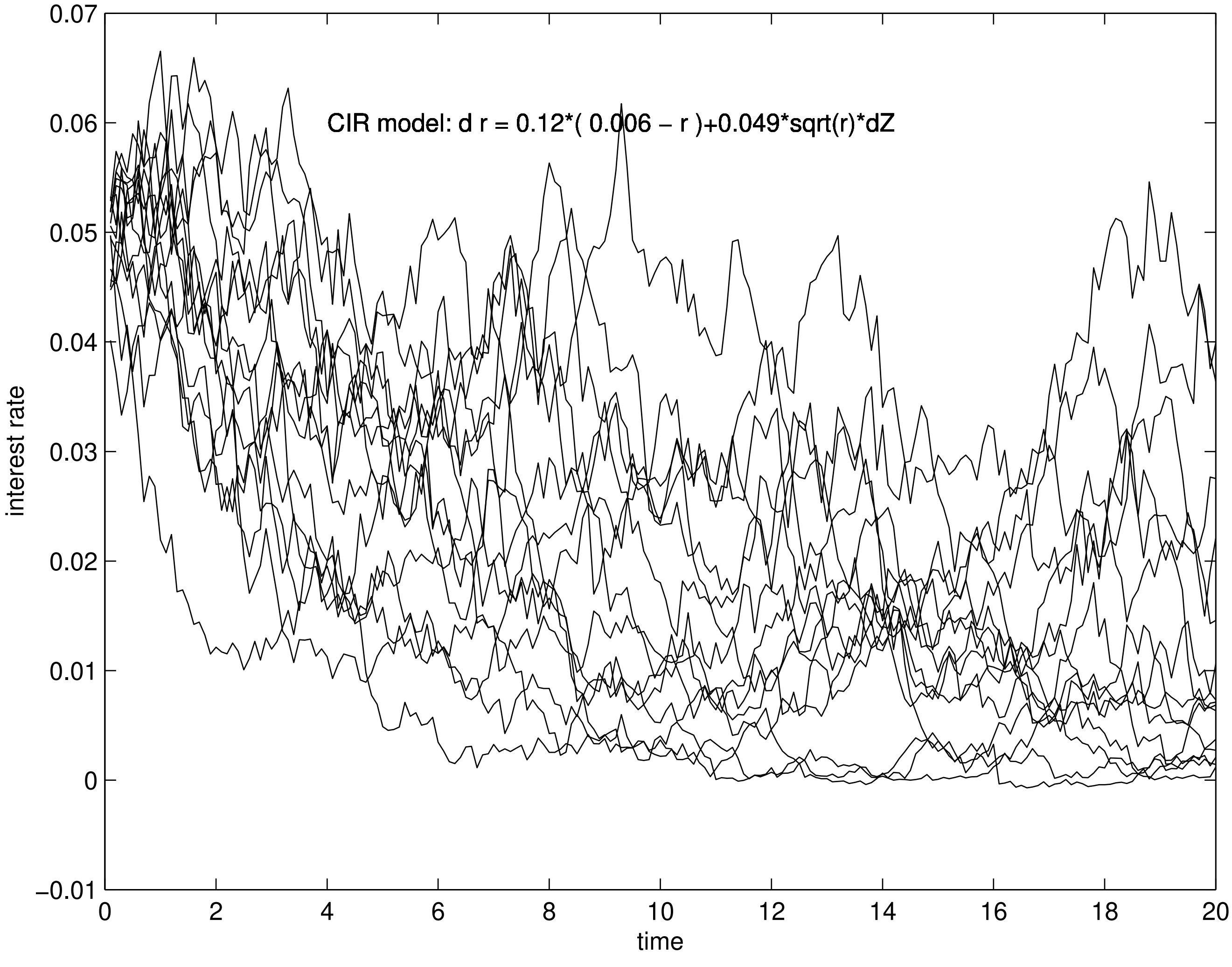


Figure 3.3: Some random sample paths generated by the CIR model. This plot shows that, when the condition $a > \sigma^2/2$ is violated, the interest rate can become zero.

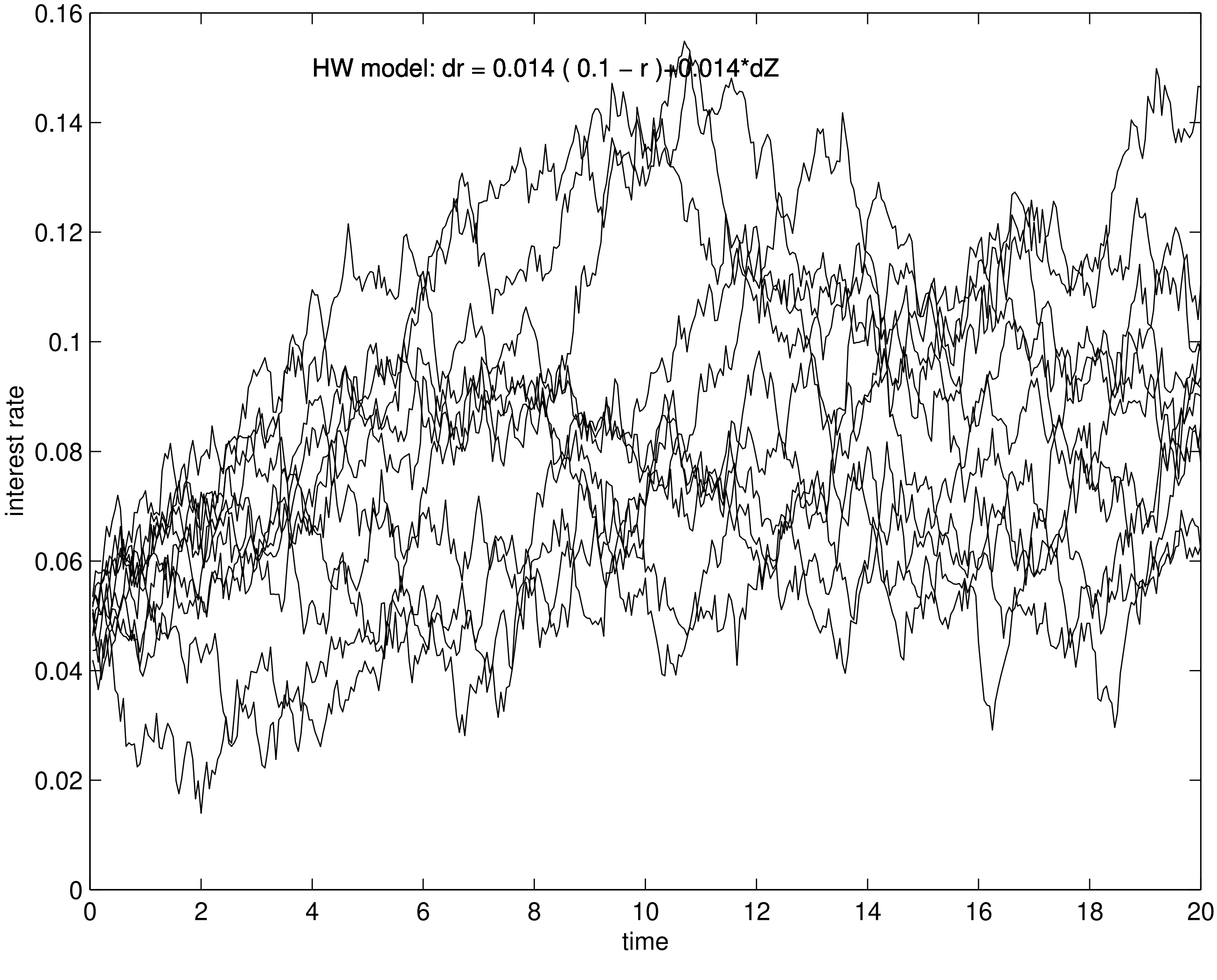


Figure 3.4: Some random sample paths generated by the Hull-White model

(3.1), we obtain the process followed by f,

$$\frac{df}{f} = \mu_f \cdot dt + \sigma_f \cdot dZ \tag{3.2}$$

where

$$\begin{aligned} \mu_f = &\ \tfrac{1}{f} \cdot (\tfrac{\partial f}{\partial r} \cdot b + \tfrac{\partial f}{\partial t} + \tfrac{\partial^2 f}{\partial^2 r} \cdot \tfrac{c^2}{2}) \\ \sigma_f = &\ \tfrac{1}{f} \cdot \tfrac{\partial f}{\partial r} \cdot c \end{aligned} \tag{3.3}$$

From the definition of the market price of risk

$$\frac{\mu_f - r}{\sigma_f} = \lambda \tag{3.4}$$

and (3.2), (3.3), we obtain the following equation

$$\frac{\partial f}{\partial r} \cdot (b - \lambda \cdot c) + \frac{\partial f}{\partial t} + \frac{\partial^2 f}{\partial^2 r} \cdot \frac{c^2}{2} - r \cdot f = 0. \tag{3.5}$$

Bond pricing

Assume a discount bond price is a derivative on r

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

At the maturity T, the payoff is

$$P(T, T) = 1. \tag{3.6}$$

Using (3.6) as a terminal condition, we compute the price of a discount bond by solving (3.5) backwards.

Specification of the market price of risk

Unlike the Black-Scholes model for equity options, the equation for interest rate derivatives (3.5) contains the market price of risk λ. This is because the interest rate is not a traded asset while an equity is.

There are two ways to deal with the market price of risk. First, we may assume that we are in the risk-neutral world, which means $\lambda = 0$. As a result, (3.5) becomes

$$\frac{\partial f}{\partial r} \cdot b + \frac{\partial f}{\partial t} + \frac{\partial^2 f}{\partial^2 r} \cdot \frac{c^2}{2} - r \cdot f = 0 \tag{3.7}$$

Second, we can estimate the market price of risk from the initial bond prices. For example, in the CIR model, we choose $\lambda = \lambda_1(t) \cdot \sqrt{r}$ and estimate λ_1 from the initial bond prices. Recall the bond prices are obtained by solving equation (3.5) with the initial condition $f(r, T) = 1$. In contrast, estimating λ is an inverse problem to that of solving (3.5). Here we know the solutions (i.e. the initial bond prices) but not the parameter λ in equation (3.5), so we need to find λ from the solutions.

Similarly, in the Hull-White model, we assume $\lambda = \lambda_1(t)$, and in the BK model, $\lambda = \lambda_1(t) \cdot r$, and estimate $\lambda_1(t)$ from the bond prices.[2]

3.2.3 Relation to HJM models

In the term structure consistent approach, all the one-factor models in Table 3.1 are special cases of the HJM models.[3]

This can be seen as follows. From equation (3.2), the price of a discount bond $f = P(t, T)$ satisfies

$$\frac{dP}{P} = \mu_f \cdot dt + \sigma_f \cdot dZ$$

Using (2.3) to replace μ_f by $r + \lambda \cdot \sigma_f$, we see that $P(t, T)$ satisfies

$$\frac{1}{P(t,T)} \cdot dP(t,T) = (r + \lambda \cdot \sigma_f) \cdot dt + \sigma_f \cdot dZ \tag{3.8}$$

Also, from the equations (3.3), (3.6), we have

$$\sigma_f(T,T) = \frac{1}{P(T,T)} \cdot \frac{\partial P(T,T)}{\partial r} \cdot c = \frac{1}{P(T,T)} \cdot 0 \cdot c = 0,$$

which says that as the maturity is approached, the bond volatility decreases to zero ("pull to par phenomenon").

[2] In fact, the two approaches are closely related. If $b_{\text{risk neutral}}$ is the drift in the risk-neutral world of dr and $b_{\text{real world}}$ is the drift in the real world, then

$$b_{\text{risk neutral}} = b_{\text{real world}} - \lambda \cdot c$$

[3] We don't understand the claim made in [Li] to the effect that all the HJM one-factor models are normal models.

Having obtained equation (3.8), we can derive the HJM equation as in Chapter 2. Introducing a new variable $y = ln(P(t,T))$, from (3.8) we have by Ito's lemma

$$dy = (r + \lambda \cdot \sigma_f - \frac{\sigma_f^2}{2}) \cdot dt + \sigma_f \cdot dZ \tag{3.9}$$

Recall the instantaneous forward rate is $\mathcal{F}(t,T) = \partial y / \partial T$. Taking the derivative of (3.9) with respect to T, we obtain the HJM equation

$$\begin{aligned} d\mathcal{F} &= d(\tfrac{\partial y}{\partial T}) \\ &= \tfrac{\partial (dy)}{\partial T} \\ &= \tfrac{\partial}{\partial T}\{r + \lambda \cdot \sigma_f - \tfrac{\sigma_f^2}{2}\} \cdot dt + \tfrac{\partial \sigma_f}{\partial T} \cdot dZ \\ &= (\lambda \cdot \tfrac{\partial \sigma_f}{\partial T} - \sigma_f \cdot \tfrac{\partial \sigma_f}{\partial T}) \cdot dt + \tfrac{\partial \sigma_f}{\partial T} \cdot dZ. \end{aligned}$$

3.3 Individual models

3.3.1 The Ho-Lee model

In the risk-neutral world, the Ho-Lee model (see [Ho-Lee]) is given by

$$dr = \theta(t) \cdot dt + \sigma \cdot dZ$$

where

- the long-run mean $\theta(t)$ is a function of time;
- σ is a constant.

The input to this model is σ and the initial yield curve. Thus $\theta(t)$ is implied from the yield curve and σ.

In this model the instantaneous interest rate is normally distributed and can become negative, while the bond price is lognormally distributed. The volatility of the interest rate is independent of r.

The most interesting feature of the model is its simplicity. Among other things, it has a closed-form solution for the price of a discount bond.

To explain this, assume the price of a discount bond is

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

This satisfies (3.7), (3.6) when

$$A_t - \theta(t) \cdot A \cdot B + \frac{\sigma^2 \cdot A \cdot B^2}{2} = 0$$

and

$$B_t + 1 = 0$$

with

$$A(T,T) = 1; \qquad B(T,T) = 0.$$

Solving the equations, we have (see [Hull])

$$\begin{aligned} B(t,T) = & \quad T - t \\ ln[A(t,T)] = & \quad ln\tfrac{P(0,T)}{P(0,t)} - (T-t) \cdot \tfrac{\partial lnP(0,t)}{\partial t} - \tfrac{1}{2} \cdot \sigma^2 \cdot t \cdot (T-t)^2 \end{aligned}$$

The main shortcoming of this model is that it fails to capture the mean-reverting property that is observed in the movement of interest rates.

3.3.2 The Hull-White model

In the risk-neutral world, the Hull-White model (see [Hull-White(1990a)]) is [4]

$$dr = (\theta(t) - a \cdot r) \cdot dt + \sigma \cdot dZ$$

where

- a is the mean-reverting rate of r and is a constant. If $a = 0$, then it becomes the Ho-Lee model;
- the long-run mean $\theta(t)$ is a function of time;

[4]Hull and White [Hull-White(1990a)] formulate the model more generally (allowing a and σ to be functions of time). Here we choose σ and a to be constant so that the model has a stationary volatility structure.

- σ is a constant.

The input to this model is a, σ and the initial yield curve, and $\theta(t)$ is implied from the input.

Like the Ho-Lee model, the instantaneous interest rate is normally distributed, and can become negative. Unlike the Ho-Lee model, it incorporates a mean-reverting term into the model.

The Hull-White model retains the simplicity of the Ho-Lee model. Like the Ho-Lee model, it has closed-form solution for the price of a discount bond.

Assume the bond price is

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

It satisfies (3.7), (3.6) when

$$A_t - \theta(t) \cdot A \cdot B + \frac{\sigma^2 \cdot A \cdot B^2}{2} = 0$$

and

$$B_t - a \cdot B + 1 = 0$$

with

$$A(T,T) = 1; \qquad B(T,T) = 0.$$

Introducing $\hat{A}(t,T) = ln[A(t,T)]$, we obtain (see [Hull-White(1990a)])

$$\begin{aligned} B(t,T) &= \tfrac{B(0,T) - B(0,t)}{\partial B(0,t)/\partial t} \\ \hat{A}(t,T) &= \hat{A}(0,T) - \hat{A}(0,t) - B(t,T) \cdot \tfrac{\partial \hat{A}(0,t)}{\partial t} \\ &\quad + \tfrac{1}{2}(B(t,T) \cdot \tfrac{\partial B(0,t)}{\partial t})^2 \cdot \textstyle\int_0^t (\tfrac{\sigma}{\partial B(0,\tau)/\partial \tau})^2 \cdot d\tau \end{aligned}$$

3.3.3 The CIR model

The CIR model is

$$dr = (a - \theta \cdot r) \cdot dt + \sigma \sqrt{r} \cdot dZ.$$

It was first developed in an equilibrium framework where a, θ, σ are constants (see [CIR]).

This model is well-posed only if

$$a > \frac{\sigma^2}{2} \tag{3.10}$$

which can be seen as follows. Introduce a new variable

$$y = \sqrt{r}.$$

Applying Ito's lemma, we have

$$dy = (\frac{2 \cdot a - \sigma^2}{4 \cdot \sqrt{r}} - \frac{\theta \cdot \sqrt{r}}{2}) \cdot dt + \frac{\sigma \cdot dZ}{2}.$$

When r is small, the first term dominates the others. If $a < \sigma^2/2$, the first term is negative, and y will become negative quickly (see Figure 3.3 for an example in which (3.10) is violated).

Hull and White [Hull-White(1990a)] extend the model to allow a to be a function of time to fit the initial term structure. However, the model extended this way cannot fit all initial yield curves. This is because, when $a = a(t)$ is implied from the initial yield curve, the restriction (3.10) is often violated.

Thus we will extend the CIR model differently: fix a to be a constant satisfying $a > \sigma^2/2$, but allow θ to be a function of time to fit the initial term structure (see [Uhrig-Walter]). The input to the model is a, σ and the initial yield curve, and $\theta = \theta(t)$ is implied from the input.

Though the CIR model does not have the same analytical tractability as the Hull-White model, the equation for discount bonds can be reduced to a pair of ordinary differential equations. Let

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

It satisfies (3.7), (3.6) when

$$A_t - a \cdot A \cdot B = 0$$

and

$$B_t - \theta(t) \cdot B - \frac{1}{2}\sigma^2 \cdot B^2 + 1 = 0$$

with

$$A(T,T) = 1; \qquad B(T,T) = 0.$$

If $\theta(t)$ is constant, then there is a closed-form solution for the equations (see [CIR]). However, if $\theta(t)$ is not constant, no analytical solutions have been found in general.

3.3.4 The Black-Karasinski model

In the risk-neutral world, the Black-Karasinski model is [5]

$$d[ln(r)] = (\theta(t) - a \cdot ln(r)) \cdot dt + \sigma \cdot dZ$$

where

- a is the mean-reverting rate of $ln(r)$ and is assumed to be a constant;
- the long-run mean $\theta(t)$ is a function of time used to fit the initial term structure; [6]
- σ is a constant.

Unlike the Ho-Lee and Hull-White models, the interest rate is always positive. Also unlike the CIR model, there is no restriction for the model to be well-posed. These two features can be considered as the main advantages of the model.

This model does not have a closed-form solution for discount bond prices.

However, Sandmann and Sondermann [Sandmann-Sondermann] point out that the Black-Karasinski model cannot explain prices of Eurodollar future contracts[7]. This is because the higher the interest rate becomes, the higher the volatility, thus the interest rate can become so high that the payoff in an Eurodollar future is negative.

[5]Black and Karasinski [Black-Karasinski] formulate the model more generally (allowing a and σ to be functions of time). Here we choose σ and a to be constant so that the model has a stationary volatility structure.

[6]Typically it is negative.

[7]A Eurodollar future is a contract that pays off

$$1. - 0.25 \cdot L$$

at the maturity where L is the prevailing 3-month LIBOR rate.

3.3.5 Empirical evidences

As the evidence below shows, it is likely that no single one-factor interest rate model will emerge as the best choice for all the market.

Chan et al. [Chan] study a family of interest rate models of the form

$$dr = (a - b \cdot r) \cdot dt + \sigma \cdot r^{\beta} \cdot dZ$$

where a, b, σ, β are constant. Using the time series of the U.S. Treasure bill one-month rates from 1964 to 1989, they conclude that the model with $\beta = 1.5$ fits the time series best. This suggests that the volatility of the U.S. short rate is highly sensitive to the level of the interest rates. The model with $\beta = 1.5$

$$dr = (a - b \cdot r) \cdot dt + \sigma \cdot r^{3/2} \cdot dZ$$

is called the cubic-variance model.

However, the situation with the British interest rates may be different. Nowman [Nowman] studies the British short rates using the one-month sterling interbank rate, and finds that the volatility is not so sensitive to the level of the interest rate.

It is also likely that no single one-factor interest rate model will emerge as the best choice for all financial products. For example, many authors criticize interest rate models that produce negative interest rates, see, e.g. [Rogers]. However, some banks recently began to sell interest rates options that pay off only if the Japaneses interest rate is negative (see [Reed]). In this case, a model that can produce negative interest rates is certainly desirable.

3.3.6 Comparison of the underlying assumptions of one-factor models

To summarize, we list the various assumptions underlying the interest rate models in Table 3.2.

Table 3.2. Comparison of the underlying assumptions of one-factor models				
Name of Model	Positive interest rates ?	No restriction on parameters?	Closed-form solution for bonds?	Mean-reverting?
BK	Yes	Yes	No	Yes
cubic-variance	Yes	Yes	No	Yes
CIR	Yes	No	In some cases	Yes
Ho-Lee	No	Yes	Yes	No
HW	No	Yes	Yes	Yes

3.3.7 Limitation of one-factor models

All one-factor interest rate models assume that the short rates are perfectly correlated with the long rates, which is unrealistic. Principal analysis has shown that one-factor models can only explain no more than 87% [8] of the movement of the yield curve ([Dybvig], [Rebonato]). Some authors suggest that, while one-factor models can price bond options, caps/floors and swaptions, they cannot price diff swaps whose payoffs depend on the slope of the yield curve ([Canabarro], [Dybvig], [Hull-White(1995)], [Rebonato]).

[8] As pointed out by Mark Reimers [Reimers], this fraction varies from market to market, for example, for the British market, it is only 51%.

Chapter 4

Tree Methods for One-factor Interest Rate Models

In the last chapter we introduced a number of one-factor interest rate models, which in general do not have closed-form solutions and thus require numerical methods. In this chapter we will consider numerical methods for one-factor models.

4.1 Introduction

There are three kinds of numerical methods for option models (see, for example, [Ames], [Hull], [GKO])

1. Monte Carlo methods,

2. PDE based numerical methods that include finite-difference methods, finite-element methods and finite-volume methods, though only finite-difference methods will be considered here,

3. tree methods.

Monte Carlo methods are easy to implement, but are computationally expensive and cannot deal with American options. Finite-difference methods are classified as either implicit or explicit. While implicit finite-difference methods are unconditionally stable, they have the disadvantage that boundary conditions need to be introduced. In contrast, explicit finite-difference/tree methods[1] are more convenient for problems without boundary conditions.[2] In this section we only consider explicit finite-difference/tree methods.

The main problem with the tree methods is that they can be unstable and, as a result, the numerical solutions need not converge to the true solutions. This problem is further complicated by the special properties of interest rates models. Unlike the Black-Scholes model for equity options,

- interest rate models are time-dependent (e.g., in the term structure consistent approach), which makes the stability difficult to archive as the construction of the tree changes over time;
- many models incorporate the mean-reverting property of the interest rates, which means that the construction of the tree changes over the space as well.

Rather than developing special numerical methods for individual models, we are more interested in developing a method that can be applied to a wide range of models, since implementing different numerical methods for different models is a complicated and error prone process. If a numerical method can be applied to all the models, then we can use an object-oriented approach (see Figure 4.1) in which the abstract base class implements the numerical method, while the derived classes are only concerned with the initialization of the parameters in a particular model. This approach not only avoids reprogramming the software system if one changes a model, but also reduces possible errors in the implementation, since adding a new model requires a few lines of code only. Motivated by these considerations, we will consider tree methods for term-structure consistent interest rate models based on the following criteria:

1. whether the method is stable for practical input data including non-smooth yield curves;

[1]Since explicit finite-difference methods are similar to tree methods, we will make no distinction between the two methods.

[2]Here we make a distinction between a boundary condition and an initial condition such as that in a payoff.

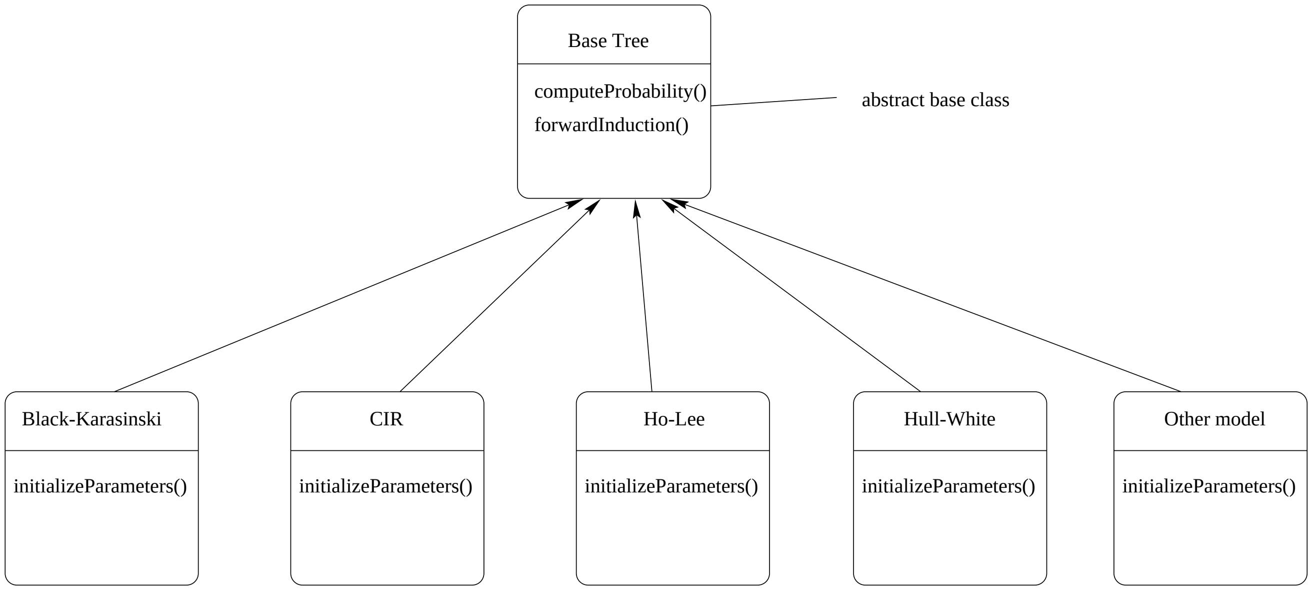


Figure 4.1: An object-oriented appproach to implementing different interest rate models

2. whether the method can be applied to any one-factor interest rate model without major modification;

3. whether the method is easy to implement;

4. whether the method has fast convergence.

We consider both explicit finite-difference methods and probability based tree methods. After extensive numerical experiments, we conclude that adaptive tree methods, namely tree methods that include adaptive branching rules to incorporate the mean-reverting property of the interest rates, are very stable and efficient, and can be applied to a wide range of interest rate models.

We also observe that if the initial yield curves are non-smooth functions of time (e.g. those obtained by piecewise linear interpolation), then the parameters of the models implied from the yield curves can have spurious fluctuations, which affects the accuracy of the numerical solutions. This is discussed in Section 4.2.9.

4.2 A generic tree method for one-factor models

4.2.1 Change of variable

Consider a generic one-factor interest rate model of the form

$$dr = b(r,t) \cdot dt + c(r,t) \cdot dZ \tag{4.1}$$

where $b = b(r,t)$, $c = c(r,t)$ are functions of interest rate r, time t, and possibly some other parameters. First we transform (4.1) into one with constant volatility. Introducing a new variable

$$y = g(r,t), \tag{4.2}$$

where

$$g(r,t) = c_1 \cdot \int_0^r \frac{dr_1}{c(r_1,t)}$$

and c_1 is a constant, we can rewrite (4.1) as

$$dy = b_1 \cdot dt + c_1 \cdot dZ, \tag{4.3}$$

where b_1 is a function of r, t:

$$b_1 = \frac{b \cdot c_1}{c} + \frac{c^2}{2} \cdot \frac{\partial^2 g}{\partial r^2} + \frac{\partial g}{\partial t}.$$

We assume that b_1 depends on a parameter, $\theta = \theta(t)$, which is a function of time to fit the initial term structure.

In the rest of the chapter we will develop tree methods for the model described by (4.3).

4.2.2 Construction of the tree

Throughout this section we will consider a trinomial tree that has three branches emanating from each node (Figure 4.2). In order for the tree to recombine, the tree is evenly spaced in y. For simplicity we assume the tree is evenly spaced in t as well, though this can be changed easily. The time step is denoted by Δt. The value of y on the tree at

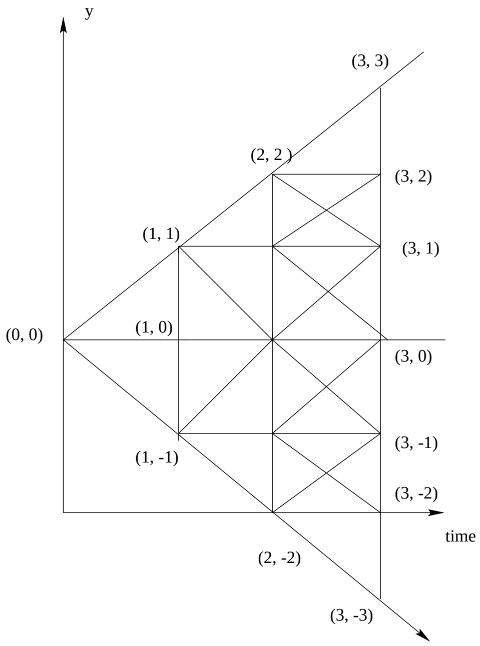


Figure 4.2: A trinomial tree

time zero is equal to $y_0 = g(r_0, 0)$, where r_0 is the initial Δt-rate.[3] The value of y at other nodes has the form $y_0 + k \cdot \Delta y$, where k is a positive or negative integer. The time step Δt is set to

$$\Delta t = \frac{(\Delta y)^2}{m \cdot c_1^2},$$

where m is a constant to be chosen. The bigger m is, the more stable the tree method. Usually $m = 3$ is a good choice.

A node (i, j) in the tree will denote the value of $y = y_0 + j \cdot \Delta y$ at time $t = i \cdot \Delta t$, where i, j are integers, and $f(i, j)$ will denote the value of a derivative f at node (i, j).

The probabilities on the branches vary in the tree. There are two methods for specifying these probabilities, one based on the standard explicit finite-difference scheme, the other based on the probability property of the model, see [Hull-White(1993)]. The resulting probabilities from the two approaches can be quite different.

[3]If $B(0, \Delta)$ is the present value of a discount bond that matures at Δt, then $r_0 = -log(B(0, \Delta t))/\Delta t$.

Methods based on the standard explicit finite-difference scheme

Recall the PDE associated with the model described by (4.3) is

$$\frac{\partial f}{\partial t} + b_1 \cdot \frac{\partial f}{\partial y} + \frac{c_1^2}{2} \cdot \frac{\partial^2 f}{\partial y^2} = r \cdot f. \tag{4.4}$$

The partial derivatives of f in (4.4) can be approximated by the standard finite-differences

$$\begin{aligned} \frac{\partial f}{\partial t} - r \cdot f &\approx \tfrac{f(i+1,j) - exp(r(i,j)\cdot\Delta t)\cdot f(i,j)}{\Delta t} \\ \frac{\partial f}{\partial y} &\approx \tfrac{f(i+1,j+1) - f(i+1,j-1)}{2\cdot\Delta y} \\ \frac{\partial^2 f}{\partial y^2} &\approx \tfrac{f(i+1,j+1) - 2\cdot f(i+1,j) + f(i+1,j-1)}{\Delta y^2} \end{aligned} \tag{4.5}$$

The last two are second-order centered finite-differences. Substituting (4.5) into (4.4), we obtain

$$f(i,j) = exp(-r(i,j)\cdot\Delta t)\cdot(p_u \cdot f(i+1,j+1) + p_m \cdot f(i+1,j) + p_d \cdot f(i+1,j-1)) \tag{4.6}$$

where

$$\begin{aligned} p_u &= \tfrac{b_1(i,j)\cdot\Delta t}{2\cdot\Delta y} + \tfrac{c_1^2\cdot\Delta t}{2\cdot(\Delta y)^2} \\ p_m &= 1 - \tfrac{c_1^2\cdot\Delta t}{(\Delta y)^2} \\ p_d &= -\tfrac{b_1(i,j)\cdot\Delta t}{2\cdot\Delta y} + \tfrac{c_1^2\cdot\Delta t}{2\cdot(\Delta y)^2} \end{aligned} \tag{4.7}$$

and $b_1(i,j)$ is the value of b_1 at node (i,j). Since $p_u + p_m + p_d = 1$, p_u, p_m, p_d can be considered as the probabilities on the branches.

Method based on probability considerations

Suppose at time $t = i \cdot \Delta t$, the value of y is $y_1 = y_0 + j \cdot \Delta y$. At the next time step $t + \Delta t$, there are three possible movements of the value of y. It can either go down to $y_1 - \Delta y$, or stay at y_1, or move up to $y_1 + \Delta y$. Thus there are three branches joining the nodes $(i+1, j-1)$, $(i+1, j)$, $(i+1, j+1)$ to the node (i,j).

The probabilities on the three branches are chosen so that the changes in y have the correct mean and standard deviation over each time interval. The equations that must be satisfied are

$$\begin{aligned} p_u + p_m + p_d &= 1 \\ p_u \cdot (j+1) \cdot \Delta y + p_m \cdot j \cdot \Delta y + p_d \cdot (j-1) \cdot \Delta y &= E(y) \\ p_u \cdot (j+1)^2 \cdot \Delta y^2 + p_m \cdot j^2 \cdot \Delta y^2 + p_d \cdot (j-1)^2 \cdot \Delta y^2 &= c_1^2 \cdot \Delta t + E(y)^2 \end{aligned} \tag{4.8}$$

where $E(y) = j \cdot \Delta y + E(\delta y)$, $E(\delta y) = b_1 \cdot \Delta t$. The 2nd and 3rd equations just say that the first and second moments in y are matched.

Solving equations (4.8), we have

$$\begin{aligned} p_u &= \frac{b_1(i,j)\cdot\Delta t}{2\cdot\Delta y} + \frac{b_1^2(i,j)\cdot\Delta t^2}{2\cdot(\Delta y)^2} + \frac{c_1^2\cdot\Delta t}{2\cdot(\Delta y)^2} \\ p_m &= 1 - \frac{b_1^2(i,j)\cdot\Delta t^2}{(\Delta y)^2} - \frac{c_1^2\cdot\Delta t}{(\Delta y)^2} \\ p_d &= -\frac{b_1(i,j)\cdot\Delta t}{2\cdot\Delta y} + \frac{b_1^2(i,j)\cdot\Delta t^2}{2\cdot(\Delta y)^2} + \frac{c_1^2\cdot\Delta t}{2\cdot(\Delta y)^2} \end{aligned} \tag{4.9}$$

Note that the probabilities (4.9) are similar to, but different from, those in the standard central finite-difference approach (4.7). (Actually, they agree up to the Δt terms, but the Δt^2 terms differ.)

It is worth mentioning that the probability based method can be interpreted as a finite-difference method. If the finite-difference approximations in (4.4) are replaced by

$$\begin{aligned} \frac{\partial f}{\partial t} - r \cdot f &\approx \frac{f(i+1,j) - exp(r(i,j)\cdot\Delta t)\cdot f(i,j)}{\Delta t} \\ \frac{\partial f}{\partial y} &\approx \frac{f(i+1,j+1)-f(i+1,j-1)}{2\cdot\Delta y} + \frac{f(i+1,j+1)-2\cdot f(i+1,j)+f(i+1,j-1)}{2\cdot(\Delta y)^2}\cdot b_1\cdot\Delta t \\ \frac{\partial^2 f}{\partial y^2} &\approx \frac{f(i+1,j+1)-2\cdot f(i+1,j)+f(i+1,j-1)}{\Delta y^2} \end{aligned} \tag{4.10}$$

then we obtain the probabilities (4.9). The additional term in (4.10) can be considered as an artificial dissipation. As we will see in Lemma 4.1 and Lemma 4.2, the artificial dissipation term improves the stability of the tree method; however, it can introduce an artifact and sometimes does not work well for the model

$$dr = (a - \theta(t) \cdot r) \cdot dt + \sigma \cdot r^{3/2} \cdot dZ.$$

If the values of b_1, c_1 have been determined at each node, we can build a tree as above and compute the probabilities p_u, p_m, p_d at each node. Given the future payoff of

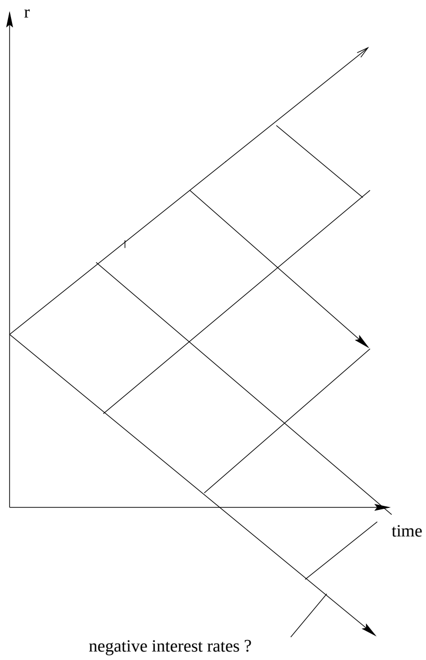


Figure 4.3: A Cox-Ross-Ruberstein tree method can lead to negative interest rates

a derivative, we can use (4.6) to compute its value by starting at the maturity and then moving back through the tree.

4.2.3 Incorporating the mean-reverting property into the tree

Many interest rate models incorporate the mean-reverting property that is observed in the movement of interest rates. In this subsection we shall show that this property can be used to improve both the efficiency and the stability of the tree method by using *adaptive branching rules.*

The basic idea behind the adaptive branching rules is very simple. If r is very small, there is a strong tendency for r to rise. In this case, the tree should branch upward (Figure 4.4.A). Similarly, if r is very large, then the tree should branch downward (Figure 4.4.B).

An example in which the adaptive branching rules are necessary is the CIR model:

$$dr = (a - \theta(t) \cdot r) \cdot dt + \sigma \cdot \sqrt{r} \cdot dZ.$$

After introducing a new variable $y = \sqrt{r}$, the above equation can be written as

$$dy = (\frac{2 \cdot a - \sigma^2}{4 \cdot \sqrt{r}} - \frac{\theta \cdot \sqrt{r}}{2}) \cdot dt + \frac{\sigma \cdot dZ}{2}.$$

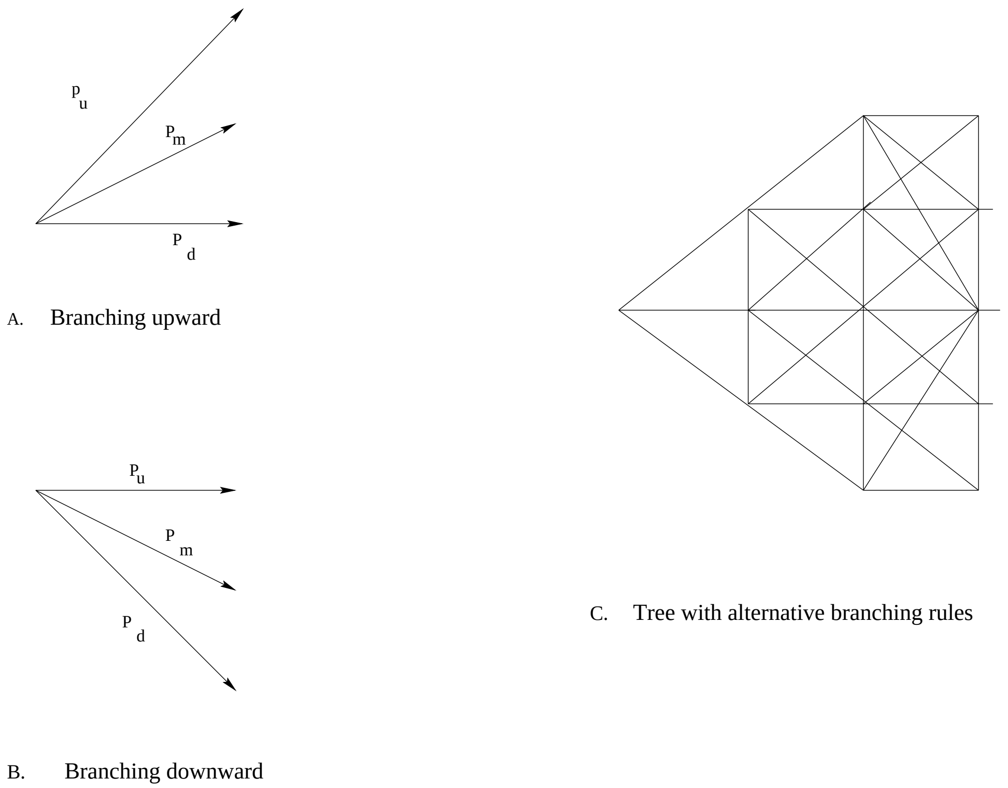


Figure 4.4: Adaptive branching rules in a trinomial tree

Note that y will remain positive if $2 \cdot a > \sigma^2$. In this case, an ordinary tree is not appropriate since it leads to negative interest rates (see Figure 4.3).

Again, there are two approaches to derive the probabilities for the adaptive branching rules. These are discussed below.

Finite-difference scheme

In the case of branching upward (Figure 4.4.A), the finite-difference approximations to the partial derivatives of f in (4.5) should be modified to

$$\begin{aligned} \frac{\partial f}{\partial t} - r \cdot f &\approx \frac{f(i+1,j) - exp(r(i,j) \cdot \Delta t) \cdot f(i,j)}{\Delta t} \\ \frac{\partial f}{\partial y} &\approx \frac{f(i+1,j+1) - f(i+1,j)}{\Delta y} \\ \frac{\partial^2 f}{\partial y^2} &\approx \frac{f(i+1,j+2) - 2 \cdot f(i+1,j+1) + f(i+1,j)}{\Delta y^2} \end{aligned} \tag{4.11}$$

which leads to the following probabilities

$$\begin{aligned} p_u = & \quad \Delta t \cdot \{\tfrac{c_1^2}{2\cdot(\Delta y)^2}\} \\ p_m = & \quad \Delta t \cdot \{\tfrac{b_1(i,j)}{\Delta y} - \tfrac{c_1^2}{(\Delta y)^2}\}; \\ p_d = & \quad 1 + \Delta t \cdot \{-\tfrac{b_1(i,j)}{\Delta y} + \tfrac{c_1^2}{2\cdot(\Delta y)^2}\} \end{aligned} \tag{4.12}$$

The second finite difference approximation in (4.11) is a up-wind difference scheme ([Ames]).

Similarly, in the case of branching down (Figure 4.4.B), we have

$$\begin{aligned} p_u = & \quad 1 + \Delta t \cdot \{\tfrac{b_1(i,j)}{\Delta y} + \tfrac{c_1^2}{2\cdot(\Delta y)^2}\} \\ p_m = & \quad \Delta t \cdot \{-\tfrac{b_1(i,j)}{\Delta y} - \tfrac{c_1^2}{(\Delta y)^2}\} \\ p_d = & \quad \Delta t \cdot \{\tfrac{c_1^2}{2\cdot(\Delta y)^2}\} \end{aligned} \tag{4.13}$$

How does one decide which branching rule to use? The answer is that one should choose a branching rule that will result in positive probabilities to assure the tree method is stable.[4] To be more precise, assume $\Delta y = c_1 \cdot \sqrt{m \cdot \Delta t}$. If

$$-(1 + \frac{1}{2 \cdot m}) < b_1(i,j) \cdot \frac{\Delta t}{\Delta y} < -\frac{1}{m} \tag{4.14}$$

then the branching down rule (4.13) should be used. On the other hand, if

$$\frac{1}{m} < b_1(i,j) \cdot \frac{\Delta t}{\Delta y} < (1 + \frac{1}{2 \cdot m}) \tag{4.15}$$

then the branching upward rule (4.12) should be used. Finally, if

$$-\frac{1}{m} < b_1(i,j) \cdot \frac{\Delta t}{\Delta y} < \frac{1}{m}, \tag{4.16}$$

then the normal branching rule (4.7) should be used.

The following lemma shows that if we follow the above rules in choosing branching rules, then the probabilities are positive. This gives a sufficient condition for the tree method to be stable.

[4] A sufficient condition for stability is that all the probabilities are positive, see Ames[Ames], page 75.

Lemma 4.1 *If $\Delta y = c_1 \cdot \sqrt{m \cdot \Delta t}$, $m > 1$, and at all the nodes of the tree, the following condition is satisfied,*

$$-(1 + \frac{1}{2 \cdot m}) < \frac{b_1 \cdot \Delta t}{\Delta y} < 1 + \frac{1}{2 \cdot m}, \tag{4.17}$$

then using the adaptive branching rules (4.12) ~ (4.16) implies that all transition probabilities p_u, p_m, p_d are positive.

Proof. Denote

$$b_2 = \frac{b_1 \cdot \Delta t}{\Delta y}.$$

If

$$-(1 + \frac{1}{2 \cdot m}) < b_2 < -\frac{1}{m},$$

then the adaptive branching down rule (4.13) is used. In this case,

$$\begin{aligned} p_u &= 1 + \Delta t \cdot \{\tfrac{b_1}{\Delta y} + \tfrac{c_1^2}{2(\Delta y)^2}\} \\ &= 1 + b_2 + \tfrac{1}{2 \cdot m} \\ &> 0 \end{aligned}$$

and

$$\begin{aligned} p_m &= \Delta t \cdot \{-\tfrac{b_1}{\Delta y} - \tfrac{c_1^2}{(\Delta y)^2}\} \\ &= -b_2 - \tfrac{1}{m} \\ &> 0 \end{aligned}$$

and

$$p_d = \Delta t \cdot \{\tfrac{c_1^2}{2 \cdot (\Delta y)^2}\} > 0$$

If

$$-\frac{1}{m} < b_2 < \frac{1}{m},$$

then the normal branching rule (4.7) is used. In this case, the positivity of p_u, p_m, p_d can be checked as follows

$$\begin{aligned} p_u &= \tfrac{b_1(i,j)\cdot\Delta t}{2\cdot\Delta y} + \tfrac{c_1^2\cdot\Delta t}{2\cdot(\Delta y)^2} \\ &= \tfrac{b_2}{2} + \tfrac{1}{2\cdot m} \\ &> 0 \end{aligned}$$

and

$$\begin{aligned} p_m &= 1 - \tfrac{c_1^2\cdot\Delta t}{(\Delta y)^2} \\ &= 1 - \tfrac{1}{m} \\ &> 0 \end{aligned}$$

and

$$\begin{aligned} p_d &= -\tfrac{b_1(i,j)\cdot\Delta t}{2\cdot\Delta y} + \tfrac{c_1^2\cdot\Delta t}{2\cdot(\Delta y)^2} \\ &= -\tfrac{b_2}{2} + \tfrac{1}{2\cdot m} \\ &> 0 \end{aligned}$$

If

$$\frac{1}{m} < b_2 < 1 + \frac{1}{2\cdot m},$$

then the branching upward rule (4.12) is used. The positivity of p_u, p_m, p_d can be checked similarly as in the case

$$-\frac{1}{m} > b_2 > -(1 + \frac{1}{2\cdot m}).$$

Remark. Since $\Delta y = c_1 \cdot \sqrt{m\cdot\Delta t}$, the condition (4.17) simplifies to

$$-c_1^2 \cdot m \cdot (1 + \frac{1}{2\cdot m}) < b_1 \cdot \Delta y < c_1^2 \cdot m \cdot (1 + \frac{1}{2\cdot m}),$$

thus we see that the bigger the value m is, the more stable the method.

If the interest rate model is time-independent, then the above condition is easy to meet. However, if the interest rate model is time-dependent, then the condition is more difficult to enforce. This is because the conditions (4.14) - (4.16) depend on the value of $\theta(i \cdot \Delta t)$ that is yet to be determined. In practice, when implementing the tree methods, we use an approximate value of θ in deciding which branching rule to use.

Method based on probability considerations

In the case of branching up (Figure 4.4.A), equations (4.8) are modified to

$$p_u + p_m + p_d = 1$$

$$p_u \cdot (j) \cdot \Delta y + p_m \cdot (j+1) \cdot \Delta y + p_d \cdot (j+2) \cdot \Delta y = E(y)$$

$$p_u \cdot (j)^2 \cdot \Delta y^2 + p_m \cdot (j+1)^2 \cdot \Delta y^2 + p_d \cdot (j+2)^2 \cdot \Delta y^2 = c_1^2 \cdot \Delta t + E(y)^2$$

where $E(y)$ remains the same. This leads to

$$\begin{aligned} p_u &= \frac{\beta(i,j)\cdot\Delta t}{2\cdot\Delta y} + \frac{\beta^2(i,j)\cdot\Delta t^2}{2\cdot(\Delta y)^2} + \frac{c_1^2\cdot\Delta t}{2\cdot(\Delta y)^2} \\ p_m &= 1 - \frac{\beta^2(i,j)\cdot\Delta t^2}{(\Delta y)^2} - \frac{c_1^2\cdot\Delta t}{(\Delta y)^2} \\ p_d &= -\frac{\beta(i,j)\cdot\Delta t}{2\cdot\Delta y} + \frac{\beta^2(i,j)\cdot\Delta t^2}{2\cdot(\Delta y)^2} + \frac{c_1^2\cdot\Delta t}{2\cdot(\Delta y)^2} \end{aligned} \tag{4.18}$$

where $\beta(i,j) = b_1(i,j) - \Delta y$.

Similarly, in the case of branching downward, the probabilities are given by (4.18) where $\beta = b_1(i,j) + \Delta y$.

As in the finite-difference method, the adaptive branching rules are chosen so that the transitive probabilities are positive. As a result, the branching downward rule (4.13) should be used if

$$-\frac{3}{2} < b_1(i,j) \cdot \frac{\Delta t}{\Delta y} < -\frac{1}{2} \tag{4.19}$$

and the branching upward rule (4.12) should be used if

$$\frac{1}{2} < b_1(i,j) \cdot \frac{\Delta t}{\Delta y} < \frac{3}{2}. \tag{4.20}$$

The normal branching should be used when

$$-\frac{1}{2} < b_1(i,j) \cdot \frac{\Delta t}{\Delta y} < \frac{1}{2}. \tag{4.21}$$

The following result shows that the probability based method is slightly more stable than the finite-difference method if one chooses the same Δy and Δt.

Figure 4.5: More adaptive branching rules in a trinomial tree

Lemma 4.2 *If $\Delta y = c_1 \cdot \sqrt{3 \cdot \Delta t}$, and at all the nodes of the tree, the following condition is satisfied,*

$$-\frac{3}{2} < \frac{b_1 \cdot \Delta t}{\Delta y} < \frac{3}{2}.$$

then using the adaptive branching rules (4.18) implies that all transition probabilities p_u, p_m, p_d are positive.

The proof of this result is similar to that of Lemma 4.1 and will not be repeated.

4.2.4 More adaptive branchings to increase the stability of the tree methods

We can introduce more adaptive branching rules (see Figure 4.5) to improve the stability of the tree methods. For example, in the probability based method, if

$$-\frac{5}{2} < \frac{b_1 \cdot \Delta t}{\Delta y} < -\frac{3}{2}$$

then one uses the branching upward two-step rule (Figure 4.5.B). In this case, the probabilities in Section 4.2.3 are no longer applicable as they are negative. Instead, we shall use the formula (4.18) with $\beta = b \cdot \Delta t + 2 \cdot \Delta y$ to compute the transition probabilities. On the other hand, if

$$\frac{3}{2} < \frac{b_1 \cdot \Delta t}{\Delta y} < \frac{5}{2}$$

then one can use the branching downward two-step rule (Figure 4.5.A) and the formula (4.18) with $\beta = b \cdot \Delta t - 2 \cdot \Delta y$ to specify the transition probabilities.

With the use of the additional branching rules, the stability of the tree method is increased, as the transition probabilities are positive as long as

$$-\frac{5}{2} < \frac{b_1 \cdot \Delta t}{\Delta y} < \frac{5}{2}.$$

As for the finite-difference method, if

$$1 + \frac{1}{2 \cdot m} < \frac{b_1 \cdot \Delta t}{\Delta y} < \frac{3}{2} + \frac{1}{2 \cdot m}$$

then the formula in Section 4.2.3 will give negative probabilities. Thus we need to use the following probabilities

$$\begin{aligned} p_u &= \Delta t \cdot \{\tfrac{c_1}{2\cdot(\Delta y)^2}\} \\ p_m &= -\tfrac{1}{2} + \Delta t \cdot \{\tfrac{b_1(i,j)}{\Delta y} - \tfrac{c_1}{(\Delta y)^2}\} \\ p_d &= \tfrac{3}{2} + \Delta t \cdot \{-\tfrac{b_1(i,j)}{\Delta y} + \tfrac{c_1}{2\cdot(\Delta y)^2}\} \end{aligned} \tag{4.22}$$

The probabilities are chosen so that they are positive. Similarly, if

$$-(1 + \frac{1}{2 \cdot m}) > \frac{b_1 \cdot \Delta t}{\Delta y} > -(\frac{3}{2} + \frac{1}{2 \cdot m})$$

then we use the following probabilities

$$\begin{aligned} p_u &= \tfrac{3}{2} + \Delta t \cdot \{\tfrac{b_1(i,j)}{\Delta y} + \tfrac{c_1}{2\cdot(\Delta y)^2}\} \\ p_m &= -\tfrac{1}{2} + \Delta t \cdot \{-\tfrac{b_1(i,j)}{\Delta y} - \tfrac{c_1}{(\Delta y)^2}\} \\ p_d &= \Delta t \cdot \{\tfrac{c_1}{2\cdot(\Delta y)^2}\} \end{aligned} \tag{4.23}$$

4.2.5 Accuracy of the methods

In this subsection we consider the accuracy of the tree methods. For simplicity we assume that b_1 is a continuous and bounded function.

Lemma 4.3 *Let f be the numerical solution obtained from the tree method.*
(1) At the nodes where the normal branching rule is used, f is an approximate solution to (4.4) with second order accuracy in Δy and first order accuracy in Δt,

$$\frac{\partial f}{\partial t} + b_1 \cdot \frac{\partial f}{\partial y} + \frac{c_1^2}{2} \cdot \frac{\partial^2 f}{\partial y^2} = r \cdot f + O((\Delta y)^2 + \Delta t), \tag{4.24}$$

(2) at the nodes where the adaptive branching rules are used, f has first order accuracy both in Δy and Δt

$$\frac{\partial f}{\partial t} + b_1 \cdot \frac{\partial f}{\partial y} + \frac{c_1^2}{2} \cdot \frac{\partial^2 f}{\partial y^2} = r \cdot f + O(\Delta y + \Delta t) \tag{4.25}$$

Proof. The relation (4.24) follows from the fact that the finite-difference approximations in (4.5) (or, (4.10) in the probability-based method) have second-order accuracy in Δy and first-order accuracy in Δt. When the adaptive branching rules are used, the accuracy of the finite-difference approximations (4.11) (or the equivalent in the probability based method) reduces to first order in both Δy and Δt.

4.2.6 Forward induction method

In this subsection we will describe a forward induction method first developed by Jamshidian [Jamshidian] that fits the initial term structure based on the concept of Arrow-Debreau prices.

Define Q_{ij} as the present value of a security that pays off \$1 if node (i, j) is reached and zero otherwise.[5] In other words, if f is a derivative that satisfies the finite-difference equation

$$f(i,j) = exp(-r(i,j) \cdot \Delta t) \cdot (p_u \cdot f(i+1, j+1) + p_m \cdot f(i+1, j) + p_d \cdot f(i+1, j-1)) \tag{4.26}$$

with the initial condition

$$f(i,k) = \begin{cases} 0 & \text{if } k \neq j; \\ 1 & \text{if } k = j \end{cases}$$

then $Q_{ij} = f(0,0)$. As we shall show, the values $\{Q_{ij}\}$ can be computed inductively.

Obviously $Q_{0,0} = 1$. If we have computed all the values of $\{Q_{ij}\}$ at step i (namely, at time $i \cdot \Delta t$), the values at step $(i+1)$ can be computed by the following lemma due to Jamshidian [Jamshidian].

Lemma 4.4 *Let $q(k, j)$ be the probability of moving from node (i, k) to node $(i+1, j)$, then*

$$Q_{i+1,j} = \sum_k Q_{i,k} \cdot q(k,j) \cdot exp(-r(i,k) \cdot \Delta t) \tag{4.27}$$

[5] In mathematics, they are called Green's functions.

where the summation is taken over all values of k for which this is not zero.

For the sake of completeness, we include the proof.

Proof. Let $\hat{f}$ be the derivative that pays off \$1 if $(i+1, j)$ is reached and zero otherwise, then $\hat{f}$ is a solution to equation (4.26) with the initial condition

$$\hat{f}(i+1,k) = \begin{cases} 0 & \text{if } k \neq j; \\ 1 & \text{if } k = j. \end{cases}$$

The present value of $\hat{f}$ is equal to $Q_{i+1,j}$, i.e. $\hat{f}(0,0) = Q_{i+1,j}$.

At node (i, k), by (4.26), the value of $\hat{f}$ is

$$\hat{f}(i,k) = q(k,j) \cdot exp(-r(i,k)).$$

Let $g_{i,k}$ be the derivative that pays off \$1 if (i, k) is reached and zero otherwise, then $g_{i,k}$ is a solution to equation (4.26) with

$$g_{i,k}(i,l) = \begin{cases} 0 & \text{if } l \neq k; \\ 1 & \text{if } l = k. \end{cases}$$

That is, at time $t = i \cdot \Delta t$, $\hat{f}$ is a linear combinations of $g_{i,k}$

$$\hat{f} = \sum q(k,j) \cdot exp(-r(i,k)) \cdot g_{i,k}.$$

Because both $\hat{f}$ and $\sum q(k,j) \cdot exp(-r(i,k)) \cdot g_{i,k}$ are the solutions to equation (4.26), if they are equal at time $i \cdot \Delta t$, they must be equal everywhere. In particular, they are equal at time 0, that is,

$$\hat{f}(0,0) = \sum q(k,j) \cdot exp(-r(i,k)) \cdot g_{i,k}(0,0)$$

which is just the relation (4.27).

Lemma 4.5 *The discount bond maturing at time $(i+1) \cdot \Delta t$ can be expressed as*

$$P_{i+1} = \sum_j Q_{i,j} \cdot exp(-r(i,j) \cdot \Delta t) \tag{4.28}$$

Proof. The present value of any security that pays off at time $(i+1)\cdot\Delta t$ can be expressed as a portfolio of $\{Q_{i+1,j}\}$. In particular, by the definition of the discount bond price,

$$P_{i+1} = \sum_j Q_{(i+1),j}. \tag{4.29}$$

In (4.29) substituting $\{Q_{i+1,j}\}$ with $\{Q_{i,j}\}$ by using (4.27), we have

$$\begin{aligned}
P_{i+1} &= \sum_j Q_{i+1,j} \\
&= \sum_j \sum_k Q_{i,k} \cdot q(k,j) \cdot exp(-r(i,k)\cdot\Delta t) \\
&= \sum_k \sum_j Q_{i,k} \cdot q(k,j) \cdot exp(-r(i,k)\cdot\Delta t) \\
&= \sum_k Q_{i,k} \cdot \left(\sum_j q(k,j)\right) \cdot exp(-r(i,k)\cdot\Delta t) \\
&= \sum_k Q_{i,k} \cdot exp(-r(i,k)\cdot\Delta t)
\end{aligned}$$

where we have used the fact that

$$\sum_j q(k,j) = 1.$$

Roughly speaking, the forward induction method is to inductively compute the value of θ from (4.27), (4.28).

First, we find $\theta(0)$ by solving (4.28) where $i = 1$, which can be accomplished by a root-finding method. Having found $\theta(0)$, we next extract the value of $\theta(\Delta t)$ by solving (4.28) where $i = 2$ and $Q_{1,j}$ are computed from (4.27). In general, having found $\theta(i\cdot\Delta t)$, we determine $\theta((i+1)\cdot\Delta t)$ from equation (4.28), where $Q_{i+1,j}$ are computed from (4.27). This induction process is continued until we reach the end of the tree under construction.

4.2.7 Numerical results

First we list in Tables 4.1 to 4.4 some numerical results for the Hull-White model using different tree methods that have been discussed. They show that, although adaptive branching rules do not necessarily improve the accuracy, they do improve the efficiency. They also show that the methods using various adaptive branching rules in general have the same accuracy.

As the second test, we look at the CIR model

$$dr = (a - \theta(t) \cdot r) \cdot dt + \sigma \cdot \sqrt{r} \cdot dZ$$

where the mean-reversion rate $\theta(t)$ is chosen to be a function of time to fit the initial term structure. We choose the initial bond prices to be

$$P(0, T) = A(T) \cdot exp(-B(T) \cdot r)$$

where

$$B(T) = \frac{2 \cdot (exp(\gamma \cdot T) - 1)}{(\gamma + a) \cdot (exp(\gamma \cdot T) - 1) + 2 \cdot \gamma}$$

$$A(T) = (\frac{2 \cdot \gamma \cdot exp((a + \gamma) \cdot T/2)}{(\gamma + a) \cdot (exp(\gamma \cdot T) - 1) + 2 \cdot \gamma})^{2 \cdot a \cdot b/\sigma^2}$$

with $a = 0.5, b = 0.06, \gamma = \sqrt{\theta^2 + 2 \cdot \sigma^2}$.[6] We use the tree method to extract the values of $\theta(t)$, and find no appreciable differences between the numerical values and the analytical ones that are equal to 0.5.

In the third example, we look at the cubic variance model

$$dr = (a - \theta(t) \cdot r) \cdot dt + r^{3/2} \cdot \sigma \cdot dZ. \tag{4.30}$$

Introducing $y = 1/\sqrt{r}$, we obtain

$$dy = (-\frac{a \cdot y^3}{2} + \frac{3 \cdot \sigma^2}{8 \cdot y} + \frac{\theta(t) \cdot y(t)}{2}) \cdot dt - \frac{\sigma}{2} \cdot dZ$$

Surprisingly enough, the probability-based tree method (4.9) does not work well for some values of a, σ.

The correctness of the implementation of the model (4.30) is tested using Monte-Carlo simulation.

[6] This is a special solution to the CIR model, see [CIR].

Table 4.1. The price of a 3 year European put option on a 7 year discount bond, using the finite-difference method (4.6), (4.7) without adaptive branching rules. The initial yield curve is $y(t) = 0.08 - 0.05 \cdot exp(-0.18 \cdot t)$, the strike is the forward bond price, $a = 0.1, \sigma = 0.06$.

Time Step	Price	Error
0.160000	0.075887	1.86 %
0.080000	0.074674	0.23 %
0.040000	0.074917	0.56 %
0.020000	0.074684	0.25 %
0.010000	0.074479	0.028 %
0.005000	0.074531	0.042 %
0.002500	0.074501	0

Table 4.2. The price of a 3 year European put option on a 7 year discount bond using finite-difference based method (4.6), (4.7) with adaptive branching rules (4.12) and (4.13). The input is the same as in Table 4.1.

Time Step	Price	Error
0.160000	0.075891	1.87 %
0.080000	0.074678	0.24 %
0.040000	0.074920	0.56 %
0.020000	0.074688	0.24%
0.010000	0.074483	0.023 %
0.005000	0.074535	0.047 %
0.002500	0.074505	0.006 %
0.001250	0.074514	0.018 %

Table 4.3. The price of a 3 year European put option on a 7 year discount bond using the probability based method (4.8) with adaptive branching rules (4.18). The input is as in Table 4.1.

Time Step	Price	Error
0.160000	0.075818	1.7 %
0.080000	0.074914	0.55570 %
0.040000	0.074615	0.154 %
0.020000	0.074701	0.269 %
0.010000	0.074568	0.0912 %
0.005000	0.074523	0.031 %
0.002500	0.074526	0.035 %
0.001250	0.074511	0.0147 %
0.000625	0.074502	0.0027 %
0.000313	0.074503	0.004 %

Table 4.4. CPU time (in minutes) for tree methods with and without adaptive branching rules

Time Step	with adaptive branching rules	without
0.02	0:17	0:25
0.01	1:07	1:39
0.005	5:05	6:23

4.2.8 Computational costs

In this subsection we estimate the computational cost of the tree method in terms of Δy and the length of the tree, T. Throughout this subsection, C_0 to C_3 denote generic constants.

Since $\Delta t \approx C_0 \cdot (\Delta y)^2$, the number of time intervals in the tree is

$$n = \frac{T}{\Delta t} \approx \frac{T \cdot C_1}{(\Delta y)^2}.$$

At the time $k \cdot \Delta t$, the forward induction process takes approximately $2 \cdot k + 1$ floating point calculations. Thus the total cost is about

$$\begin{aligned} \text{Cost} \approx\ & C_1 \cdot (1 + 3 + 5 + \cdots + (2 \cdot n + 1)) \\ \approx\ & C_2 \cdot n^2 \\ \approx\ & \tfrac{C_3 \cdot T^2}{(\Delta y)^4} \end{aligned}$$

So in terms of the number of floating point calculations, the computation cost is equal to

$$\text{Cost} \approx C_3 \cdot \frac{T^2}{(\Delta y)^4}. \tag{4.31}$$

4.2.9 Effect of non-smooth yield curves

If the yield curve is obtained by linear interpolation, then the forward yield curve is not continuous, which may cause serious problems for the accuracy of the numerical solutions.

For example, consider the Hull-White model

$$dr = (\theta(t) - a \cdot r) \cdot dt + \sigma \cdot dZ,$$

where

$$\theta(t) = F_t(0,t) + a \cdot F(0,t) + \frac{\sigma^2}{2 \cdot a} \cdot (1 - exp(-a \cdot t)),$$

and $F(0,t)$ is the instantaneous forward rate and $F_t(0,t) = \frac{\partial F(0,t)}{\partial t}$. The problem is that $F_t(0,t)$ and thus $\theta(t)$ are not well defined at the nodes of the linear interpolation. As a result, the values of θ computed from the tree method have spurious oscillations near the nodes of the interpolation (Figure 4.6). The accuracy of the numerical solution can be compromised and a 5% error can persist even for small time steps such as $\Delta t = 0.01$ for the finite-difference method. However, the error is less significant for the probability based method.

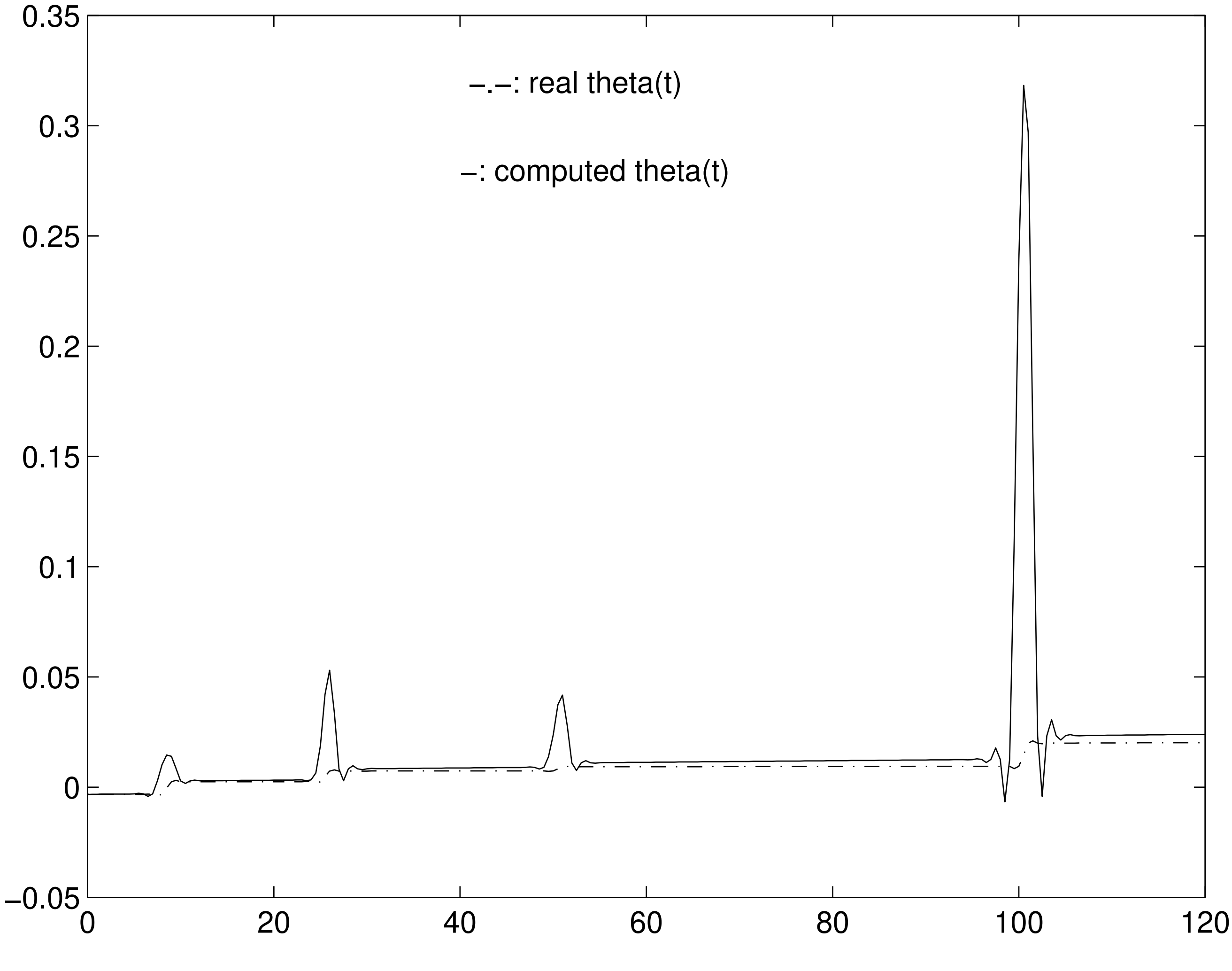


Figure 4.6: The computed parameter θ has spurious oscillations near the nodes

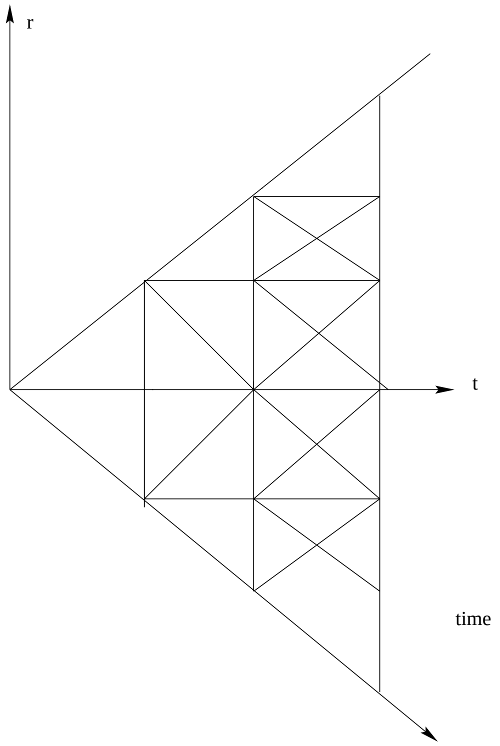


Figure 4.7: The tree in the first step of the Hull-White tree method

4.3 Other methods

4.3.1 Hull-White tree method

Hull and White [Hull-White(1996)] propose a tree method that exploits the special structure of the Hull-White model.

The construction is accomplished in two steps. In the first step, one constructs a tree for the model

$$dr^* = -a \cdot r^* \cdot dt + \sigma \cdot dZ$$

as in Figure 4.7. The probabilities are given by

$$\begin{aligned} p_u &= \tfrac{1}{6} + \tfrac{a^2 \cdot j^2 \cdot \Delta t^2 - a \cdot j \cdot \Delta t}{2} \\ p_m &= \tfrac{2}{3} - a^2 \cdot j^2 \cdot \Delta t \\ p_d &= \tfrac{1}{6} + \tfrac{a^2 \cdot j^2 \cdot \Delta t^2 + a \cdot j \cdot \Delta t}{2} \end{aligned}$$

To make the method more stable, one can introduce adaptive branching rules as described before.

The second step is to adjust the shape of the tree, but not the probabilities, such that the initial term structure is matched by the model. This is done by finding $\alpha(t) = r^*(t) - r(t)$ from the initial yield curve by using the forward induction method that has been discussed earlier.

Since

$$dr = (\theta(t) - a \cdot r) \cdot dt + \sigma \cdot dZ,$$

we have

$$d\alpha = (\theta(t) - a \cdot \alpha) \cdot dt.$$

As α is deterministic, the processes for r and r^* have the same probabilities.

Since $r(m, j) = r^*(j) + \alpha_m$, (4.28) becomes

$$P_{m+1} = \sum_j Q_{m,j} \cdot exp(-(r^*(j) + \alpha_m) \cdot \Delta t),$$

or

$$\alpha_m = -\frac{ln[P_{m+1} \cdot \sum Q_{m,j} \cdot exp(-r^*(j) \cdot \Delta t)]}{\Delta t}$$

Since this gives an explicit formula for α_m, there is no need to solve (4.28) by a root-finding numerical method.

A major advantage of the Hull-White tree is that the probabilities are independent of time (if a, σ are constant) and thus the method is very stable. It can also deal with non-smooth initial forward yield curve $F(0, t)$ better, since it avoids the computation of $\theta(t)$ that is related to the derivative of $F(0, t)$. Instead, it computes $\alpha(t)$ that involves $F(0, t)$ and not its derivative.

This method can also be generalized to the Black-Karasinski model

$$d[ln(r)] = (\theta(t) - a \cdot ln(r)) \cdot dt + \sigma \cdot dZ.$$

Table 4.5 shows that the Hull-White method has the same accuracy as the tree methods discussed earlier.

The main limitation of the Hull-White tree method is that it can only be applied to models of the form of

$$d[f(r)] = (\theta(t) - a \cdot f(r)) \cdot dt + \sigma \cdot dZ$$

Table 4.5. The price of a 3 year put option on a 7 year discount bond using the Hull-White tree method. The input is the same as in Table 4.1.

Time Step	Price	Error
0.160000	0.075771	1.71%
0.080000	0.074726	0.3 %
0.040000	0.074758	0.35 %
0.020000	0.074650	0.2 %
0.010000	0.074598	0.12 %
0.005000	0.074547	0.063%
0.002500	0.074502	0.027 %
0.001250	0.074498	0.0027 %
0.000625	0.074500	0 %

that includes the Hull-White model and the Black-Karasinski model. It does not apply to any model of the form

$$dr = (\theta(t) - a \cdot r) \cdot dt + \sigma(t) \cdot r^{\beta} \cdot dZ.$$

where $\beta \neq 0$.

4.3.2 Other special methods

There are other special numerical procedures for the Hull-White model. For example, [Reisman] develops a tree whose shape is independent of the initial yield curve, and [Kijima-Nagayama] proposes a procedure to construct a trinomial tree by making use of symmetric properties of Wiener processes. While these methods are very efficient, they do not apply to other one-factor models.

Chapter 5

Calibrations of One-factor Models and Numerical Examples

The aim of this chapter is two-fold. First we consider the issue of calibration, that is, the estimation of the parameters in an interest rate model, and survey different approaches to calibration. Second we use a set of market data (zero rates and caplet forward-forward volatilities) to calibrate and test various one-factor interest rate models. Though the test is limited by the sample size, we hope it will lead to suggestions for further studies.

Some authors suggest that interest rate models, once calibrated, produce similar prices (see, e.g. [Lochoff]). We find that, while this is true for some interest rate models, it is not the case for others. Even for at-the-money options or slightly-out-of-the-money put options, different models can produce quite different prices.

5.1 Calibration

5.1.1 What is calibration?

By *calibration* we mean the estimation of parameters of an interest rate model. For example, for the Hull-White model, this means the estimation of a and σ.[1]

Given an interest rate model, there are many different ways to calibrate it (Figure 5.1). A recent survey by the Bank of England reports that many financial institutions calibrate the same model in vastly different ways and, as a result, produce quite different

[1] The volatility structure in an one-factor model such as the Hull-White model is usually determined by two parameters, a and σ. See [Hull]

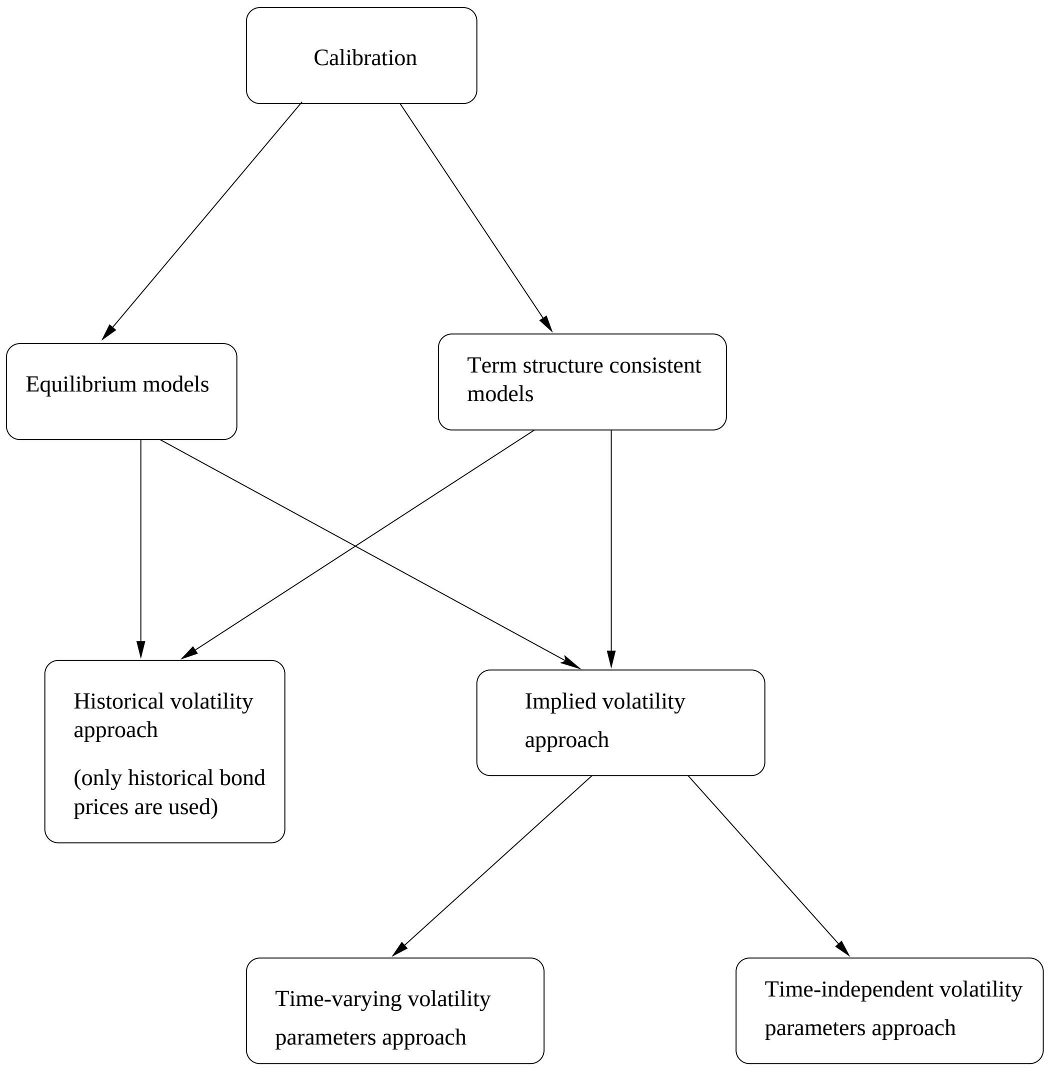


Figure 5.1: Different approaches to calibration

prices for the same interest rate product (see [Weston-Cooper], page 25). It also finds that the situation for equity options is not so serious.

With so much diversity in calibration, no wonder many problems can arise. For example, one may seriously mis-estimate the volatilities and consequently suffer from a loss. Indeed, in a recent article in *Risk*, [Paul-Choudhury], page 19, some losses are attributed to "poor volatility estimations."

We will not discuss the calibration of equilibrium models, whose parameters are all time-independent. Instead, we will concentrate on the so-called term structure consistent models, in which at least one parameter is allowed to be a function of time to fit the initial term structure.

5.1.2 Implied volatility approach versus historical volatility approach

Traditionally there are two approaches to the calibration:

1. *Historical volatility approach:* volatility parameters are inferred from historical data;

2. *Implied volatility approach:* volatility parameters are inferred from current market prices of some options (*benchmark options*). In this approach, in addition to the initial term structure of bond prices, some market prices of options are needed as inputs to the models.

Previous studies have shown that the first approach is inadequate for pricing options (see [Hull]). In addition to its problems with data collecting and estimation, the historical volatility approach can lead to underestimating the volatilities. It is well known that, for equity options, the implied volatilities are much higher than the historical volatilities.

In contrast, the implied volatility approach allows one to price an option consistently with the market expectations, giving one the hope that, if the benchmark options are priced correctly, so will be the option that is to be priced.

5.1.3 Implied volatility approach

When the implied volatility approach is used, one is faced with several questions. In particular,

1. Which options should be chosen as the benchmark options?

2. Which parameters should be time-varying and which ones should be constant?

There is some general consensus about the first question. Some criteria for choosing benchmark options are listed below.

- The benchmark options should be similar to the product to be priced. (However, what constitutes "similar" is subject to one's judgement.)
- The benchmark options should be liquid.

Some commonly used benchmarks are caplets, options on U.S. Treasury products and swaptions.

So far there is little consensus about the second question. Some practitioners use time-varying volatility parameters to fit, for example, an initial term structure of volatilities (see, e.g. [Canabarro]). However, Hull ([Hull], page 450) cautions about the use of time-varying volatility parameters, arguing that it could lead to mis-pricing of compound options such as captions. Hull-White ([Hull-White(1996)]) proposes that only one parameter, e.g. the long-run mean, should be time-varying to fit the initial term structure.

Time-independent volatility parameters approach

If we postulate that the volatility parameters in a model are constant, it is desirable that they are chosen to fit a group of options prices of *different maturities* as closely as possible. For example, if $P_{1,market}, \cdots, P_{n,market}$ are the market prices of caplets of different maturities, and $P_{1,model}, P_{2,model}, \cdots, P_{n,model}$ the corresponding prices generated from the model, then the parameters in the model should be chosen so that $P_{1,model}, P_{2,model}, \cdots, P_{n,model}$ are as close to $P_{1,market}, \cdots, P_{n,market}$ as possible. For example, the parameters can be chosen so that the mean squared relative error

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{P_{i,model}}{P_{i,market}}-1\right)^2 \tag{5.1}$$

is minimized.

This approach offers the advantage of safety. If the mean squared relative error is small, then one is more confident that the model can be used to price an option similar to the benchmark options.

A numerical optimization package is needed to minimize the mean square error (5.1).

Time-varying volatility parameters approach

In this approach one allows all the parameters in the model to be functions of time to fit the present market values of some benchmark options of various maturities. This approach is more flexible than the constant volatility parameter approach; however, one should be aware of its danger of overfitting. Whenever it is used, one should check if the volatility parameters exhibit severe non-stationarity.

To illustrate this approach, first assume that the interest rate model has two volatility parameters $a = a(t)$, $\sigma = \sigma(t)$ as functions of time. To estimate them, we choose two caps as the benchmark options. For simplicity, further assume that the two caps have the same reset dates $T_1 < T_2 < \cdots < T_n$, but different cap rates. (A cap consists of a sequence of caplets maturing at $T_1 < T_2 < \cdots < T_n$. A caplet maturing at T_i pays off an amount equal to

$$max(0, L_i - capRate)$$

at T_{i+1}. Here $capRate$ is a cap rate, and L_i is the LIBOR rate observed at T_i.) Let $P_{i,market}, Q_{i,market}$ be the prices of the caplets maturing at T_i (with different cap rates).

We first choose $a(t)$, $\sigma(t)$ to be constant over $[0, T_1]$ such that market prices $P_{1,market}, Q_{1,market}$ maturing at T_1 are matched exactly. Having determined $a(t)$, $\sigma(t)$ over time $[0, T_1]$, we then choose $a(t)$, $\sigma(t)$ to be constant over $[T_1, T_2]$ such that market prices $P_{2,market}, Q_{2,market}$ maturing at T_2 are matched. We continue until all $P_{i,market}, Q_{i,market}$ are matched. In this way, all the caplet prices are fit exactly at the expense of introducing non-stationary volatilities.

5.2 Numerical examples

In this section we will use market data (yield curve and caplet prices) to calibrate the following five interest rate models: [2]

[2]Black and Karasinski [Black-Karasinski], Hull and White [Hull-White(1990a)] formulate their models more generally (allowing a and σ to be functions of time). Here we choose σ and a to be constant so that the models have stationary volatility structure.

1. The Black-Karasinski model (BK)

$$d[log(r)] = (\theta(t) - a \cdot log(r)) \cdot dt + \sigma \cdot dZ$$

2. The CIR model (CIR)

$$dr = (a - \theta(t) \cdot r) \cdot dt + \sigma \cdot \sqrt{r} \cdot dZ$$

3. The cubic-variance model

$$dr = (a - \theta(t) \cdot r) \cdot dt + \sigma \cdot \sqrt{r} \cdot dZ$$

4. The Ho-Lee model (Ho-Lee)

$$dr = \theta(t) \cdot dt + \sigma \cdot dZ$$

5. The Hull-White model (HW)

$$dr = (\theta(t) - a \cdot r) \cdot dt + \sigma \cdot dZ$$

5.2.1 Description of the data

The data is obtained from the U.S. market on June 30, 1997. We first list the zero spot rates in Table 5.1:

Table 5.1. U.S. Term Structure on June 30, 1997

Maturity (in years)	spot rate
0.008219	0.047662
0.027397	0.049207
0.065753	0.048846
0.257534	0.051542
0.413699	0.051855
0.509589	0.052089

U.S. Term Structure (continued)	
Maturity (in years)	spot rate
0.756164	0.053908
0.986301	0.055320
1.125425	0.057346
2.127773	0.059245
3.125700	0.061303
4.124532	0.061799
5.128154	0.062290
6.126391	0.062876
7.118730	0.063656
8.127626	0.064528
9.135703	0.064819
10.124016	0.065307

The yield curve is obtained by piecewise linear interpolation between the above rates.

The *forward-forward volatilities* are the volatilities to be used in the Black's formula for individual caplets (see Appendix A). We collect their data near the end of June 1997 in Table 5.2 (see also Figure 5.2).

Table 5.2. Forward forward volatilities on June 30, 1997	
Maturity (in years)	vol
0	0.0857
0.25	0.1281
0.5	0.1531
0.75	0.1680
1	0.1785
1.25	0.1849

Forward forward volatilities (continued)	
Maturity (in years)	vol
1.50	0.1878
1.75	0.1883
2.00	0.1867
2.25	0.1835
2.50	0.1789
2.75	0.1734
3.00	0.1674
3.25	0.1612
3.50	0.1551
3.75	0.1493
4.00	0.1441
4.25	0.1393
4.50	0.1361
4.75	0.1337
5.00	0.1321
5.25	0.1313
5.50	0.1310
5.75	0.1312
6.00	0.1317
6.25	0.1324
6.50	0.1332
6.75	0.1339
7.00	0.1346
7.25	0.1352
7.50	0.1357

Forward forward volatilities (continued)	
Maturity (in years)	vol
7.75	0.1362
8.00	0.1366
8.25	0.1370
8.50	0.1374
8.75	0.1377
9.00	0.1380
9.25	0.1383
9.50	0.1386

Given a cap rate R_X, we can generate caplet prices using the Black's formula and Table 5.2. More precisely, for a caplet maturing at $T = i \cdot 0.25$, we first look up the spot rate with maturity at the next reset date, $y((i+1) \cdot 0.25)$, and the corresponding forward-forward volatility $\sigma(i)$, then compute the caplet price

$$P_{i,market} = 0.25 \cdot exp(-y((i+1) \cdot 0.25)) \cdot (F \cdot N(d_1) - R_X \cdot N(d_2))$$

where

$$d_1 = \frac{ln(F/R_x) + \sigma(i)^2 \cdot T/2}{\sigma(i) \cdot \sqrt{T}}$$

$$d_2 = \frac{ln(F/R_x) - \sigma(i)^2 \cdot T/2}{\sigma(i) \cdot \sqrt{T}}$$

and F is the forward rate for the period between $i \cdot 0.25$ and $(i+1) \cdot 0.25$. In this way, we obtain about forty caplet prices with maturities ranging from 0.25 years to 10 years.

5.2.2 Fitting the models to the caplets prices

After having generated market caplet prices $P_{1,market}, \cdots, P_{40,market}$ from the forward-forward volatilities, we test how closely the model prices produced by various one-factor interest rate models match the caplet prices. We keep σ and a constant and minimize

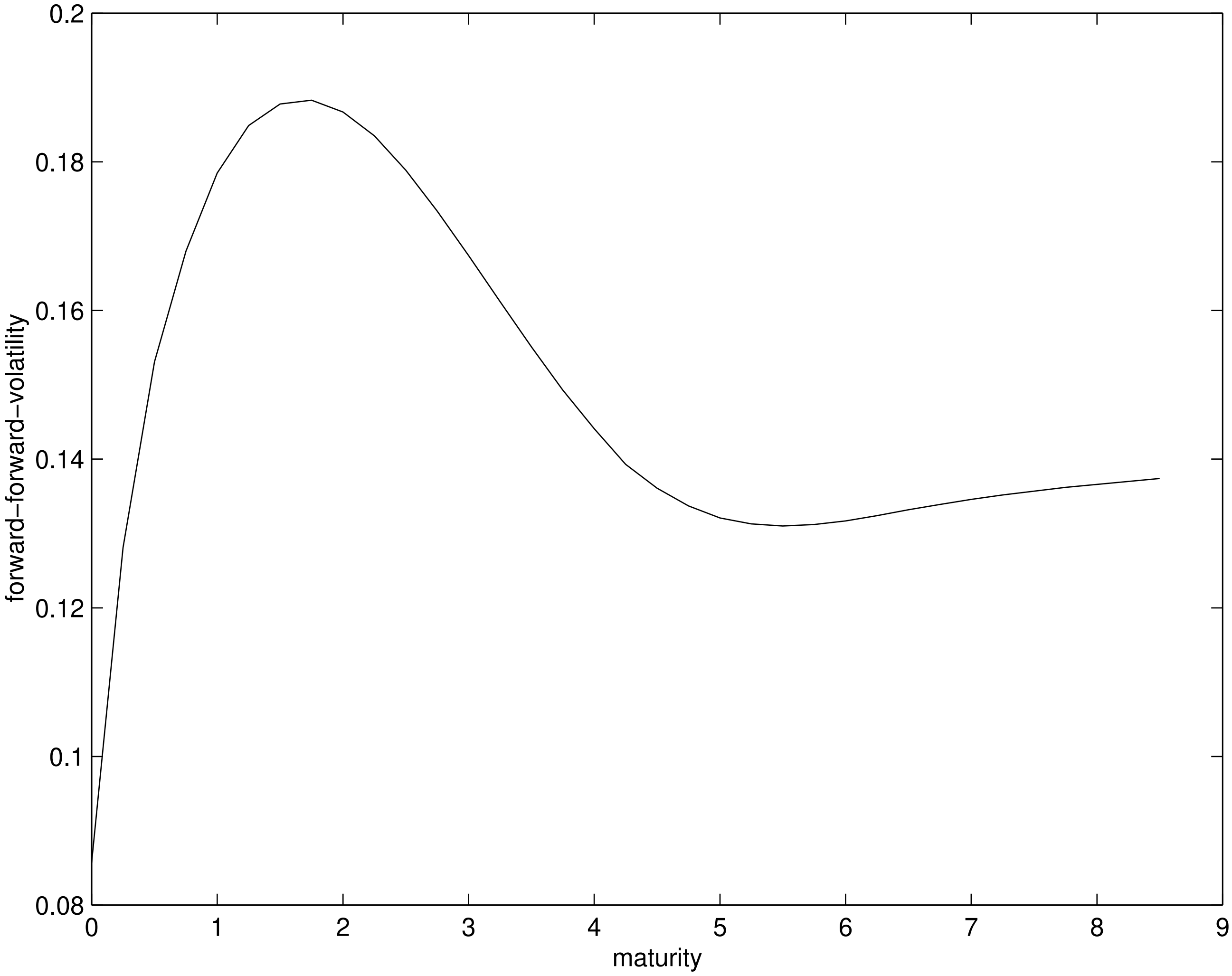


Figure 5.2: The forward-forward volatility

either the mean-square relative error norm

$$\frac{1}{k_1 - k_0 + 1} \cdot \sum_{k=k_0}^{k_1} \left(\frac{P_{i,model}}{P_{i,market}} - 1\right)^2 \tag{5.2}$$

or the mean-square absolute error norm

$$\frac{1}{k_1 - k_0 + 1} \cdot \sum_{k=k_0}^{k_1} \cdot (P_{i,model} - P_{i,market})^2$$

where

- $P_{i,model}$ is the price generated by the model for the caplet maturing at T_i,

- $P_{i,market}$ is the market price for the caplet maturing at T_i.

After experimenting with two error norms, we decide to use the mean-square relative error norm as it seems to give sharper estimates. Since the mean-square relative error norm allows smaller market caplet prices to have a bigger weight than larger ones, we disregard the first few caplet prices which are usually very small. Thus, in (5.2) we consider only those caplet prices having maturity of at least 6 months, $k_0 \geq 2$.

The fitting results are summarized in Table 5.3. The Black-Karasinski, CIR, cubic variance and Hull-White models fit the caplet prices quite well for caplets of maturity bigger than 1 year. In contrast, the Ho-Lee model does not fit the caplet prices well.

However, it can be seen from the last row of Table 5.3 that, for caplets starting with maturity of 0.5 year ($k_0 = 2$), the fitting error becomes much larger. This is because one-factor models cannot reproduce the hump in the forward-forward volatilities in Figure 5.2 (compare Figure 5.3). Figure 5.4 shows that the relative error of the fitting increases sharply for caplets of maturity $0.5 - 0.75$ years.

Table 5.3. Implied volatility parameters and fitting errors (cap rate = 6 %)

Model	k_0	k_1	# of caplets	σ	a	$e/\sqrt{\# \text{ of caplets}}$
Black-Karasinski	4	34	30	0.20	0.14	0.008115
CIR	4	34	30	0.049	0.0090	0.008045
cubic variance (CU)	4	34	30	0.90	0.011	0.008408
Ho-Lee	4	34	30	0.00900	0	0.014455
Hull-White	4	34	30	0.012	0.12	0.008680
Hull-White	4	16	12	0.013	0.17	0.008742
Hull-White	2	16	14	0.012	0.17	0.012195

where e is

$$e = \{\frac{1}{\# \text{ of caplets}} \sum_{i=k_0}^{k_1} (\frac{P_{i,model}}{P_{i,market}} - 1)^2\}^{1/2}$$

$$\# \text{ of caplets} = k_1 - k_0 + 1$$

5.2.3 Comparison of prices of put options by different models

Having determined the parameters of the five interest rate models in the previous subsection, we consider the prices produced by the five models for a 3-year put option on a 8-year discount bond (Figure 5.5). The results can be summarized as follows:

1. All the models, except that of Ho-Lee, produce similar prices for the at-the-money option. This is consistent with the conclusion of [Hull-White(1993)], but significantly different from the conclusion of [Uhrig-Walter] which uses the historical volatility approach in the calibration.

2. The Ho-Lee model generates a price for the at-the-money option that is much higher than those generated by the other models. This is somewhat surprising as it is considered easy to price an at-the-money option.

3. The Black-Karasinski model and the CIR model produce similar prices for both in-the-money and out-the-money options. This somehow confirms the conclusion of

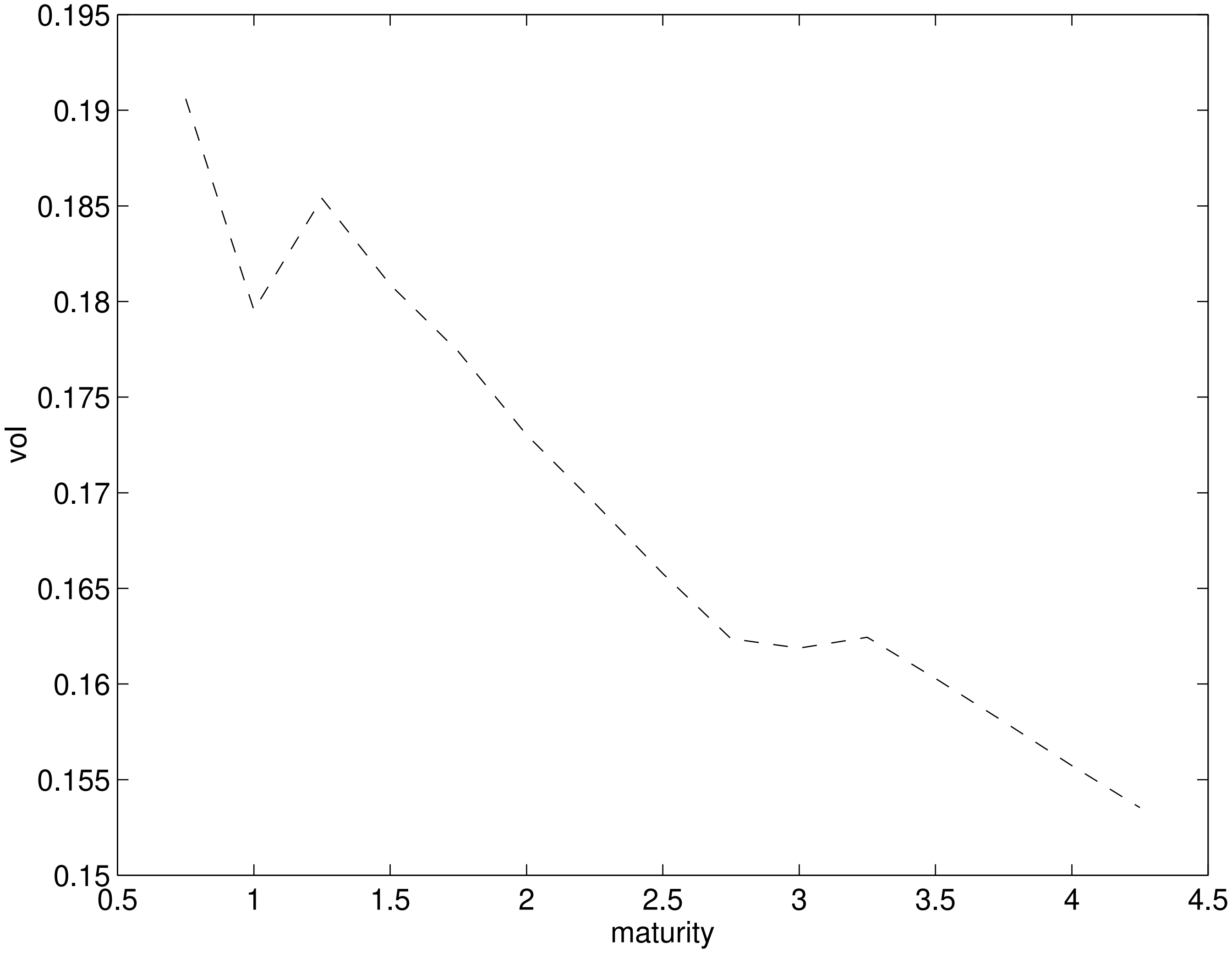


Figure 5.3: The forward-forward vol. implied by the prices produced by the HW model

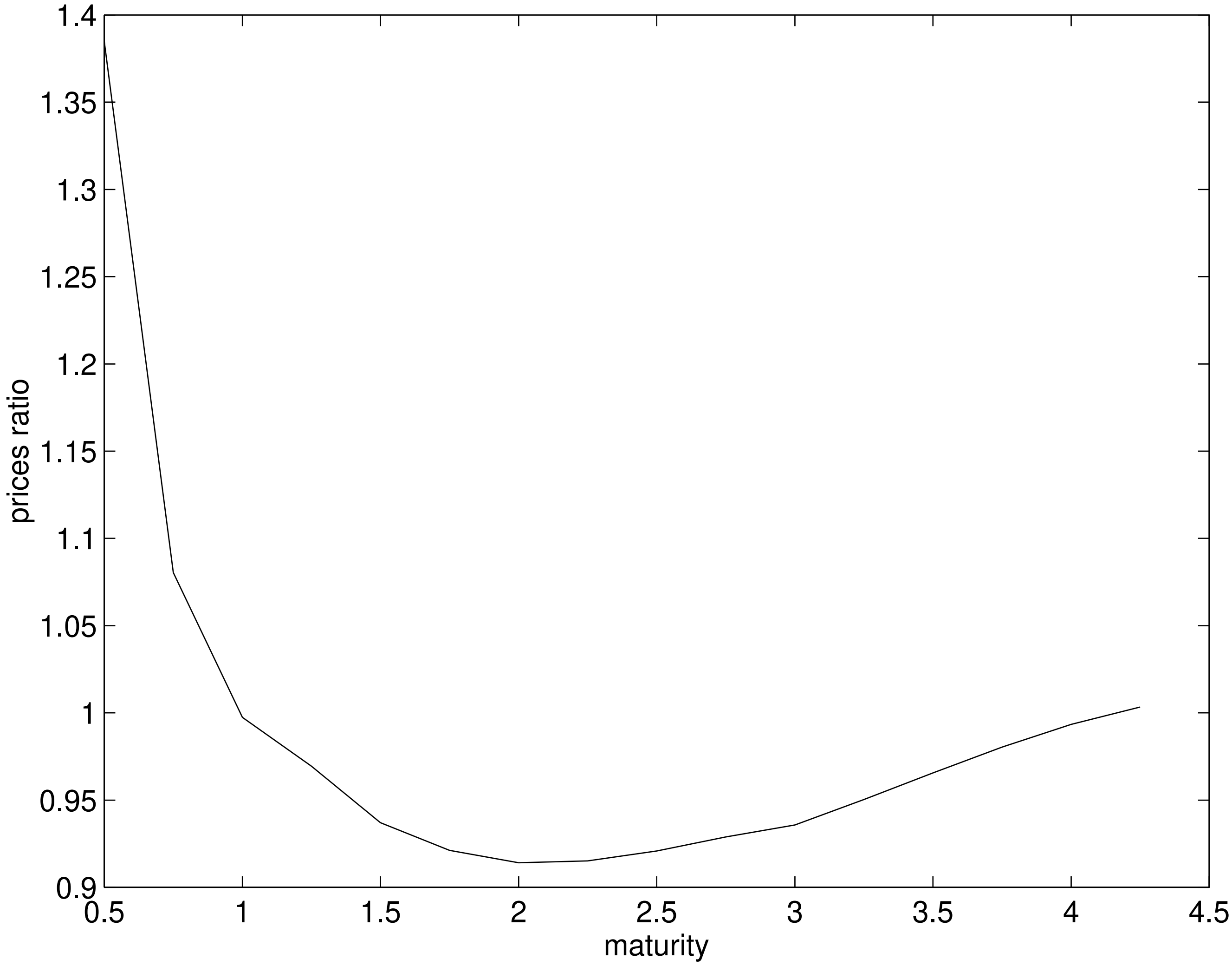


Figure 5.4: The caplet prices generated by the HW model are shown as a propotion of the market caplet prices

Canabarro [Canabarro], who compares the CIR model and the Black-Derman-Toy model (which is similar to the Black-Karasinski model), and Lochoff [Lochoff].

4. However, even for slightly out-of-the-money options, both the cubic-variance and the Hull-White models produce prices that are quite different from thoses generated by the Black-Karasinski and CIR models.

5. Except the Ho-Lee model, for deep-out-of-the-money put options, the more sensitive the volatility of a model is to the level of the interest rate, the higher the prices the model produces.

Table 5.4. Comparison of prices given by different models on 3-year put option on a 8-year discount bond

Strike price	BK model	CIR model	Ho-Lee model	Hull-White model	CU model
0.80	0.000117	0.000095	0.000088	0.000008	0.000545
0.85	0.000500	0.000462	0.000517	0.000127	0.001222
0.90	0.001894	0.001877	0.002214	0.001039	0.002788
0.93	0.003885	0.003985	0.004676	0.002847	0.004651
0.95	0.006055	0.006265	0.007326	0.005079	0.006518
0.97	0.009169	0.009496	0.010970	0.008396	0.009149
1.00	0.016012	0.016555	0.018620	0.015857	0.015142
1.05	0.034207	0.034808	0.037262	0.035024	0.032760
1.10	0.059668	0.059949	0.061669	0.060418	0.059793
1.15	0.088452	0.088492	0.089232	0.088718	0.089594
1.20	0.117883	0.117883	0.118110	0.117940	0.119459
1.30	0.176824	0.176823	0.176853	0.176835	0.179188
1.40	0.235765	0.235764	0.235784	0.235776	0.238918

Similar conclusions hold for a 4-year put option on a 8-year discount bond (see Figure 5.6).

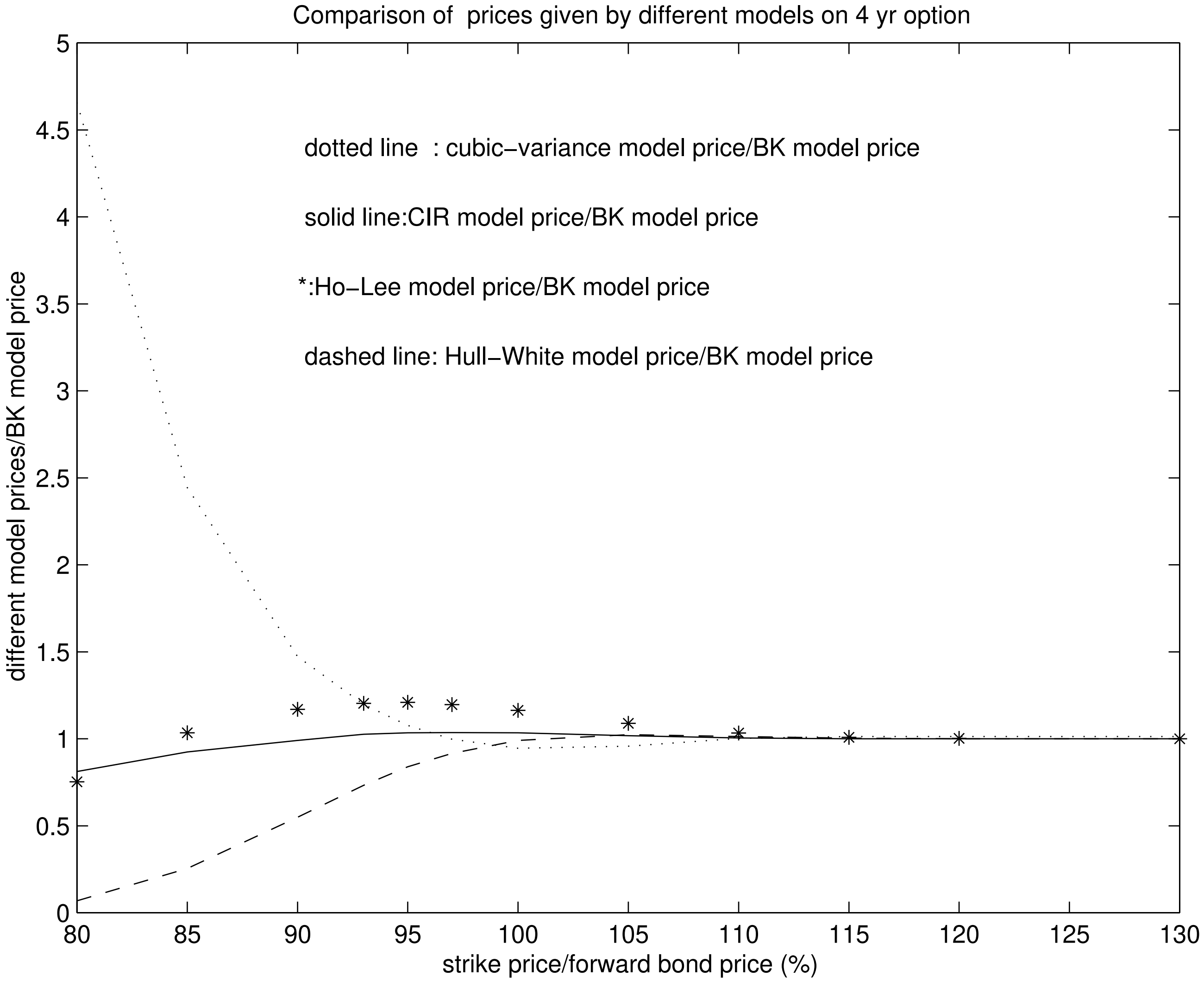


Figure 5.5: Comparison of prices of 3-year put option on a 8-year discount bond given by different models. The prices shown are the ratios of the prices given by different models divided by the price given by the Black-Karasinski model. The strike price is shown as a proportion of the forward bond price.

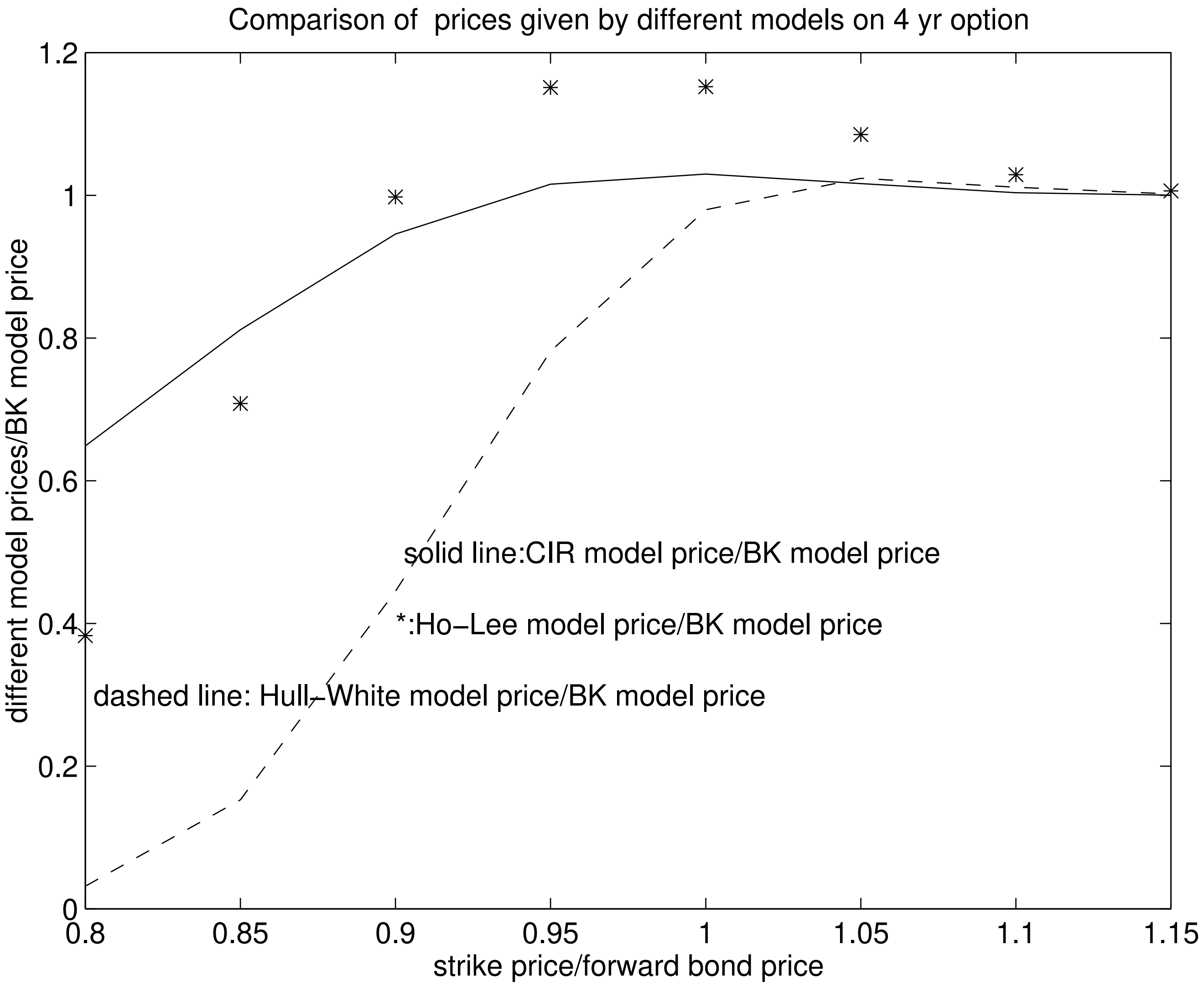


Figure 5.6: Comparison of prices of 4-year put option on a 8-year discount bond given by different models. The prices shown are the ratios of the prices given by different models divided by the price given by the Black-Karasinski model. The strike price is shown as a proportion of the forward bond price.

Chapter 6

Negative Interest Rates in the Hull-White Model

Since the Hull-White model assumes that the interest rate is normally distributed, there is a positive possibility that the interest rate will become negative. As a result, the Hull-White model has often been criticized for producing negative interest rates (see, e.g. [Rogers]). In this chapter, using the volatility parameters implied from the U.S. market data (yield curves and caplet prices) near the end of June 1997, we compute the probability that the interest rate will become negative. Our result suggests that the negative interest rates in the Hull-White model are *insignificant* for today's U. S. market.

Next we consider a modification of the Hull-White model using an idea of Black [Black] that truncates the negative interest rates to zero. Using the implied volatility parameters, the modified model produces almost the same prices as the original model. This somehow confirms our conclusion about the insignificance of the negative interest rates in the Hull-White model in today's U.S. market.

6.1 The probability of negative interest rates

In this section, we study the probability of negative interest rates and the contribution of negative interest rates to bond prices.

First we make clear what we mean by the contribution of negative interest rates to bond prices. As before, let $P(t, T)$ denote the price at time t of a discount bond maturing at T. We decompose $P(t,T)$ into the sum of two derivatives P_{po}, P_{ne},

$$P(t,T) = P_{po}(t,T) + P_{ne}(t,T) \tag{6.1}$$

where

- $P_{ne}(t,T)$ is the price at time t of a derivative that pays off \$1 at T if and only if the interest rate becomes negative sometimes before T,
- $P_{po}(t,T)$ the price at time t of a derivative that pays off \$1 at T if and only if the interest rate stays non-negative until T.

We take *the contribution of negative interest rates to the bond price* to be the ratio $P_{ne}(t,T)/P(t,T)$.

The decomposition (6.1) is best explained in terms of a Monte-Carlo simulation. Recall that in a Monte-Carlo simulation, we generate N random sample paths for the Hull-White interest rate model. A discount bond price is computed as

$$P(t,T) \approx \frac{1}{N}\sum_{i=1}^{N} exp(-\int_t^T r(s)\cdot ds) \tag{6.2}$$

where

- the integral $\int_t^T r(s)\cdot ds$ is the integral of r along a sample path,
- the summation $\sum_{i=1}^{N}$ is over all sample paths.

Now P_{ne} is the part of summation in (6.2) over sample paths along which r becomes negative

$$P_{ne}(t,T) = \frac{1}{N}\sum_{r(s)<0 \text{ for some } s} exp(-\int_t^T r(s)\cdot ds),$$

and P_{po} is the part of summation (6.2) over sample paths along which r remains non-negative

$$P_{po}(t,T) = \frac{1}{N}\sum_{r(s)\geq 0 \text{ for all } s} exp(-\int_t^T r(s)\cdot ds).$$

The probability that the interest rate becomes negative is N_{ne}/N, where N_{ne} is the number of sample paths along which r becomes negative at some point.

Obviously, the probability of negative interest rates depends on the values of the inputs a, σ, and $\theta(t)$. For example, if

- θ becomes smaller;
- σ becomes larger;
- a becomes smaller;

the possibility of negative interest rates becomes larger. Therefore, it is necessary that we compute the probability of negative interest rates not only for the values of a and σ implied from the U.S. market at the end of June 1997, but also for various nearby values of a and σ as well.

Table 6.1 lists the results. It shows that both the probability of negative interest rates and the contribution to the bond price from the negative interest rates are small ($\approx 2-5$ %). Table 6.2 shows a few values of a, σ that do produce significant probability of negative interest rates. (See also Figure 6.1.)

We also price a floor at zero that consists of 40 floorlets of maturities ranging from 0.25 year to 10 year. At each resetting date $T = i \cdot 0.25$, the *floor* pays off an amount equal to $L \cdot max(-Libor_i, 0) \cdot 0.25$, where $Libor_i$ is the LIBOR rate observed at the previous resetting date, L is the notional principal. We find that, for a notional principal of \$ 100, the price of the floor produced by the Hull-White model is about \$ 0.0015.

6.2 Truncating negative interest rates in the Hull-White model

If one is worried about the negative interest rates in the Hull-White model, one can modify the model so that the modified model only produce non-negative interest rates. Rogers ([Rogers], page 44) proposes reflecting the negative interest rates to obtain non-negative interest rates. In this section we consider another modification of the Hull-White model.

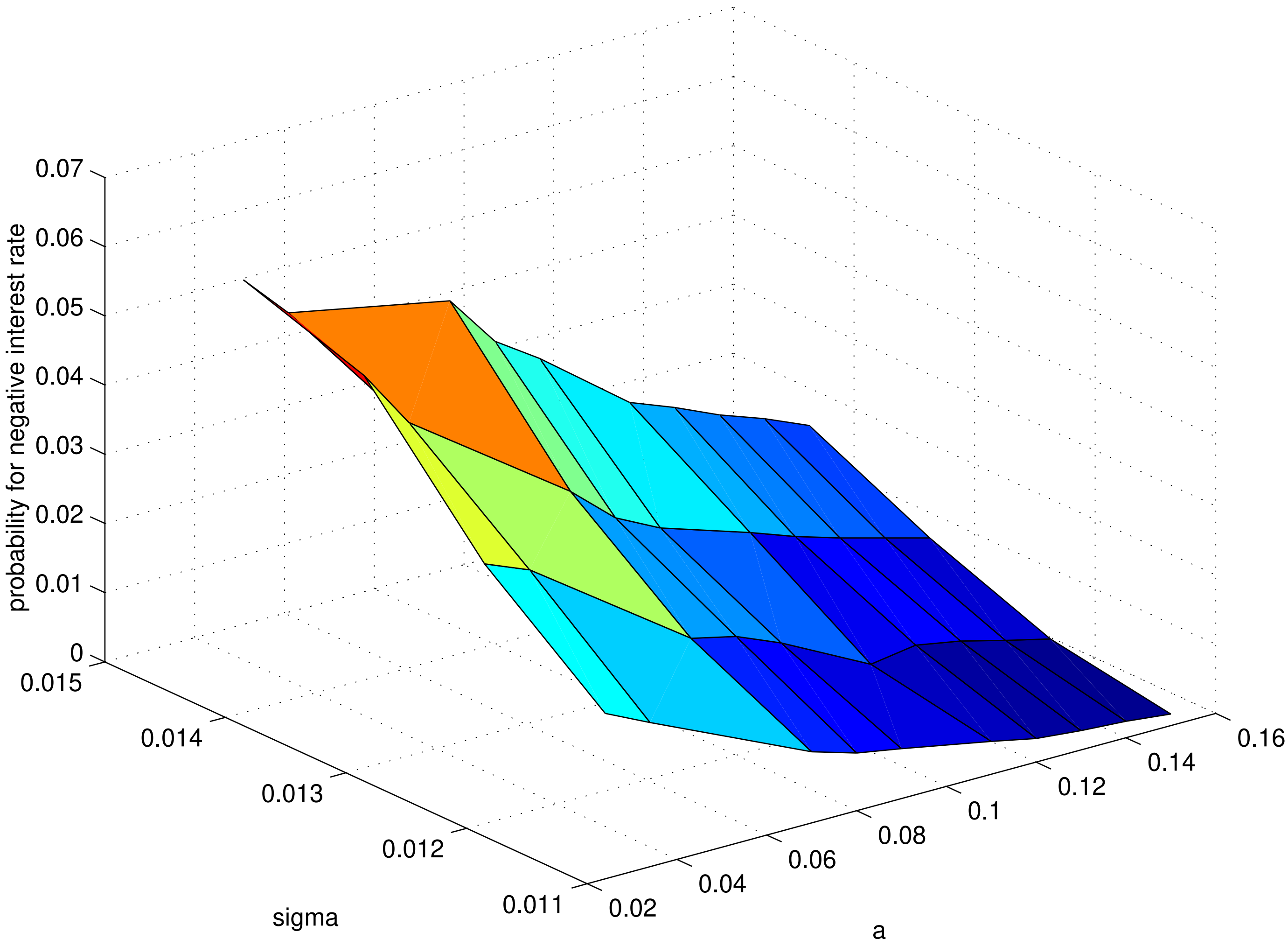


Figure 6.1: The probability of negative interest rates for various values of a and σ

The modified model is obtained by truncating the negative interest rates in the Hull-White model using an idea of Black [Black]. Black argues that interest rates can be considered as an option. When the interest rates offered by a bank become negative, people would prefer to put their money under the mattress. Thus we have

$$r_{investor} = max(r_{bank}, 0),$$

where r_{bank} is the interest rate offered by a bank, and $r_{investor}$ is the interest rate on an investor's money.

Table 6.1. The probability of negative interest rate in 10 year P_{ne}/P is the contribution of the negative interest rates to the bond price

a	σ	P_{ne}/P	probability of negative r
0.100000	0.014000	0.05	0.035950
0.100000	0.016000	0.10	0.072050
0.070000	0.011000	0.014	0.010275
0.070000	0.012000	0.02	0.018675
0.070000	0.013000	0.0461	0.031875
0.070000	0.014000	0.07	0.051425
0.080000	0.011000	0.012	0.008350
0.080000	0.012000	0.0248	0.017175
0.080000	0.013000	0.0379	0.026300
0.080000	0.014000	0.0632	0.043800
0.090000	0.011000	0.01	0.007225
0.090000	0.012000	0.02	0.014500
0.090000	0.014000	0.0568	0.039550
0.090000	0.013000	0.033	0.023075
0.100000	0.014000	0.0484	0.034025

Table 6.1 (Continued)			
a	σ	P_{ne}/P	probability of negative r
0.110000	0.011000	0.0068	0.004750
0.110000	0.013000	0.027	0.018900
0.110000	0.014000	0.042	0.029700
0.120000	0.011000	0.0049	0.003425
0.120000	0.012000	0.0126	0.008900
0.120000	0.013000	0.0234	0.016575
0.120000	0.014000	0.038	0.027200
0.130000	0.011000	0.004	0.002800
0.130000	0.012000	0.011	0.007725
0.130000	0.013000	0.02	0.014625
0.130000	0.014000	0.034	0.024375
0.140000	0.011000	0.003	0.002400
0.140000	0.012000	0.0085	0.006050
0.140000	0.013000	0.0179	0.012850
0.140000	0.014000	0.0308	0.022075
0.150000	0.011000	0.002	0.001675
0.150000	0.012000	0.0063	0.004525
0.150000	0.013000	0.015	0.011100
0.150000	0.014000	0.027	0.019325
0.024000	0.011000	0.0352	0.023900
0.024000	0.012000	0.0556	0.037525
0.034000	0.011000	0.031	0.020825
0.034000	0.012000	0.05	0.034875
0.034000	0.013000	0.07	0.048150
0.024000	0.013000	0.0840	0.056725

Table 6.1 (Continued)			
a	σ	P_{ne}/P	probability of negative r
0.100000	0.012000	0.0178	0.012450
p 0.100000	0.013000	0.031	0.021750

Table 6.2. The values of a and σ necessary to produce significant probability of negative interest rates			
a	σ	P_{ne}/P	probability of negative r
0.024000	0.014000	0.1144	0.076950
0.034000	0.014000	0.1	0.070850
0.100000	0.020000	0.23	0.164600

Now we modify the Hull-White model as follows. This modification is best explained by considering the trinomial tree version of the Hull-White model. Suppose the branching rule in the Hull-White model is as in Figure 6.2. Then the modified model (for lack of terminology, we call it the *HWB model*) has the following branching rules:

1. When the interest rate is positive, the branching rule is the same as that in the Hull-White model;

2. When the interest rate r is zero, the branching rule is shown in Figure 6.2, and the possibility of r moving up is

$$\hat{p}_u = \frac{\theta \cdot \Delta t}{\Delta r} \tag{6.3}$$

while the possibility of r staying at zero is

$$\hat{p}_m = 1 - \frac{\theta \cdot \Delta t}{\Delta r} \tag{6.4}$$

and the possibility of r becoming negative is

$$\hat{p}_d = 0.$$

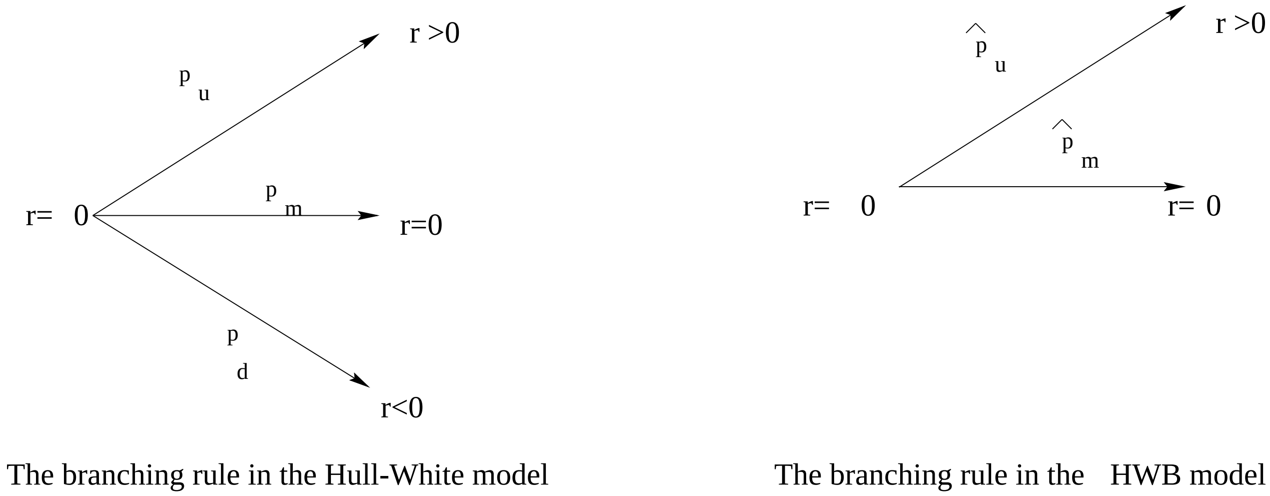


Figure 6.2: Modify the branching rule in the Hull-White model

Since $\Delta t/\Delta r$ is very small, when the interest rate r is zero, (6.3) and (6.4) say that there is a small possibility of r becoming positive, and a larger possibility of r staying at zero. This is consistent with the experience of U. S. interest rates in the 1930s ([Black]).

Roughly speaking, the HWB model is obtained by truncating the negative interest rates in the Hull-White model to zero.

Taking the continuous time limit of the branching rules for the HWB model, we can write down the PDE and the boundary condition for the HWB model. When the interest rate is positive, the PDE for the HWB model is the same as that for the Hull-White model:

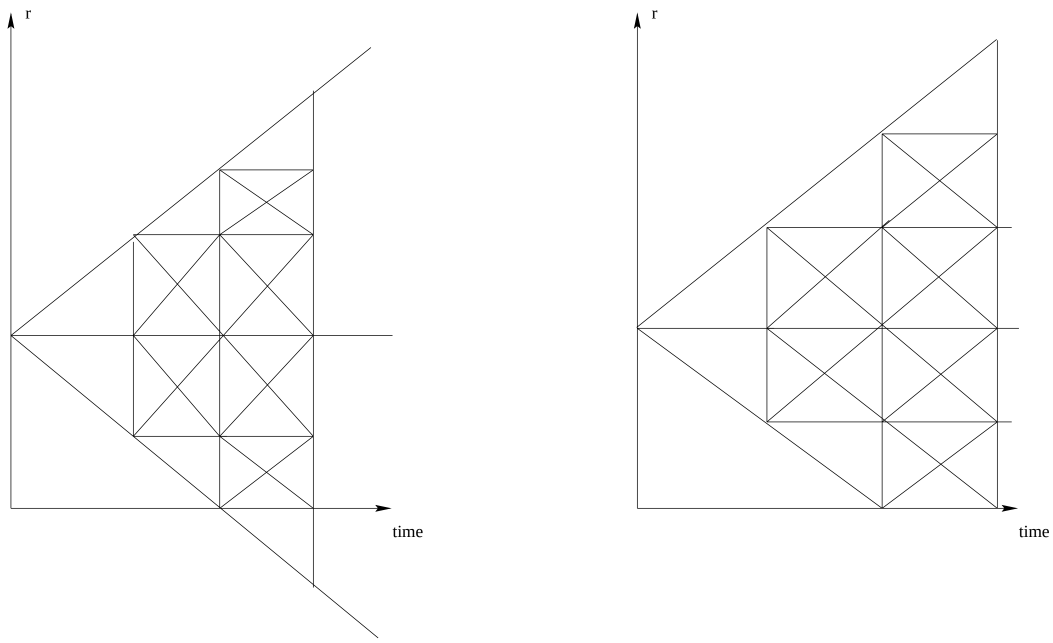
$$\frac{\partial f}{\partial t} + (\theta(t) - a \cdot r) \cdot \frac{\partial f}{\partial r} + \frac{\sigma^2}{2} \cdot \frac{\partial^2 f}{\partial r^2} = r \cdot f, \qquad r > 0.$$

However, the HWB model has a boundary condition at $r = 0$:

$$\frac{\partial f}{\partial t} + \theta(t) \cdot \frac{\partial f}{\partial r} = 0, \qquad r = 0. \tag{6.5}$$

In fact, (6.5) can be proved as follows. From the branching rule for the case $r = 0$, we have

$$f(i,j) = \hat{p}_u \cdot f(i+1, j+1) + \hat{p}_m \cdot f(i+1, j),$$



Trinomial tree for HW model

Trinomial tree for HWB model

Figure 6.3: Comparison of the Hull-White model and the HWB model

which can be rewritten as

$$\frac{f(i+1,j)-f(i,j)}{\Delta t}+\theta\cdot\frac{f(i+1,j+1)-f(i+1,j)}{\Delta r}=0.$$

Taking the limit $\Delta t \to 0, \Delta r \to 0$, we obtain (6.5).

How different is the Hull-White model from the HWB model? Obviously the answer depends on the values of a, σ and θ, as they affect the possibility of negative interest rates in the Hull-White model. We compare the two models using the implied values of a, σ from the U.S. market at the end of June, 1997 (see Table 6.3). The prices produced by the HWB model are very similar to those produced by the Hull-White model. However, if σ is increased to 0.04, then the difference between the prices produced by the two models increases to more than 20%.

Table 6.3. A comparison of caplet prices from the Hull-White model and from the HWB model with the same input (the yield curve on June 30, 1997, $a = 0.12$ and $\sigma = 0.012$)

Maturity (years)	HWB model	HW model	relative difference (%)
7.5	0.002346	0.002390	1.8
7.75	0.002360	0.002400	1.69
8.00	0.002133	0.002172	1.83
8.25	0.001879	0.001931	1.02

Appendix A. Application: Black's model for caps

The *LIBOR rate* for the time period δ is

$$L(t,T,\delta) = \frac{1}{\delta} \cdot \left(\frac{P(t,T)}{P(t,T+\delta)} - 1\right)$$

This is the simply compounded forward rate between T and $T+\delta$, implied from the bond prices maturing at T and $T+\delta$ respectively:

$$\frac{P(t,T+\delta)}{P(t,T)} = \frac{1}{1+\delta \cdot L(t,T,\delta)}.$$

A cap is a portfolio of caplets maturing at $T_1 < T_2 < \cdots < T_n$, $T_i = T_1 + (i-1)\cdot\delta$. A *caplet* maturing at T_i is a contract that pays off

$$max(0, L_i - X)$$

at T_{i+1}. Here X is a predetermined cap rate, and $L_i = L(T_i, T_i, \delta)$ is the LIBOR rate observed at T_i.

Let $c_i(t)$ be the price of the caplet that pays off at T_{i+1}. Since the payoff time is T_{i+1}, we choose $g(t) = P(t, T_{i+1})$ as the numeraire. By (2.24) we have

$$\begin{aligned} \frac{c_i(0)}{P(0,T_{i+1})} &= E_{P(\cdot,T_{i+1})}\left(\frac{c_i(T_{i+1})}{P(T_{i+1},T_{i+1})}\right) \\ &= E_{P(\cdot,T_{i+1})}(max(0, L(T_i,T_i,\delta) - X)) \end{aligned} \tag{6.6}$$

If we further assume that $L(T_i,T_i,\delta)$ is lognormally distributed in the forward risk-neutral world, then we can calculate (6.6) explicitly, which leads to *Black's formula for caps.*

Lemma 6.1

$$E_{P(\cdot,T_{i+1})}(max(0, L(T_i,T_i,\delta) - X)) = exp(r \cdot T_i) \cdot m_1 \cdot N(d_1) - X \cdot N(d_2).$$

where

$$m_1 = L(0,T_i,\delta) \tag{6.7}$$

$$d_1 = \frac{ln[m_1/X] + \sigma^2 \cdot T_i/2}{\sigma \cdot \sqrt{T_i}}. \tag{6.8}$$

$$d_2 = \frac{ln[m_1/X] - \sigma^2 \cdot T_i/2}{\sigma \cdot \sqrt{T_i}}. \tag{6.9}$$

and $N(x)$ is the cumulative normal distribution function.

Proof. Let

$$L = L(T_i, T_i, \delta),$$
$$m = E_{P(\cdot, T_{i+1})}(L).$$

Since L is lognormal we may assume

$$ln[L] = \phi(\ ln(m) - \frac{\sigma^2 \cdot T_i}{2}, \sigma \cdot \sqrt{T_i} \ ) \tag{6.10}$$

where $\phi(\mu, \sigma)$ denotes a normal distribution with mean μ and standard deviation σ. Then (6.6) can be written as

$$\int_X^\infty (L - X) \cdot g(L) \cdot d[L],$$

where g is the probability density function of L. Introducing $w = ln(L)$, we can rewrite the formula above as

$$\begin{aligned}
&\int_{ln(X)}^\infty (exp(w) - X) \cdot exp\{-\frac{1}{2 \cdot \sqrt{\pi}} \cdot (\frac{w - (ln(m) - \sigma^2 \cdot T_i/2)}{\sigma \cdot \sqrt{T_i}})^2\} \cdot d(\frac{w}{\sigma \cdot \sqrt{T_i}}) \\
&= \int_{ln(X)}^\infty exp(w) \cdot exp\{-\frac{1}{2 \cdot \sqrt{\pi}} \cdot (\frac{w - (ln(m) - \sigma^2 \cdot T_i/2)}{\sigma \cdot \sqrt{T_i}})^2\} \cdot d(\frac{w}{\sigma \cdot \sqrt{T_i}}) \\
&- \int_{ln(X)}^\infty X \cdot exp\{-\frac{1}{2 \cdot \sqrt{\pi}} \cdot (\frac{w - (ln(m) - \sigma^2 \cdot T_i/2)}{\sigma \cdot \sqrt{T_i}})^2\} \cdot d(\frac{w}{\sigma \cdot \sqrt{T_i}})
\end{aligned} \tag{6.11}$$

We will first compute the second term above.

Introducing

$$\hat{w} = -\frac{w - (ln(m) - \sigma^2 \cdot T_i/2)}{\sigma \cdot \sqrt{T_i}}$$

we have

$$\begin{aligned}
&\int_{ln(X)}^\infty X \cdot exp\{-\frac{1}{2 \cdot \sqrt{\pi}} \cdot (\frac{w - (ln(m) - \sigma^2 \cdot T_i/2)}{\sigma \cdot \sqrt{T_i}})^2\} \cdot d(\frac{w}{\sigma \cdot \sqrt{T_i}}) \\
&= X \cdot \int_{-\infty}^{d_2} exp\{-\frac{\hat{w}^2}{2 \cdot \sqrt{\pi}}\} \cdot d\hat{w} \\
&= X \cdot N(d_2)
\end{aligned}$$

where d_2 is from (6.9).

Now we compute the first term in (6.11),

$$\int_{ln(X)}^\infty exp(w) \cdot exp\{-\frac{1}{2 \cdot \sqrt{\pi}} \cdot (\frac{w - (ln(m) - \sigma^2 \cdot T_i/2)}{\sigma \cdot \sqrt{T_i}})^2\} \cdot d(\frac{w}{\sigma \cdot \sqrt{T_i}})$$

By a straight forward calculation, the above is equal to

$$\int_{ln(X)}^{\infty} m \cdot exp(r \cdot T_i) \cdot exp\{-\frac{1}{2 \cdot \sqrt{\pi}} \cdot (\frac{w - (ln(m) + \sigma^2 \cdot T_i/2)}{\sigma \cdot \sqrt{T_i}})^2\} \cdot d(\frac{w}{\sigma \cdot \sqrt{T_i}})$$

Further introducing a new variable

$$\hat{w}_1 = -\frac{w - ln(m) - \sigma^2 \cdot T_i/2}{\sigma \cdot \sqrt{T_i}}$$

we have

$$\begin{aligned}
&\int_{ln(X)}^{\infty} m \cdot exp(r \cdot T_i) \cdot exp\{-\frac{1}{2 \cdot \sqrt{\pi}} \cdot (\frac{w - (ln(m) - \sigma^2 \cdot T_i/2)}{\sigma \cdot \sqrt{T_i}})^2\} \cdot d(\frac{w}{\sigma \cdot \sqrt{T_i}}) \\
&= m \cdot \int_{-\infty}^{d_1} exp(r \cdot T_i) \cdot exp\{-\frac{\hat{w}_1{}^2}{2 \cdot \sqrt{\pi}}\} \cdot d\hat{w}_1 \\
&= m \cdot exp(r \cdot T_i) \cdot N(d_1)
\end{aligned}$$

where d_1 is from (6.8). Thus we have

$$E_{P(\cdot,T_{i+1})}(max(0, L(T_i, T_i, \delta) - X)) = exp(r \cdot T_i) \cdot m \cdot N(d_1) - X \cdot N(d_2).$$

Now we need to determine $m = E_{P(\cdot,T_{i+1})}(L)$. In (2.24) choose $f(t) = P(t, T_i)$, $g = P(t, T_{i+1})$, then

$$\frac{P(0, T_i)}{P(0, T_{i+1})} = E_{P(\cdot,T_{i+1})}(\frac{P(T_i, T_i)}{P(T_i, T_{i+1})})$$

By the definition of the LIBOR rate, the above equation can be written as

$$1 + \delta \cdot L(0, T_i, \delta) = E_{P(\cdot,T_{i+1})}(1 + \delta \cdot L(T_i, T_i, \delta))$$

or,

$$L(0, T_i, \delta) = E_{P(\cdot,T_{i+1})}(L(T_i, T_i, \delta)),$$

Thus

$$m = L(0, T_i, \delta). \tag{6.12}$$

Appendix B. Convexity adjustment

In Appendix A, the payoff of a caplet is delayed δ days after the δ-period LIBOR rate is observed. Surprising enough, if the payoff is made at exactly the same time as the the LIBOR rate is observed, then a *convexity adjustment* needs to be made to the computation in Appendix A.

The reason that the convexity adjustment is needed is that (6.12) is incorrect if the payoff time is T_i rather than T_{i+1}. In this case, since the payoff is made at time T_i, we take $g(t) = P(t, T_i)$ rather than $P(t, T_{i+1})$ as the numeraire.

Choose $f(t) = P(t, T_{i+1})$ in (2.24)

$$\frac{P(0,T_{i+1})}{P(0,T_i)} = E_{P(\cdot,T_i)}(\frac{P(T_i,T_{i+1})}{P(T_i,T_i)})$$

or,

$$\frac{1}{1+\delta \cdot L(0,T_i,\delta)} = E_{P(\cdot,T_i)}(\frac{1}{1+\delta \cdot L(T_i,T_i,\delta)}). \tag{6.13}$$

Introducing

$$f(x) = 1/(1+\delta \cdot x),$$

$$L = L(T_i,T_i,\delta), \qquad L_0 = L(0,T_i,\delta),$$

we can write (6.13) as

$$f(L_0) = E_{P(\cdot,T_i)}(f(L)). \tag{6.14}$$

By a Taylor expansion of f at L_0,

$$f(L) \approx f(L_0) + f'(L_0)\cdot(L-L_0) + \frac{f''(L_0)\cdot(L-L_0)^2}{2}.$$

Taking the expectation of both sides, we have

$$E_{P(\cdot,T_i)}(f(L)) \approx f(L_0) + f'(L_0)\cdot(E_{P(\cdot,T_i)}(L)-L_0) + \frac{f''(L_0)\cdot E_{P(\cdot,T_i)}((L-L_0)^2)}{2}.$$

Using (6.14), we obtain

$$E_{P(\cdot,T_i)}(L) \approx L_0 - \frac{f''(L_0)\cdot E_{P(\cdot,T_i)}((L-L_0)^2)}{2\cdot f'(L)}. \tag{6.15}$$

This relation is similar to, but different from, (6.12). The extra term

$$\frac{f''(L_0)\cdot E_{P(\cdot,T_i)}((L-L_0)^2)}{2\cdot f'(L_0)}$$

is called the convexity adjustment.

Bibliography

[Ames] Ames, W., Numerical Methods for Partial Differential Equations, 3rd Ed., Academic Press, Boston, 1993.

[Baxter-Rennie] Baxter, M., & A., Rennie, *Financial calculus: An introduction to derivative pricing,* Cambridge University Press, 1996.

[Black] Black, F., *Interest rates as options,* Journal of Finance, 50, 5(1995), 1371-76.

[Black-Derman-Toy] Black, F., E. Derman, & W. Toy, *A one-factor model of interest rates and its application to treasury bond options*, Financial Analyst Journal, 46, January-February(1990), 33-39.

[Black-Karasinski] Black, F. and P. Karasinski, *Bond and option pricing when short rates are lognormal*, Financial Analyst Journal, 47, July-August(1991), 52-59.

[BGM] Bruce, A., D. Gatarek and M. Musiela, *The market model of interest rate dynamics*, preprint.

[Canabarro] Canabarro, E., *Where do one-factor interest rate models fail?* The Journal of Fixed Income, 5, September(1995), p31-52.

[Chan] Chan, K. C. and G. A. Karolyi, F. Longstaff, A. Sanders, *An empirical comparison of alternative models of the short-term interest rate,* Journal of Finance, 47(1992), 1209-1227.

[Carverhill] Carverhill, A., *A note on the models of Hull and White for pricing options on the term structure,* The Journal of Fixed Income, 5, September (1995), 89-96.

[CIR] Cox, J. C., J. E. Ingersoll, and S. A. Ross, *A theory of the term structure of interest rates,* Econometrica, 53(1985), 385-407.

[Dybvig] Dybvig, P. H. *Bond and option pricing based on the current term structure,* working paper, Washington University, St. Louis, Missouri, 1989.

[Dybvig-Ross] Dybvig, P. H. and S. A. Ross, *Arbitrage*, in *The New Palgrave: Finance* edited by J. Eatwell, M. Milgate and P. Newman, Macmillian, 1989.

[Duffie-Kan] Duffie, D. and R. Kan, A yield-factor model of interest rates, *Mathematical Finance,* 6(1996), 379-406.

[Ho-Lee] Ho, T. S.Y. and S. B. Lee, *Term structure movements and pricing interest rate contingent claims,* Journal of Finance, 41(1986), pages 1011-1029.

[HJM] Heath, D., R. Jarrow, and A. Morton, *Bond pricing and the term structure of the interest rates: a new methodology,* Econometrica, 60, 1(1992), 77-105.

[HJMS] Heath, D., R. Jarrow, A. Morton and M. Spindel, *Easier done than said,* Risk, May(1993), 77-80.

[Hull] Hull, J., *Options, Futures, and Other Derivative Securities,* Prentice Hall, 1997.

[Hull-White(1990a)] Hull, J. and A. White, *Pricing interest rate derivative securities,* Review of Financial Studies, 3(1990), pages 573-592.

[Hull-White(1990b)] Hull, J. and A. White, *Valuing derivative securities using the explicit finite difference method,* Journal of Financial and Quantitative Analysis, 25(1990), 87-100.

[Hull-White(1993)] Hull, J. and A. White, *One-factor interest rate models and the valuation of interest rate derivative securities,* Journal of Financial and Quantitative Analysis, 28(1993), 235-254.

[Hull-White(1993)] Hull, J. and A. White, *One-factor interest rate models and the valuation of interest rate derivative securities,* Journal of Financial and Quantitative Analysis, 28(1993), 235-254.

[Hull-White(1994a)] Hull, J. and A. White, *Numerical procedures for implementing term structure models I: single-factor models,* The Journal of Derivatives, 2, Fall(1994), 7-15.

[Hull-White(1994b)] Hull, J. and A. White, *Numerical procedures for implementing term structure models II: two-factor models,* The Journal of Derivatives, 2, Winter(1994), 37-48.

[Hull-White(1995)] Hull, J. and A. White, *" A note on the models of Hull and White for pricing options on the term structure:" A response,* The Journal of Fixed Income, 5, September(1995), 97-102.

[Hull-White(1996)] Hull, J. and Alan White, *Using Hull-White interest rate trees,* Journal of Derivatives, 3(1996). 26-36.

[Hull-White(1997)] Hull, J. and A. White, *A note on Martingale, risk-neutrality, and Black's model,* preprint, 1997.

[GKO] Gustafsson, B., H. O. Kreiss, and J. Oliger, Time dependent problems and difference methods, John Wiley & Sons, Inc., New York, 1995.

[Kijima-Nagayama] Kijima M. and I. Nagayama, *Efficient numerical procedures for the Hull-White extended Vasicek model,* The Journal of Financial Engineering, 3(1995), 275-292.

[Jamshidian] Jamshidian, F. *Forward induction and construction of yield curve diffusion models,* Merrill Lynch working paper.

[Vetzal] Vetzal, K. *PDE models of the term structure: matching yield and volatility curves,* preprint 1996.

[Li] Li, A., P. Ritchken, & Sankarasubramanian, *Lattice models for pricing American interest rate claims,* Journal of Finance, 50(1995), 719-37.

[Nowman] Nowman, K. B., *Gaussian estimation of single-factor continuous time models of the term structure of interest rates,* Journal of Finance, 52(1997), 1695-1706.

[Lochoff] Lochoff, R., *The Contingent-claim arms race*, Journal of Portfolio Management, 20, 1(1993), 88-99.

[Paul-Choudhury] Paul-Choudhury, Sumit, *This year's model,* Risk, 10, May(1997), 18-23.

[Reed] Reed, N., *Floored by negative rates,* Risk, 8(1995), page 8.

[Rebonato] Rebonato, R., *Interest rate option models,* John Wiley & Sons, 1996.

[Reisman] Reisman, H., *A binomial tree for the Hull and White model with probabilities independent of the initial term structure,* The Journal of Fixed Income, 6, December(1996), 92-96.

[Reimers] Reimers, M., Private communication, 1998.

[Rogers] Rogers, C., *Gaussian errors,* Risk, 9(1996), pages 42-45.

[Sandmann-Sondermann] Sandmann K. and D. Sondermann, *On the stability of log-normal interest rate models and the pricing of Eurodollor futures,* preprint.

[Schmidt] Schmidt, W.M., *On a general class of one-factor models for the term structure of interest rates,* Finance and Stochastics, 1(1997), 1-23.

[Weston-Cooper] Weston, S. and S. Cooper, *Bank checks,* Risk, 9, February(1996), 23-27.

[Uhrig-Walter] Uhrig, M., & U. Walter, *A new numerical approach for fitting the initial yield curve,* The Journal of Fixed Income, 6, March(1996), 82-90.

[Vasicek] Vasicek, O.A., *An equilibrium characterisation of the term structure,* Journal of Financial Economics, 5(1977), pages 177-188.