

A Novel Term Structure Model
Based on
Tsallis Entropy and Information Geometry

YANG, Yiping

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
in
Automation and Computer-Aided Engineering

The Chinese University of Hong Kong
September 2010

UMI Number: 3484738

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent on the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3484738

Copyright 2011 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Thesis/Assessment Committee

Professor Li Duan (Chair)

Professor Kwong Chung Ping (Thesis Supervisor)

Professor Wong Po Shing (Committee Member)

Professor Brody Dorje C (External Examiner)

Abstract

Following the initial study of Brody and Hughston on applying information geometry to interest rate modeling, we propose a novel term structure model and investigate its application in the US swap market. Different from the traditional term structure models that impose assumptions on either bonds or rates, the newly proposed model is characterized by the evolution of a density function which is obtained from the derivative of the discount function with respect to the time left till maturity. We prove that such a density function can be interpreted as interest return on the discount bond.

The introduction of the term structure density turns the problem of yield curve dynamics into a problem of the evolution of a density distribution. There are at least three steps to model the dynamics of the density function: calibrate the initial term structure density, specify the market risk premium, and choose a proper volatility structure. First, we introduce two initial calibration methods, one by maximizing the Tsallis entropy and the other by the notion of superstatistics. By use of either method, we deduce a power-law distribution for the initial term structure density function. The entropy index q in this function, which is a well-known physics quantity, now finds its financial interpretation as the measure of departure of the current term structure from flatness on a continuously compounded basis. Our empirical experiments in the US swap market fully demonstrate this observation. Next, given

the calibrated initial density, we develop the term structure dynamics in the risk-neutral world and prove that the market risk premium is immaterial. To deduce a concise martingale representation for the bond pricing formula, we choose a density volatility that possesses zero mean. Finally, as an illustration of the importance of volatility structure, the HJM volatilities are redesigned for interest rate positivity under the framework of the current model.

An important application of term structure models is to measure the difference between the evolutions of two yield curves starting from the same initial point. Such a geometric problem can be tackled by use of the notion of information geometry after the mapping of yield curves to density functions on a Hilbert space. We prove that a pair of yield curves with large initial Bhattacharyya spherical distance would diverge from each other with a significant probability.

Finally, we implement the proposed model with initial data in the US swap market for 15 Feb, 2007. To test our model improvements over the traditional models, we also run the simulation with the Hull-White model and compare these two no-arbitrage models in various major characteristics. It shows that the proposed model forms a bridge linking interest rates and discount bonds, namely, given the initial term structure density and the volatility structure, we are able to reconstruct the short rate process and the bond price process. Our term structure density model is thus a unification of traditional models each having its own advantage.

摘要

本文基于Brody和Hughston的工作（应用信息几何研究利率模型）提出了一个新的利率结构模型并将其应用于美国外汇掉期市场。与传统模型不同，该模型并非对债券价格或利率直接提出假设，而是刻画了一个概率密度函数的演化过程。此密度函数正是通过对贴现函数关于期限求导而得。我们证明了该期限密度函数可表示为贴现函数的年利息。

通过引进期限密度函数，我们将对收益率曲线的动力学研究转化为对该密度函数的动力学研究。要刻画该期限密度函数的动态过程，至少需要三个步骤：确定初始期限结构、给出市场风险溢价、以及选择合适的波动率期限结构。首先，我们提出了两种估计初始结构的方法，一种是基于Tsallis熵的最大化方法，另一种则受到超统计这一概念的启发。无论运用哪种方法，我们得到的初始期限密度函数都服从幂率分布。而此分布中的一个重要物理统计量——熵指数，在我们模型中则用于度量当前期限结构与连续复利计算的平坦期限结构之间的距离。我们在美国外汇掉期市场上的实证研究充分证明了熵指数的这一金融涵义。其次，给定初始期限函数之后，我们研究了利率期限结构在风险中性世界中的动态过程，并证明了整个期限结构的演变与风险溢价无关。此外，通过限制波动率过程均值为零，我们得到了一个简洁的鞅表示作为债券定价公式。最后，为了说明波动率结构在利率期限结构研究中的重要性，我们在当前新模型的构架下重新设计了HJM模型中的波动率结构，以确保其所得利率非负。

期限结构模型不仅可用于单一收益率曲线的动力学研究，还可以用来刻画两条不同收益率曲线间距离的演变过程，因而提出了一个几何学问题。由于收益率曲线已经转化为一个希尔伯特空间上的期限密度函数，我们可以应用统计学领域的信息几何方法解决这个距离度量问题。我们证明了两条收益率曲线如果其初始Bhattacharyya球面距离过大，则他们必然以大概率渐行渐远。

最后，我们以美国外汇掉期市场2007年2月15日的数据为初始数据运行了新的

模型，并与传统模型之一的Hull-White 模型进行了比较。实验证明给定初始期限密度函数和波动率期限结构，通过刻画期限密度函数的动态过程，我们可以得到短期利率和债券价格的演变过程。因此，新的期限结构模型可以对传统模型取长补短，并将传统模型统一在了同一框架之下。

Acknowledgement

I would like to express my sincere gratitude to people who make this thesis possible.

First of all, I am indebted to my supervisor, Professor Kwong Chung Ping, for his continuous support and patient guidance throughout the whole stage of my Ph.D study. Prof. Kwong is the light tower, when I were at sea, that I could rely on and find out the direction. His insight for research, his enthusiasm for education, and his strong sense of responsibility to students all impress me deeply.

I am thankful to my defense committee: Prof. Brody Dorje C, Prof. Li Duan, and Prof. Wong Po Shing, for their invaluable suggestions to improve this thesis.

My thanks also go to my lab mates and friends. I thank my lab mates in CUHK: Dr. Li Dan Dan, Dr. Yan Yin Hui, and Mr. Xu Yang. I would never forget the support and encouragement from Dan Dan when I was depressed. I also enjoy the discussion with Yang, who brings the inspiration for my research. I also feel lucky to make a lot of friends in Hong Kong. I am particularly grateful to Dr. Yu Min Jie and Mr. Lu Wen Jie. Without Min Jie's providing of market data, the model implementation cannot go so smoothly. Special thanks must go to my dear friend Wen Jie. He proofreads the first chapter of my thesis. Far more than that, he gives me valuable advice on both my work and life.

Last but not least, I would like to dedicate this thesis to my grandparents, my parents, and Ken for their unreserved support and love. Although my grandmother

is now in another world, I believe she stays with me all the time, sharing in my frustrations when I get lost and also my joys when the sun shines the road ahead.

Contents

Abstract	i
Acknowledgement	v
1 Introduction	1
1.1 Background of Research	1
1.1.1 Bonds, Rates, and Term Structures	1
1.1.2 Notations in Interest Rate Theory	6
1.2 Motivations of Research	8
1.3 Objectives, Results, and Contributions	13
1.3.1 Objectives and Results	13
1.3.2 Contributions	20
1.4 Organization of Thesis	21
2 Literature Review	25
2.1 Models of Short Rates	26
2.1.1 Equilibrium Models	27
2.1.2 Martingale Model	29
2.1.3 No-arbitrage Models	30
2.2 Models of Discount Bonds and Forward Rates	31
2.3 Comparison Among Traditional Term Structure Models	33

2.4	Term Structure Model Proposed by Brody and Hughston	34
2.4.1	Term Structure Densities	36
2.4.2	Information Geometry Applied to Statistical Models	37
2.4.3	Dynamics of the Term Structure Density	40
2.4.4	Formulas for Processes	42
2.5	Summary and Discussion	45
3	Interpretation of Term Structure Density	47
3.1	Interpretation With a Flat Term Structure	48
3.1.1	Term Structure Densities	48
3.1.2	Normalization Condition	49
3.2	Interpretation With a Non-flat Term Structure	53
3.2.1	Term Structure Densities	53
3.2.2	Normalization Condition	53
3.3	Summary and Discussion	55
4	Analysis of the Proposed Term Structure Model	57
4.1	Dynamics of the Term Structure Density	58
4.1.1	Dynamics in the Risk-neutral World	58
4.1.2	Dynamics in the Real World	65
4.2	Martingale Representations	66
4.3	Properties of the Proposed Model	69
4.4	Comparison With Traditional Models	70
5	Design of the HJM Volatility Structure	73
5.1	HJM Volatilities in the Risk-neutral World	74
5.2	Volatility Dynamics in the Risk-neutral World	79

5.3	HJM Volatilities and Volatility Dynamics in the Specified World with $\lambda_t = -\bar{\nu}_t$	80
5.4	Summary and Discussion	84
6	Distance Between Yield Curves	87
6.1	Distance Dynamics for a Pair of Yield Curves	88
6.2	Divergence of Yield Curves With Large Initial Distance	91
6.3	Summary and Discussion	97
7	Initial Calibration of the Proposed Model	99
7.1	Introduction to Tsallis Entropy	101
7.1.1	History of Entropy	101
7.1.2	Maximization of Shannon and Tsallis Entropies	105
7.1.3	Why Tsallis Entropy	112
7.2	An Initial Calibration Approach Based on Tsallis Entropy Maximization	114
7.2.1	Maximization of Tsallis Entropy With Market Data	115
7.2.2	Interpretation of Entropy Index in the Proposed Model	116
7.2.3	Initial Calibration Algorithm	119
7.2.4	Comparison With the Brody-Hughston Calibration Approach	124
7.3	An Initial Calibration Approach Based on Superstatistics	127
7.3.1	Clarification of Spatio Parameter	128
7.3.2	Basic Idea of Superstatistics	130
7.3.3	Initial Calibration Algorithm	131
7.3.4	Observations and Properties	136
7.4	Summary and Discussion	138

8	Implementation of the Proposed Model	141
8.1	Data Description	141
8.1.1	US Term Structure Data	142
8.1.2	US Derivatives Data	147
8.2	Initial Calibration	149
8.2.1	Initial Calibration Based on Shannon Entropy Maximization	150
8.2.2	Initial Calibration Based on Tsallis Entropy Maximization	154
8.2.3	Connection Between the Two Calibration Algorithms	157
8.3	Model Implementation in the Risk-Neutral World	159
8.3.1	Implementation of the Proposed Model	160
8.3.2	Implementation of the Hull-White Model	170
8.3.3	Comparison and Discussion	177
9	Conclusions and Future Works	181
9.1	Conclusions	181
9.2	Future Works	183
	Bibliography	187

Chapter 1

Introduction

1.1 Background of Research

1.1.1 Bonds, Rates, and Term Structures

People have come to realize the time value of money in a world of excess liquidity and stubbornly high oil prices. One hundred dollars in ten years is worth only seventy-four in today's dollars in the context of purchasing power assuming a 3% interest rate per annum. This drives people to seek for more sophisticated protection in the capital markets rather than simple bank deposits. In the following paragraphs, we will introduce an important investment vehicle — bonds, together with the role of interest rates in bond pricing, and the benefits and limitations of traditional rate models.

There are plenty of investment instruments depending on the risk. A risk-seeking investor tends to buy high-yield instruments, such as stocks and equity derivatives. Whereas a risk-averse investor prefers safer instruments with lower yields, such as Treasury bonds. Two types of bonds are most active in the US market, US Treasury bills and US Treasury bonds. The difference is their maturities and coupons — a US Treasury bill is a short-term and zero-coupon instrument, whereas a US Treasury bond is long-term and coupon-bearing.

A zero-coupon bond is a paper to guarantee the holder the face value at maturity. Therefore, the price of the bond at any time before the maturity, or the initial value, should be lower than its face value. Investors earn the price difference between the initial value and the face value.

In real markets, most bonds are, however, coupon-bearing. In addition to the principal returned at maturity, a coupon-bearing bond also pays "coupons" periodically during its life. Therefore, the initial cash price for purchasing such a bond may exceed its face value. A coupon-bearing bond can always be stripped as a portfolio of zero-coupon bonds. It is therefore sufficient to consider only zero-coupon bonds for our theoretical study.

The bond valuation is a hard problem because of the dependence of a bond price on various kinds of rates. In principle, the price of a bond is simply the sum of the present value of future cash flows. However, the rate to discount the future flows is usually not a constant and may follow a complex time series. It is thus difficult to predict the future value of a bond based on the current price.

In financial markets, an interest rate is defined essentially in two ways:

1. The first way to define an interest rate is related to its credit risk, namely, the higher the credit risk, the higher the interest rate. By this definition interest rates fall into three categories: Treasury rates, LIBOR rates, and repo rates. First, Treasury rates are rates at which a government borrows in its own currency. Second, LIBOR rates are rates at which a large international bank lends its money to another. Finally, the repo rate is quoted in a repurchase agreement in which an investment dealer agrees to sell its securities and then buy them back later at a slightly higher price. A government, backed by its power to collect taxes and to issue money, usually has the highest credit

grading and should, in theory, never default. Therefore, Treasury rates are often treated as risk-free rates and have in general lower value than the other two rates.

2. The second way to define an interest rate is related to its settlement date. In this way interest rates are classified as spot rates or forward rates. The t -year spot rate (also called the t -year zero rate) is the rate of interest earned on a zero-coupon bond that is settled today and lasts for t years. Implied by current spot rates, forward rates are the rates for future periods of time whilst currently contracted. According to the expectation theory, a forward interest rate for a future period is equal to the expected future spot rate for that period. However, this theory could result in excessive interest rate risk because investors would tend to deposit their funds for periods as short as possible, whereas borrowers would tend to borrow for periods as long as possible. The liquidity preference theory has been proposed to fix this problem by requiring that forward rates should always be higher than expected future spot rates.

Interest rates, no matter how they are defined, are taken usually not as constant when one values bonds of different maturities. The concept of-term structure is thus introduced to describe the relationship between interest rates and bond maturities. By plotting spot rates against the time to maturity, one can derive a yield curve. A term structure model describes the evolution of the yield curve through time and thus depicts the evolution of spot rates. In academic research [30,32,37,42,52], term structure models always focus on the evolution of the instantaneous short-term risk-free rate r_t (or short rate for simplicity), which is the annualized interest rate at which an entity can borrow money for an infinitesimally short period between t and $t + \Delta t$. In some academic research [25], short rates and spot rates are sometimes

exchangeable. The difference is subtle — a t -year spot rate is the constant rate over 0 to t years whilst a short rate is the rate at year t .

The study on interest rate models has flourished since Merton's initial investigation in 1973. Researchers in this area are broadly divided into two camps. The first camp builds term structure models with simple assumptions on rates (for example, a short rate is assumed to behave like a stock price). The other camp builds models with strong pre-conditions on rates, trying to characterize the economic properties of rates as completely as possible. In general, the interest rates generated by a good term structure model should possess the following four features:

1. The fluctuation of the interest rates should be restricted within a reasonable range. The behavior of an individual rate is more complex than that of a stock price. It can neither infinitely grow nor endlessly decline.
2. The interest rates should converge towards a long-term average level as time evolves. Such a phenomenon, known as mean reversion, can be observed frequently in our economy. When rates are too high, the economy tends to slow down and the model will have a negative drift that can pull the rates back to an appropriate level. Conversely when rates are too low, more people would like to borrow funds and the government would implement tight monetary policies, for example, rising the rates, to fight inflation. This property is a major difference between the behavior of interest rates and stock prices.
3. The volatilities of different points on a yield curve are different. This is because interest rates always reflect borrowers' uncertainty on inflation expectations.
4. An interest rate model should not be too complicated to perform real-time pricing calculation.

Traditional interest rate models fall into two groups — the equilibrium models and the no-arbitrage models, each having its own merits and limitations. In the family of equilibrium models, a pioneer model was proposed by Vasicek [37] in 1977. This is a mean-reverting model and provides an explicit form for the short rate process. The major problem of the Vasicek model is that it allows negative interest rates. The Rendleman-Bartter (RB) model [42] is built on a very simple assumption — the short rate process is the same as a stock price process. But this model ensures neither interest rate positivity nor mean reversion. In the wake of Vasicek's and RB's work, Cox-Ingersoll-Ross [32] modified the diffusion coefficient in the Vasicek model to guarantee interest rate positivity. A major problem with these models is, however, that they cannot provide a good fit to today's term structure.

In contrast, no-arbitrage models are designed to be consistent with today's term structure. In 1986, Ho and Lee [52] proposed the first no-arbitrage model, which automatically fits the initial term structure by following the arbitrage-free principle. However, this model fails to incorporate mean reversion. In 1990, the Hull-White model [30] finally solved this problem by applying an analogous drift as in the Vasicek model to the Ho-Lee model.

Both the equilibrium models and the no-arbitrage models address the evolution of short rates. The Heath-Jarrow-Morton (HJM) model [15], however, focuses on forward rates. It proves that the instantaneous forward rate process can be fully and only determined by the initial term structure and the volatility structure of the associated discount bonds. The disadvantage of the HJM model is that the rates may be non-positive if one freely specifies the volatility structure. Besides, the calculation of the HJM model is quite time-consuming because the rate process is non-Markov and one has to use the Monte Carlo simulation.

1.1.2 Notations in Interest Rate Theory

Before discussing the details of interest rate theory, we first introduce a set of notations for easier reading. In this subsection, we first express interest rates on a simple or continuous basis, next introduce the principle of risk-neutral valuation, and finally reveal the relationships among bonds, short rates, and forward rates.

We write P_{tT} for the price at time t of a discount bond with principal \$1 maturing at time T ($T \geq t \geq 0$). Let L_{tT} denote the *simple interest rate* at time t for a term $T - t$. Then the relation between the bond price and the simple interest rate is given by

$$P_{tT} = \frac{1}{1 + (T - t)L_{tT}}. \quad (1.1)$$

When the bond is continuously compounded, the relation between the bond price and the *continuously compounded interest rate* R_{tT} (also called the *yield to maturity*) is given by

$$P_{tT} = e^{-R_{tT}(T-t)}. \quad (1.2)$$

A simple case, assuming the yield is a constant r over the bond life, gives rise to

$$P_{tT} = e^{-r(T-t)}, \quad (1.3)$$

which determines a *flat* term structure. Here we have to point to a fact: Eqn. (1.3) is actually a special case of bond pricing formulas based on the so called *risk-neutral valuation principle*.

The principle of risk-neutral valuation states that, in a risk-neutral world, a derivative can be valued by discounting its expected payoff at the risk-free short rate r_t . A risk-neutral world is an environment in which all individuals are indifferent to risk and all investors require no compensation for risk. In such a world, the average

return in a very short time period between t and $t + \Delta t$ is expected to be $r_t \Delta t$. As an illustration of the valuation principle, Eqn. (1.3) is obtained based on three assumptions: 1) the short rate is a constant r over the life of a bond; 2) the bond is a derivative of the rate; and 3) the expected payoff of the bond at maturity equals one $P_{TT} = 1$.

In reality, the interest rates are most often stochastic and there exist various risk-neutral worlds, each defined by a *market price of risk* λ (also called the *market risk premium*) of interest rate. λ functions as the excess return over the risk-free interest rate, expressed by per unit of risk in bonds. In fact, the risk-neutral world in which bonds are valued by (1.3) is referred to as the *traditional risk-neutral world* with the market risk premium being zero. Unless specified otherwise, in this thesis we will use the risk-neutral world to indicate the world with zero market risk premium, and call other risk-neutral worlds with non-vanishing risk premiums the original worlds or the real worlds.

Now we introduce forward rate and its relationship with short rate. We write $P_{tT_1T_2}$ for the forward price at time t of a contract for delivering a discount bond at time T_1 that matures at time T_2 ($T_2 \geq T_1 \geq t \geq 0$). According to the no-arbitrage condition,

$$P_{tT_1T_2} = \frac{P_{tT_2}}{P_{tT_1}}.$$

Define $F_{tT_1T_2}$ as the *forward interest rate* contracted at time t for the period between T_1 and T_2 and expressed on a simple basis. Its relation with the forward price $P_{tT_1T_2}$ is given by

$$P_{tT_1T_2} = \frac{1}{1 + (T_2 - T_1)F_{tT_1T_2}}, \quad (1.4)$$

so that

$$F_{tT_1T_2} = \frac{P_{tT_1} - P_{tT_2}}{P_{tT_2}(T_2 - T_1)}. \quad (1.5)$$

If we set $T_1 = T$, $T_2 = T + \Delta t$, and let $\Delta t \rightarrow 0$ in the above equations, then the *instantaneous forward short rate* f_{tT} is defined by

$$f_{tT} = \lim_{\Delta t \rightarrow 0} F_{tT, T+\Delta t}.$$

f_{tT} is related to P_{tT} by

$$f_{tT} = -\frac{\partial \ln P_{tT}}{\partial T}, \quad (1.6)$$

or equivalently

$$P_{tT} = \exp\left(-\int_t^T f_{ts} ds\right). \quad (1.7)$$

Moreover, if we set $t \rightarrow T$ in (1.1) and $t \rightarrow T = T_1 = T_2$ in (1.4), respectively, then we find the short rate r_t is related to L_{tT} and f_{tT} by

$$r_t = L_{tt}, \quad (1.8)$$

$$r_t = f_{tt}. \quad (1.9)$$

1.2 Motivations of Research

Almost all the term structure models target at a single issue — modeling the evolution of yield curves (**Dynamical Problem**). The yield curve is of tremendous importance both in concept and in practice. From a conceptual viewpoint, the yield curve is the base for asset pricing and economic policy decisions since the curve determines the value that investors place today on nominal payments at all future dates. From a practical viewpoint, the yield curve of US Treasuries are extensively used as the benchmark for the asset pricing around the world.

An important application of term structure models is to measure the difference between the evolutions of two yield curves starting from the same initial point (**Distance Problem**). Such a difference arises in the following three cases. First, the yield curves are depicted by different term structure models. Second, the yield

curves are characterized by the same model but different set of parameters, such as the volatility structure and the market risk premium process. Finally and most subtly, the initial term structure is calibrated with different methods.

The first problem — **Dynamical Problem** — has been studied extensively in various models. The most intuitive idea is to model the evolution of bond prices directly by a stochastic differential equation. However, in such models the bond prices can inconveniently become negative or even be larger than their face value. To tackle these problems, researchers try to model rates rather than bond prices, such as short rates [30, 32, 37, 42, 52] or instantaneous forward rates [15]. But these models are still far from perfect. For example, among the models of short rates, the equilibrium models [32, 37, 42] generate the initial term structure rather than take the actual initial term structure as the input. Although the no-arbitrage models [30, 52] solve this problem by setting the drift coefficient as a time variable in the underlying stochastic equation, other desirable properties of rates, for instance, mean reversion or interest rate positivity, cannot be incorporated simultaneously. The HJM model [15] is carefully built to fulfill most of these requirements. However, the Achilles heel of HJM is the resulting non-positive rates if the volatility structure is freely specified.

An intuitive thought is to place all the models above into a uniform framework. The starting point is to translate the financial variables, such as bond prices, short rates, and instantaneous forward rates, to the same mathematic variable. Our idea is inspired by Brody and Hughston's contribution [17–19], namely, to translate a family of discount functions to a term structure density process:

$$\rho_t(x) = -\frac{\partial B_t(x)}{\partial x}, \quad (1.10)$$

where $B_t(x) \triangleq P_{t,t+x}$ denotes the bond price at time t with time x left till maturity

(thus $T = t + x$ is the maturity date). Then we study the evolution of the term structure density with the help of some powerful tools such as stochastic calculus and information geometry. Finally, we translate the dynamics of the term structure density back to the dynamics of the financial variables. With this method, we could link the above discussed models together and fix their drawbacks.

The second problem — **Distance Problem** — is also addressed in Brody and Hughston's study. The key idea is to find a space with proper metric such that the difference between two yield curves becomes a distance measure between the points in the space that associate with the yield curves. In their study, the space chosen to accommodate such a comparison is the Hilbert space $\mathcal{H} = \mathcal{L}^2(\mathcal{R}^1)$ of square-integrable functions. The metric is defined by the spherical distance function of Bhattacharyya.

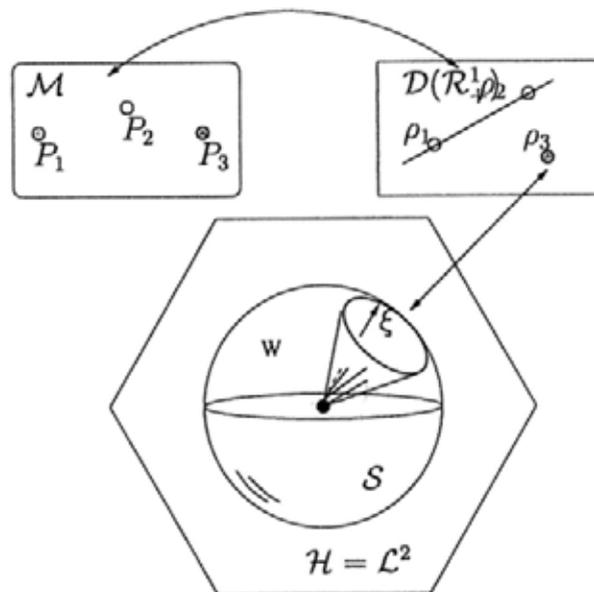


Figure 1.1: Connections of spaces

Figure 1.1 illustrates how we measure the distance between two yield curves (points) in \mathcal{H} . First, in the original space \mathcal{M} of all the yield curves, each yield curve

could be regarded as one point P . Second, by virtue of (1.10), each point P in \mathcal{M} could be mapped into a positive density function ρ on the positive real line. We denote the space of density functions as $\mathcal{D}(\mathcal{R}_+^1)$. Finally, each density function ρ in $\mathcal{D}(\mathcal{R}_+^1)$ could be mapped into a unit vector $\xi = \sqrt{\rho}$ in \mathcal{H} or equivalently a point on the unit sphere \mathcal{S} in \mathcal{H} . These mapping processes convert the distance between two yield curves in \mathcal{M} to the angle between the corresponding two vectors ξ_1 and ξ_2 in \mathcal{S} . Since \mathcal{S} is a unit sphere, the value of the angle equals the spherical distance between the points on \mathcal{S} determined by the vectors ξ_1 and ξ_2 .

The discussion on the term structure density is not the end of our study. A term structure model should eventually tell us the knowledge of the financial variables bond prices, short rates, and forward rates — rather than the features of a density function. It is thus desirable to show how the term structure density is converted back to those financial variables. That is why all the arrows in Figure 1.1 are in both ways.

The backward mapping is presented in Figure 1.2. The node at the center represents the term structure density function. We will prove in Chapter 3 that it stands for the annualized interest return of the bond under consideration. Note that all the arrows except the upward one in Figure 1.2 are one way. In Step 1 of our research, we translate an individual discount function to a density function. By interpreting this density function in financial backsets, in Step 2 we reformulate the value of the instruments like bonds, short rates, and instantaneous forward rates as functions of the term structure density.

Given the equivalence between the term structure density and the bond value, Brody and Hughston have constructed a geometric measure for the difference between two term structures and characterized the evolutionary trajectory of the term

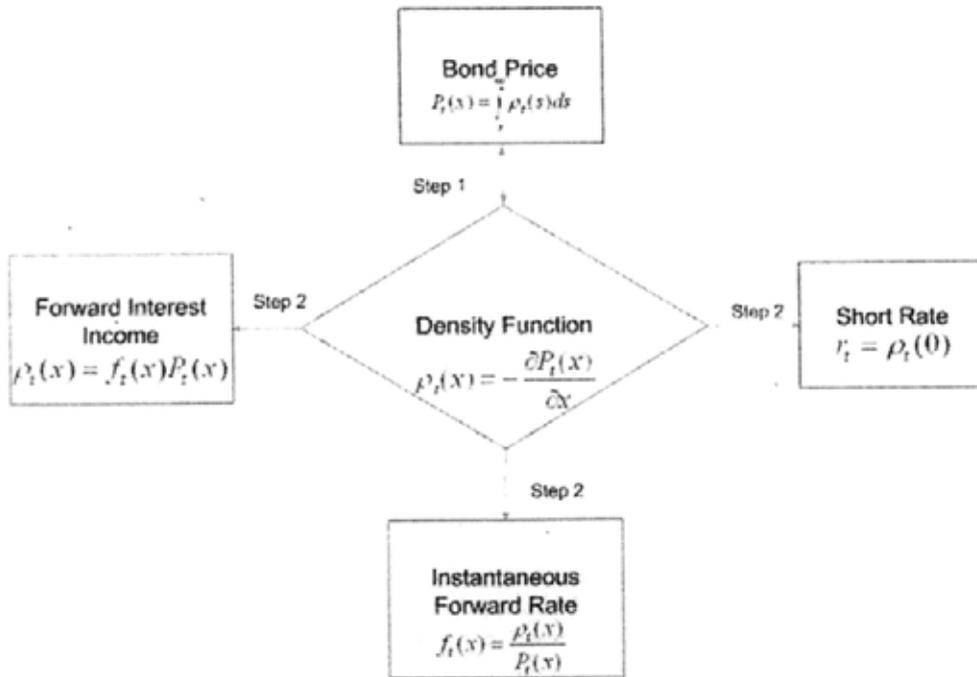


Figure 1.2: The term structure density and interest rate derivatives

structure as a measure-valued process. However, some problems in the Brody-Hughston model deserve our further study:

1. The term structure density defined via (1.10) has not been interpreted from the financial viewpoint.
2. Brody and Hughston have introduced a term structure calibration method [19] based on maximizing the Shannon entropy. But is the Shannon entropy the best candidate for the initial calibration?
3. When Brody and Hughston solved the dynamical equation that depicts the evolution of the term structure density, they eliminated the market risk premium by incorporating it into the volatility structure. But we are still interested in the dynamics with a different risk premium process.
4. Brody and Hughston's study on the distance problem focuses on a static state,

i.e. at a fixed time. It is a challenge to develop a dynamical theory of the yield curve distance.

5. Brody and Hughston have neither attempted nor mentioned an implementation of their model.

1.3 Objectives, Results, and Contributions

1.3.1 Objectives and Results

Following the initial study [17–19] of Brody and Hughston, our research aims to further develop the model of the density function (1.10) that generates the term structure. We summarize the roadmap of our research in Figure 1.3.

Each node in the figure represents an objective of our research and is labeled with the chapter number where it is considered. The numbers enclosed in the boxes correspond to the item numbers of the objective list in the below. There are four major objectives:

0. to interpret the term structure density in the language of interest rate theory;
1. Dynamical Problem — to characterize the evolutionary trajectory of the term structure:
 - (a) to calibrate the initial term structure;
 - (b) to specify the market risk premium process; and
 - (c) to choose a proper volatility structure;
2. Distance Problem — to detect the distance evolution for a pair of yield curves with the current model; and
3. to implement the model in real markets.

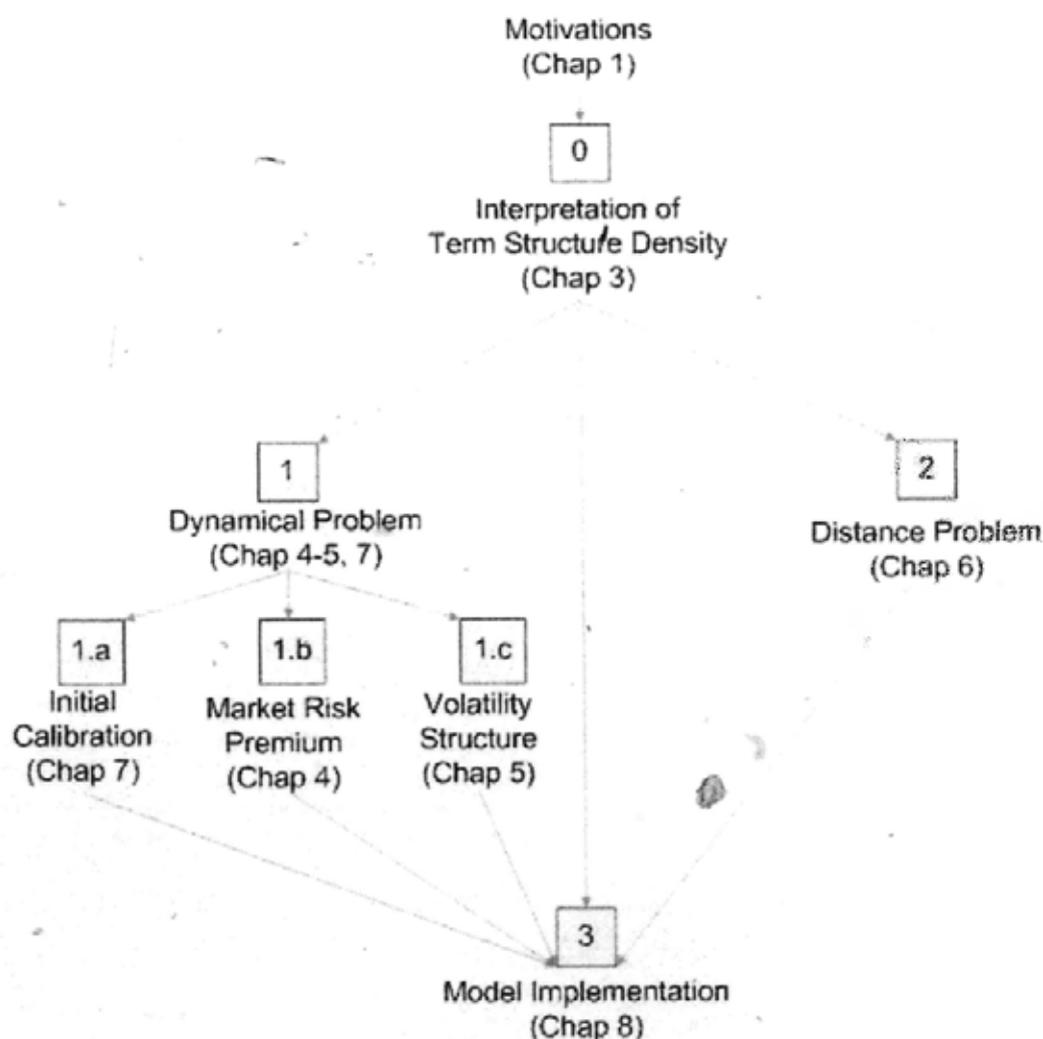


Figure 1.3: Development of the proposed model

The key methods and results in this thesis are summarized in the same order as in the objective list.

- Obj 0.** At the first outset of our research, we need to interpret both the term structure density (defined via (1.10)) and its normalization condition in the language of interest rate theory. The interpretation is discussed in Chapter 3 by use of series expansion and other calculus theories. The discussion simply starts with a flat term structure and then generalizes to a non-flat term structure.

We prove in the general case that the term structure density ρ associated with a discount bond represents the annualized interest earned on the bond. The integral $\int_0^\infty \rho(x) dx$ provides the interest earned during successive periods of one bond that matures in an infinity future. Therefore, the normalization condition is satisfied because the whole return of interest on a bond purchased initially at no cost (a bond with an infinity maturity is assumed to possess a vanishing initial value) should equal the face value one.

Obj 1-a. For Objective 1 — Dynamical Problem, the first step is the initial calibration, namely, to translate the current market information to the initial term structure density. This is accomplished in Chapter 7 by use of two approaches — Tsallis entropy maximization and superstatistics.

- (i) Based on maximizing the Tsallis entropy, we determine the initial term structure density that is consistent with multiple bond price data and the value of a perpetual annuity. The idea is to treat the Tsallis entropy as a functional of the term structure density and express the known data as constraints on the density function. With the calculus of variations and Lagrange multipliers, we obtain a piecewise power-law distribution for the initial density.

The initial distribution is parameterized by $1 - q$, where q is the entropy index, a physical measurement originally used in thermodynamics. In the power-law distributed term structure density, the power-law exponent N is defined via $\frac{1}{N} = 1 - q$. We prove that N is nothing new but the compounding frequency of the observed bonds. When we are given the prices of continuously compounded bonds, $N \rightarrow \infty$ or equivalently $q \rightarrow 1$. At this time, the calibrated density function reduces to the

piecewise exponential form. In this sense, the entropy index q essentially measures the departure of the current term structure from flatness on a continuously compounded basis.

- (ii) Inspired by the concept of superstatistics, we initially suppose that the term structure in a short term is flat associated with a constant continuously compounded rate β , and further assume that the rate follows a χ^2 -distribution. Therefore, the whole term structure could be regarded as a superposition of local flat structures, and proves to follow the same power-law distribution as the entropic method indicates if the only source of information available is the existence of a perpetual annuity.

The power-law exponent N in the initial term structure density is defined via $\frac{1}{N} = \frac{2}{n+2}$, where n is the degree of the χ^2 -distribution for the local short rate. We prove that N accounts for the compounding frequency of the underlying bonds. Many interesting properties of the current model are observed when we are given the prices of continuously compounded bonds, namely, $N \rightarrow \infty$. First, the calibrated initial term structure becomes flatter on a continuously compounded basis as N increases. Second, in the limit $N \rightarrow \infty$ the mean value β_0 of the χ^2 -distributed local short rate tends to be the long-term rate. Third, for an arbitrary N the perpetual annuity price does not equal but differs a little from the reciprocal of the mean value β_0 of short rate. In the limit $N \rightarrow \infty$, however, the annuity is valued precisely at $\frac{1}{\beta_0}$.

Obj 1-b. For Objective 1 — Dynamical Problem, the second step is to specify the world where the yield curve evolves — the risk-neutral world with a vanishing market risk premium or the real world with a specified risk premium? Our study

begins in the risk-neutral world and we apply stochastic calculus to develop the dynamics of the term structure density. The results could be immediately extended to the real world with the aid of Girsanov's Theorem, a theory of change of probability measure.

We prove that the term structure dynamics in the risk-neutral world is fully and only determined by two processes: the initial term structure and the volatility structure. To obtain the bond pricing formula in a concise martingale representation, the volatility of the term structure density that is parameterized by the tenor variable should possess zero mean. The market risk premium proves to be irrelevant when the model is used to price interest rate derivatives in the risk-neutral world. However, in the real world we should prominently specify the risk premium process.

Obj 1-c. For Objective 1 — Dynamical Problem, the third step is to choose a proper volatility structure. The volatility structure is of tremendous importance in determining the ultimate properties of the resulting rates. For example, the bond volatility under the current framework should be vanishing both at the initial time and in the infinity future. It leads to the zero-mean constraint on the density volatility and thus guarantees interest rate positivity. On the contrary, an improper volatility structure may result in undesirable properties for the underlying model. For example, if the bond volatility of the HJM model is freely specified, the resulting rates may be non-positive.

This inspires us to impose certain constraints on the HJM volatilities to ensure interest rate positivity. First, we show that the HJM bond volatility Ω_{tT}^{HJM} in the risk-neutral world can be regarded as the "normalized" weighted average of the density volatility σ_{tT} of the new model. Second, the HJM instantaneous

forward rate volatility σ_{tT}^{HJM} is dominated by the difference between σ_{tT} and Ω_{tT}^{HJM} . It follows that for interest rate positivity of the HJM model we only need to define the HJM volatilities in terms of the initial term structure density and the density volatility under the current framework.

Besides, for convenience of simulation, we also develop the dynamics of the HJM bond volatility and ours, both under the current framework. It provides a way to update the volatility structures timely and precisely in the two models so that the models can promptly reflect the latest market information.

Obj 2. After solving the Dynamical Problem, we come to tackle the Distance Problem. Since the input of our term structure model is a density function rather than any raw data in real markets, the initial term structure density is different, no matter how small, from the real distribution. For this reason, we need to study whether the initial error in term structure densities would disappear over time.

To begin with, we provide a derivation to the stochastic differential equation that governs the distance evolution for two yield curves. This equation was first proposed by Brody and Hughston in [17] without proof.

As an illustration, we consider the relative dynamics of two yield curves with different initial flat term structures but the same volatility structure and market risk premium. Each yield curve is initially dominated by a constant continuously compounded rate. By use of confidence interval, we prove that the given yield curves tend to diverge with a significant probability under two conditions: 1) the initial difference is large enough (such that the initial Bhattacharyya spherical distance is larger than the arc cosine of the ratio of the geometric and arithmetic means of the two short rates); 2) the market risk

premium is bounded within a certain range. In other words, if we want the evolution of the yield curve to be indifferent from the initial calibration error, the initial density should be as close to the real distribution as possible. This in return requires a more precise calibration algorithm.

Obj 3. Based on the above theoretical study, we implement the proposed model with initial data in the US swap market for 15 Feb, 2007.

- (i) The initial zero curve indicates an expectation of interest rate cuts (an economic decline) in 2008 and rate hikes (an economic recovery) in 2012.
- (ii) We calibrate the initial term structure density with two approaches, one based on maximizing the Shannon entropy as Brody and Hughston suggested [19], the other based on maximizing the Tsallis entropy as we propose. By performing experiments with the compounding frequency of 1, 2, and 6 months, we observe that the Tsallis density function approaches the Shannon exponential distribution as the frequency increases. This observation confirms our theoretical results in Chapter 7, that the entropy index which parameterizes the Tsallis entropy actually measures the departure of the current term structure from flatness on a continuously compounded basis.
- (iii) We implement the proposed model using the Monte Carlo simulation. Except for a time-consuming calculation, the implementation has satisfactory performance on the resulting rate values and bond prices. In order to find out the improvements over classical models, we also implement the Hull-White model by constructing the Black-Karasinski tree.

1.3.2 Contributions

Our work is developed along the line of Brody and Hughston's contribution to the term structure theory. However, our extension to the proposed model is unprecedented in the following aspects.

1. The financial roles of the term structure density (defined via (1.10)) and its integrals over various time intervals have not been explored systematically in the literature. Our work is the first attempt to interpret the term structure density as the annualized interest return of the underlying bonds.
2. We develop an initial calibration approach based on maximizing the Tsallis entropy. The involved entropy index q is a physical measurement traditionally used in thermodynamics. We are the first to introduce this physical variable to measure the departure of the current term structure from flatness on a continuously compounded basis. As a result, the piecewise power-law distribution parameterized by q provides the most general model for the initial term structure density. Besides, our work is original in explaining the power-law exponent $N = \frac{1}{1-q}$ as the compounding frequency of the observed bonds.
3. The application of superstatistics in the initial term structure calibration is original. To our best knowledge of literature, it is the first attempt to link superstatistics and entropy maximization under the framework of interest rate theory. Most properties of the resulting term structure density, such as the relationship between the long-term rate and the mean of the local short rate, are thus new findings.
4. The study on the term structure dynamics in the risk-neutral world and the design for the HJM volatility structure under the current framework are sup-

plements to the work of Brody and Hughston. By comparing our results with those of Brody and Hughston, our work fully demonstrates the role of the market risk premium in modeling the term structure dynamics.

5. We explore the uncharted territory of the relative dynamics of a pair of yield curves, though our study on the distance evolution problem is preliminary and deserves further study.
6. Our simulation with real market data is the first attempt to show how this model calibrates with the initial market information and predicts the future bond prices or yields.

1.4 Organization of Thesis

This thesis consists of four main parts, each aiming to accomplish one of the four major objectives listed in Section 1.3.1. The first part, Chapter 3, interprets the term structure density (defined via (1.10)) from the financial viewpoint. The second part, consisting of Chapter 4, 5 and 7, addresses the theoretical study of the term structure dynamics. The third part, Chapter 6, studies the distance evolution for a pair of yield curves. The fourth part, Chapter 8, introduces an implementation scheme of the proposed model.

Chapter 2 We briefly review traditional models, and introduce in detail a new term structure model on a density function proposed by Brody and Hughston.

Chapter 3 We explore the financial roles of the term structure density (1.10) together with its integrals over different time intervals. The investigation starts with a flat term structure in Section 3.1 and then extends to a non-flat term structure in Section 3.2. A comparison between these two cases, as shown in Table 3.1 in Section 3.3, reveals a broader connection between the term structure density and

financial instruments.

Chapter 4 We develop first in Section 4.1.1 the term structure dynamics in the risk-neutral world and then extend our discussion to the real world in Section 4.1.2. A comparison with the results obtained in a world where $\lambda_t = -\bar{\nu}_t$ (as Brody and Hughston proposed [17, 19]; $\bar{\nu}_t$ denotes the expectation of a freely specified process $\nu_t(x)$ with respect to (w.r.t) the term structure density) is presented in Table 4.1 in Section 4.1.1. The properties of the proposed model are elaborated in Section 4.2 and 4.3. Finally, we compare the new model with the traditional models in Table 4.2 in Section 4.4.

Chapter 5 We redesign the HJM volatility structure for interest rate positivity. In Section 5.1, by showing up the connections between the HJM volatilities and ours, we impose direct conditions in the risk-neutral world on the HJM volatilities to ensure interest rate positivity. Besides, for convenience of simulation, we also develop in Section 5.2 the dynamics of the HJM bond volatility and ours, both under the current framework. Parallel results in the world with $\lambda_t = -\bar{\nu}_t$ (as Brody and Hughston proposed [17, 19]) are obtained in Section 5.3. A comparison of observations in these two worlds is given in Table 5.1 in Section 5.4.

Chapter 6 We get down to the distance evolution problem. Our aim is to depict the influence of initial error from term structure calibration on the subsequent yield curve evolution. To begin with, in Section 6.1 we supplement a proof to a key proposition in the study of distance evolution. As an illustration of the proposition, we consider in Section 6.2 the relative dynamics of two yield curves with different initial flat term structures but the same volatility structure and market risk premium.

Chapter 7 We design two calibration algorithms for the initial term structure density, one based on the Tsallis entropy maximization and the other based on su-

perstatistics. To begin with, we introduce in Section 7.1 several types of entropy and highlight the superiority of the Tsallis entropy to others. In Section 7.2, we present an iterative algorithm, which is based on the Tsallis entropy maximization, to determine the initial density in terms of the short rate and multiple bond price data. In Section 7.3, we propose another initial calibration approach by use of the superstatistics concept. Many interesting properties concerned with the entropy index, the mean of the short rate, and the annuity price are elaborated in Section 7.3.4.

Chapter 8 We implement the proposed model using the US swap market data for 15 Feb, 2007. First, we analyze in Section 8.1 the raw data and the bootstrapped zero rates. Next, in Section 8.2 we calibrate the initial term structure by maximizing, respectively, the Shannon entropy and the Tsallis entropy. Finally, we implement the proposed model in Section 8.3 and obtain the evolutions of short rates and bond prices over a long term. To test our model improvements over the traditional models, we also run the simulation with the Hull-White model. A comparison of these two no-arbitrage models is presented in Table 8.7 in Section 8.3.3.

Chapter 9 We conclude our research and present future research directions.

Guides for reading At the beginning of each chapter (except Chapter 1), we briefly explain the significance of the study in the current chapter and provide a roadmap. At the end of each chapter (except Chapter 1), Section **Summary and Discussion** highlights the key results in that chapter. An exception is Chapter 4, which summarizes its key results in the last section. **Comparison With Traditional Models**. Such an arrangement is for readers to have a quick and complete idea of each chapter and soonest establish the whole picture of our thesis.

□ **End of chapter.**



Chapter 2

Literature Review

The problem of valuation of bonds and their derivatives is a big deal for investors. At the first onset, researchers tend to study the evolution of bond prices directly by a stochastic differential equation. However, in such models the bond prices can inconveniently become negative or even be larger than their face value. To overcome these problems, models of various kinds of rates have been proposed when researchers recognize the relationship between the rates and bond prices.

A yield curve provides the rates of return of zero-coupon bonds for different investment periods. Every day this curve changes as bond prices change. Figure 2.1 shows the possible shapes of a yield curve: upward sloping, downward sloping,

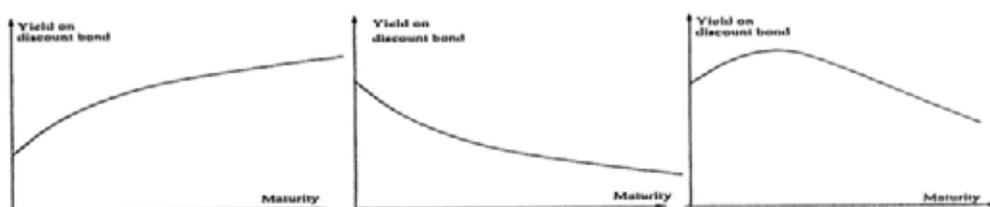


Figure 2.1: Yield curves

or slightly “humped”. A normal yield curve would be upward sloping, showing that yields rise as maturity lengthens. This pattern reflects the liquidity preference theory that long-term securities are general more risky than short-term securities and

thus demand higher yields. The inverted yield curve with downward slope appears usually when investors believe the economy will slow or even decline in the future and prefer longer-term securities. This pattern reflects the market segmentation theory conjecturing that the supply and demand in the markets for short-term and long-term instruments is determined largely independently. In practice, a yield curve is most often humped especially since the US government began to issue bonds with maturities larger than 20 years. The downward tilt to yields at long horizons most likely reflects the nonlinear relation between bond prices and yields: the price increase induced by a decline in the yield is larger than the price decrease induced by an equal-sized increase in the yield.

Term structure models are designed to explain these changes in yield curves. To begin with, in Section 2.1-2.3 we will briefly review several well-known models of short rates or forward rates. Following them in Section 2.4 is an introduction to a new model put forward by D.C. Brody and L.P. Hughston [18]. Different from the classical models, this new term structure model focuses on the evolution of a density function which is derived from the derivative of the discount function w.r.t the time left till maturity.

2.1 Models of Short Rates

According to the principle of risk-neutral valuation, the price at time t of a discount bond with principle \$1 maturing at T is determined by the short rate process r_t via

$$P_{tT} = \hat{E}_Q \left[e^{-\int_t^T r_s ds} \middle| \mathcal{F}_t \right], \quad (2.1)$$

where Q is a martingale measure and \hat{E}_Q denotes the conditional expectation w.r.t the filtration \mathcal{F}_t in the risk-neutral world [29]. Considering the relation (1.2) between the bond price and the yield to maturity, we can express the yield R_{tT} in terms of

the short rate r_t as

$$R_{tT} = -\frac{1}{T-t} \ln \hat{E}_Q \left[e^{-\int_t^T r_s ds} \middle| \mathcal{F}_t \right]. \quad (2.2)$$

It shows that the short rate process r_t defines everything about the initial zero curve and the term structure of interest rates at any time. In this section we will review some classical models of short rates, including equilibrium models, martingale model, and no-arbitrage models, and discuss the benefits and limitations of each.

2.1.1 Equilibrium Models

The short rate process r_t is usually described by a stochastic differential equation

$$dr_t = m(r) dt + s(r) dW_t,$$

where W_t denotes the Wiener process (or Brownian motion as some physicists prefer). The drift coefficient $m(r)$ and the diffusion coefficient $s(r)$ are functions of r but independent of time t . In financial research, $s(r)$ is called the standard deviation of the underlying security r . Specifically when $s(r) = \sigma r$, σ is called the volatility of r .

We will illustrate three one-factor equilibrium models in the below:

1. The Vasicek Model [37]:

$$dr_t = a(b - r_t) dt + \sigma dW_t. \quad (2.3)$$

2. The Rendleman and Bartter Model [42]:

$$dr_t = \mu r_t dt + \sigma r_t dW_t. \quad (2.4)$$

3. The Cox, Ingersoll, and Ross Model [32]:

$$dr_t = a(b - r_t) dt + \sigma \sqrt{r_t} dW_t. \quad (2.5)$$

In the Vasicek model, Eqn. (2.3) is in the form of the Langevin equation and the Ornstein-Uhlenbeck process. We solve the equation and obtain

$$\begin{aligned} r_t &= b - e^{-at}(b - r_0) + \sigma \int_0^t e^{-a(t-s)} dW_s, \\ E[r_t] &= b - e^{-at}(b - r_0) \rightarrow b. \end{aligned}$$

It shows that the short rate r_t is pulled to a mean level b at rate a . If we take account of the market risk premium λ_t , Eqn. (2.3) becomes

$$\begin{aligned} dr_t &= a \left[b - \left(r_t + \frac{\sigma \lambda_t}{a} \right) \right] dt + \sigma (dW_t + \lambda_t dt) \\ &= a (b^* - r_t) dt + \sigma dW_t^*, \end{aligned}$$

where $W_t^* = W_t + \int_0^t \lambda_s ds$ denotes the Wiener process corresponding to the risk-neutral measure. Since $b^* = b - \frac{\sigma \lambda_t}{a}$ involves λ_t , it shows that the market risk premium process λ_t needs to be specified independently and exogenously when we price interest rate derivatives in the risk-neutral world. The major disadvantage of this model is that it allows negative interest rates.

Vasicek also proved that the bond price P_{tT} can be expressed as

$$P_{tT} = A_{tT} e^{-B_{tT} r_t}, \quad (2.6)$$

where functions A_{tT} and B_{tT} are determined by the parameters a , b , and σ only but irrelevant to the initial term structure. By use of (2.2), we obtain the yield process

$$R_{tT} = -\frac{1}{T-t} \ln A_{tT} + \frac{1}{T-t} B_{tT} r_t,$$

indicating that the entire term structure is defined by the short rate process, and the initial term structure is generated as an output.

In the RB model (2.4), both the drift and the diffusion coefficients are proportional to r — the same assumption as for stock prices. Therefore we obtain an

explicit expression for the short rate process:

$$r_t = r_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}$$

However, such a concise expression comes at a price: this model neither ensures interest rate positivity nor incorporates mean reversion.

In the CIR model (2.5), the diffusion coefficient is proportional to \sqrt{r} and thus interest rate positivity is guaranteed. Cox, Ingersoll, and Ross also verified that the bond P_{tT} is priced by the same formula (2.6) as in the Vasicek model but with different forms for A_{tT} and B_{tT} . Besides, the initial term structure is still generated as an output.

2.1.2 Martingale Model

All the models listed above could be uniformly generalized into the so-called martingale model (2.1). There Q is a martingale measure in the risk-neutral world which is equivalent to the measure P in the real world. Applying the technique of change of measure, we can transform the bond price process from the risk-neutral world to the real world and obtain the following pricing formula for P_{tT} :

$$P_{tT} = \frac{1}{\Lambda_t} \hat{E}_P \left[\Lambda_T \exp \left(- \int_t^T r_s ds \right) \right], \quad (2.7)$$

$$\Lambda_t = \exp \left(- \int_0^t \lambda_s dW_s - \frac{1}{2} \int_0^t \lambda_s^2 ds \right). \quad (2.8)$$

Here \hat{E}_P denotes the conditional expectation w.r.t the original measure P . The density martingale Λ_t is not only an exponential martingale in terms of $-\int_0^t \lambda_s dW_s$ but is also the Radon-Nikodym derivative of measure Q w.r.t measure P [25].

With the martingale model, the bond price process in the real world is determined by two factors only: the short rate process r_t and the market risk premium process

λ_t . The disadvantage of this model is obvious — the conditional expectation is generally difficult to calculate.

2.1.3 No-arbitrage Models

In Section 2.1.1 we find that the equilibrium models cannot provide a good fit to today's term structure. This shortcoming can be successfully overcome by the choice of a time function as the drift coefficient in the stochastic equation that depicts the interest rate process. Following this line, no-arbitrage models are developed. In this subsection we will introduce two representatives of this model family: the Ho-Lee Model [52] and the Hull-White Model [30].

The Ho-Lee model makes assumptions on the short rate process:

$$\begin{aligned} dr_t &= \theta_t dt + \sigma dW_t \\ \theta_t &= \dot{f}_{0t} + \sigma^2 t, \end{aligned} \tag{2.9}$$

where f_{0t} is the instantaneous forward rate as seen at the present time 0 for a maturity t , and \dot{f} denotes a partial derivative w.r.t to t . The drift coefficient θ_t in Eqn. (2.9) is a function of time chosen to ensure the fitting of initial term structure.

There are two major disadvantages of the Ho-Lee model: the volatility structure is not flexible and mean reversion is not incorporated. They are both due to the drift expression in (2.9). The slope of forward curve defines the average direction of the short rate movement at any time. Therefore, the changes in both short rates and forward rates during a short period have the same standard deviation, which gives model users very little flexibility in choosing the volatility structure.

To incorporate mean reversion of the short rate, Hull and White extended the

Vasicek model by setting the long-term rate level to be a time-dependent function:

$$\begin{aligned} dr_t &= (\theta_t - ar_t) dt + \sigma dW_t \\ &= a \left(\frac{\theta_t}{a} - r_t \right) dt + \sigma dW_t \\ \theta_t &= \dot{f}_{0t} + af_{0t} + \frac{\sigma^2}{2a}(1 - e^{-2at}). \end{aligned} \tag{2.10}$$

Compared with the Vasicek model (2.3), the modified drift coefficient in the Hull-White model guarantees the fitting of initial term structure. Compared with the Ho-Lee model (2.9), the drift coefficient in (2.10) enables the short rate to converge towards a long-term average level at rate a .

The other improvement of the Hull-White model over the Ho-Lee model is that the former involves a richer volatility structure determined by both σ and a . Specifically when the parameter a equals zero, the Hull-White model reduces to the Ho-Lee model.

2.2 Models of Discount Bonds and Forward Rates

We now present the HJM model [15], which intends to model the process followed by instantaneous forward rates.

Assume the process for P_{tT} in the risk-neutral world follows the stochastic equation

$$\frac{dP_{tT}}{P_{tT}} = r_t dt + \Omega_{tT}^{HJM} (dW_t + \lambda_t dt), \tag{2.11}$$

where Ω_{tT}^{HJM} denotes the discount bond volatility and the superscript HJM is used to distinguish the HJM volatilities from those applied in other term structure models. If we let $W_t^* = W_t + \int_0^t \lambda_s ds$, it proves to be a Wiener process in the risk-neutral world after a change of measure from the real world.

By the relation (1.7) between the discount bond price P_{tT} and the instantaneous forward rate f_{tT} , the risk-neutral process for f_{tT} is characterized by

$$df_{tT} = \Omega_{tT}^{HJM} \dot{\Omega}_T^{HJM} dt - \dot{\Omega}_T^{HJM} dW_t^*, \quad (2.12)$$

where $\dot{\Omega}_T^{HJM}$ denotes the partial derivative of Ω^{HJM} w.r.t T .

Eqn. (2.12) shows that there is a close relation between the drift and the standard deviation of f_{tT} . If we define the standard deviation σ_{tT}^{HJM} of f_{tT} via the bond volatility Ω_{tT}^{HJM} as

$$\Omega_{tT}^{HJM} = - \int_t^T \sigma_{ts}^{HJM} ds, \quad (2.13)$$

then Eqn. (2.12) becomes

$$df_{tT} = \sigma_{tT}^{HJM} \int_t^T \sigma_{ts}^{HJM} ds + \sigma_{tT}^{HJM} dW_t^*. \quad (2.14)$$

The HJM model possesses several ideal properties. First, as a consequence of (2.13), the discount bond volatility Ω_{tT}^{HJM} tends to zero in the limit $t \rightarrow T$ as the bond matures. Second, the HJM model is consistent with the initial term structure P_{0T} and the forward rate process is dominated by the bond volatility Ω_{tT}^{HJM} . This can be clearly observed if we rewrite (2.14) as

$$f_{tT} = - \frac{\partial \ln P_{0T}}{\partial T} - \int_0^t \sigma_{sT}^{HJM} \Omega_{sT}^{HJM} ds + \int_0^t \sigma_{tT}^{HJM} dW_t^*.$$

Finally, the HJM framework has no reference to the market risk premium λ when we price interest rate derivatives in the risk-neutral world. This is a major difference between the HJM model and other models of short rates (for example, recall the Vasicek model (2.3) for comparison).

Yet the Achilles heel of HJM is the resulting non-positive interest rates if the volatility structure is freely specified. In Chapter 5 we will impose a constraint on the bond volatility Ω_{tT}^{HJM} of the HJM model to ensure interest rate positivity.

	Equilibrium Models			No-arbitrage Models		HJM
	R.B.	Vasick	C.I.R.	Ho-Lee	Hull-White	
Mean Reversion	×	✓	✓	×	✓	
Degrees of Freedom	λ, r	λ, r	λ, r	λ, r P_{0x}	λ, r P_{0x}	$\Omega_{tT}^{HJM}, P_{0x}$
Initial Term Structure	×	×	×	✓	✓	✓
Positivity Definite	×	×	✓	×	×	×
Others				Markov	Markov	Non-Markov

Table 2.1: Comparison among traditional term structure models

2.3 Comparison Among Traditional Term Structure Models

Table 2.1 unfolds a comparison among the various models discussed in Section 2.1-2.3. Our focus is on the fitting of initial term structure, the role of volatility structure, and the function of market risk premium in determining the term structure dynamics.

First, the essential difference between an equilibrium model and a no-arbitrage model consists in the fitting of initial term structure. In an equilibrium model, the initial term structure is an output, whereas a no-arbitrage model takes the actual initial term structure as an input.

Next, The HJM model is the only one that discount bond volatility Ω_{tT}^{HJM} enters as an independent parameter when we model the interest rate process.

Finally, when we price interest rate derivatives in the risk-neutral world, the HJM model is the only one that has no reference to the market risk premium process λ_t .

The traditional models discussed here all make assumptions on the short rate

or the forward rate. In the next section 2.4 we will present a new model proposed by Brody and Hughston [18], which focuses on the evolution of a density function that is derived from the derivative of the discount function w.r.t the time left till maturity.

2.4 Term Structure Model Proposed by Brody and Hughston

All the traditional models reviewed in Section 2.1-2.3 target at a single issue — modeling the evolution of yield curves (Dynamical Problem). Yet another important application of term structure models is to tell how differently one yield curve is from the other (Distance Problem). Both the two problems were tackled by Brody and Hughston [17–19] using the notion of information geometry.

Information geometry is the study of probability and information from the geometric viewpoint, simply by considering the statistical models as geometric objects. As early as in the 1930s, P.C. Mahalanobis [38, 39], an Indian physicist and statistician, had applied the geometric methods to define a measure of mutual separation in the study of statistical data arising from anthropometric measurements. Later till the 1980s, information geometry reached maturity through the work of Amari [48, 49]. The key idea is to regard a parametric statistical model as a differential manifold equipped with a metric and then study the structure of this manifold by way of differential geometry. We take a term structure model as an example. As illustrated in Figure 2.2, the starting point of the information geometry approach is to map every yield curve into a probability density function (pdf) in a Hilbert space we view as a manifold. The correspondence between the two spaces — the space of yield curves and the space of pdfs — is indicated by the both-way arrow in

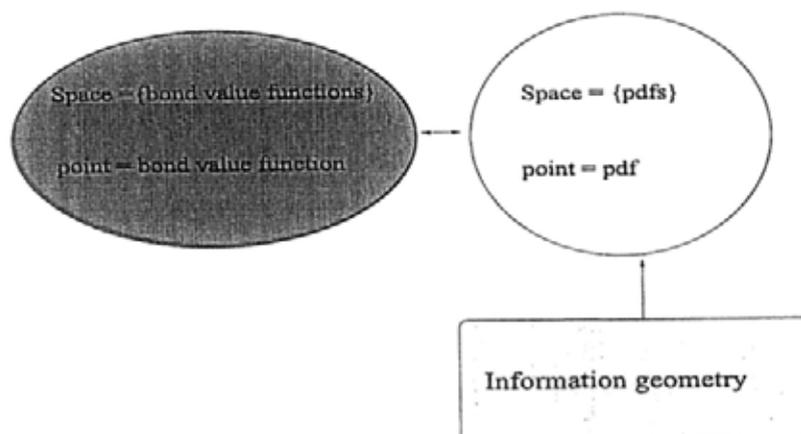


Figure 2.2: Space transformation between discount functions and pdfs

Figure 2.2. Then we apply information geometry to the space of density functions, as the upward arrow shows in the figure. The evolution of the yield curve is thus depicted by a process in the Hilbert space. The difference between two yield curves becomes a metric defined on the Hilbert space for two distribution functions.

Brody and Hughston fixed both the Dynamical Problem and the Distance Problem by following three steps:

- Step 1.** Deduce a pdf from the derivative of the discount function w.r.t the time left to maturity. This will be shown in Section 2.4.1.
- Step 2.** View each density function as a point embedded in the space of all probability distributions. Define a geometric measure on this space for every two points (pdfs). This will be discussed in Section 2.4.2.
- Step 3.** Given an initial point (initial term structure) in the structured space, model the dynamics of the yield curve as a random trajectory. We will outline the major procedures in Section 2.4.3 and present the key results in Section 2.4.4.

2.4.1 Term Structure Densities

The starting point of the Brody-Hughston model is to define a term structure density process:

$$\rho_t(x) = -\frac{\partial B_t(x)}{\partial x}, \tag{2.15}$$

where $B_t(x) \triangleq P_{t,t+x}$ denotes a family of bond prices at time t with time x left till maturity (thus $T = t + x$ is the maturity date). The tenor variable x is introduced according to the ‘‘Musielia parametrization’’ [35]. $\rho_t(x)$ proves to be a density function, i.e.

$$\rho_t(x) > 0, \quad \int_0^\infty \rho_t(x) dx = 1,$$

if we consider an admissible term structure with which interest rates should always be positive:

$$\lim_{x \rightarrow \infty} B_t(x) = 0, \quad 0 < B_t(x) \leq 1, \quad \frac{\partial B_t(x)}{\partial x} < 0. \tag{2.16}$$

It follows that each yield curve is associated with a density function on the positive real line, as illustrated in Figure 2.3. We denote the space of density functions as $\mathcal{D}(\mathcal{R}_+^1)$ and have the following characterization of this space [17, 19].

Proposition 2.1. *The system of admissible term structures is isomorphic to the convex space $\mathcal{D}(\mathcal{R}_+^1)$ of everywhere positive smooth density functions on the positive real line.*

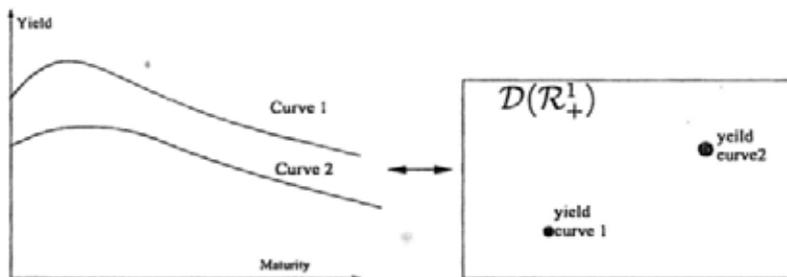


Figure 2.3: Mapping from yield curves to density functions

2.4.2 Information Geometry Applied to Statistical Models

In the previous subsection we have shown that the derivative of the discount function w.r.t the time left till maturity gives rise to a density function. The difference between two yield curves is thus turned to a comparison between the associated density functions. The theory of information geometry proves to be crucial in comparing the term structure densities.

In mathematics and especially in statistical inference, information geometry is the study of statistical models by way of differential geometry. A statistical model is a set of probability distributions to which we believe the true distribution belongs. Usually statistical models fall into two classes — parametric and non-parametric, each having its own geometric measure.

In the parametric family of pdfs, the parameter space has a Riemannian structure induced by the embedding of the family into the Hilbert space of square-integrable functions, and is characterized by the Fisher-Rao metric [18, 50]. In the non-parametric family of pdfs, the geometry structure is determined by the spherical distance function of Bhattacharyya. In the Brody-Hughston model, our focus is on the non-parametric family since the term structure density defined via (2.15) is a non-parametric pdf.

Let X be a continuous random variable taking values on the positive real line \mathcal{R}^1 , and $\rho(x)$ be a density function w.r.t X . The square-root likelihood function is defined by

$$\xi(x) = \sqrt{\rho(x)}. \quad (2.17)$$

Since $\rho(x)$ is non-negative and has integral unity, $\xi(x)$ satisfies the following condi-

tions:

$$\xi(x) \geq 0 \quad \int_{-\infty}^{\infty} (\xi(x))^2 dx = 1.$$

We see that $\xi(x)$ can be regarded as a unit vector in the Hilbert space $\mathcal{H} = \mathcal{L}^2(\mathcal{R}^1)$ or equivalently a point lying on the unit sphere \mathcal{S} of \mathcal{H} . This is illustrated in Figure 2.4.

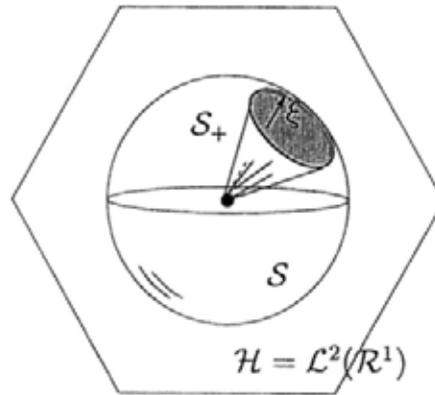


Figure 2.4: Unit sphere in the Hilbert space $\mathcal{H} = \mathcal{L}^2(\mathcal{R}^1)$

Given two yield curves, we let $\rho_1(x), \rho_2(x)$ denote the associated term structure densities on \mathcal{R}^1 , and $\xi_1(x), \xi_2(x)$ the corresponding Hilbert space elements. The inner product between ξ_1 and ξ_2 can be calculated via two ways:

$$\langle \xi_1, \xi_2 \rangle = \|\xi_1\| \|\xi_2\| \cos \phi = \cos \phi$$

where ϕ is the angle between the two vectors ξ_1 and ξ_2 , or

$$\langle \xi_1, \xi_2 \rangle = \int_{-\infty}^{\infty} \xi_1(x) \xi_2(x) dx.$$

Therefore

$$\phi = \arccos \left(\int_{-\infty}^{\infty} \xi_1(x) \xi_2(x) dx \right). \tag{2.18}$$

defines an angle ϕ that can be interpreted as the distance between ρ_1 and ρ_2 . More precisely, since S is a unit sphere, the value of ϕ equals the spherical distance between

the points on S determined by the vectors ξ_1 and ξ_2 :

$$\widehat{\xi_1 \xi_2} = \|\xi_1\| \phi = \phi. \quad (2.19)$$

This is illustrated in Figure 2.5. We call the angle ϕ the Bhattacharyya distance

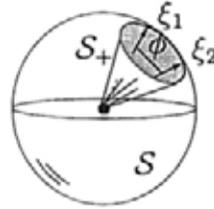


Figure 2.5: Bhattacharyya spherical distance

between the given yield curves.

As an illustration of the principles set forth above, we will discuss several examples to show the calculation of distance between two yield curves. Consider a family of discount bonds given by

$$P_{0T} = \left(1 + \frac{rT}{k}\right)^{-k}, \quad (2.20)$$

which determines a “flat” term structure for a constant annualized interest rate r compounded at the frequency k over the life of each bond.

Example 2.1. Let $k \rightarrow +\infty$ in (2.20). Hence r becomes a continuously compounded rate. By virtue of (2.15) and (2.17), we obtain the density function and square-root function

$$\rho(T) = re^{-rT}, \quad \xi(T) = \sqrt{re^{-rT}}.$$

Given two yield curves dominated by, respectively, r_1 and r_2 , we calculate the spherical distance to be

$$\begin{aligned} \cos \phi_{12} &= \int_{-\infty}^{\infty} \xi_1(T) \xi_2(T) dT, \\ \implies \phi_{12} &= \cos^{-1} \left(\frac{2\sqrt{r_1 r_2}}{r_1 + r_2} \right). \end{aligned}$$

It shows that the Bhattacharyya distance between the continuously compounded flat yield curves is determined by the arc cosine value of the ratio of the geometric and arithmetic means of the flat rates.

Example 2.2. Let $k = 1$ in (2.20). Hence r reduces to a simple interest rate. The density function becomes

$$\rho(T) = r(1 + rT)^{-2}.$$

The spherical distance between two yield curves in this family is calculated to be

$$\phi_{12} = \cos^{-1} \left(\frac{\sqrt{r_1 r_2}}{r_1 - r_2} \ln \frac{r_1}{r_2} \right).$$

2.4.3 Dynamics of the Term Structure Density

The distance measure between two yield curves as discussed above is a static problem. Now we turn to develop the dynamics of the term structure density. We give in the below a brief outline of Brody and Hughston's work [17–19]. For easier reading, we divide the modeling process into four steps.

Step 1. Assume that the dynamics of the bond price P_{tT} follow the stochastic differential equation

$$dP_{tT} = \mu_{tT} dt + \Sigma_{tT} dW_t, \quad (2.21)$$

where W_t is a Wiener process and μ_{tT} and Σ_{tT} are, respectively, the drift and the diffusion processes. Then, by virtue of $B_{tx} = P_{t,t+x}$, Eqn. (2.15), and Eqn. (2.21), we obtain the dynamics of $\rho_t(x)$ as

$$\begin{aligned} d\rho_t(x) &= \left(-\frac{\partial \mu_{t,t+x}}{\partial x} + \frac{\partial \rho_t(x)}{\partial x} \right) dt - \frac{\partial \Sigma_{t,t+x}}{\partial x} dW_t \\ &= \left(\beta_t(x) + \frac{\partial \rho_t(x)}{\partial x} \right) dt + \omega_t(x) dW_t, \end{aligned} \quad (2.22)$$

where we write $\beta_t(x) = -\frac{\partial \mu_{t,t+x}}{\partial x}$ and $\omega_t(x) = -\frac{\partial \Sigma_{t,t+x}}{\partial x}$ to simplify the notations.

Step 2. Impose the arbitrage-free condition by constraining the drift coefficient such that

$$\mu_{tT} = r_t P_{tT} + \Sigma_{tT} \lambda_t, \quad (2.23)$$

where r_t is the short rate process and λ_t is the market risk premium process. Under this condition, Eqn. (2.22) becomes

$$d\rho_t(x) = \left(r_t \rho_t(x) + \frac{\partial \rho_t(x)}{\partial x} \right) dt + \omega_t(x) (dW_t + \lambda_t dt). \quad (2.24)$$

Step 3. Verify that the normalization condition $\int_0^\infty \rho_t(x) dx = 1$ on $\rho_t(x)$ is preserved in the density dynamics (2.24). We just need to integrate the drift and the diffusion terms of (2.24) (w.r.t x) and see if both the integrals exactly equal zero:

$$r_t + \int_0^\infty \frac{\partial \rho_t(x)}{\partial x} dx = 0 \quad (2.25)$$

and

$$\int_0^\infty \omega_t(x) dx = 0. \quad (2.26)$$

Actually Eqn. (2.25) is satisfied because $\rho_t(x) \rightarrow 0$ as $x \rightarrow \infty$ and

$$r_t = \rho_t(0), \quad (2.27)$$

which will be clarified in Section 4 after we interpret $\rho_t(x)$ from the financial viewpoint. Additionally, condition (2.26) is fulfilled by virtue of the definition for $\omega_t(x)$ and observing that $\Sigma_{t,t+x}$ vanishes both as $x \rightarrow 0$ and as $x \rightarrow \infty$. The conditions (2.25) and (2.26) lead further to the following results:

1. r_t can be expressed as either (2.27) or

$$r_t = - \int_0^\infty \rho_t(x) \frac{\partial \ln \rho_t(x)}{\partial x} dx = -E_\rho \left[\frac{\partial \ln \rho_t(x)}{\partial x} \right], \quad (2.28)$$

where we use E_ρ to denote the expectation w.r.t $\rho_t(x)$.

2. $\omega_t(x)$ can be expressed in terms of an exogenously specified process $\nu_t(x)$:

$$\omega_t(x) = \rho_t(x) (\nu_t(x) - E_\rho[\nu_t(x)]). \quad (2.29)$$

Step 4. Eqn. (2.24) can be written as

$$\frac{d\rho_t(x)}{\rho_t(x)} = \left(\frac{\partial \ln \rho_t(x)}{\partial x} + r_t \right) dt + (\nu_t(x) - E_\rho[\nu_t(x)]) (dW_t + \lambda_t dt) \quad (2.30)$$

with r_t expressed as either (2.27) or (2.28). This completes the derivation of the Brody-Hughston model.

2.4.4 Formulas for Processes

As a consequence of (2.29), the standard deviation $\Sigma_{t,t+x}$ of $\rho_t(x)$ is unchanged under the transformation $\nu_t(x) \rightarrow \nu_t(x) + \vartheta(t)$, where $\vartheta(t)$ is independent of x . This freedom allows λ_t to be specified as $-\bar{\nu}_t$ [17, 19]. With this substitution, Brody and Hughston have derived the formulas for the density process, the bond price process, and some related rate processes [17, 19]. We summarize the major results as follows.

Proposition 2.2. *The general admissible term structure evolution based on the filtration generated by a Brownian motion W_t on the Hilbert space $\mathcal{H} = \mathcal{L}^2(\mathcal{R}^1)$ is a measure-valued process $\rho_t(x)$ on $\mathcal{D}(\mathcal{R}_+^1)$ that satisfies*

$$\frac{d\rho_t(x)}{\rho_t(x)} = \left(\frac{\partial \ln \rho_t(x)}{\partial x} + \rho_t(0) \right) dt + (\nu_t(x) - \bar{\nu}_t) (dW_t - \bar{\nu}_t dt), \quad (2.31)$$

where $\bar{\nu}_t = E_\rho[\nu_t(x)]$. The volatility structure $\nu_t(x)$ can be specified exogenously along with the initial term structure density $\rho_0(x)$.

The associated short rate process $r_t = \rho_t(0)$ satisfies

$$dr_t = \left(r_t^2 + \frac{\partial \rho_t(x)}{\partial x} \Big|_{x=0} \right) dt + r_t (\nu_t(0) - \bar{\nu}_t) (dW_t - \bar{\nu}_t dt). \quad (2.32)$$

Proof. See [19]. □

Strictly speaking, the volatility of $\rho_t(x)$ is $\nu_t(x) - \bar{\nu}_t$ rather than $\nu_t(x)$. However, for convenience in the Brody-Hughston model we call $\nu_t(x)$ the volatility of $\rho_t(x)$.

An alternative expression for (2.31) is given by

$$\frac{d\rho_t(x)}{\rho_t(x)} = \left(\frac{\partial \ln \rho_t(x)}{\partial x} - E_\rho \left[\frac{\partial \ln \rho_t(x)}{\partial x} \right] \right) dt + (\nu_t(x) - \bar{\nu}_t) (dW_t - \bar{\nu}_t dt). \quad (2.33)$$

This expression is sometimes more suggestive since we can evidently inspect the normalization condition on $\rho_t(x)$ by observing that

$$E_\rho \left[\frac{d\rho_t(x)}{\rho_t(x)} \right] = 0. \quad (2.34)$$

The appearance of r^2 in the drift of (2.32) might seem abnormal. As pointed out by Brody and Hughston [17], this is compensated by the second term in the drift bracket and ensures the mean reversion behavior, for example, in the CIR model [32].

Proposition 2.3. *The solution of the dynamical equation for $\rho_t(x)$ in terms of the volatility structure $\nu_t(x)$ and the initial term structure density $\rho_0(x)$ is*

$$\rho_t(T-t) = \rho_0(T) \frac{\exp \left(\int_0^t V_{sT} dW_s - \frac{1}{2} \int_0^t V_{sT}^2 ds \right)}{\int_t^\infty \rho_0(u) \exp \left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds \right) du}, \quad (2.35)$$

where $V_{tu} = \nu_t(u-t)$. The corresponding formula for the bond price process is

$$P_{tT} = \frac{\int_T^\infty \rho_0(u) \exp \left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds \right) du}{\int_t^\infty \rho_0(u) \exp \left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds \right) du}. \quad (2.36)$$

Proof. See [19]. □

We observe that both the term structure density process and the bond price process are dominated by two factors only: the initial term structure density $\rho_0(x)$ and the volatility structure $\nu_t(x)$. In particular, by setting $T = t$ in (2.35) we obtain the formula for the short rate process.

Proposition 2.4. *The short rate process $r_t = \rho_t(0)$ is given by*

$$r_t = \rho_0(t) \frac{\exp\left(\int_0^t V_{st} dW_s - \frac{1}{2} \int_0^t V_{st}^2 ds\right)}{\int_t^\infty \rho_0(u) \exp\left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds\right) du}. \quad (2.37)$$

In particular, in a deterministic model with $V_{st} = 0$, we obtain

$$r_t = \rho_0(t) / \int_t^\infty \rho_0(u) du = f_{0t}. \quad (2.38)$$

We know that in a stochastic model, $r_t = f_{tt}$, which means the instantaneous forward rate at time t for an infinitely near future is exactly the short rate at that time. However, Eqn. (2.38) shows that in a deterministic model, $f_{tt} = f_{0t}$, owing to the vanishing of randomness.

Proposition 2.5. *By virtue of $\lambda_t = -\bar{\nu}_t = \int_0^\infty \rho_t(x) \nu_t(x) dx$, the market risk premium process is given by*

$$\lambda_t = - \frac{\int_t^\infty \rho_0(u) V_{tu} \exp\left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds\right) du}{\int_t^\infty \rho_0(u) \exp\left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds\right) du}. \quad (2.39)$$

Proposition 2.4 and Proposition 2.5 show that, given the initial term structure density $\rho_0(x)$ and the volatility structure $\nu_t(x)$, we can reconstruct the short rate process and the market risk premium process.

Proposition 2.6. *The money market account B_t satisfying $dB_t = r_t B_t dt$ and $B_0 = 1$ follows the dynamical process*

$$B_t = \frac{\exp\left(\int_0^t \bar{\nu}_s dW_s - \frac{1}{2} \int_0^t \bar{\nu}_s^2 ds\right)}{\int_t^\infty \rho_0(u) \exp\left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds\right) du}. \quad (2.40)$$

Proposition 2.7. *By the relationship (1.6) $f_{tT} = -\frac{\partial \ln P_{tT}}{\partial T}$, the instantaneous forward rate process is given by*

$$f_{tT} = \rho_0(T) \frac{\exp\left(\int_0^t V_{sT} dW_s - \frac{1}{2} \int_0^t V_{sT}^2 ds\right)}{\int_T^\infty \rho_0(u) \exp\left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds\right) du}. \quad (2.41)$$

Note that when $T = t$, Eqn. (2.41) becomes (2.37). In other words, the instantaneous forward rate process reduces to the short rate process as $T \rightarrow t$.

2.5 Summary and Discussion

Before proceeding to present our own work, let us summarize other researchers' study we have reviewed in this chapter. To begin with, we have introduced the models of short rates in Section 2.1. Among them equilibrium models stand out usually for the simple assumptions imposed on short rates but suffer from the inconsistency of initial term structure. A 1% error in the price of the underlying bond may lead to a 25% error in an option price [29]. On the contrary, no-arbitrage models earn a reputation for automatically fitting the initial term structure.

Next, we have presented the HJM model in Section 2.2, which focuses on the evolution of instantaneous forward rates. The HJM model proves to excel the short rate models in some aspects. First, the HJM model is consistent with the initial term structure. Second, the rate process is determined by two factors only: the initial term structure and the volatility structure. Finally, the HJM model has no reference to the market risk premium when we price interest rate derivatives in the risk-neutral world.

All the traditional models attempt to characterize the shape change of yield curves. Yet another application of term structure models is to tell how differently a yield curve evolves from the other. With this end in mind, Brody and Hughston proposed a new term structure model of a density function which is derived from the derivative of the discount function w.r.t the time left till maturity (see Section 2.4). By embedding the family of probability distributions into the Hilbert space of square-integrable functions, the Bhattacharyya spherical distance is defined to measure the

difference between two yield curves. Besides, Brody and Hughston have also derived a dynamical equation to depict the evolution of the density process. By solving this stochastic equation, Brody and Hughston obtained the formulas for the density process, the bond price process, and other related rate processes.

End of chapter.

Chapter 3

Interpretation of Term Structure Density

Associated with every positive interest term structure there is a probability density function over the positive real line. This makes the Brody-Hughston model stand out. Brody and Hughston have explored the “mathematical role” of the term structure density: it is obtained from the derivative of the discount function w.r.t the time left till maturity, and proves to satisfy the normalization condition (the integral of the density over the positive real line equals one). However, neither the term structure density itself nor its normalization condition has been endowed with a “financial role” in practice. Practitioners in finance would hence have little confidence in applying this model to practice. For this reason, we aim to interpret the abstract density and its normalization condition in the language of interest rate theory.

The interpretation will start with a flat term structure in Section 3.1, where we will consider a family of bonds that are continuously compounded at a constant interest rate. Then our discussion will generalize to a non-flat term structure in Section 3.2, where the bonds are priced via an instantaneous forward rate process. A comparison between these two cases, as shown in Table 3.1 in Section 3.3, reveals a broader connection between the term structure density and the financial instruments

like bonds and rates.

3.1 Interpretation With a Flat Term Structure

To investigate the physical meaning of term structure densities from the financial viewpoint, we start with the simplest case — a flat term structure for which a bond is continuously compounded at a constant rate. The yield curve corresponding to such a bond may be seldom used in practical modeling. Nevertheless, it is sufficient for our preliminary research.

3.1.1 Term Structure Densities

Consider a family of discount bonds represented by

$$P_0(T) = e^{-rT}, \quad (3.1)$$

which determines a flat term structure with a constant continuously compounded rate r for each maturity date T . Here we have chosen the notation $P_0(T)$ to make the tenor variable, T , more prominent. Definition (2.15) of the term structure density, $\rho_0(T) = -\frac{\partial P_0(T)}{\partial T}$, gives rise to

$$\rho_0(T) = rP_0(T). \quad (3.2)$$

For the people working in finance, Eqn. (3.2) indicates everything — the resulting density function means nothing but the annualized interest earned on the discount bond. However, this may not be evident for those having not been working in the financial industry. It is therefore appropriate here to analyze (3.2) purely from the technical viewpoint.

By virtue of Taylor's expansion and (3.1), the short rate can be treated as the

rate of return of the bond:

$$\Delta t \cdot r \simeq \frac{P_{\Delta t}(T - \Delta t) - P_0(T)}{P_0(T)}, \quad (3.3)$$

if here Δt is an infinitesimal. This is obvious in the risk-neutral world, where investors are assumed to earn on average $r\Delta t$ in a very short time period Δt . Now suppose that the maturity T is also an infinitesimal. Substitution of (3.3) into (3.2) gives rise to

$$\begin{aligned} T\rho_0(T) &= rTP_0(T) = P_T(0) - P_0(T) \\ \implies \rho_0(T) &= \frac{P_T(0) - P_0(T)}{T}, \end{aligned} \quad (3.4)$$

by which $\rho_0(T)$ can be interpreted as the annualized interest earned on the bond.

Most often the maturity T takes ten years or above for long-term bonds. In this case, the term structure density given by (3.4) will be interpreted as the averagely annualized interest on account of flatness of the term structure.

3.1.2 Normalization Condition

With the physical meaning of $\rho_0(T)$ being revealed, we come to prove that the normalization condition is fulfilled not only because of the mathematical nature of $\rho_0(T)$ but also because of the financial nature of the density integral. To interpret the normalization condition

$$\int_0^\infty \rho_0(x) dx = 1 \quad (3.5)$$

from the financial viewpoint, we suggest to discretize it as

$$\lim_{\Delta t \rightarrow 0} \sum_{n=0}^{\infty} \rho_0(n\Delta t) \Delta t = 1. \quad (3.6)$$

In the following analysis, we just need to keep in mind the financial function of ρ as annualized interest and temporarily forget its mathematical role as a density function.

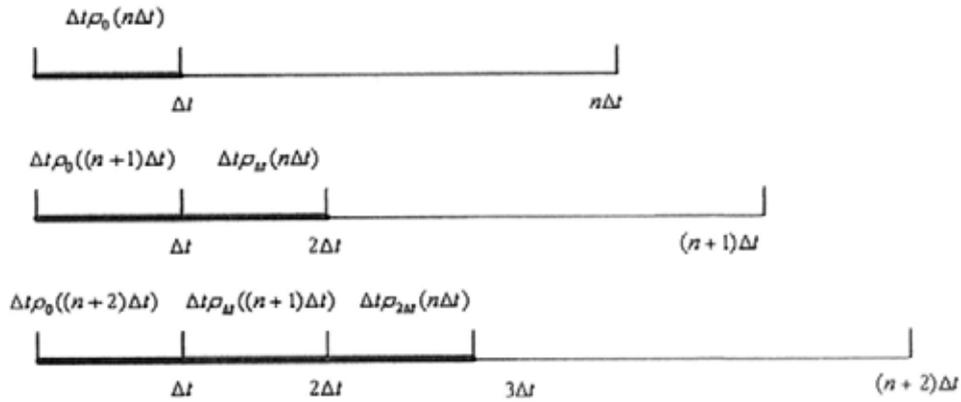


Figure 3.1: Illustration of interest earned on different bonds

infinity future:

$$\lim_{\Delta t \rightarrow 0} \sum_{n=0}^{\infty} \rho_0(n\Delta t) \Delta t = \lim_{\substack{\Delta t \rightarrow 0 \\ n \rightarrow \infty}} \sum_{i=0}^n \rho_i \Delta t ((n-i)\Delta t) \Delta t.$$

The arbitrage-free condition requires that all the interest one earns on a discount bond which is purchased initially at no cost should equal the face value one. Thus we complete the interpretation of the normalization condition (3.5).

In addition, we consider the density integral on a specific time interval $[T_1, T_2]$. Taking the discretization form (3.6) into consideration, we obtain

$$\begin{aligned} \int_{T_1}^{T_2} \rho_0(x) dx &= \lim_{\Delta t \rightarrow 0} \sum_{n=a}^b \rho_0(n\Delta t) \Delta t \\ &= \lim_{\Delta t \rightarrow 0} \sum_{i=0}^{b-a} \rho_i \Delta t ((b-i)\Delta t) \Delta t \\ &= P_{T_2-T_1}(T_1) - P_{\emptyset}(T_2), \end{aligned} \tag{3.9}$$

where $a\Delta t = T_1$ and $b\Delta t = T_2$. The last equality in (3.9) is not plain to see and needs to be illustrated with the help of Figure 3.2. The segments with the same color

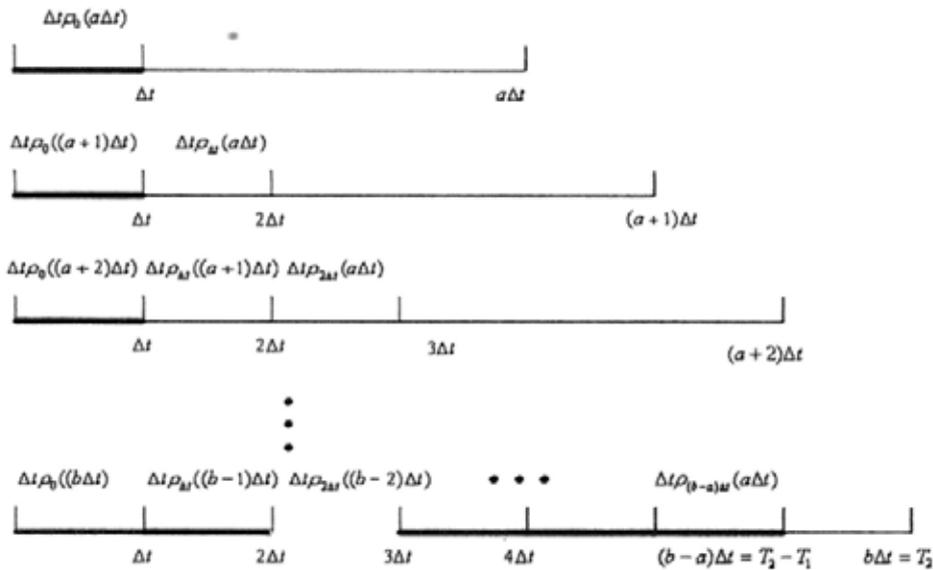


Figure 3.2: Illustration of density integral

indicate the equivalent amount of interest earned on different bonds. Note that the l.h.s of the second equality in (3.9) means to add up the first-period interest earned on a portfolio of bonds maturing differently from T_1 to T_2 — just as the red segment for the first bond maturing at $a\Delta t = T_1$, the blue for the second bond maturing at $(a + 1)\Delta t$, and so on. Whereas the r.h.s of the last equality represents the interest earned on an individual bond maturing at T_2 , during successive periods from the very beginning till $T_2 - T_1$. This is indicated by the colorful segments at the bottom in Figure 3.2. Such a “transfer” of interest from a portfolio of bonds to an individual bond is caused by the relation between forward rates and forward bond prices.

Specifically, by setting $T_1 = 0$ in (3.9), we obtain

$$\int_0^{T_2} \rho_0(x) dx = 1 - P_0(T_2),$$

which stands obviously for the reason that the interest accumulated during the life of a bond (T_2 is arbitrary) is what one expects to earn by investing this bond.

3.2 Interpretation With a Non-flat Term Structure

The analysis presented in the previous section shows that if the discount bond possesses a flat term structure, then the term structure density (2.15) can find its financial interpretation, so does the normalization condition (3.5). In this section, we will handle a general case in which the bond is valued via an instantaneous forward rate process f_{tT} :

$$P_0(T) = e^{-\int_0^T f_{0s} ds}. \quad (3.10)$$

3.2.1 Term Structure Densities

According to the law of variable upper limit integral, for the term structure density $\rho_0(T) = -\frac{\partial P_0(T)}{\partial T}$ we obtain

$$\rho_0(T) = f_{0T}P_0(T). \quad (3.11)$$

Comparing (3.11) with the flat term structure density (3.2), we see the only difference lies in the rate — its value is currently not a constant but depends on time. In fact, the instantaneous forward rate f_{0T} , as seen at the present time 0 for a contract maturing at time T , is precisely the T -year spot rate used in real markets. Therefore the term structure density given by (3.11) shares an analogous meaning as in the flat case, namely, that it represents the annualized interest earned on the bond.

3.2.2 Normalization Condition

With the term structure density $\rho_0(x)$ being interpreted, we turn to find an explanation for the density integral just as what we have discussed for the flat term structure. Integrate $\rho_0(x)$ over time interval $[T_1, T_2]$ and we obtain

$$\int_{T_1}^{T_2} \rho_0(x) dx = P_0(T_1) - P_0(T_2). \quad (3.12)$$

On the other hand, based on the relation (1.5) between forward rates and bond prices, we obtain the expression for $P_0(T_1) - P_0(T_2)$ as

$$P_0(T_1) - P_0(T_2) = F_{0T_1T_2}P_0(T_2)(T_2 - T_1). \quad (3.13)$$

In connection with (3.12), Eqn. (3.13) shows that the integral of term structure density over $[T_1, T_2]$ provides the interest earned during T_1 and T_2 on a discount bond maturing at T_2 .

Recall the flat case, in which the integral of term structure density provides, however, the interest gained during the initial time 0 and $T_2 - T_1$. It indicates that if the interest rate is a constant independent of time t , then the interest received during $[T_1, T_2]$ is equivalent to the amount collected from the beginning till $T_2 - T_1$. Such a "parallel shift" of interest income occurs only in the case of flat term structure since the rate there is independent of time t and so does the term structure density.

Specifically, we set $T_1 = 0$ and $T_2 = \infty$ in Eqn. (3.12) and Eqn. (3.13). For a bond maturing in an infinity future, its initial price equals zero. The whole return of interest on such a bond should equal the face value one. Thus we complete the interpretation of the normalization condition (3.5) in the case of non-flat term structure.

An alternative explanation for the normalization condition arises when we realize the relation between the instantaneous forward rate and the term structure density:

$$\begin{aligned} f_{0x} &= -\frac{\partial \ln P_0(x)}{\partial x} = \frac{\rho_0(x)}{P_0(x)} \\ \implies \rho_0(x) &= f_{0x}P_0(x). \end{aligned} \quad (3.14)$$

Hence the normalization condition (3.5) turns to

$$\int_0^\infty P_0(x)f_{0x} dx = 1, \quad (3.15)$$

which implies that the initial value of a continuous cash flow that pays the small amount $f_{0x}dx$ during $[x, x + dx]$ is unity.

	Flat Term Structure	Non-flat Term Structure
Bond Value Function	$P_0(T) = e^{-rT}$	$P_0(T) = e^{-\int_0^T f_{0s} ds}$
Term Structure Density	$\rho_0(T) = rP_0(T)$	$\rho_0(T) = f_{0T}P_0(T)$
Financial Interpretation	Annualized interest	Annualized interest
$\int_{T_1}^{T_2} \rho_0(x) dx$	$P_{T_2-T_1}(T_1) - P_0(T_2)$	$F_{0T_1T_2}P_0(T_2)(T_2 - T_1)$

Table 3.1: Financial interpretation of term structure density and its integrals: a comparison between the flat term structure and the non-flat term structure

3.3 Summary and Discussion

In this chapter we have interpreted the term structure density and its integrals in the language of interest rate theory. The discussion simply starts with a flat term structure and then generalizes to a non-flat term structure. Table 3.1 unfolds a comparison between these two cases.

First, the bond value function differs from one case to the other. For a flat term structure, the bond is continuously compounded at a constant rate r . For a non-flat term structure, the bond is valued via an instantaneous forward rate process f_{tT} .

Furthermore, because of the different expressions for the discount function, the resulting term structure density function $\rho_0(T)$ possesses different forms in each case. However, in both cases $\rho_0(T)$ represents the annualized interest earned on the corresponding bond,

Last of all, as an illustration of the financial nature of ρ , we try to interpret its integral over $[T_1, T_2]$. For a flat term structure, the integral provides a series of interest accumulated from the beginning until $T_2 - T_1$ on a bond maturing at T_2 . For a non-flat term structure, however, the integral gives the interest earned during T_1 and T_2 of a discount bond maturing at T_2 . We observe that when the interest

rate becomes a constant in the flat case, the interest gained during T_1 and T_2 is equivalent to the amount earned during 0 and $T_2 - T_1$. Such a “parallel shift” of interest income occurs only in the flat case since the rate there is independent of time t and so does the term structure density.

An interesting phenomenon is worth mentioning. The integral $\int_{T_1}^{T_2} \rho_0(x) dx$ itself, in whatever case, is seemingly to add up the first-period interest of a portfolio of differently-maturing bonds. However, the integral results actually account the interest earned during successive periods on an individual bond maturing at T_2 . Such a “transfer” of interest from a portfolio of bonds to an individual bond is resulted by the relationship between forward rates and forward bond prices.

Specifically, let the integration interval to be $[0, \infty]$. The integral result should be one since the whole return of interest on a bond that is purchased initially at no cost (a bond with an infinity maturity is assumed to possess a vanishing initial value) should equal the face value one. Thus we have interpreted the normalization condition from the financial viewpoint.

□ End of chapter.

Chapter 4

Analysis of the Proposed Term Structure Model

In Chapter 2 we have briefly introduced the Brody-Hughston model, which makes assumptions on a density function $\rho_t(x)$ that is obtained from the derivative of the discount function w.r.t the time left till maturity. In the process of characterizing the evolution of $\rho_t(x)$, Brody and Hughston have introduced a freely specified volatility process $\nu_t(x)$. By the incorporation of market risk premium into the volatility structure $\lambda_t = -\bar{\nu}_t$ ($\bar{\nu}_t$ denotes the expectation of $\nu_t(x)$ w.r.t $\rho_t(x)$), the resulting dynamics proves to be dominated by two factors only: the initial term structure density $\rho_0(x)$ and the volatility structure $\nu_t(x)$.

There are two reasons for the choice $\lambda_t = -\bar{\nu}_t$. First, the volatility process $\nu_t(x)$ is freely defined, which gives model users a flexibility in specifying the risk premium process. More important, when we solve the stochastic equation for the term structure density, the choice of $-\bar{\nu}_t$ could cancel the integral of some cross terms like $\int_0^t \nu_s(T-s)\bar{\nu}_s ds$, leaving a concise martingale representation for discount bond prices.

However, several problems also arise with the choice $\lambda_t = -\bar{\nu}_t$. First, the relation $\lambda_t = -\bar{\nu}_t$ does not always hold. When we price risk-free securities such as US Trea-

surely bills, the risk-neutral world is preferred and $\lambda_t = 0$. At this time, however, the volatility should not be vanishing; otherwise the evolution of the term structure will follow a deterministic trajectory. Second, study in the risk-neutral world provides a common ground for the comparison with traditional term structure models — see if the market risk premium is involved (just like in the models of short rates) or not (just like in the HJM model).

To fix the problems mentioned above, we will develop first in Section 4.1.1 the dynamics of the term structure density in the risk-neutral world, and compare our results with those obtained in the world where $\lambda_t = -\bar{\nu}_t$ (as Brody and Hughston proposed) in Table 4.1. Next, in Section 4.1.2 we extend our discussion to the real world where the market risk premium λ_t is prominently specified. The properties of the proposed model are elaborated in Section 4.2 and 4.3. Finally, we compare the new term structure model with traditional models in Table 4.2 in Section 4.4.

4.1 Dynamics of the Term Structure Density

4.1.1 Dynamics in the Risk-neutral World

To begin with, we restate the key notations and definitions used in the Brody-Hughston model to refresh readers' memory.

1. We write P_{tT} ($T \geq t \geq 0$) or $B_t(x) \triangleq P_{t,t+x}$ ($T = t + x$) for the random value at time t of a discount bond with principal \$1 maturing at time T . The initial time is set to be 0.
2. Consider an admissible term structure for which the bond price satisfies

$$\lim_{x \rightarrow \infty} B_t(x) = 0, \quad 0 < B_t(x) \leq 1, \quad \frac{\partial B_t(x)}{\partial x} < 0.$$

3. Additionally, we impose the asymptotic condition $\lim_{x \rightarrow \infty} \frac{\partial B_t(x)}{\partial x} = 0$.

4. Associated with every term structure, we define a probability density function

$$\rho_t(x) = -\frac{\partial B_t(x)}{\partial x} \quad (\text{Eqn. (2.15)}) \quad \text{w.r.t the tenor variable } x = T - t .$$

Now we develop the dynamics of the term structure density $\rho_t(x)$ in the risk-neutral world. To avoid a repeat of expressions used in Section 2.4.3, we just give their equation numbers when necessary.

Assume for each maturity T that P_{tT} is an Itô process on the interval $t \in [0, T]$. In the underlying stochastic equation (2.21), the diffusion process Σ_{tT} denotes the standard deviation of the bond and the drift process m_{tT} is subject to the arbitrage-free condition (2.23). Hence the bond price under this condition follows the equation

$$dP_{tT} = r_t P_{tT} dt + \Sigma_{tT} dW_t^*, \quad (4.1)$$

where r_t is the short rate process and $W_t^* = W_t + \int_0^t \lambda_s ds$ is a Wiener process in the risk-neutral world after a change of measure from the real world.

By virtue of Eqn. (3.14) $f_{t,t+x} = \frac{\rho_t(x)}{B_t(x)}$ and Eqn. (4.1), the stochastic equation for $B_t(x)$ is given by

$$dB_t(x) = (dP_{tT})|_{T=t+x} + \frac{\partial B_t(x)}{\partial x} dt, \quad (4.2)$$

$$= (r_t - f_{t,t+x})B_t(x) dt + \Sigma_{t,t+x} dW_t^*. \quad (4.3)$$

We observe that the drift coefficient is the difference between the short rate r_t at time t and the instantaneous forward rate $f_{t,t+x}$ as seen at time t but for a future time $t+x$. As the bond approaches its maturity, namely, in the limit $t \rightarrow T$, the drift vanishes as a consequence of the relation (1.9) $r_t = f_{tt}$. Moreover, the standard deviation $\Sigma_{t,t+x}$ also tends to vanish since no uncertainty exists for a maturing bond with a definite value. Therefore both sides of (4.3) equal zero when the bond matures.

For the term structure density $\rho_t(x) = -\frac{\partial B_t(x)}{\partial x}$, we obtain

$$d\rho_t(x) = \left(r_t \rho_t(x) + \frac{\partial \rho_t(x)}{\partial x} \right) dt + \omega_t(x) dW_t^*, \quad (4.4)$$

where we have written

$$\omega_t(x) = -\frac{\partial \Sigma_{t,t+x}}{\partial x} \quad (4.5)$$

to simplify the notation. Because of the normalization condition on $\rho_t(x)$, the drift and the diffusion processes are subject to the conditions (2.25) and (2.26), respectively. As a result, r_t can be expressed as either (2.27) or (2.28), and $\omega_t(x)$ can be expressed in terms of an exogenously specified volatility process $\nu_t(x)$ as (2.29) shows. Alternatively, we recommend to introduce the volatility process $\sigma_t(x)$ such that

$$\omega_t(x) = \rho_t(x) \sigma_t(x). \quad (4.6)$$

By virtue of the condition (2.26) $\int_0^\infty \omega_t(x) dx = 0$, the volatility $\sigma_t(x)$ is constrained to zero mean w.r.t $\rho_t(x)$:

$$E_\rho[\sigma_t(x)] = 0. \quad (4.7)$$

Proposition 4.1. *The general admissible term structure evolution based on the filtration generated by a Brownian motion W_t^* in the risk-neutral world is given by a measure-valued process $\rho_t(x)$ on $\mathcal{D}(\mathcal{R}_+^1)$ that satisfies*

$$d\rho_t(x) = \left(r_t \rho_t(x) + \frac{\partial \rho_t(x)}{\partial x} \right) dt + \rho_t(x) \sigma_t(x) dW_t^*, \quad (4.8)$$

where the volatility structure $\sigma_t(x)$ is specified exogenously subject to the zero-mean constraint (4.7) and the initial term structure density $\rho_0(x)$ is determined by the current yield curve.

The associated short rate process $r_t = \rho_t(0)$ satisfies

$$dr_t = \left(r_t^2 + \frac{\partial \rho_t(x)}{\partial x} \Big|_{x=0} \right) dt + r_t \sigma_t(x) dW_t^*. \quad (4.9)$$

Proposition 4.2. *The solution of the dynamical equation in the risk-neutral world for $\rho_t(x)$ in terms of the volatility structure $\sigma_t(x)$ and the initial term structure density $\rho_0(x)$ is*

$$\rho_t(T-t) = \rho_0(T) \frac{\exp\left(\int_0^t \sigma_{sT} dW_s^* - \frac{1}{2} \int_0^t \sigma_{sT}^2 ds\right)}{\int_t^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s^* - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}, \quad (4.10)$$

where $\sigma_{tu} = \sigma_t(u-t)$. The corresponding formula for the bond price process is

$$P_{tT} = \frac{\int_T^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s^* - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}{\int_t^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s^* - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}. \quad (4.11)$$

To avoid overuse of notations, we apply the same symbol for a term that is parameterized by the maturity T sometimes or by the time left to maturity x elsewhere. For instance, $\sigma_t(x)$, $\sigma_{t,t+x}$, and σ_{tT} are all referred to the same volatility when $T = t + x$.

Proof. The second term in the drift on the right of (4.8) can be eliminated by setting $x = T - t$, which gives us

$$\begin{aligned} d\rho_t(T-t) &= r_t \rho_t(T-t) dt + \rho_t(T-t) \sigma_t(T-t) dW_t^*, \\ &\triangleq \rho_t(T-t) dX_t, \end{aligned} \quad (4.12)$$

where $dX_t = r_t dt + \sigma_t(T-t) dW_t^*$. We observe that $\rho_t(T-t)$ is the stochastic exponential of X with initial value $\rho_0(T)$ and thus can be solved as

$$\rho_t(T-t) = \rho_0(T) \exp\left(\int_0^t r_s ds + \int_0^t \sigma_s(T-s) dW_s^* - \frac{1}{2} \int_0^t \sigma_s^2(T-s) ds\right). \quad (4.13)$$

If we rewrite (4.13) as $\rho_t(T-t) = \rho_0(T) \frac{M_{tT}}{Z_t}$, where

$$M_{tT} = \exp\left(\int_0^t \sigma_s(T-s) dW_s^* - \frac{1}{2} \int_0^t \sigma_s^2(T-s) ds\right), \quad (4.14)$$

then Z_t here plays a role as the normalization factor, indicating that

$$\exp\left(-\int_0^t r_s ds\right) = \int_t^\infty \rho_0(u) \exp\left(\int_0^t \sigma_s(u-s) dW_s^* - \frac{1}{2} \int_0^t \sigma_s(u-s)^2 ds\right) du. \quad (4.15)$$

	The World With Specified Market Risk Premium $\lambda_t = -\bar{\nu}_t$	The Risk-neutral World
Market Risk Premium	$\lambda_t = -\bar{\nu}_t$	$\lambda_t = 0 = -\bar{\sigma}_t$
Degrees of Freedom	$\nu_t(x), \rho_0(x)$	$\sigma_t(x), \rho_0(x)$
M_{tT}	$\exp\left(\int_0^t V_{sT} dW_s - \frac{1}{2} \int_0^t V_{sT}^2 ds\right)$	$\exp\left(\int_0^t \sigma_{sT} dW_s^* - \frac{1}{2} \int_0^t \sigma_{sT}^2 ds\right)$
Volatility Structure	$\nu_t(x)$ with no constraint	$\sigma_t(x)$ s.t. $E_\rho[\sigma_t(x)] = 0$

Table 4.1: Formulas for processes: a comparison between the world with specified market risk premium $\lambda_t = -\bar{\nu}_t$ and the risk-neutral world

Substituting (4.15) into (4.13) and noting that $\sigma_{tu} = \sigma_t(u - t)$, we immediately obtain (4.10). As a consequence of the relationship between bond prices and term structure densities

$$P_{tT} = \int_T^\infty \rho_t(u - t) du, \quad (4.16)$$

the formula (4.11) for the bond price process is instantly deduced. \square

Comparing Proposition 4.1 ($\lambda_t = 0$) with Proposition 2.2 ($\lambda_t = -\bar{\nu}_t$), we see the density volatilities in the two worlds — the risk-neutral world and the world with $\lambda_t = -\bar{\nu}_t$ are linked via $\sigma_t(x) = \nu_t(x) - \bar{\nu}_t$, which also verifies the zero-mean constraint on $\sigma_t(x)$ in the risk-neutral world. Comparing Proposition 4.2 ($\lambda_t = 0$) with Proposition 2.3 ($\lambda_t = -\bar{\nu}_t$), we list our findings in Table 4.1.

First, in whatever pricing world, the resulting processes for the density function and the bond price are both determined by two factors only: the initial term structure density $\rho_0(x)$ and the volatility structure $\nu_t(x)$ or $\sigma_t(x)$.

Second, the bond pricing formulas in these two worlds are both expressed in a concise martingale representation, dominated by the exponential martingale M_{tT} but associated with different volatility structures. In the world with $\lambda_t = -\bar{\nu}_t$, the volatility structure $\nu_t(x)$ is freely specified. Whereas in the risk neutral world, the volatility structure $\sigma_t(x)$ should possess zero mean w.r.t $\rho_t(x)$.

Last of all, the market risk premium process λ_t proves to be irrelevant in both cases when we price interest rate derivatives but because of different reasons. In Brody and Hughston's work, λ_t is eliminated by incorporating it into the volatility structure. In our work, λ_t is immaterial because all individual are indifferent to risk in the risk-neutral world and all investors require no compensation for risk.

The differences and similarities between density processes in the two worlds, as revealed in Table 4.1, also exist for the short rate process and the forward rate process. We will derive in the below the rate processes in the risk-neutral world.

Proposition 4.3. *The short rate process $r_t = \rho_t(0)$ is given by*

$$r_t = \rho_0(t) \frac{\exp\left(\int_0^t \sigma_{st} dW_s^* - \frac{1}{2} \int_0^t \sigma_{st}^2 ds\right)}{\int_t^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s^* - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}. \quad (4.17)$$

In particular, in a deterministic model with $\sigma_{st} = 0$, we obtain

$$r_t = \rho_0(t) / \int_t^\infty \rho_0(u) du = f_{0t}. \quad (4.18)$$

We know that in a stochastic model, $r_t = f_{tt}$, which means the instantaneous forward rate at time t for an infinitely near future is exactly the short rate at that time. However, it follows from (4.18) that in a deterministic model, $f_{tt} = f_{0t}$, owing to the vanishing of randomness.

Proposition 4.4. *The money market account B_t satisfying $dB_t = r_t B_t dt$ and $B_0 =$*

1 follows the dynamic process

$$B_t = \frac{1}{\int_t^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s^* - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}. \quad (4.19)$$

Proof. This formula follows directly from (4.15). \square

Recall the definition (2.8) of density martingale Λ_t , which is introduced when we discuss the martingale model in Section 2.1.2. It is known that $Z_t B_t = \Lambda_t$, where Z_t is the state price density. Following this line and surveying Proposition 2.6, we obtain the density martingale Λ_t in Brody and Hughston's work as

$$\Lambda_t = \exp\left(\int_0^t \bar{\nu}_s dW_s - \frac{1}{2} \int_0^t \bar{\nu}_s^2 ds\right). \quad (4.20)$$

It verifies that the market risk premium in Brody and Hughston's preference is chosen to be $-\bar{\nu}_t$. On the other hand, indicated by (4.19), the density martingale Λ_t in the risk-neutral world is unity. This follows as a result of the vanishing risk premium in the risk-neutral world. In fact, our density martingale can also be expressed in an analogous form as (4.20):

$$\Lambda_t = \exp\left(\int_0^t \bar{\sigma}_s dW_s^* - \frac{1}{2} \int_0^t \bar{\sigma}_s^2 ds\right), \quad (4.21)$$

where $\bar{\sigma}_t = E_\rho[\sigma_t(x)]$ is, however, constrained to be zero.

Proposition 4.5. *By the relation $f_{tT} = -\frac{\partial \ln P_{tT}}{\partial T}$, the instantaneous forward rate process is given by*

$$f_{tT} = \rho_0(T) \frac{\exp\left(\int_0^t \sigma_{sT} dW_s^* - \frac{1}{2} \int_0^t \sigma_{sT}^2 ds\right)}{\int_T^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s^* - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}. \quad (4.22)$$

Note that when $T = t$, Eqn. (4.22) becomes (4.17). In other words, the instantaneous forward rate process reduces to the short rate process as $T \rightarrow t$. Besides, it follows from (4.10), (4.11), and (4.22) that $\rho_t(T-t) = P_{tT} f_{tT}$, which confirms our finding in Section 3.2.1 that $\rho_t(T-t)$ can be interpreted in financial practice as the annualized interest earned on a discount bond.

4.1.2 Dynamics in the Real World

With the aid of Girsanov's Theorem, a theory of change of probability measure, we immediately extend our discussion to the real world, where the market risk premium λ_t is prominently specified.

Proposition 4.6. *The solution of the dynamical equation in the real world for $\rho_t(x)$ in terms of the volatility structure $\sigma_t(x)$, the initial term structure density $\rho_0(x)$, and the market risk premium process λ_t is*

$$\rho_t(T-t) = \rho_0(T) \frac{\exp\left(\int_0^t \sigma_{sT} dW_s + \int_0^t \sigma_{sT} \lambda_s ds - \frac{1}{2} \int_0^t \sigma_{sT}^2 ds\right)}{\int_t^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s + \int_0^t \sigma_{su} \lambda_s ds - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}, \quad (4.23)$$

where $\sigma_{tu} = \sigma_t(u-t)$. The corresponding formula for the bond price process is

$$P_{tT} = \frac{\int_T^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s + \int_0^t \sigma_{su} \lambda_s ds - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}{\int_t^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s + \int_0^t \sigma_{su} \lambda_s ds - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}. \quad (4.24)$$

Proposition 4.7. *The short rate process $r_t = \rho_t(0)$ is given by*

$$r_t = \rho_0(t) \frac{\exp\left(\int_0^t \sigma_{st} dW_s + \int_0^t \sigma_{st} \lambda_s ds - \frac{1}{2} \int_0^t \sigma_{st}^2 ds\right)}{\int_t^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s + \int_0^t \sigma_{su} \lambda_s ds - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}. \quad (4.25)$$

In particular, in a deterministic model with $\sigma_{st} = 0$, we obtain

$$r_t = \rho_0(t) / \int_t^\infty \rho_0(u) du = f_{0t}. \quad (4.26)$$

Proposition 4.8. *The money market account B_t satisfying $dB_t = r_t B_t dt$ and $B_0 = 1$ follows the dynamic process*

$$B_t = \frac{1}{\int_t^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s + \int_0^t \sigma_{su} \lambda_s ds - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}. \quad (4.27)$$

Proposition 4.9. *By the relationship $f_{tT} = -\frac{\partial \ln P_{tT}}{\partial T}$, the instantaneous forward rate process is given by*

$$f_{tT} = \rho_0(T) \frac{\exp\left(\int_0^t \sigma_{sT} dW_s + \int_0^t \sigma_{sT} \lambda_s ds - \frac{1}{2} \int_0^t \sigma_{sT}^2 ds\right)}{\int_T^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s + \int_0^t \sigma_{su} \lambda_s ds - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}. \quad (4.28)$$

Compare each process in the real world with the corresponding process in the risk-neutral world, for example, the bond price processes (4.24) and (4.11). We observe that in addition to the initial term structure density $\rho_0(x)$ and the volatility structure $\sigma_t(x)$, the market risk premium λ_t also plays a crucial role in determining the term structure dynamics.

4.2 Martingale Representations

Now let us turn to study the representations for the term structure density (4.10), the bond price (4.11), and the related rates (4.17) (4.22) in the risk-neutral world.

The term M_{tT} , which is defined by (4.14), plays a crucial role in all the formulas. For each maturity T , the process M_{tT} ($0 \leq t \leq T$) is the exponential martingale of the volatility structure σ_{tT} with initial value $M_{0T} = 1$, satisfying

$$dM_{tT} = M_{tT}\sigma_{tT} dW_t^*.$$

With this martingale family, the formula (4.10) for $\rho_t(T-t)$ can be alternatively expressed as

$$\rho_t(T-t) = \rho_0(T) \frac{M_{tT}}{\int_t^\infty \rho_0(u) M_{tu} du}, \quad (4.29)$$

and the discount bond family has the representation

$$P_{tT} = \frac{\int_T^\infty \rho_0(u) M_{tu} du}{\int_t^\infty \rho_0(u) M_{tu} du}. \quad (4.30)$$

Obviously, if we can specify the initial term structure density $\rho_0(x)$ and choose certain martingales M_{tT} such that the integrals in (4.29) and (4.30) can be carried out explicitly, then we will obtain the closed forms for $\rho_t(T-t)$ and P_{tT} . Brody and Hughston have provided an example in [17, 19], semilinear models, to illustrate this idea. However, we should point out that the form of the initial density given

in [17, 19] is short of rationality. Only a litter modification on it deserves a more reasonable explanation.

In [17, 19] Brody and Hughston wrote

$$\rho_0(u) = \int_0^\infty e^{-ur} \phi(r) dr \quad (4.31)$$

for the initial term structure density, where $\phi(r)$ is the inverse Laplace transform of $\rho_0(u)$. Nevertheless, we suggest to write

$$\rho_0(u) = \int_0^\infty r e^{-ur} \phi(r) dr \quad (4.32)$$

instead, where we use $r e^{-ur}$ to replace e^{-ur} in (4.31). This is because associated with $\rho_0(u, r) = r e^{-ur}$ there is a flat discount bond $P_0(u, r) = e^{-ur}$ with maturity u for a constant continuously compounded rate r . In other words, the initial discount function corresponding to (4.32) is given by

$$P_0(u) = \int_0^\infty e^{-ur} \phi(r) dr, \quad (4.33)$$

which can be regarded as a weighted superposition of elementary discount functions e^{-ur} . Taking $u \rightarrow 0$, we find the function $\phi(r)$ should satisfy the normalization condition as a density behaves. Indeed, if we restrict our consideration to non-negative functions, then $\phi(r)$ can be interpreted as a density function to characterize the distribution of the short rate r . Such a notion inspires us to treat the whole term structure as a superposition of “local” flat term structures on a continuously compounded basis. We will develop this idea further in Chapter 7, specifically when we try to explain why the Tsallis entropy is superior to the Shannon entropy in initial calibration.

In order to complete the example on semilinear models, we will follow a similar procedure as shown in [17, 19]. In the risk-neutral world we choose the martingale

family

$$M_{tT} = \exp \left((A + BT)M_t - \frac{1}{2}(A + BT)^2 Q_t \right), \quad (4.34)$$

where Q_t is the associated quadratic variation satisfying $(dM_t)^2 = dQ_t$. This model is obtained by setting $\sigma_{tT} = (A + BT)\vartheta_t$ in Proposition 4.2, where the process ϑ_t is defined by $dM_t = \vartheta_t dW_t^*$.

Then the u -integrals in (4.29) and (4.30) can be carried out explicitly for $\rho_t(T-t)$ and P_{tT} :

$$\rho_t(T-t) = \frac{\int_0^\infty \phi(r)r e^{-Tr} M_{tT} dr}{\int_0^\infty \phi(r)r \left(\int_t^\infty e^{-ur} M_{tu} du \right) dr} \quad (4.35)$$

$$P_{tT} = \frac{\int_0^\infty \phi(r)r \left(\int_T^\infty e^{-ur} M_{tu} du \right) dr}{\int_0^\infty \phi(r)r \left(\int_t^\infty e^{-ur} M_{tu} du \right) dr}. \quad (4.36)$$

Here the bracket expression in the integrand in the numerator is given by

$$\int_T^\infty e^{-ur} M_{tu} du = \frac{1}{|B|\sqrt{Q_t}} \exp \left(\frac{(M_t - B^{-1}r)^2}{2Q_t} + AB^{-1}r \right) \times \mathcal{N}(Y_{tT}),$$

where $Y_{tT} = \pm \frac{M_t - B^{-1}r}{\sqrt{Q_t}} \mp (A + BT)\sqrt{Q_t}$, and $\mathcal{N}(\cdot)$ denotes the standard normal distribution. The sign \pm in Y_{tT} is chosen according to the sign of B .

As an illustration of semilinear models, we take a Dirac function $\phi(r) = \delta(r - R)$ as the distribution function for r and obtain the initial discount function $P_0(T) = e^{-RT}$, corresponding to a flat term structure with a constant continuously compounded rate R . As time t passes, the initial flat discount function evolves, however, randomly:

$$P_{tT} = \frac{\mathcal{N} \left(\pm \frac{M_t - B^{-1}R}{\sqrt{Q_t}} \mp (A + BT)\sqrt{Q_t} \right)}{\mathcal{N} \left(\pm \frac{M_t - B^{-1}R}{\sqrt{Q_t}} \mp (A + Bt)\sqrt{Q_t} \right)}. \quad (4.37)$$

Finally, we have to write a few words about the similarity between our results and those obtained in [17, 19]. Although the martingale family M_{tT} we choose here appears in an analogous form as Brody and Hughston's, they are essentially different. M_{tT} , defined by (4.14) in our work, is the exponential martingale associated with

the volatility structure σ_{tT} in the risk-neutral world, whereas Brody and Hughston's N_{tT} , defined via

$$N_{tT} = \exp \left(\int_0^t V_{sT} dW_s - \frac{1}{2} \int_0^t V_{sT}^2 ds \right),$$

is the exponential martingale associated with the volatility structure V_{tT} in the specified world with $\lambda_t = -\bar{v}_t$. These two martingales are linked by

$$M_{tT} = \frac{N_{tT}}{N_t}, \quad (4.38)$$

where $N_t = \exp \left(\int_0^t \bar{v}_s dW_s - \frac{1}{2} \int_0^t \bar{v}_s^2 ds \right)$ is the exponential martingale associated with \bar{v}_t in the world with $\lambda_t = -\bar{v}_t$.

4.3 Properties of the Proposed Model

Some properties of the proposed model have been briefly discussed in Chapter 2 but far from enough. In this section, some intrinsic properties will be further explored based on the formulas we have deduced in Section 4.1. Our focus is on the expressions and features of the short rate process.

First, let us answer the question left behind in Chapter 2 — why the relation $r_t = \rho_t(0)$ stands. In Section 2.4.3 when we verify the preservation of the normalization condition in density dynamics given by (2.24), the relation (2.27) $r_t = \rho_t(0)$ has already been taken into use but without explanation. Here we attempt to analyze this expression based on our discussion in Chapter 3. We have proved in Section 3.2.1 that the term structure density $\rho_t(x)$, associated with a bond priced at time t and maturing in x units of time, is actually referred to the annualized interest accumulated on the bond just after time t till the maturity. Mathematically,

$$\rho_t(x) = f_{t,t+x} P_t(x). \quad (4.39)$$

In particular, when the bond is maturing, $\rho_t(0)$ represents the instantaneous interest of the maturing bond, i.e. $\rho_t(x) = f_{t,t}$. By virtue of the relation (1.9) $r_t = f_{t,t}$, we clarify the relation $r_t = \rho_t(0)$.

Second, an alternative expression $r_t = -E_\rho \left[\frac{\partial \ln \rho_t(x)}{\partial x} \right]$ for the short rate process is more suggestive when we depict the evolution of the density process. In this form, r_t is actually minus the expectation of the gradient of the log-likelihood function w.r.t $\rho_t(x)$. Following this line, the evolution of the density function is alternatively characterized by

$$\frac{d\rho_t(x)}{\rho_t(x)} = \left(\frac{\partial \ln \rho_t(x)}{\partial x} - E_\rho \left[\frac{\partial \ln \rho_t(x)}{\partial x} \right] \right) dt + \sigma_t(x) dW_t^*. \quad (4.40)$$

An advantage of this representation is that the normalization condition on $\rho_t(x)$ can be evidently inspected by observing

$$E_\rho \left[\frac{d\rho_t(x)}{\rho_t(x)} \right] = 0.$$

Finally, by virtue of either expression for r_t , Eqn. (2.27) or Eqn. (2.28), we find that the modeled short rates under the current framework are positive. Furthermore, the expression (2.28) $r_t = \rho_t(0)$ ensures a reasonable range for the fluctuation of short rates. This is because $\rho_t(0)$ is derived from an admissible term structure and its value is thus restricted within $[0, 1]$.

4.4 Comparison With Traditional Models

In order to find out the improvements over traditional term structure models, we compare the proposed model with the traditional models and present the key results in Table 4.2. Our focus is on the object of study, the role of market risk premium, the choice of volatility structure, and interest rate positivity of each model.

	Equilibrium Models			No-arbitrage Models		HJM	New Model
	R.B.	Vasick	C.I.R.	Ho-Lee	Hull-White		
Object of Study	r_t	r_t	r_t	r_t	r_t	f_{tT}	$\rho_t(x)$
Mean Reversion	×	✓	✓	×	✓		✓
Degrees of Freedom	λ, r	λ, r	λ, r	λ, r P_{0x}	λ, r P_{0x}	Ω_{tT}^{HJM} P_{0x}	σ_{tT} $\rho_0(x)$
Initial Term Structure	×	×	×	✓	✓	✓	✓
Positivity Definite	×	×	✓	×	×	×	✓
Others				Markov	Markov	Non-Markov	

Table 4.2: Comparison among the proposed model and traditional models

To begin with, the proposed model neither focuses on the short rate or on the forward rate as traditional models have tried to do. Instead, it intends to depict the evolution of annualized interest earned on a discount bond. Such interest income can be represented in the mathematical language by a density function that is obtained from the derivative of the discount function w.r.t the time left till maturity. There are several advantages of studying the term structure density $\rho_t(x)$. First, the model on $\rho_t(x)$ provides us with a novel angle to look into the fixed-income market. Some conventional interest rate derivatives are now endowed with “mathematical roles”. For example, a discount function can be regarded as an integral of the associated term structure density. Second, many powerful tools in statistics such as information geometry are applicable to the study on $\rho_t(x)$. Finally, many economic properties of rates, such as the proper fluctuation range and interest rate positivity, are guaranteed by the mathematical nature of the term structure density.

Next, with the proposed model, the term structure dynamics is fully and only

determined by two factors: the initial term structure density $\rho_0(x)$ and the volatility structure $\sigma_t(x)$, whereas the market risk premium λ_t is immaterial. Recall in the HJM model, the risk premium also proves to be irrelevant when we price interest rate derivatives in the risk-neutral world. This is a major difference between the HJM model and the models of short rates. It follows that more common grounds exist between the proposed model and the HJM model.

Moreover, the volatility structure $\sigma_t(x)$ we recommend to use in the risk-neutral world possesses zero mean. By use of this volatility structure, we can obtain a concise martingale representation for discount bonds.

Finally, by virtue of (2.27) $r_t = \rho_t(0)$, we can incorporate r_t directly into the dynamics of $\rho_t(x)$ and thus guarantee interest rate positivity in the proposed model.

It follows from the above discussion that it is the dual role of the term structure density — as the probability density function in mathematical research and as the annualized interest in financial practice — that makes the proposed model stand out. One more thing to be mentioned is that the initial term structure density cannot be straightly told from real markets. In Chapter 7, we will introduce an initial calibration method, which is based on the maximization of the Tsallis entropy, to translate the current market information to the initial term structure density.

□ End of chapter.

Chapter 5*

Design of the HJM Volatility Structure

In a term structure model, the volatility structure is of tremendous importance in determining the ultimate properties of the resulting rates. For example, the bond volatility under the current framework should be vanishing both at the initial time and in the infinity future. It leads to the zero-mean constraint on the density volatility and thus guarantees interest rate positivity. On the contrary, an improper volatility structure may result in undesirable properties for the underlying model. For example, if the bond volatility of the HJM model is freely specified, the resulting rates may be non-positive.

With this end in mind, in Section 5.1 we redesign the HJM volatility structure in the risk-neutral world for interest rate positivity. Besides, for convenience of simulation, we also develop in Section 5.2 the dynamics of the HJM bond volatility and ours, both under the current framework. Parallel results in the world with $\lambda_t = -\bar{\nu}_t$ (as Brody and Hughston proposed [17, 19]) are obtained in Section 5.3. A comparison of observations in these two worlds is given in Table 5.1 in Section 5.4.

5.1 HJM Volatilities in the Risk-neutral World

Comparing the HJM model and the proposed model, we find they both start with an assumption on the bond price process following the arbitrage-free condition. However, by use of different volatility structures, the proposed model successfully ensures interest rate positivity, whereas the HJM model fails. This inspires us to seek for the relationship between the HJM volatilities and ours, and then use it to redesign the HJM volatility structure for interest rate positivity. Our discussion will start in the risk-neutral world.

To begin with, we review the key notations and stochastic equations (SDEs) in both models. All of them have appeared in the previous chapters and are listed below only for convenient reference. The number on the left of each item is the equation number for each reviewed process.

In the HJM model:

(2.11) SDE for the bond value process P_{tT} :

$$\frac{dP_{tT}}{P_{tT}} = r_t dt + \Omega_{tT}^{HJM} dW_t^*$$

(2.13) The bond volatility Ω_{tT}^{HJM} and the instantaneous forward rate volatility σ_{ts}^{HJM} :

$$\Omega_{tT}^{HJM} = - \int_t^T \sigma_{ts}^{HJM} ds,$$

(2.14) SDE for the instantaneous forward rate process f_{tT} :

$$df_{tT} = \sigma_{tT}^{HJM} \int_t^T \sigma_{ts}^{HJM} ds + \sigma_{tT}^{HJM} dW_t^*,$$

In the proposed model:

(4.1) SDE for the bond value process P_{tT} :

$$dP_{tT} = r_t P_{tT} dt + \Sigma_{tT} dW_t^*,$$

(4.5) & (4.6) The bond volatility $\Sigma_{t,t+x}$ and the term structure density volatility $\sigma_t(x)$:

$$\begin{aligned}\omega_t(x) &= -\frac{\partial \Sigma_{t,t+x}}{\partial x} \\ \omega_t(x) &= \rho_t(x) \sigma_t(x).\end{aligned}$$

(4.8) SDE for the term structure density process $\rho_t(x)$:

$$d\rho_t(x) = \left(r_t \rho_t(x) + \frac{\partial \rho_t(x)}{\partial x} \right) dt + \rho_t(x) \sigma_t(x) dW_t^*,$$

(4.16) The relation between the bond price and the term structure density:

$$P_{tT} = \int_T^\infty \rho_t(u-t) du,$$

In the HJM model, the dynamics for P_{tT} in the risk neutral-world follow SDE (2.11), where Ω_{tT}^{HJM} denotes the discount bond volatility. The risk-neutral process for f_{tT} is characterized by SDE (2.14). The forward rate volatility σ_{tT}^{HJM} is connected to the bond volatility Ω_{tT}^{HJM} via the integral (2.13), or equivalently,

$$\sigma_{tT}^{HJM} = -\frac{\partial \Omega_{tT}^{HJM}}{\partial T}. \quad (5.1)$$

In the proposed model, the dynamics for P_{tT} in the risk neutral-world follow SDE (4.1), where Σ_{tT} denotes the standard deviation of the bond. However, for convenience in our thesis we call Σ_{tT} the bond volatility. The risk-neutral process for $\rho_t(x)$ is depicted by SDE (4.8), where $\sigma_t(x)$ denotes the density volatility. By virtue of (4.5) and (4.6), we find the relation between Σ_{tT} and σ_{tu} ($\sigma_{tT} = \sigma_t(T-t)$):

$$\Sigma_{tT} = \int_T^\infty \rho_t(u-t) \sigma_{tu} du. \quad (5.2)$$

Equations (2.11) and (4.1) are used respectively in the two models to govern the bond price processes. If we assume the same random source in both models, we can link their bond volatilities via

$$\Omega_{tT}^{HJM} = \frac{\Sigma_{tT}}{P_{tT}}. \quad (5.3)$$

Summing up matters so far, we illustrate in Figure 5.1 the connections between

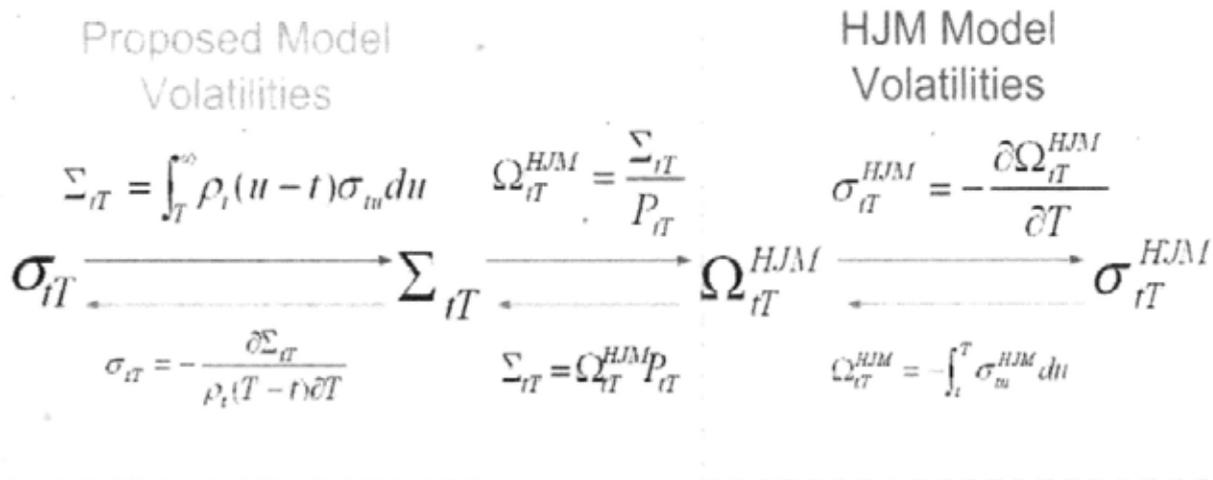


Figure 5.1: Volatility connections

the volatilities employed in the propose model and the HJM model. The left part enclosed in the yellow box exhibits the volatility connections within the proposed model, whereas the right part enclosed in the green box reveals the volatility connections within the HJM model. The bridge linking both sides is equation (5.3). Substituting (4.16) and (5.2) into, respectively, the denominator and the numerator of (5.3), we obtain

$$\Omega_{tT}^{HJM} = \frac{\Sigma_{tT}}{P_{tT}} = \frac{\int_T^\infty \rho_t(u-t)\sigma_{tu} du}{\int_T^\infty \rho_t(u-t) du}, \quad (5.4)$$

which expresses the HJM bond volatility Ω_{tT}^{HJM} in terms of the term structure density process $\rho_t(x)$ and the density volatility process σ_{tT} under the current framework.

Furthermore, applying the formula (4.10) for $\rho_t(x)$, we can define Ω_{tT}^{HJM} via the initial term structure density $\rho_0(x)$ and the density volatility structure σ_{tT} :

$$\begin{aligned}\Omega_{tT}^{HJM} &= \frac{\int_T^\infty \rho_0(u) \sigma_{tu} \exp\left(\int_0^t \sigma_{su} dW_s^* - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}{\int_T^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s^* - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du} \\ &\triangleq g_{tT}(\rho_0, \sigma).\end{aligned}\quad (5.5)$$

Here we use $g_{tT}(\rho_0, \sigma)$ to denote a process under the current framework that is parameterized by ρ_0 and σ . The notation $g_{tT}(\rho_0, \sigma)$ is chosen to make the (ρ_0, σ) dependence more prominently.

It follows from Eqn. (5.4) that the HJM bond volatility Ω_{tT}^{HJM} can be regarded as the “normalized” weighted average of the density volatility σ_{tT} w.r.t $\rho_t(x)$ under the framework of the proposed model. We use quotation marks here to indicate that it is not a standard normalization since the integral starts from T rather than t . As a consequence, for interest rate positivity the proposed model accommodates a richer volatility structure than the HJM model. This is because σ_{tT} employed in the proposed model does not need to follow the asymptotic condition $\sigma_{tT} \rightarrow 0$ as the bond matures $T - t \rightarrow 0$. Nevertheless, owing to the zero-mean constraint (4.7) on σ_{tT} , Ω_{tT}^{HJM} in the HJM model is restricted to this asymptotic condition.

So far we have derived the expression for the HJM bond volatility Ω_{tT}^{HJM} under the current framework. Next, We will probe the HJM instantaneous forward rate volatility σ_{tT}^{HJM} and seek for its relationship with the proposed model. At the first step, for σ_{tT}^{HJM} we need to take the derivative of Ω_{tT}^{HJM} w.r.t T according to (5.1). The calculation would be complicated if we directly take the derivative of g in (5.5). Instead, we consider (5.3) and obtain

$$\frac{\partial \Omega_{tT}^{HJM}}{\partial T} = \frac{-\sigma_{tT} P_{tT} + \Sigma_{tT}}{P_{tT}^2} \rho_t(T - t). \quad (5.6)$$

Noting the relation (3.14) $f_{tT} = \frac{\rho_t(T-t)}{P_{tT}}$, we deduce the representation for σ_{tT}^{HJM} as

$$\begin{aligned}\sigma_{tT}^{HJM} &= f_{tT} (\sigma_{tT} - \Omega_{tT}^{HJM}) \\ &= f_{tT} (\sigma_{tT} - g_{tT}(\rho_0, \sigma)).\end{aligned}\tag{5.7}$$

The results established above can be summarized as follows.

Proposition 5.1. *For interest rate positivity of the HJM model, we define the general HJM bond volatility in terms of the initial term structure density $\rho_0(x)$ and the density volatility $\sigma_t(x)$ under the framework of the proposed model:*

$$\Omega_{tT}^{HJM} \triangleq g_{tT}(\rho_0, \sigma) = \frac{\int_T^\infty \rho_0(u) \sigma_{tu} \exp\left(\int_0^t \sigma_{su} dW_s^* - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du}{\int_T^\infty \rho_0(u) \exp\left(\int_0^t \sigma_{su} dW_s^* - \frac{1}{2} \int_0^t \sigma_{su}^2 ds\right) du},\tag{5.8}$$

where $\sigma_{tu} = \sigma_t(u-t)$ is subject to the zero-mean constraint (4.7). $g_{tT}(\rho_0, \sigma)$ denotes a process under the current framework parameterized by ρ_0 and σ .

The corresponding HJM instantaneous forward rate volatility structure is

$$\sigma_{tT}^{HJM} = f_{tT} (\sigma_{tT} - g_{tT}(\rho_0, \sigma)),\tag{5.9}$$

where the instantaneous forward rate process f_{tT} under the current framework is given by (4.22).

As an application of Proposition 5.1, we will derive the risk-neutral process for f_{tT} under the current framework. Recall in Chapter 3 we only provide the formula for f_{tT} . The stochastic equation for f_{tT} given in the below is thus more suggestive in simulation.

Proposition 5.2. *The evolutionary process for the instantaneous forward rate, which is associated with an admissible term structure and based on the filtration generated by a Brownian motion W_t^* in the risk-neutral world, is governed by the stochastic differential equation*

$$\frac{df_{tT}}{f_{tT}} = -g_{tT}(\sigma_{tT} - g_{tT}) dt + (\sigma_{tT} - g_{tT}) dW_t^*,\tag{5.10}$$

where the density volatility $\sigma_{tT} = \sigma_t(T - t)$ can be specified exogenously subject to the constraint (4.7) and $g_{tT}(\rho_0, \sigma)$ is given by (5.8).

Proof. This follows directly from equation (2.14) and Proposition 5.1. \square

5.2 Volatility Dynamics in the Risk-neutral World

Because of Eqn. (5.2), the bond volatility Σ_{tT} under the current framework is actually a stochastic process even when the density volatility process σ_{tT} is deterministic. Besides, based on the relation between Ω_{tT}^{HJM} and Σ_{tT} , the HJM bond volatility Ω_{tT}^{HJM} also follows a stochastic trajectory as the term structure density randomly evolves. For convenience of simulation, we develop the dynamics of Σ_{tT} and Ω_{tT}^{HJM} as follows.

Proposition 5.3. *Assume in the proposed model the density volatility σ_{tT} follows a deterministic process. Let $\dot{\sigma}_{tu}$ denotes the partial derivative of σ_{tu} w.r.t time t . Then the evolutionary process for the general bond volatility structure, which is associated with an admissible term structure and based on the filtration generated by a Brownian motion W_t^* in the risk-neutral world, satisfies*

$$d\Sigma_{tT} = \left(r_t \Sigma_{tT} + \int_T^\infty \rho_t(u - t) \dot{\sigma}_{tu} du \right) dt + \left(\int_T^\infty \rho_t(u - t) \sigma_{tu}^2 du \right) dW_t^*, \quad (5.11)$$

where the term structure density $\rho_t(T - t)$ follows the dynamical equation (4.12).

The associated HJM bond volatility process $\Omega_{tT}^{HJM} \triangleq g_{tT}$ satisfies

$$\begin{aligned} d\Omega_{tT}^{HJM} \triangleq dg_{tT} &= \left[g_{tT} \left(g_{tT}^2 - \frac{\int_T^\infty \rho_t(u - t) \sigma_{tu}^2 du}{P_{tT}} \right) + \frac{\int_T^\infty \rho_t(u - t) \dot{\sigma}_{tu} du}{P_{tT}} \right] dt \\ &\quad - \left(g_{tT}^2 - \frac{\int_T^\infty \rho_t(u - t) \sigma_{tu}^2 du}{P_{tT}} \right) dW_t^*, \end{aligned} \quad (5.12)$$

where g_{tT} is given by (5.8).

Proof. Taking the operation of the differentiation under the integral sign in (5.2), we obtain

$$d\Sigma_{tT} = \int_T^\infty d(\rho_t(u-t)\sigma_{tu}) du. \tag{5.13}$$

Applying Itô's lemma and substituting the dynamical equation (4.12) for $d\rho_t(u-t)$ into (5.13), we immediately deduce Eqn. (5.11).

For the process $\Omega_{tT}^{HJM} \triangleq g_{tT}$, we have

$$g_{tT} = \frac{\Sigma_{tT}}{P_{tT}},$$

by (5.4). For the risk-neutral process $\frac{1}{P_{tT}}$, we obtain

$$d\left(\frac{1}{P_{tT}}\right) = \left(-\frac{r_t}{P_{tT}} + \frac{\Sigma_{tT}^2}{P_{tT}^3}\right) dt - \frac{\Sigma_{tT}}{P_{tT}^2} dW_t^*. \tag{5.14}$$

by Itô's lemma. Therefore, Eqn. (5.12) follows as a consequence of Itô's lemma and equations (5.11) and (5.14). □

Proposition 5.3 provides a way to update the volatility structures timely and precisely in the two models so that the models can promptly reflect the latest market information.

5.3 HJM Volatilities and Volatility Dynamics in the Specified World with $\lambda_t = -\bar{\nu}_t$

Proposition 5.1-5.3 are all presented in the risk-neutral world, where investors require no compensation for risk and thus the risk premium $\lambda_t = 0$. In Brody and Hughston's work [17,19], however, λ_t is specified by incorporating it into the volatility structure $\lambda_t = -\bar{\nu}_t$. Here the density volatility process $\nu_t(x)$ is freely specified and $\bar{\nu}_t$ denotes the expectation of $\nu_t(x)$ w.r.t the term structure density $\rho_t(x)$. It is therefore natural to extend our discussion on volatility into the specified world

with $\lambda_t = -\bar{\nu}_t$. In the paragraphs that follow, we will develop a parallel theory of Proposition 5.1-5.3. First, for interest rate positivity we redesign the HJM volatility structure in the specified world. Next, for convenience of simulation, we develop the dynamics of the HJM bond volatility and ours, both under the current framework.

As pointed out in Table 4.1 in Chapter 4, a major difference between the term structure dynamics in the two worlds — with $\lambda_t = -\bar{\nu}_t$ and $\lambda_t = 0$ — lies in the volatility structure. If we use $\nu_t(x)$ to replace $\sigma_t(x)$ in the formulas resulted in the risk-neutral world and change the risk-neutral measure to the original measure, then we will immediately obtain parallel results in the specified world. This inspire us to make necessary modifications only on the volatility structure and the risk premium in Section 5.1-5.2. To avoid a repeat of expressions appeared in the previous sections, we just restate the key expressions in the below and divide our discussion into four steps.

Step 1. In the HJM model the instantaneous forward rate volatility σ_{tT}^{HJM} is still connected to the bond volatility Ω_{tT}^{HJM} via (5.1)

$$\sigma_{tT}^{HJM} = -\frac{\partial \Omega_{tT}^{HJM}}{\partial T}.$$

In the proposed model Eqn. (5.2), which is used in the risk-neutral world to describe the relationship between the bond volatility Σ_{tT} and the density volatility σ_{tT} , is now revised to be

$$\Sigma_{tT} = \int_T^\infty \rho_t(u-t)(V_{tu} - \bar{\nu}_t) du, \quad (5.15)$$

where $V_{tu} = \nu_t(u)$. This is due to the equivalence $\sigma_t(x) = \nu_t(x) - \bar{\nu}_t$, as pointed out in Section 4.1.1.

Step 2. The bridge linking both models, as shown in Figure 5.1, is still (5.3)

$$\Omega_{tT}^{HJM} = \frac{\Sigma_{tT}}{P_{tT}}.$$

By virtue of Eqn. (5.15) for Σ_{tT} , Eqn. (2.36) for P_{tT} , and Eqn. (2.35) for $\rho_t(T-t)$, we obtain the formula for Ω_{tT}^{HJM} in the specified world:

$$\begin{aligned} \Omega_{tT}^{HJM} &= \frac{\int_T^\infty \rho_0(u) V_{tu} \exp\left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds\right) du}{\int_T^\infty \rho_0(u) \exp\left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds\right) du} - \bar{\nu}_t \\ &\triangleq U_{tT} - \bar{\nu}_t, \end{aligned} \tag{5.16}$$

where

$$U_{tT} = \frac{\int_T^\infty \rho_0(u) V_{tu} \exp\left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds\right) du}{\int_T^\infty \rho_0(u) \exp\left(\int_0^t V_{su} dW_s - \frac{1}{2} \int_0^t V_{su}^2 ds\right) du}. \tag{5.17}$$

Step 3. For the HJM instantaneous forward rate volatility σ_{tT}^{HJM} , we take the derivative of the bond volatility Ω_{tT}^{HJM} w.r.t time T and deduce (5.7)

$$\sigma_{tT}^{HJM} = f_{tT} (\sigma_{tT} - \Omega_{tT}^{HJM})$$

Substituting $\sigma_t(x) = \nu_t(x) - \bar{\nu}_t$ and (5.16) into the above equation, we deduce the formula for σ_{tT}^{HJM} in the specified world:

$$\sigma_{tT}^{HJM} = f_{tT} (V_{tT} - U_{tT}). \tag{5.18}$$

Step 4. We follow an analogous procedure as in Section 5.2 to develop the volatility dynamics of both models in the specified world with $\lambda_t = -\bar{\nu}_t$. To simplify some notations such as the derivative of $V_{tT} - \bar{\nu}_t$ w.r.t t , we remain the use of σ_{tT} as the density volatility.

The results established above are summarized as follows.

Proposition 5.4. *For interest rate positivity of the HJM model, we define the general HJM bond volatility in terms of the initial term structure density $\rho_0(x)$ and the density volatility $\nu_t(x)$ under the framework of the proposed model:*

$$\Omega_{tT}^{HJM} = U_{tT} - \bar{\nu}_t, \tag{5.19}$$

where U_{tT} is given by (5.17) and $\bar{\nu}_t$ denotes the expectation of $\nu_t(x)$ w.r.t $\rho_t(x)$.

The corresponding HJM instantaneous forward rate volatility structure is

$$\sigma_{tT}^{HJM} = f_{tT} (V_{tT} - U_{tT}), \quad (5.20)$$

where f_{tT} is given by (2.41) and $V_{tT} = \nu_t(T - t)$.

Proposition 5.5. *The evolutionary process for the instantaneous forward rate, which is associated with an admissible term structure and based on the filtration generated by a Brownian motion W_t in the specified world with $\lambda_t = -\bar{\nu}_t$, is governed by the stochastic differential equation*

$$\frac{df_{tT}}{f_{tT}} = (\bar{\nu}_t - U_{tT})(V_{tT} - U_{tT}) dt + (V_{tT} - U_{tT})(dW_t - \bar{\nu}_t dt), \quad (5.21)$$

where the density volatility $V_{tT} = \nu_t(T - t)$ is freely specified and $\bar{\nu}_t$ denotes the expectation of $\nu_t(x)$ w.r.t $\rho_t(x)$. U_{tT} is given by (5.17).

Proof. This follows directly from (2.14) and Proposition 5.4. \square

Proposition 5.6. *Assume in the proposed model the density volatility σ_{tT} follows a deterministic process. Let $\dot{\sigma}_{tu}$ denotes the partial derivative of σ_{tu} w.r.t time t . Then the evolutionary process for the general bond volatility structure, which is associated with an admissible term structure and based on the filtration generated by a Brownian motion W_t in the specified world with $\lambda_t = -\bar{\nu}_t$, satisfies*

$$d\Sigma_{tT} = (\tau_t \Sigma_{tT} + A_{tT}) dt + B_{tT} (dW_t - \bar{\nu}_t dt). \quad (5.22)$$

where $A_{tT} = \int_T^\infty \rho_t(u - t) \dot{\sigma}_{tu} du$ and $B_{tT} = \int_T^\infty \rho_t(u - t) \sigma_{tu}^2 du$. The term structure density $\rho_t(T - t)$ follows the dynamical equation (2.31). The density volatility $\sigma_{tu} = \sigma_t(u - t)$ is subject to the zero-mean constraint (4.7).

The associated HJM bond volatility structure process $\Omega_{tT}^{HJM} = U_{tT} - \bar{\nu}_t$ satisfies

$$\begin{aligned} d\Omega_{tT}^{HJM} &= \left[(U_{tT} - \bar{\nu}_t) \left((U_{tT} - \bar{\nu}_t)^2 - \frac{B_{tT}}{P_{tT}} \right) + \frac{A_{tT}}{P_{tT}} \right] dt \\ &\quad - \left((U_{tT} - \bar{\nu}_t)^2 - \frac{B_{tT}}{P_{tT}} \right) (dW_t - \bar{\nu}_t dt), \end{aligned} \quad (5.23)$$

where U_{tT} is given by (5.17) and $\bar{\nu}_t$ denotes the expectation of a freely specified process $\nu_t(x)$ w.r.t $\rho_t(x)$.

Equation (5.20) in Proposition 5.4 was first proposed by Brody and Hughston [19] without proof. Proposition 5.6 provides a way to update the volatility structures timely and precisely in the two models so that the models can promptly reflect the latest market information.

5.4 Summary and Discussion

In this chapter we have redesigned the HJM volatility structure to ensure interest rate positivity. For convenience of simulation, we have also developed the dynamics of the HJM bond volatility and ours, both under the current framework. Our discussions starts in the risk-neutral world and then extends to the world with specified market risk premium $\lambda_t = -\bar{\nu}_t$ (as Brody and Hughston proposed [17,19]), where $\bar{\nu}_t$ denotes the expectation of a freely specified process $\nu_t(x)$ w.r.t $\rho_t(x)$. Table 5.1 unfolds a comparison between observations in the specified world and the risk-neutral world. This table is completed based on Table 4.1 in Chapter 4 by adding our new discoveries to the last three rows.

First, we compare Eqn. (5.19) in Proposition 5.4 with Eqn. (5.8) in Proposition 5.1. It shows that in either world the HJM bond volatility structure Ω_{tT}^{HJM} is determined by a process under the current framework: U_{tT} corresponding the specified world or g_{tT} corresponding to the risk-neutral world. We call U_{tT} or g_{tT} the

	The World With Specified Market Risk Premium $\lambda_t = -\bar{\nu}_t$	The Risk-neutral World
Market Risk Premium	$\lambda_t = -\bar{\nu}_t$	0
Degrees of Freedom	$\nu_t(x), \rho_0(x)$	$\sigma_t(x), \rho_0(x)$
M_{tT}	$\exp\left(\int_0^t V_{sT} dW_s - \frac{1}{2} \int_0^t V_{sT}^2 ds\right)$	$\exp\left(\int_0^t \sigma_{sT} dW_s^* - \frac{1}{2} \int_0^t \sigma_{sT}^2 ds\right)$
Volatility Structure	$\nu_t(x)$ with no constraint	$\sigma_t(x)$ s.t. $E_\rho[\sigma_t(x)] = 0$
Ω_{tT}^{HJM}	$U_{tT} - \bar{\nu}_t$ ($U_{tT} = \frac{\int_T^\infty \rho_t(u-t)V_{tu} du}{\int_T^\infty \rho_t(u-t) du}$)	g_{tT} ($g_{tT} = \frac{\int_T^\infty \rho_t(u-t)\sigma_{tu} du}{\int_T^\infty \rho_t(u-t) du}$)
σ_{tT}^{HJM}	$f_{tT}(V_{tT} - U_{tT})$	$f_{tT}(\sigma_{tT} - g_{tT})$
df_{tT}	V_{tT}, U_{tT}	σ_{tT}, g_{tT}

Table 5.1: HJM volatilities for positive rates: a comparison between the world with specified market risk premium $\lambda_t = -\bar{\nu}_t$ and the risk-neutral world

dominant process. As shown in the bracket beside either U_{tT} or g_{tT} , the dominant process can be regarded as the “normalized” weighted average of the associated density volatility. In the specified world the density volatility process V_{tT} is freely specified, whereas in the risk-neutral world the density volatility σ_{tT} is restricted to possess zero mean. It is due to the zero-mean constraint on σ_{tT} that we can rewrite g_{tT} as $g_{tT} - \bar{\sigma}_t$ for Ω_{tT}^{HJM} in the risk-neutral world, which appears in an analogous form as in the specified world.

Next, we compare Eqn. (5.20) in Proposition 5.4 with Eqn. (5.9) in Proposition 5.1. For the HJM instantaneous forward rate volatility σ_{tT}^{HJM} , its process in either world is dominated by the difference between the associated density volatility (V_{tT} or σ_{tT}) and the dominant process (U_{tT} or g_{tT}).

Finally, we compare Proposition 5.5 with Proposition 5.2. For convenience of simulation, we have developed the dynamics for the instantaneous forward rate under the current framework. In either world, the forward rate process is also determined by the associated density volatility (V_{tT} or σ_{tT}) and the dominant process (U_{tT} or g_{tT}).

To sum up, we find that in either world the dominant process and the density volatility process under the current framework are of most importance in redesigning the HJM volatility structure. Since both the dominant processes U_{tT} and g_{tT} are further determined by the initial term structure density and the density volatility structure of the proposed model, the whole HJM volatility structure is thus controlled by two factors under the current framework: the initial term structure density and the density volatility structure.

□ End of chapter.

Chapter 6

Distance Between Yield Curves

Since Chapter 3 our focus has been on the Dynamical Problem of the proposed model — to characterize the evolutionary trajectory of the term structure with a given initial point (initial term structure density $\rho_0(x)$). In this chapter, we will get down to the Distance Problem — to detect the distance evolution for a pair of yield curves.

Typically there are three factors — models, parameters, and initial calibration methods — that would cause a pair of yield curves starting in the same market environment to progress differently as time evolves.

First, when we apply different term structure models, a distance process would arise from the modeled yield curves. In this case, we study the distance evolution mainly for comparing the employed models and finding out the advantages of each. For example, as we have discussed in Table 4.2 in Chapter 4, models of short rates, the HJM model, and the proposed model all differ from one another in some aspects.

Second, the distance process occurs when the yield curve dynamics are depicted by the same model but different set of parameters. Our discussion in Chapter 4 and 5 has fully demonstrated the roles of volatility structure and market risk premium in determining the term structure dynamics. In this case, by studying the distance

evolution we target at selecting a proper volatility structure in accordance with the risk premium process.

Finally and most subtly, the distance process arises when the initial term structure needs to be specified as an input and is calibrated with different methods. For example, the input of our term structure model is a density function rather than any raw data in real markets. Therefore, no matter what approach we use to translate the current market data to the initial density, the input will be different from the real distribution. For this reason, we need to study whether the initial error in term structure densities would disappear over time.

For the choices of models and parameters (like volatility structures and risk premiums), we have studied their influence on the term structure dynamics in Chapter 4 and 5, respectively. In this chapter our aim is to depict the influence of different initial term structures on the subsequent evolution under the framework of the proposed model. To begin with, in Section 6.1 we supplement a proof to a key proposition (first proposed by Brody and Hughston in [17]) in the study of distance evolution. As an illustration of this proposition, we consider in Section 6.2 the relative dynamics of two yield curves with different initial flat term structures but the same volatility structure and market risk premium.

6.1 Distance Dynamics for a Pair of Yield Curves

In Section 2.4.2 we have defined the spherical distance function of Bhattacharyya to measure the difference between two yield curves. Given a pair of yield curves, we let $\rho_1(x)$ and $\rho_2(x)$ denote the associated term structure densities on \mathcal{R}_+^1 . The Bhattacharyya spherical distance is defined by

$$\phi_{12} = \arccos \left(\int_{-\infty}^{\infty} \xi_1(x) \xi_2(x) dx \right), \quad (6.1)$$

where $\xi_i(x) = \sqrt{\rho_i(x)}$. The map $\rho(x) \rightarrow \xi(x)$ associates with every yield curve in the real world a unit vector in the Hilbert space $\mathcal{H} = \mathcal{L}^2(\mathcal{R}^1)$ or equivalently a point in the positive orthant \mathcal{S}^+ of the unit sphere in \mathcal{H} . ϕ_{12} is actually the angle between ξ_1 and ξ_2 . Because both the vectors have norm one, the value of ϕ_{12} equals the spherical distance between the points on \mathcal{S}^+ determined by the vectors ξ_1 and ξ_2 .

At any time t , formula (6.1) provides the distance measure between the given yield curves. However, we are more interested in the dynamics of the distance, namely, see how differently one yield curve evolves from the other. We thus need to generate a process for the cosine of the spherical distance ϕ_{12} between the corresponding yield curves given in formula (6.1). Obviously, an increase in $\cos \phi_{12}$ means a decrease in the distance.

Proposition 6.1. *The evolution of the Bhattacharyya spherical distance ϕ_{12} between a pair of yield curves, which is based on the filtration generated by a Brownian motion W_t in the world with market risk premium λ_t , is given by a process for $\cos \phi_{12}$ defined via (6.1) that satisfies*

$$\begin{aligned} d \cos \phi_{12} &= \left[\frac{1}{2}(r_1 + r_2) \cos \phi_{12} - \sqrt{r_1 r_2} - \frac{1}{8} \int_0^\infty (\sigma_1(x) - \sigma_2(x))^2 \xi_1(x) \xi_2(x) dx \right] dt \\ &+ \frac{1}{2} \left[\int_0^\infty (\sigma_1(x) + \sigma_2(x)) \xi_1(x) \xi_2(x) dx \right] (dW_t + \lambda_t dt), \end{aligned} \quad (6.2)$$

where $r_1 = \rho_1(0)$ and $r_2 = \rho_2(0)$ denote, respectively, the short rate processes for the given yield curves. σ_i ($i = 1, 2$) is the volatility structure of the term structure density process ρ_i , which is constrained to the zero-mean condition (4.7). $\xi_i = \sqrt{\rho_i}$ is regarded as a point evolving on the unit sphere \mathcal{S} in the Hilbert space $\mathcal{H} = \mathcal{L}^2(\mathcal{R}^1)$.

Note that the parameters in Proposition 6.1, such as r_i , σ_i , and ξ_i ($i = 1, 2$), are not constants but evolving with time t . The distance function ϕ_{12} itself is also a time-dependent variable. Here we omit the subscript t to simplify the notations.

As we have discussed in the opening words of this chapter, our aim is to depict the influence of different initial conditions, rather than parameters, on the subsequent yield curve dynamics. Hence we employ the same volatility structure and risk premium for the yield curves. Specifically, as a consequence of the zero-mean condition (4.7) on σ_t , we set

$$\sigma_t = \nu_t(x) - \bar{\nu}_t, \quad (6.3)$$

where $\bar{\nu}_t$ denotes the expectation of a freely specified process $\nu_t(x)$ w.r.t $\rho_t(x)$. Moreover, we assume $\nu_1(x) = \nu_2(x) \triangleq \nu(x)$ and define

$$\bar{\nu}_{12} = \int_t^\infty \xi_1(x)\xi_2(x)\nu(x) dx. \quad (6.4)$$

Proposition 6.2. *The Bhattacharyya distance process for two yield curves subject to the same underlying interest rate dynamics satisfies*

$$\begin{aligned} d \cos \phi_{12} &= \left(\frac{1}{2}(r_1 + r_2) \cos \phi_{12} - \sqrt{r_1 r_2} - \frac{1}{8}(\bar{\nu}_1 - \bar{\nu}_2)^2 \cos \phi_{12} \right) dt \\ &+ \left(\bar{\nu}_{12} - \frac{1}{2}(\bar{\nu}_1 + \bar{\nu}_2) \cos \phi_{12} \right) (dW_t + \lambda_t dt). \end{aligned} \quad (6.5)$$

Proof. Since Proposition 6.2 is a special case of Proposition 6.1, it is appropriate to prove the more general result in Proposition 6.1.

First, we rewrite the $\cos \phi_{12}$ definition (6.1) in terms of the term structure densities ρ_i ($i = 1, 2$) as

$$\cos \phi_{12} = \int_0^\infty \sqrt{\rho_1(x)\rho_2(x)} dx. \quad (6.6)$$

By virtue of differentiation under the integral sign, we obtain

$$d \cos \phi_{12} = \int_0^\infty d\sqrt{\rho_1(x)\rho_2(x)} dx. \quad (6.7)$$

Let $f(\rho_1, \rho_2) \triangleq \sqrt{\rho_1(x)\rho_2(x)}$ to simplify the notation. It follows from Itô's lemma

that the stochastic differential equation for $df(\rho_1, \rho_2)$ is given by

$$\begin{aligned} df(\rho_1, \rho_2) &= \frac{\partial f}{\partial \rho_1} d\rho_1 + \frac{\partial f}{\partial \rho_2} d\rho_2 \\ &+ \frac{\partial^2 f}{\partial \rho_1 \partial \rho_2} [d\rho_1, d\rho_2] + \frac{1}{2} \frac{\partial^2 f}{\partial \rho_1^2} [d\rho_1, d\rho_1] + \frac{1}{2} \frac{\partial^2 f}{\partial \rho_2^2} [d\rho_2, d\rho_2], \end{aligned} \quad (6.8)$$

where $[X, Y]$ is calculated as the quadratic variation of Itô processes X and Y .

Next, for the term structure density ρ_i ($i = 1, 2$), our results in Chapter 4 show that its process in the real world follows the stochastic equation

$$d\rho_i(x) = \left(r_i \rho_i(x) + \frac{\partial \rho_i(x)}{\partial x} \right) dt + \rho_i(x) \sigma_i(x) (dW_t + \lambda_t dt), \quad (6.9)$$

where the risk premium process λ_t assumes to be identical for both density processes ρ_1 and ρ_2 . Substituting SDE (6.9) ($i = 1, 2$) into (6.8), we deduce

$$\begin{aligned} df &= \left[\frac{1}{2} (r_1 + r_2) f - \frac{1}{8} f (\sigma_1(x) - \sigma_2(x))^2 \right] dt + \frac{1}{2} \left(\frac{\sqrt{\rho_2}}{\sqrt{\rho_1}} \frac{\partial \rho_1(x)}{\partial x} + \frac{\sqrt{\rho_1}}{\sqrt{\rho_2}} \frac{\partial \rho_2(x)}{\partial x} \right) dt \\ &+ \frac{1}{2} (\sigma_1(x) + \sigma_2(x)) f (dW_t + \lambda_t dt), \end{aligned} \quad (6.10)$$

where the integral of $\frac{1}{2} \left(\frac{\sqrt{\rho_2}}{\sqrt{\rho_1}} \frac{\partial \rho_1(x)}{\partial x} + \frac{\sqrt{\rho_1}}{\sqrt{\rho_2}} \frac{\partial \rho_2(x)}{\partial x} \right)$ (w.r.t x) over the positive real line proves to be $-\sqrt{r_1 r_2}$ after an integration by parts. Substituting this result along with (6.10) into (6.7), we immediately obtain (6.2) and thus prove Proposition 6.1.

Specifically, consider a pair of yield curves with different initial conditions but governed by the same volatility structure and market risk premium. A direct calculation leads to (6.5) in Proposition 6.2. \square

6.2 Divergence of Yield Curves With Large Initial Distance

Observing Proposition 6.2, we find that the ratio of the geometric and arithmetic means of the short rates plays a critical role in determining the behavior of $\cos \phi_{12}$.

If the initial yield curves are not so close to each other that the spherical distance in between satisfies $\cos \phi_{12} \leq \frac{\sqrt{r_1 r_2}}{\frac{1}{2}(r_1 + r_2)}$, then the drift coefficient in (6.5) is non-positive. In particular, for the deterministic process for $\cos \phi_{12}$, namely, without the interference of the random source $dW_t + \lambda_t dt$, its evolution will be characterized by the ordinary differential equation

$$d \cos \phi_{12} = \left(\frac{1}{2}(r_1 + r_2) \cos \phi_{12} - \sqrt{r_1 r_2} - \frac{1}{8}(\bar{\nu}_1 - \bar{\nu}_2)^2 \cos \phi_{12} \right) dt$$

which indicates an increase in the distance.

However, the real challenge arises when the random source is taken up and adds uncertainty to the deterministic evolution. Specifically when we specify a market risk premium λ_t such that the product $(\bar{\nu}_{12} - \frac{1}{2}(\bar{\nu}_1 + \bar{\nu}_2) \cos \phi_{12}) \lambda_t$ is positive, it will be hard to determine the sign of the real drift in (6.5). In what follows, we will investigate the complicated stochastic behavior of $\cos \phi_{12}$. As the first attempt in this area, we will narrow our focus to a pair of flat initial yield curves, each initially dominated by a continuously compounded rate r_i ($i=1, 2$) but governed by the same volatility structure, i.e.

$$\rho_i(x) = r_i e^{-r_i x}, \quad i = 1, 2; \quad r_1 \neq r_2 \quad (6.11)$$

and

$$\nu_1(x) = \nu_2(x) = a e^{-bx}, \quad (6.12)$$

where a and b are positive constants.

First, we obtain a lemma about the sign of the diffusion coefficient in (6.5).

Lemma 6.1. *Given two continuously compounded flat yield curves, for which the initial term structure densities are given in the form of (6.11) and the volatility structure is set via (6.12), suppose that the initial Bhattacharyya spherical distance*

ϕ_{12} in between is not sufficiently small such that

$$\cos \phi_{12} \leq \frac{\sqrt{r_1 r_2}}{\frac{1}{2}(r_1 + r_2)}. \quad (6.13)$$

Then the diffusion coefficient in the distance equation (6.5) is non-negative. Specifically, it vanishes if and only if the given yield curves (6.11) are identical, i.e. $r_1 = r_2$.

Proof. A simple calculation from (6.11) and (6.12) gives rise to

$$\begin{aligned} \bar{\nu}_{12} &= \frac{a\sqrt{r_1 r_2}}{b + \frac{1}{2}(r_1 + r_2)} \\ \bar{\nu}_1 &= \frac{ar_1}{b + r_1} \\ \bar{\nu}_2 &= \frac{ar_2}{b + r_2}. \end{aligned} \quad (6.14)$$

Substituting (6.14) into the diffusion term of (6.5), we obtain

$$\frac{1}{2}(\bar{\nu}_1 + \bar{\nu}_2) \cos \phi_{12} = \frac{a}{2} \left(\frac{r_1}{b + r_1} + \frac{r_2}{b + r_2} \right) \cos \phi_{12}. \quad (6.15)$$

Note that function $f(r) = \frac{r}{b+r}$ ($b > 0$) is concave for $r > 0$. Then we deduce

$$\begin{aligned} \frac{1}{2}(\bar{\nu}_1 + \bar{\nu}_2) \cos \phi_{12} &= a \left(\frac{1}{2}f(r_1) + \frac{1}{2}f(r_2) \right) \cos \phi_{12} \\ &\leq a \cdot f \left(\frac{1}{2}r_1 + \frac{1}{2}r_2 \right) \cos \phi_{12} \\ &= a \cdot \frac{\frac{1}{2}(r_1 + r_2) \cos \phi_{12}}{b + \frac{1}{2}(r_1 + r_2)} \\ &\leq \frac{a\sqrt{r_1 r_2}}{b + \frac{1}{2}(r_1 + r_2)} = \bar{\nu}_{12}, \end{aligned} \quad (6.16)$$

where the first inequality is due to the concave property of function $f(\cdot)$ and the second inequality holds upon the assumption. It follows that the diffusion coefficient $(\bar{\nu}_{12} - \frac{1}{2}(\bar{\nu}_1 + \bar{\nu}_2) \cos \phi_{12})$ in (6.5) is non-negative. It equals zero if and only if both the two inequalities become equalities simultaneously, namely, $r_1 = r_2$ for two identical yield curves. \square

The result given in Lemma 6.1 is only applicable to the initial state since the yield curves afterwards will not be continuously compounded flat any more. Now let us analyze the evolution of $\cos \phi_{12}$ when it just leaves the initial state, namely, during the first time interval $[0, dt]$.

First, we let

$$\mu(\phi) = \frac{1}{2}(\tau_1 + \tau_2) \cos \phi_{12} - \sqrt{\tau_1 \tau_2} - \frac{1}{8}(\bar{\nu}_1 - \bar{\nu}_2)^2 \cos \phi_{12}, \quad (6.17)$$

$$h(\phi) = \bar{\nu}_{12} - \frac{1}{2}(\bar{\nu}_1 + \bar{\nu}_2) \cos \phi_{12}, \quad (6.18)$$

denote, respectively, the drift and the diffusion coefficients in (6.5). At the initial time $t = 0$, the drift coefficient $\mu(\phi)$ can be further simplified as

$$\mu(\phi) = -\frac{1}{8}(\bar{\nu}_1 - \bar{\nu}_2)^2 \cos \phi_{12}, \quad (6.19)$$

since the initial spherical distance is calculated to be $\cos \phi_{12} = \frac{\sqrt{\tau_1 \tau_2}}{\frac{1}{2}(\tau_1 + \tau_2)}$. This indicates a fulfillment of condition (6.13) in Lemma 6.1 with the equality sign. The diffusion coefficient $h(\phi)$ is strictly positive and will be vanishing when $\tau_1 = \tau_2$.

Next, it is known that dW_t , the infinitesimal displacement of a standard Brownian motion, is normally distributed with mean zero and variance dt . For the distance equation $d \cos \phi_{12} = \mu(\phi)dt + h(\phi)(dW_t + \lambda_t dt)$, we hence obtain

$$dW_t = \frac{d \cos \phi_{12} - \mu(\phi)dt - h(\phi)\lambda_t dt}{h(\phi)} \sim \mathcal{N}(0, dt). \quad (6.20)$$

Furthermore, by dividing dW_t by the standard deviation \sqrt{dt} , we normalize dW_t and deduce a standard normally distributed term

$$U = \frac{d \cos \phi_{12} - \mu(\phi)dt - h(\phi)\lambda_t dt}{h(\phi)\sqrt{dt}} \sim \mathcal{N}(0, 1). \quad (6.21)$$

Given a probability $1 - \alpha$, where α usually takes a small value over the interval $(0, 1)$, we can always find the numbers $-u_{\frac{\alpha}{2}}$ and $u_{\frac{\alpha}{2}}$ such that U lies in between

these two with probability $1 - \alpha$:

$$P\{|U| < u_{\frac{\alpha}{2}}\} = 1 - \alpha. \quad (6.22)$$

In fact, the positive number $u_{\frac{\alpha}{2}}$ is the upper α quantile of the standard normal distribution satisfying

$$P\{U > u_{\frac{\alpha}{2}}\} = \frac{\alpha}{2}. \quad (6.23)$$

As is often the case, $u_{\frac{\alpha}{2}} = 1.96$ for $\alpha = 0.05$ and we call the interval $[-1.96, 1.96]$ the 95% confidence interval of U . This means that with probability 0.95 the standard normally distributed variable U lies in the interval $[-1.96, 1.96]$, or in statistics 95% of the sample following the standard normal distribution will fall into the interval $[-1.96, 1.96]$. Following this line, we conclude that the inequality below holds with probability $1 - \alpha$:

$$|d \cos \phi_{12} - \mu(\phi)dt - h(\phi)\lambda_t dt| < u_{\frac{\alpha}{2}} h(\phi) \sqrt{dt}, \quad (6.24)$$

or equivalently,

$$(\mu(\phi) + h(\phi)\lambda_t) dt - u_{\frac{\alpha}{2}} h(\phi) \sqrt{dt} < d \cos \phi_{12} < (\mu(\phi) + h(\phi)\lambda_t) dt + u_{\frac{\alpha}{2}} h(\phi) \sqrt{dt}. \quad (6.25)$$

Moreover, if the market risk premium is bounded such that

$$\lambda_t \leq -\frac{\mu(\phi)}{h(\phi)} - \frac{u_{\frac{\alpha}{2}}}{\sqrt{dt}}, \quad (6.26)$$

then $d \cos \phi_{12} < 0$ with a probability no less than $1 - \alpha$. It follows that two yield curves with large initial distance (such that condition (6.13) is satisfied) would tend to diverge in a second after the initial time with a significant probability.

From the above discussion we find that condition (6.13) in Lemma 6.1 still holds at time dt since $\cos \phi_{12}$ decreases during the first time interval $[0, dt]$. The diffusion

coefficient $h(\phi)$ in SDE (6.5) at time dt is, however, not guaranteed to be positive any more. This is because Lemma 6.1 is unavailable at arbitrary time when the term structure densities ρ_i ($i = 1, 2$) evolve away from the initial flat term structures (6.11). At any time $t > 0$, the $1 - \alpha$ confidence interval for $d \cos \phi$ becomes

$$\left[(\mu(\phi) + h(\phi)\lambda_t) dt - u_{\frac{\alpha}{2}} |h(\phi)| \sqrt{dt}, (\mu(\phi) + h(\phi)\lambda_t) dt + u_{\frac{\alpha}{2}} |h(\phi)| \sqrt{dt} \right]. \quad (6.27)$$

If the market risk premium λ_t is bounded such that

$$\begin{aligned} \lambda_t &\leq -\frac{\mu(\phi)}{h(\phi)} - \frac{u_{\frac{\alpha}{2}}}{\sqrt{dt}}, \quad \text{for } h(\phi) > 0, \\ \lambda_t &\geq -\frac{\mu(\phi)}{h(\phi)} + \frac{u_{\frac{\alpha}{2}}}{\sqrt{dt}}, \quad \text{for } h(\phi) < 0, \end{aligned} \quad (6.28)$$

then the yield curves will continue to diverge from each other with a probability no less than $1 - \alpha$. The results established above will be summarized as follows.

Proposition 6.3. *Consider two continuously compounded flat yield curves with initial term structure densities (6.11) and the same volatility structure (6.12). Assume that the initial Bhattacharyya spherical distance ϕ_{12} in between is not sufficiently small such that condition (6.13) is fulfilled:*

$$\cos \phi_{12} \leq \frac{\sqrt{r_1 r_2}}{\frac{1}{2}(r_1 + r_2)},$$

and that the market risk premium λ_t is bounded such that

$$\begin{aligned} \lambda_t &\leq -\frac{\mu(\phi)}{h(\phi)} - \frac{u_{\frac{\alpha}{2}}}{\sqrt{dt}}, \quad \text{for } t = 0; \\ \lambda_t &\leq -\frac{\mu(\phi)}{h(\phi)} - \frac{u_{\frac{\alpha}{2}}}{\sqrt{dt}}, \quad \text{for } h(\phi) > 0 \text{ and } t > 0; \\ \lambda_t &\geq -\frac{\mu(\phi)}{h(\phi)} + \frac{u_{\frac{\alpha}{2}}}{\sqrt{dt}}, \quad \text{for } h(\phi) < 0 \text{ and } t > 0. \end{aligned} \quad (6.29)$$

Here $\mu(\phi)$ and $h(\phi)$ are given by, respectively, (6.17) and (6.18). The parameter α is an arbitrary constant in $(0, 1)$ and $u_{\frac{\alpha}{2}}$ is the upper α quantile of the standard normal distribution. Then the given yield curves will tend to diverge as time passes with a significant probability no less than $1 - \alpha$.

6.3 Summary and Discussion

In this chapter we have investigated the relative dynamics of two yield curves with different initial flat term structures but the same volatility structure and market risk premium. Each yield curve is initially dominated by a constant continuously compounded rate. By use of confidence interval, we have proved that the given yield curves tend to diverge with a significant probability under two conditions: 1) the initial difference is large enough (such that the initial Bhattacharyya spherical distance is larger than the arc cosine of the ratio of the geometric and arithmetic means of the two short rates; see condition (6.13) in Lemma 6.13); 2) the risk premium process λ_t is bounded within a certain range (see condition (6.29) in Proposition 6.3). For the given flat initial yield curves, the initial distance condition is always fulfilled. As for the boundary condition on λ_t , we have to admit that it is not easy to fulfill in either theory or practice. A particular case arises in the risk-neutral world where $\lambda_t = 0$ and the condition (6.29) in Proposition 6.3 is thus satisfied. As a consequence, two flat yield curves in the risk-neutral world would diverge as long as their initial rates differ (no matter how little) from each other.

The phenomenon of divergence between two yield curves indicates that the initial error in term structure densities will lead to a significant bias in the subsequent predication for bond prices. For this reason, the initial term structure calibration is of tremendous importance in the whole term structure modeling. In the next chapter we will exert ourselves to developing a new initial calibration algorithm based on the maximization of the Tsallis entropy.

□ End of chapter.

Chapter 7

Initial Calibration of the Proposed Model

The first step to implement a term structure model is initial calibration — in our proposed model we need to translate the current market data to the initial term structure density. Brody and Hughston [19] have introduced a model-independent calibration method based on maximization of the Shannon entropy. The idea is to treat the Shannon entropy as a functional of the term structure density and express the given market data as constraints on the density function. The advantage of such an entropic method is that it avoids over fitting of model parameters by imposing only minimal assumptions on the initial density. However, the use of the logarithmic entropy measure of Shannon leads to a grave drawback — if the only source of information used to maximize entropy is prices of multiple bonds with different maturities together with the value of a perpetual annuity, then the resulting density function is necessarily of exponential form. Actually, the exponential form exactly describes the initial term structure density only when the underlying bonds are continuously compounded. Otherwise, it only approximates the initial density when the bonds are compounded at other specific frequencies. This shows that the Shannon entropy is not a good candidate for the initial calibration.

In order to provide a general (no matter how frequently the observed bonds are compounded) description for the initial term structure density, we propose here use of the Tsallis entropy as the basis for entropy maximization. The form of the Tsallis entropy was first introduced by Havrda and Charvat in 1967 [28], parameterized by a constant α (also called the α -order entropy), in the context of quantifying classification schemes. It is initially applied to define the distance between parametric probability densities on a statistical manifold, and gives rise to the Fisher-Rao metric [43]. In the current problem of initial calibration, the maximum Tsallis entropy distribution proves to be power-distributed and to be a general attribute in the sense that it reduces to the exponential distribution as $\alpha \rightarrow 1$ [14].

Interestingly, the power-law distribution for the initial term structure density can be identified from another viewpoint — the notion of superstatistics. We initially suppose that the term structure in a short term is flat associated with a constant continuously compounded rate, and further assume that the rate follows a χ^2 -distribution. Therefore, the whole term structure could be regarded as a superposition of local flat structures, and proves to follow the same power-law distribution as the entropic method indicates if the only source of information available is the existence of a perpetual annuity.

This chapter is organized as follows. To begin with, we introduce in Section 7.1 several types of entropy and highlight the superiority of the Tsallis entropy to the others. In Section 7.2, we present an iterative algorithm, which is based on the Tsallis entropy maximization, to determine the initial density in terms of the short rate and multiple bond price data. More observations from the comparison between the calibration algorithms proposed by Brody-Hughston and us could be found in Table 7.1 in Section 7.2.4. In Section 7.3, we propose another initial calibration

approach by use of the superstatistics concept. Many interesting properties concerned with the entropy index, the mean of the short rate, and the annuity price are elaborated in Section 7.3.4.

7.1 Introduction to Tsallis Entropy

The family of Tsallis entropies of the density function $\rho(x)$ is defined by

$$S_q^{(T)}[\rho] = \frac{1}{q-1} \left(1 - \int_{-\infty}^{\infty} \rho^q(x) dx \right). \quad (7.1)$$

Here q is the *entropy index*, a physical measurement originally used in thermodynamics. In the limit $q \rightarrow 1$, the Tsallis entropy (7.1) reduces to

$$S[\rho] = - \int_{-\infty}^{\infty} \rho(x) \ln \rho(x) dx, \quad (7.2)$$

which defines the Shannon entropy. In this sense, q actually quantifies the departure from the logarithmic entropy measure.

In what follows we will first review the history of entropy in Section 7.1.1, particularly introducing the background of development of the Tsallis entropy. Then in Section 7.1.2 we will outline the general procedures for the maximization of the Shannon and the Tsallis entropies under simple constraints. Finally in Section 7.1.3 we will explain why the Tsallis entropy stands out from the others that also produce power-law distributions.

7.1.1 History of Entropy

The concept of entropy has developed in a history of more than two hundred years, originated from a measure of lost energy in physical systems to a functional of density function in statistical studies. Roughly speaking, the history of entropy can be divided into four periods:

1. 1800s — 1860s: entropy in classical thermodynamics;
2. 1870s — 1890s: entropy in statistical thermodynamics;
3. 1940s — 1950s: entropy in information theory;
4. 1960s — present: entropy in statistics and cybernetics.

The concept of entropy was originally defined in the context of classical thermodynamics, to measure the amount of energy in a thermodynamic system that cannot be used to do work. Based on the work of Lazare Carnot and his son Sadi Carnot [1], Rudolf Clausius [2] was the first in history to name the quantity of lost energy S “entropy” and presented the first-ever mathematical formulation of entropy in a paper he published in 1865 [41]. Clausius continued to develop his ideas on entropy and stated in 1862 the *second law of thermodynamics*, that in any irreversible process a small amount of heat energy ΔQ is incrementally dissipated across the system boundary. Quantitatively, let ΔQ denote the changed heat of a body in the system, T the absolute temperature of the body, and define the entropy by

$$S = \frac{Q}{T}. \quad (7.3)$$

Then the equation

$$\int \frac{\Delta Q}{T} \geq 0 \quad (7.4)$$

must hold for any cyclical process, where the equality holds for reversible cycles. We see that Clausius’s definition for entropy is given purely from the macroscopic aspect.

Nearly a half century later, an alternative definition of entropy was given in the context of statistical mechanics from the microscopic aspect. The most general

form for the statistical entropy S of a thermodynamic system is the *Boltzmann-Gibbs entropy*, which is defined by J. Willard Gibbs [34] in 1878 after earlier work of Boltzmann:

$$S = -k_B \sum_i \rho_i \ln \rho_i. \quad (7.5)$$

Here ρ_i is the probability of the microstate i taken from an equilibrium ensemble and k_B is a physical constant known as *Boltzmann's constant*. If all the microstates are equiprobable in the thermodynamic system, Eqn. (7.5) will reduce to the form [22,24] of

$$S = k_B \ln \Omega, \quad (7.6)$$

which was defined by Ludwig Boltzmann as early as in 1977. Here Ω represents the number of microstates consistent with the observed macrostate. The statistical entropy (7.6) corresponds to the equilibrium configuration of the thermodynamic system and is therefore the parallel expression of the classical entropy (7.3) at the thermodynamic equilibrium from the microscopic aspect.

An analog to thermodynamic entropy is *information entropy*. It was defined in 1948 by Claude E. Shannon [21], an electrical engineer at Bell Telephone Laboratories:

$$H(X) = -K \sum_i \rho(x_i) \log_b \rho(x_i). \quad (7.7)$$

Here $\rho(x_i)$ is the probability of identifying an outcome value x_i out of a set of possibilities X , and b is the base of the logarithm, commonly taken as 2 if the entropy is measured in bits, e for nats, and 10 for digits. K here is merely a constant corresponding to the choice of measurement units. Consider a coding scheme in which a message is coded to identify a value x_i with probability

$$\rho(x_i) = b^{-l_i}$$

over the whole phrase space X , where l_i is the length of the code for x_i in units b . Then the *Shannon entropy* (7.7) can be interpreted as the average message-length per datum that is needed to encode.

An application of the Shannon entropy (7.7) in modern statistics is to measure loss of information when the exact value of a random variable X is unknown. At this time, $\log_b \frac{1}{\rho(x_i)}$ in (7.7) is regarded as the information content (or uncertainty) of X w.r.t output x_i . To extend this idea to the continuous case, the *continuous entropy* (also called the *differential entropy*) is defined [51], just as in the form of (7.2). There $\rho(x)$ denotes a probability function w.r.t a continuous random variable X , and the logarithm is taken with the nature base unless otherwise specified (the choice of base does not affect the nature of entropy). Although the extension is carried out simply by replacing summation with integral, we should point out that the differential entropy loses some properties that the Shannon discrete entropy possesses. For example, the differential entropy (7.2) can be negative. Besides, it is not invariant under continuous coordinate transformations. However, such imperfections cannot reduce people's attachment for the differential entropy, and most often it is still called the Shannon entropy without confusion.

It is worth noting that the other scientist who makes a big contribution to information theory is Norbert Wiener, a contemporary mathematician of Shannon. Norbert Wiener invented the field of cybernetics, and articulated in his 1948 book *Cybernetics* [36] the close relation between communication and control. Although the cybernetic theory was first attached to biology, in Wiener's later work he attempted to bring together Shannon's concept of entropy as a measure of uncertainty in communication [16], and defined the amount of information as "the negative of the quantity usually defined as entropy in similar situations" [36].

Up to now, we have briefly introduced the entropy definition from three viewpoints: classical thermodynamic views, statistical thermodynamic views, and information theory views. A common ground among them is to assume entropy quantities like energy or information as *extensive* variables [3], namely, that the total energy of a system or information about an event is proportional to the system size. Such an assumption is reasonable most often, for example, when the energy exists among short-range interactions which hold matters together. However, suppose that we deal with long-range interactions, for example, gravity, and we can then find that energy is not extensive. It is therefore necessary to define another kind of entropy that can also cover the nature of non-extensive systems. The most eminent type of those desirable entropies is the Tsallis entropy (7.1), named after the physicist Tsallis.

7.1.2 Maximization of Shannon and Tsallis Entropies

We have shown that entropy, whatever definition is applied, is nothing more than the lost energy or lost information within a system. It follows that entropy is maximized when the information at hand is least. This is justified in an example where we employ the discrete Shannon entropy. Assume a gambler is trying to make money by toiling a coin and betting on the outcome. Let X be the random variable denoting the possible outputs $\{Head, Tail\}$. If the coin is fair, the gambler would be in a hard plight since he could make no preference on the outcome. The information content of the outcome is null at this time and the entropy arrives at its maximum value one if the logarithm base is taken to be 2 and $K = 1$ as we usually do in practice. On the contrary, if the coin is unfair, the gambler would bet preferentially on the most frequent result. In this case, the uncertainty is lower and correspondingly the Shannon entropy is lower. Just as G.N. Lewis pointed out in

1930, "Gain in entropy always means loss of information, and nothing new."

The most eminent principle that reflects the information nature of entropy is the *principle of maximum entropy*. The principle was first explained by E.T. Jaynes in two papers in 1957 [22, 23], where he explored a correspondence between the statistical entropy and the information entropy. It states that the probability distribution that maximizes the information entropy is the only true distribution reflecting the information prescribed. Inspired by this idea, we attempt to determine the initial term structure density of our proposed model by maximizing some entropy. To choose a distribution with a lower entropy would be to assume excessive information we do not possess; to choose a distribution with a higher entropy would violate the constraints of information we do possess. Thus the maximum entropy distribution would precisely reflect the current market information and is chosen to be as uninformative as possible.

As an illustration of the principle of maximum entropy, we will outline the general procedures for the maximization of the Shannon and the Tsallis entropies under simple constraints.

Shannon Entropy Maximization

Given a set of functions $\{f_j(X)\}$ for $j = 1, \dots, n$ of a random variable X , we are told that the expectation of $f_j(X)$ w.r.t to an unknown distribution ρ equals a_j :

$$\int_{-\infty}^{\infty} f_j(x)\rho(x) dx = a_j, \quad \text{for } j = 1, \dots, n. \quad (7.8)$$

Our aim is to determine the density $\rho(x)$ that is consistent with the information given in the form of (7.8). Additionally, we also have the normalization condition

$$\int_{-\infty}^{\infty} \rho(x) dx = 1. \quad (7.9)$$

Subject to these constraints we intend to maximize the Shannon entropy (7.2). This is accomplished straightforward by introducing Lagrange multipliers $\{\lambda_j\}$ and ν , and solving the variational equation

$$\frac{\delta}{\delta \rho} \left[- \int_{-\infty}^{\infty} \rho(x) \ln \rho(x) dx - \sum_j \lambda_j \left(\int_{-\infty}^{\infty} f_j(x) \rho(x) dx - a_j \right) - \nu \left(\int_{-\infty}^{\infty} \rho(x) dx - 1 \right) \right] = 0. \quad (7.10)$$

The solution gives rise to an exponential density function

$$\begin{aligned} \rho(x) &= \frac{1}{\exp(\nu + 1)} \exp \left(- \sum_j \lambda_j f_j(x) \right) \\ &\triangleq \frac{1}{Z(\lambda_1, \dots, \lambda_n)} \exp \left(- \sum_j \lambda_j f_j(x) \right), \end{aligned} \quad (7.11)$$

where Z is known as the partition function and acts as the normalization factor:

$$Z(\lambda_1, \dots, \lambda_n) = \int_{-\infty}^{\infty} \exp \left(- \sum_j \lambda_j f_j(x) \right) dx, \quad (7.12)$$

and $\{\lambda_j\}$ are determined implicitly by

$$-\frac{\partial \ln Z}{\partial \lambda_j} = a_j. \quad (7.13)$$

All the analysis presented above are summarized in the following theorem by Boltzmann.

Theorem 7.1. *Suppose U is a closed subset of the real numbers R and we are given n measurable functions f_1, \dots, f_n with n numbers a_1, \dots, a_n . We consider the class C of all continuous random variables which are supported on U and satisfy the n expected value conditions*

$$E[f_j(X)] = a_j, \quad \text{for } j = 1, \dots, n.$$

If there is a member in \mathcal{C} whose density function is positive everywhere in U , and if there exists a maximum entropy distribution for \mathcal{C} , then its probability density function $\rho(x)$ is given in the form of (7.11), where the parameters Z and λ_j are determined by, respectively, (7.12) and (7.13). Conversely, if constants λ_j and Z like these can be found, then $\rho(x)$ is indeed the density of the unique maximum entropy distribution for our class \mathcal{C} .

We observe that the resulting maximum Shannon entropy distribution is of exponential form.

Tsallis Entropy Maximization

Now we show how a power-law distribution is deduced by maximizing the Tsallis entropy. Note that during the procedure of the maximization of the Shannon entropy, the available information, as shown in (7.8), is expressed as expectation constraints w.r.t the density function $\rho(x)$. Under the framework of the Tsallis entropy, however, the single expectation constraint is expressed as the “generalized mean” value

$$\int_{-\infty}^{\infty} f(x)\rho_q(x) dx = a_q, \quad (7.14)$$

w.r.t a new “escort” distribution ρ_q [14, 45] defined by

$$\rho_q(x) = \frac{\rho^q(x)}{\int_{-\infty}^{\infty} \rho^q(x) dx}. \quad (7.15)$$

In conjunction with the normalization condition (7.9) on the original distribution $\rho(x)$, our aim is to find the density $\rho(x)$ that maximizes the Tsallis entropy (7.1).

Introducing Lagrange multipliers α and β , we obtain the variational equation

$$\begin{aligned} \frac{\delta}{\delta\rho} \left[\frac{1}{q-1} \left(1 - \int_{-\infty}^{\infty} \rho^q(x) dx \right) - \alpha \left(\int_{-\infty}^{\infty} \rho(x) dx - 1 \right) \right. \\ \left. - \beta \left(\int_{-\infty}^{\infty} f(x)\rho_q(x) dx - a_q \right) \right] = 0. \end{aligned} \quad (7.16)$$

The solution of Eqn. (7.16) is

$$\begin{aligned}\rho(x) &= \left(\frac{\alpha(1-q)}{q}\right)^{\frac{1}{q-1}} [1 - \beta(q-1)f(x)]^{\frac{1}{q-1}} \\ &\triangleq \frac{1}{Z_q(\beta)} [1 - \beta(q-1)f(x)]^{\frac{1}{q-1}},\end{aligned}\quad (7.17)$$

where the partition function $Z_q(\beta)$ is determined by the normalization condition as

$$Z_q(\beta) = \int_{-\infty}^{\infty} [1 - \beta(q-1)f(x)]^{\frac{1}{q-1}} dx, \quad (7.18)$$

and the Lagrange multiplier β is implicitly determined by the expectation constraint (7.14). It is immediately verified that in the limit $q \rightarrow 1$ the power-law distribution (7.17) reduces to an exponential distribution

$$\rho(x) = \frac{1}{Z_q} \exp(-\beta f(x)) \quad (7.19)$$

with

$$Z_q = \int_{-\infty}^{\infty} \exp(-\beta f(x)) dx.$$

Obviously, the limit case is exactly the maximum Shannon entropy distribution.

The use of the escort density ρ_q here seems peculiar since the information we obtain from real markets, such as the prices of bonds or perpetual annuity, are mostly expressed in the form of traditional expectations w.r.t the original density ρ . Therefore, we introduce in the below an alternative manner for the Tsallis entropy maximization subject to the traditional expectation constraint

$$\int_{-\infty}^{\infty} f(x)\rho(x) dx = a. \quad (7.20)$$

Here we have used the original density ρ to replace the escort density ρ_q in (7.14). Then we follow an analogous procedure as in the Shannon case by introducing the

Lagrange multipliers α , β , and defining the functional

$$\begin{aligned} \phi(\rho) = & \frac{1}{q-1} \left(1 - \int_{-\infty}^{\infty} \rho^q(x) dx \right) - \alpha \left(\int_{-\infty}^{\infty} \rho(x) dx - 1 \right) \\ & - \alpha\beta(1-q) \left(\int_{-\infty}^{\infty} f(x)\rho(x) dx - a \right), \end{aligned} \quad (7.21)$$

which is written this way for further calculation convenience. Solving the variational equation $\frac{\partial\phi}{\partial\rho} = 0$, we immediately obtain the density function in the same form as (7.17).

Since the distribution (7.17) is parameterized by the entropy index q , we often call it q -distribution. It is important to emphasize that different choices of q determine different nature of the distribution. If $f(x) = x$ in (7.17), then as $q \rightarrow \infty$, the q -distribution appears as a "window" function which is a constant over a finite support; as $q \rightarrow 1$, the q -distribution reduces to the ordinary exponential distribution; whereas for $0 < q < 1$, the distribution has power-law tails and β is used to control the width of the distribution.

It is noticeable that some recent researchers [11,12] prefer to write the q -distribution in the form of

$$\rho(x) = \frac{C_{\bar{q}}}{[1 + \beta(\bar{q} - 1)f(x)]^{\frac{1}{\bar{q}-1}}} \quad (7.22)$$

where \bar{q} is defined by

$$\bar{q} - 1 = -(q - 1). \quad (7.23)$$

Thus for $1 < \bar{q} < 2$, the \bar{q} -distribution (7.22) with $f(x) = x$ follows the power-law distribution. In the text afterwards we tend to adopt the popular form (7.22) instead of (7.17) unless specified otherwise.

Before proceeding further, let us introduce other designations and major characteristics of the maximum Tsallis entropy distribution (7.17) or (7.22).

First, the resulting power-law distribution is sometimes called the *q-exponential* distribution since it can be rewritten in an analogous form as the ordinary exponential form:

$$\rho(x) = \frac{1}{Z_q(\beta)} \exp_q(-\beta f(x)). \quad (7.24)$$

The only difference lies in the definition of the exponential function, which is now replaced by the *q-exponential* function

$$\exp_q(t) = [1 + (q - 1)t]^{\frac{1}{q-1}}, \quad (7.25)$$

which reduces to the ordinary exponential form as $q \rightarrow 1$. Accordingly, the *q-exponential* distribution (7.24) reduces to the ordinary exponential distribution as $q \rightarrow 1$, just like the Tsallis entropy includes the Shannon entropy as a special case.

Second, the *q-distribution* (7.17) with $f(x) = x^2$ is also called the *q-Gaussian distribution*. The motivation is illustrated as follows. It is known that the sum of a large number of small displacements in a Brownian motion is Gaussian distributed with a variance that grows with time. This follows as a consequence of the central limit theory and is essentially valid when the elementary displacements are sufficiently decor related. However, the Gaussian is not the only limit distribution for the sum of random variables, especially in some complex systems in which the correlations between variables are too strong to be omitted. Actually the *q-Gaussian* distribution is most often applies for the study of strongly correlated systems.

Finally, the maximum Tsallis entropy distribution and the entropy itself reflect the information prescribed in a system from, respectively, the microscopic and macroscopic aspects. In Section 7.1.1, we have elucidated that the Tsallis entropy is used as a measure of the nonlinearly growing or dissipating energy in non-extensive systems from the macroscopic aspect. Meanwhile, in the level of microstates, the maximum Tsallis entropy distribution (7.17) is applied to characterize the distribu-

tion of the sum of a number of strongly correlated variables in the non-extensive systems. In brief, the q -distribution (7.17) is the parallel expression of the Tsallis entropy in the microscopic world.

7.1.3 Why Tsallis Entropy

Treating the Tsallis entropy as a functional of the term structure density and expressing the current market data as constraints on the density function, we can calibrate the initial term structure by maximizing the Tsallis entropy. But why the Tsallis entropy, and not else?

In the preceding section we have explained why we prefer the Tsallis entropy to the Shannon entropy as the basis for our initial calibration. Depending on the value of entropy index q , the maximum Tsallis entropy distribution has quite a large fitting spectrum and specifically reduces to the maximum Shannon entropy distribution as $q \rightarrow 1$. Therefore, the Tsallis entropy allows for a richer structure for the initial term structure density.

However, the Tsallis entropy is not the only one that produces the q -distribution. There exist three well-known different entropies, listed as follows, that are maximized by the q -distribution (7.17) under the single q -expectation constraint (7.14).

1. The Rényi entropy [4, 5]:

$$S_q^{(R)}[\rho] = \frac{-1}{q-1} \ln \int_{-\infty}^{\infty} \rho^q(x) dx. \quad (7.26)$$

2. The Tsallis entropy [14]:

$$S_q^{(T)}[\rho] = \frac{1}{q-1} \left(1 - \int_{-\infty}^{\infty} \rho^q(x) dx \right). \quad (7.27)$$

3. The normalized Tsallis entropy [7, 40]:

$$S_q^{(NT)}[\rho] = \frac{1}{q-1} \left(\frac{1}{\int_{-\infty}^{\infty} \rho^q(x) dx} - 1 \right). \quad (7.28)$$

The Rényi entropy is conventionally used for the definition of the generalized dimension in multifractals [13], and the Tsallis entropy plays a central role in non-extensive statistical mechanics [46]. Obviously, the Tsallis entropy can be regarded as the truncated form of the Rényi entropy or the normalized Tsallis entropy by eliminating the higher order terms in the series expansions of the latter two. All of them converge to the Shannon entropy as the entropy index q tends to one.

Following the standard procedure of entropy maximization as presented in Section 7.1.2, the Rényi, Tsallis, and normalized Tsallis entropies all lead to the q -distribution of the same type. This means that mere fittings of observed market data into the q -distribution tells nothing about the kind of entropy we use. Thus a question arises naturally: would it make no difference to use whatever kind of entropy? The answer is no.

When we attempt to determine a density function by use of the entropic method, we should pick a “stable” entropy to maximize. Taking our proposed model as an example, we can elaborate the reason. An entropy $S[\rho]$ is said to be stable if the amount of its change under an arbitrary small deformation of the distribution remains small. In our model, let ρ denote the real distribution for the initial term structure and $\bar{\rho}$ denote an approximation to the real one. We measure the size of deformation from $\bar{\rho}$ to ρ by the l_1 norm:

$$\|\rho - \bar{\rho}\|_1 = \int_{-\infty}^{\infty} |\rho(x) - \bar{\rho}(x)| dx.$$

Suppose that the associated entropic basis in our initial calibration is unstable, namely, the functional $S[\rho]$ has to possess the following property:

$$\exists \varepsilon > 0, \forall \delta > 0, \text{ s.t. } \|\rho - \bar{\rho}\|_1 \leq \delta \Rightarrow \left| \frac{S[\rho] - S[\bar{\rho}]}{S_{max}} \right| > \varepsilon, \quad (7.29)$$

where S_{max} denotes the maximal value of the entropy. Since entropy measures the information at hand, property (7.29) indicates that the information used to obtain ρ

and \bar{p} differs quite a lot from each other. However, the market information (assume that all the information in markets is free for investors) are indifferent to all the investors. Thus a contradiction occurs.

S. Abe [44] has shown that the Rényi entropy and the normalized Tsallis entropy are unstable and therefore cannot reproduce experimentally observable quantities, whereas the Tsallis entropy is stable and can provide an appropriate entropic basis for the q -distribution.

To sum up, we have proved that the Tsallis entropy produces a richer term structure (compared with the Shannon entropy) and provides a stable entropic basis (compared with the Rényi entropy and the normalized Tsallis entropy) for our initial calibration. In the next section, we will introduce an initial calibration approach based on the Tsallis entropy maximization .

7.2 An Initial Calibration Approach Based on Tsallis Entropy Maximization

In this section, we will first show in Section 7.2.1 how the Tsallis entropy is applied in our proposed model to determine the initial term structure density subject to the prices of multiple bonds and a perpetual annuity. Next, in Section 7.2.2 we will explore the characteristics of the resulting distribution and particularly interpret the entropy index from the financial viewpoint. Following it in Section 7.2.3 we will design an iterative algorithm for determining the initial term structure in terms of the short rate and the specified bond prices. Finally, in Section 7.2.4 we compare our entropic calibration approach with Brody-Hughston's and present our discussion results in Table 7.1.

7.2.1 Maximization of Tsallis Entropy With Market Data

To begin with, let us see how the principle of maximum entropy works with the proposed model when we are given a set of data points on the initial yield curve together with the value of a perpetual annuity. The problem is to calibrate the initial term structure with the given data.

Recall that the term structure density $\rho(x)$ in our proposed model is deduced by the introduction of the tenor variable x (the time left to maturity). Assume that a set of bond prices $\{P_{0T_k}\}$ with different tenors $\{T_k\}$ ($k = 1, 2, \dots, n, T_k > 0$) are observed from real markets. As a consequence of the relation (4.16) between bond values and the term structure density function, P_{0T_k} can be expressed as the expectation of step function I_{T_k} w.r.t $\rho(x)$:

$$\int_0^{\infty} \rho(x) I_{T_k}(x) dx = P_{0T_k}, \quad \text{for } k = 1, 2, \dots, n, \quad (7.30)$$

where the step function is defined by

$$I_{T_k}(x) = \begin{cases} 1 & \text{for } x \geq T_k \\ 0 & \text{for } x < T_k \end{cases}. \quad (7.31)$$

In addition, we are given the initial value ζ of the perpetual annuity, which can be expressed as the mean of $\rho(x)$:

$$\int_0^{\infty} \rho(x)x dx = \zeta. \quad (7.32)$$

Without loss of generality, we will use the expectation constraint on an abstract function $g(X)$ instead, which is given by

$$\int_0^{\infty} \rho(x)g(x) dx = U. \quad (7.33)$$

Obviously, constraint (7.33) reduces to (7.32) when $g(x) = x$.

Subject to the constraints (7.30) and (7.33), together with the normalization condition (3.5) $\int_0^{\infty} \rho(x) dx = 1$, we then determine the density $\rho(x)$ that maximizes

the Tsallis entropy (7.1). Following the similar procedure as in Section 7.1.2, we introduce Lagrange multipliers α , λ , and $\{\beta_k\}$, and define the functional

$$\begin{aligned} \phi(\rho) &= \frac{1}{q-1} \left(1 - \int_0^\infty \rho^q(x) dx \right) - \alpha \left(\int_0^\infty \rho(x) dx - 1 \right) \\ &- \alpha\lambda(1-q) \left(\int_0^\infty \rho(x)g(x) dx - U \right) \\ &- \alpha \sum_{k=1}^n \beta_k(1-q) \left(\int_0^\infty \rho(x)I_{T_k}(x) dx - P_{0x_k} \right). \end{aligned} \tag{7.34}$$

Solve the variational equation $\frac{\partial \phi}{\partial \rho} = 0$ and we obtain the density function in the form of

$$\rho(x) = C_q \left[1 - \lambda(q-1)g(x) - \sum_{k=1}^n \beta_k(q-1)I_{T_k}(x) \right]^{\frac{1}{q-1}}, \tag{7.35}$$

or equivalently,

$$\rho(x) = \frac{C_{\bar{q}}}{[1 + \lambda(\bar{q}-1)g(x) + \sum_{k=1}^n \beta_k(\bar{q}-1)I_{T_k}(x)]^{\frac{1}{\bar{q}-1}}}, \tag{7.36}$$

if we define \bar{q} by (7.23) as most researchers prefer. Here the normalization factor $C_{\bar{q}}$ is determined by

$$C_{\bar{q}} = \int_0^\infty \left[1 + \lambda(\bar{q}-1)g(x) + \sum_{k=1}^n \beta_k(\bar{q}-1)I_{T_k}(x) \right]^{\frac{1}{\bar{q}-1}} dx. \tag{7.37}$$

The Lagrange multipliers λ and $\{\beta_k\}$ are implicitly determined by, respectively, the annuity constraint (7.33) and the bond price constraint (7.30).

7.2.2 Interpretation of Entropy Index in the Proposed Model

As a consequence of (7.36), we see that the pointwise calibration to the discount bond prices along with the information of the expectation of function $g(X)$ gives a piecewise power-law distribution. Let us consider a special case in which the entropy index \bar{q} tends to one. Comparing the limiting distribution with the one given by

Brody and Hughston [19] based on the maximization of the Shannon entropy, we can find more interpretations of the Lagrange multipliers λ , $\{\beta_k\}$, and the entropy index \bar{q} from the financial viewpoint.

Let $\bar{q} \rightarrow 1$. Then we have

$$\rho(x) \rightarrow C_{\bar{q}} e^{-\sum_{k=1}^n \beta_k I_{T_k}(x)} e^{-\lambda g(x)}, \quad \text{as } \bar{q} \rightarrow 1. \quad (7.38)$$

Specifically, when $0 \leq x < T_1$,

$$\rho(x) \rightarrow C_{\bar{q}} e^{-\lambda g(x)}.$$

When $T_k \leq x < T_{k+1}$ ($k = 1, \dots, n$),

$$\rho(x) \rightarrow C_{\bar{q}} e^{-(\beta_1 + \dots + \beta_k)} e^{-\lambda g(x)} \triangleq C_k e^{-\lambda g(x)}.$$

where $C_k = C_{\bar{q}} e^{-\sum_{j=1}^k \beta_j}$. Thus we can write the limiting density in a compact form as

$$\rho(x) \rightarrow \sum_{k=0}^n I_{T_k T_{k+1}}(x) C_k e^{-\lambda g(x)}, \quad \text{as } \bar{q} \rightarrow 1, \quad (7.39)$$

where $T_0 = 0$, $T_{n+1} = \infty$, $I_{T_k T_{k+1}} = 1$ if $x \in [T_k, T_{k+1})$ and vanishes otherwise. And

$$C_k = \begin{cases} C_{\bar{q}} & \text{for } 0 \leq x < T_1 \\ C_{\bar{q}} e^{-\sum_{j=1}^k \beta_j} & \text{for } T_k \leq x < T_{k+1}, \quad k \neq 0 \end{cases} \quad (7.40)$$

Given the annuity price, i.e. $g(x) = x$ in (7.39), we observe that the limiting initial distribution gives a piecewise exponential form, just as Brody and Hughston have obtained in [19]. In what follows we will find further interpretations of the involved parameters such as C_k , λ , and \bar{q} .

First, we consider the simplest case in which we are given only the value ζ of the perpetual annuity. Then the limiting distribution (7.39) becomes

$$\rho(x) \rightarrow \lambda e^{-\lambda x}, \quad \text{as } \bar{q} \rightarrow 1, \quad (7.41)$$

where $\lambda = \frac{1}{\zeta}$ and accordingly $P_{0x} = e^{-\lambda x}$ for the discount function. It follows that when the entropy index $\bar{q} \rightarrow 1$, the maximization of the Tsallis entropy under the annuity constraint gives rise to a flat term structure with a constant continuously compounded rate λ .

Furthermore, we consider in more detail the case where the observed data consists of two pieces of information — the bond price P_{0T_1} for a fixed maturity T_1 and the value ζ of the perpetual annuity. Then (7.39) turns to be a two-piece function

$$\rho(x) \rightarrow \begin{cases} C_0 e^{-\lambda x} & \text{for } 0 \leq x < T_1 \\ C_1 e^{-\lambda x} & \text{for } T_1 \leq x < \infty \end{cases} \quad \text{as } \bar{q} \rightarrow 1, \quad (7.42)$$

which gives a continuously compounded flat term structure on each interval. Obviously, the parameter λ plays a role as the long-term rate that dominates the trend of the term structure, whereas the parameters $\{C_k\}$ ($k = 0, 1$) reflect the piecewise information indicated by the bond prices for different maturities. Specifically, set $x = 0$ and we obtain $\rho(0) = C_0$, which indicates that C_0 is actually the initial short rate. All these observations are also available in the general case (7.39), where more bond price data are taken up.

Summing up matters so far, we observe that in the limit $\bar{q} \rightarrow 1$ the maximum Tsallis entropy distribution tends to be a piecewise flat term structure density on a continuously compounded basis, and thus recovers the maximum Shannon entropy distribution as a special case. In this sense, the entropy index \bar{q} actually measures the departure of the current term structure from the piecewise flat term structure.

Now let us turn back to the original maximum Tsallis entropy distribution (7.36). As discussed in Section 7.1.2, it obeys the power-law distribution only for $1 < \bar{q} < 2$. The larger diversity between \bar{q} and 1, the larger distance from the flat term structure. Following this line of argument, we naturally impose an assumption on the value of \bar{q} such that $\bar{q} = 1 + \frac{1}{N}$, where $N > 0$ and more precisely $N > 1$ to guarantee $\bar{q} < 2$.

Then the calibrated initial term structure density (7.36), subject to the bond price constraint (7.30) and the annuity constraint (7.32), is given by

$$\rho(x) = \sum_{k=0}^n I_{T_k T_{k+1}}(x) \frac{C}{\left[1 + \frac{r_k}{N} + \frac{R}{N}x\right]^N}, \quad (7.43)$$

where $T_0 = 0$, $T_{n+1} = \infty$, $r_0 = 0$, $I_{T_k T_{k+1}} = 1$ if $x \in [T_k, T_{k+1})$ and vanishes otherwise. For the involved parameters we illustrate their physical meanings as follows.

1. Compared with (7.36), here we write $C \triangleq C_{\bar{q}}$ to simplify the notation. It acts as the normalization factor in the density function. Moreover, we set $x = 0$ in (7.43) and obtain $\rho(0) = C$, indicating that C is actually the initial short rate.
2. Compared with (7.36), here we write $r_k \triangleq \sum_{j=1}^k \beta_j$ ($k \neq 0$) to simplify the notation and more importantly reflect the piecewise information indicated by the bond prices for different maturities.
3. In (7.36), the Lagrange multiplier λ is introduced in accordance with the annuity constraint. In (7.43), however, we use the notation R to replace λ , which makes its roles as the long-term rate more prominent.
4. Most importantly, the power-law exponent N in (7.43) accounts for the compounding frequency of the bonds if all the bonds are compounded at the same frequency.

7.2.3 Initial Calibration Algorithm

In (7.43) we can determine the values of $\{r_k\}$ and R by use of bond prices for different maturities $\{T_k\}$ and the initial price ζ of the perpetual annuity. Then the initial

short rate C is inferred since it functions as the normalization factor. Alternatively, we can regard the short rate C and the bond prices as the actual independent data, and then infer the annuity price ζ . This idea leads us to an iterative algorithm for determining the term structure in terms of the short rate and the specified bond price data.

There are $n + 2$ parameters in (7.43): C , R and $\{r_k\}$ for $k = 1, \dots, n$. Corresponding to them are the $n + 2$ constraints, involving n bond price constraints (7.30), the annuity constraint (7.32), and the normalization condition (3.5). Alternatively, the constraint (7.30) for bond prices can be rewritten as

$$\int_{T_k}^{T_{k+1}} \rho(x) dx = P_{0T_k} - P_{0T_{k+1}}, \quad \text{for } k = 0, 1, \dots, n. \quad (7.44)$$

Note that the normalization condition is incorporated into the bond price constraints by adding up the integrals (7.44) over k .

Substitute the power-law density function (7.43) into the bond price constraint (7.44).

In order to guarantee the integrability of density function over the infinite interval, we assume $N > 1$. A short calculation gives rise to

$$\frac{CN}{R(N-1)} \left[\frac{1}{\left(1 + \frac{r_k}{N} + \frac{R}{N}T_k\right)^{N-1}} - \frac{1}{\left(1 + \frac{r_k}{N} + \frac{R}{N}T_{k+1}\right)^{N-1}} \right] = P_{0T_k} - P_{0T_{k+1}}, \quad (7.45)$$

for $k = 0, 1, \dots, n$. Furthermore, substitution of (7.43) in the integral $\int_{T_k}^{T_{k+1}} \rho(x)x dx$, if we assume $N > 2$ to guarantee the integrability, results in

$$\begin{aligned} & \int_{T_k}^{T_{k+1}} \rho(x)x dx \\ &= \frac{CN^2}{R^2(N-1)(N-2)} \left[\frac{1 + \frac{r_k}{N} + (N-1)\frac{R}{N}T_k}{\left(1 + \frac{r_k}{N} + \frac{R}{N}T_k\right)^{N-1}} - \frac{1 + \frac{r_k}{N} + (N-1)\frac{R}{N}T_{k+1}}{\left(1 + \frac{r_k}{N} + \frac{R}{N}T_{k+1}\right)^{N-1}} \right], \end{aligned} \quad (7.46)$$

by which the annuity price is given by

$$\frac{CN^2}{R^2(N-1)(N-2)} \sum_{k=0}^n \left[\frac{1 + \frac{r_k}{N} + (N-1)\frac{R}{N}T_k}{\left(1 + \frac{r_k}{N} + \frac{R}{N}T_k\right)^{N-1}} - \frac{1 + \frac{r_k}{N} + (N-1)\frac{R}{N}T_{k+1}}{\left(1 + \frac{r_k}{N} + \frac{R}{N}T_{k+1}\right)^{N-1}} \right] = \zeta. \quad (7.47)$$

Obviously, given the prices of multiple bonds $\{P_{0T_k}\}$ ($k = 0, 1, \dots, n$) and the price of the perpetual annuity ζ , we can determine the values of C , R , and $\{r_k\}$ by solving the equations (7.45) and (7.47). The calculation is, however, quite complicated. In our experiment, we can alternatively regard the initial short rate C and the bond prices $\{P_{0T_k}\}$ as the actual information at our disposal. Then as a particular case of (7.45), for $k = 0$ we obtain

$$\frac{CN}{R(N-1)} \left[1 - \frac{1}{\left(1 + \frac{R}{N}T_1\right)^{N-1}} \right] = 1 - P_{0T_1}, \quad (7.48)$$

which can be used to solve for R in terms of the initial short rate C and the bond price P_{0T_1} . Then, by substitution of R and further bond price data into (7.45) for general k , we can iteratively obtain $\{r_k\}$ ($k = 1, \dots, n-1$). Specifically, for $k = n$, Eqn. (7.45) becomes

$$\frac{CN}{R(N-1)} \frac{1}{\left(1 + \frac{r_n}{N} + \frac{R}{N}T_n\right)^{N-1}} = P_{0T_n}, \quad (7.49)$$

which determines r_n in terms of C , R , and P_{0T_n} . In this procedure, we just assume the existence of a perpetual annuity and could infer its implied value by the substitution of the initial short rate and values of r_k and R into formula (7.47). Finally, the discount function can be determined by use of the fact that

$$\begin{aligned} 1 - P_{0x} &= \int_0^x \rho(u) du \\ &= \int_0^{T_k} \rho(u) du + \int_{T_k}^x \rho(u) du \\ &= 1 - P_{0T_k} + \frac{CN}{R(N-1)} \left[\frac{1}{\left(1 + \frac{r_k}{N} + \frac{R}{N}T_k\right)^{N-1}} - \frac{1}{\left(1 + \frac{r_k}{N} + \frac{R}{N}x\right)^{N-1}} \right], \end{aligned} \quad (7.50)$$

when $x \in [T_k, T_{k+1})$. Thus the bond price P_{0x} for $x \in [T_k, T_{k+1})$ is given by

$$P_{0x} = P_{0T_k} - \frac{CN}{R(N-1)} \left[\frac{1}{\left(1 + \frac{r_k}{N} + \frac{R}{N}T_k\right)^{N-1}} - \frac{1}{\left(1 + \frac{r_k}{N} + \frac{R}{N}x\right)^{N-1}} \right]. \quad (7.51)$$

The results presented so far lead to an iterative algorithm for determining the initial term structure in terms of the initial short rate and the specified bond data points.

Proposition 7.2. *Given a set of bond prices $\{P_{0T_k}\}$ ($k = 1, 2, \dots, n$) and the existence of the value of the perpetual annuity, the maximum Tsallis entropy term structure density function is given by (7.43), where $T_0 = 0$, $T_{n+1} = \infty$, $r_0 = 0$, $I_{T_k T_{k+1}} = 1$ if $x \in [T_k, T_{k+1})$ and vanishes otherwise. Here $N > 1$ is the common compounding frequency of the observed bonds and defined via the entropy index as:*

$$N = \frac{1}{\bar{q} - 1}. \tag{7.52}$$

C is the initial short rate. The value of R is determined by equation (7.48). $\{r_k\}$ ($k = 1, 2, \dots, n$) are iteratively determined by equation (7.45). The corresponding discount function P_{0x} is given by (7.51) for $x \in [T_k, T_{k+1})$.

We see that the calibrated initial term structure is piecewise power-law distributed if the observed bonds are compounded more than one time each year over their lives. Interestingly, when we are given the prices of bonds that are compounded more frequently, the calibrated term structure will become more flat until it reduces to the piecewise exponential distribution attributed to Brody and Hughston [19].

Clearly, if there is further information at our disposal, then that can also be included in the system of constraints to maximize the Tsallis entropy. For example, we will consider a case where we are given a set of bond price data as well as the second moment of the term structure density, i.e. $g(x) = x^2$ in (7.33). For convenience of calculation, we specify the value of \bar{q} , for instance, $\bar{q} = \frac{3}{2}$. In other words, the observed bonds are all compounded semiannually.

Proposition 7.3. *Given the prices of a set of semiannually compounded bonds $\{P_{0T_k}\}$ ($k = 1, 2, \dots, n$) and the existence of the second moment of the term structure*

density, the maximum Tsallis entropy term structure density function is given by

$$\rho(x) = \sum_{k=0}^n I_{T_k T_{k+1}}(x) \frac{C}{[\alpha_k + \beta x^2]^2}, \quad (7.53)$$

where $T_0 = 0$, $T_{n+1} = \infty$, $\alpha_0 = 1$, $I_{T_k T_{k+1}} = 1$ if $x \in [T_k, T_{k+1})$ and vanishes otherwise. Here C is the initial short rate. The value of β is determined by

$$\frac{C}{2} \left[\frac{1}{\sqrt{\beta}} \arctan(\sqrt{\beta} T_1) + \frac{T_1}{1 + \beta T_1^2} \right] = 1 - P_{0T_1}. \quad (7.54)$$

$\{\alpha_k\}$ ($k = 1, 2, \dots, n$) are determined iteratively by

$$\begin{aligned} P_{0T_k} - P_{0T_{k+1}} &= \frac{C}{2\alpha_k} \left[\frac{1}{\sqrt{\alpha_k \beta}} \left(\arctan\left(\sqrt{\frac{\beta}{\alpha_k}} T_{k+1}\right) - \arctan\left(\sqrt{\frac{\beta}{\alpha_k}} T_k\right) \right) \right. \\ &\quad \left. + \left(\frac{T_{k+1}}{\alpha_k + \beta T_{k+1}^2} - \frac{T_k}{\alpha_k + \beta T_k^2} \right) \right]. \end{aligned} \quad (7.55)$$

The corresponding discount bond P_{0x} is valued by

$$\begin{aligned} P_{0x} &= P_{0T_k} - \frac{C}{2\alpha_k} \left[\frac{1}{\sqrt{\alpha_k \beta}} \left(\arctan\left(\sqrt{\frac{\beta}{\alpha_k}} x\right) - \arctan\left(\sqrt{\frac{\beta}{\alpha_k}} T_k\right) \right) \right. \\ &\quad \left. + \left(\frac{x}{\alpha_k + \beta x^2} - \frac{T_k}{\alpha_k + \beta T_k^2} \right) \right], \end{aligned} \quad (7.56)$$

for $x \in [T_k, T_{k+1})$.

Proof. We obtain the density function (7.53) directly by substituting $\bar{q} = \frac{3}{2}$ and $g(x) = x^2$ into (7.36).

Insert the piecewise density function (7.53) into the bond price constraints (7.44). A simple calculation leads to (7.55) for $k = 0, 1, \dots, n$. In particular, for $k = 0$, we obtain (7.54), which can be used to solve for β in terms of the initial short rate C and the bond price P_{0T_1} . Then by use of (7.55) for general k , we can iteratively solve $\{\alpha_k\}$ ($k = 1, 2, \dots, n-1$) in terms of β and further bond prices. Specifically, for $k = n$, Eqn. (7.55) becomes

$$\frac{C}{2\alpha_n} \left[\frac{1}{\sqrt{\alpha_n \beta}} \left(\frac{\pi}{2} - \arctan\left(\sqrt{\frac{\beta}{\alpha_n}} T_n\right) \right) - \frac{T_n}{\alpha_n + \beta T_n^2} \right] = P_{0T_n}, \quad (7.57)$$

which determines α_n in terms of C , β , and P_{0T_n} .

Substitution of (7.53) in the integral $\int_{T_k}^{T_{k+1}} \rho(x)x^2 dx$ results in

$$\int_{T_k}^{T_{k+1}} \rho(x)x^2 dx = \frac{C}{2\beta} \left[\frac{1}{\sqrt{\alpha_k\beta}} \left(\arctan\left(\sqrt{\frac{\beta}{\alpha_k}}T_{k+1}\right) - \arctan\left(\sqrt{\frac{\beta}{\alpha_k}}T_k\right) \right) - \left(\frac{T_{k+1}}{\alpha_k + \beta T_{k+1}^2} - \frac{T_k}{\alpha_k + \beta T_k^2} \right) \right], \quad (7.58)$$

by which the implied value σ^2 of the second moment of $\rho(x)$ is deduced in terms of the initial short rate C and the bond price data $\{P_{0T_k}\}$:

$$\sigma^2 = \frac{C}{2\beta} \sum_{k=0}^n \left[\frac{1}{\sqrt{\alpha_k\beta}} \left(\arctan\left(\sqrt{\frac{\beta}{\alpha_k}}T_{k+1}\right) - \arctan\left(\sqrt{\frac{\beta}{\alpha_k}}T_k\right) \right) - \left(\frac{T_{k+1}}{\alpha_k + \beta T_{k+1}^2} - \frac{T_k}{\alpha_k + \beta T_k^2} \right) \right]. \quad (7.59)$$

Finally, the discount function can be determined by use of the fact that

$$\begin{aligned} 1 - P_{0x} &= \int_0^x \rho(u) du \\ &= \int_0^{T_k} \rho(u) du + \int_{T_k}^x \rho(u) du \\ &= 1 - P_{0T_k} + \frac{C}{2\alpha_k} \left[\frac{1}{\sqrt{\alpha_k\beta}} \left(\arctan\left(\sqrt{\frac{\beta}{\alpha_k}}x\right) - \arctan\left(\sqrt{\frac{\beta}{\alpha_k}}T_k\right) \right) + \left(\frac{x}{\alpha_k + \beta x^2} - \frac{T_k}{\alpha_k + \beta T_k^2} \right) \right], \end{aligned} \quad (7.60)$$

when $x \in [T_k, T_{k+1})$. □

7.2.4 Comparison With the Brody-Hughston Calibration Approach

In order to find out the improvements of our initial calibration method over Brody and Hughston's [19], we compare the two methods in Table 7.1. The results of the comparison are discussed from five perspectives: 1) the entropy form; 2) the maximum entropy distribution (MED) subject to a single expectation constraint on

	The Tsallis Entropy	The Shannon Entropy	Connection
Entropy Form $S[\rho]$	$\frac{1}{q-1} \left(1 - \int_{-\infty}^{\infty} \rho^q(x) dx \right)$	$\int_{-\infty}^{\infty} \rho(x) \ln \rho(x) dx$	$S^{(T)} \rightarrow S^{(S)}$ as $q \rightarrow 1$
MED $\rho(x)$ With $E[g(X)] = U$	$\frac{C_q}{[1 + \lambda(\bar{q}-1)g(x)]^{\frac{1}{\bar{q}-1}}}$	$C \exp(-\lambda g(x))$	$\rho^{(T)} \rightarrow \rho^{(S)}$ as $\bar{q} \rightarrow 1$
Initial Calibration $\rho(x)$	$\sum_{k=0}^n I_{T_k}^{T_{k+1}}(x) \frac{C}{[1 + \frac{r_k}{N} + \frac{R}{N}x]^N}$	$\sum_{k=0}^n I_{T_k}^{T_{k+1}}(x) r_k e^{-Rx}$	$\rho^{(T)} \rightarrow \rho^{(S)}$ as $N \rightarrow \infty$
Parameters	$C^{(T)}$: initial short rate $R^{(T)}$: long-term rate	$r_0^{(S)}$: initial short rate $R^{(S)}$: long-term rate	$C^{(T)} \sim r_0^{(S)}$ $C^{(T)} e^{-r_k^{(T)}} \sim r_k^{(S)}$ $R^{(T)} \sim R^{(S)}$ as $N \rightarrow \infty$

Table 7.1: The Tsallis entropy VS The Shannon entropy

$g(X)$; 3) the MED calibrated against multiple bond price data along with the value of a perpetual annuity; 4) the interpretations of the involved parameters; and 5) the connection between each pair of results associated with the two methods. In order to avoid overuse of symbols, we use the superscripts S or T only in the last row and column to distinguish our Tsallis entropic method from Brody-Hughston's Shannon entropic method. The step function $I_{T_k}^{T_{k+1}}$ used here is the same as $I_{T_k T_{k+1}}$ defined before but for saving space.

To begin with, our calibration method is based on the maximization of the Tsallis entropy, whereas Brody and Hughston have chosen the Shannon entropy in their calibration method. As we have discussed in Section 7.1, the Tsallis entropy recovers the Shannon entropy as a special case as the entropy index $q \rightarrow 1$.

Next, by use of the principle of maximum entropy and given the only expectation

constraint on an abstract function $g(X)$ (X denotes the random variable), the Tsallis entropy produces a power-law maximum entropy distribution, called \bar{q} -distribution with a reset entropy index defined via $\bar{q} - 1 = 1 - q$. Here $C_{\bar{q}}$ functions as the normalization factor, and λ is the Lagrange multiplier introduced in accordance with the expectation constraint. On the other hand, the Shannon entropy produces an exponential function, where the normalization factor and the Lagrange multiplier are denoted by, respectively, C and λ . In the limit $\bar{q} \rightarrow 1$, the \bar{q} -distribution reduces to the ordinary exponential distribution. Since a exponential distributed term structure represents a flat term structure on a continuously compounded basis, the physical term \bar{q} can now find its financial interpretation as a measure of departure from the continuously compounded flat term structure.

Finally, assume that the information at our disposal includes multiple bond price data observed from real markets together with the value of a perpetual annuity. Based on maximizing the Tsallis entropy, the initial term structure density proves to be piecewise power-law distributed, where

1. the normalization factor C is nothing new but the initial short rate;
2. the Lagrange multiplier R in accordance with the annuity constraint proves to be the long-term rate that dominates the trend of the term structure evolution;
3. the Lagrange multipliers $\{\tau_k\}$ reflect the piecewise information indicated by the bond prices maturing at different time;
4. the power-law exponent N defined via $\frac{1}{N} = 1 - q$ actually counts the compounding frequency of the observed bonds.

On the other hand, the initial term structure density based on the Shannon entropy maximization proves to be piecewise exponentially distributed, where

1. r_0 represents the initial short rate;
2. the Lagrange multiplier R in accordance with the annuity constraint proves to be the long-term rate that dominates the trend of the term structure evolution;
3. $\{\tau_k\}$ ($k > 0$) reflect a combination of information indicated by the pointwise bond prices and the normalization condition.

Comparing the Tsallis distribution with the Shannon distribution, we see that the larger N is in the Tsallis entropy, the more frequently compounded bonds we are observing. When we are given the prices of continuously compounded bonds, $N \rightarrow \infty$ and thus the resulting piecewise power-law distribution converges to the piecewise exponential distribution. Since the piecewise exponential form (derived from the Shannon entropy) actually represents a piecewise flat term structure with a constant continuously compounded rate for each time interval $[T_k, T_{k+1}]$, the entropy index \bar{q} (or equivalently q) measures the departure of our calibrated term structure from flatness on a continuously compounded basis.

7.3 An Initial Calibration Approach Based on Superstatistics

The entropic method we present in the preceding section suggests a research direction from the general term structure (with a compounding frequency N) to the continuously compounded flat term structure as the entropy index $\bar{q} \rightarrow 1$, or equivalently, $N \rightarrow \infty$. A question arises naturally: what if we reverse the study direction? for example, we initially suppose that the term structure in a short term is flat associated with a constant continuously compounded rate and impose further assumption on the rate. With this end in mind, we assume that the initial term

structure density is “locally” characterized by a ordinary Boltzmann factor:

$$\rho(x) = \frac{1}{Z(\beta)} \exp(-\beta f(x)), \quad (7.61)$$

where $f(X)$ denotes an abstract function and is specified when we are given some market information. Here the short rate β is taken as a constant in a short term but obeys a certain distribution in a long term. Our aim is to determine the whole term structure over a long time. This can be accomplished by use of the notion of superstatistics.

In what follows we will start with an identification of the role of the “local” parameter β in Section 7.3.1. Subsequently, we will introduce the general idea of superstatistics in Section 7.3.2. Following it in Section 7.3.3 we will propose another initial calibration algorithm with the information of the price of a perpetual annuity. Finally, many interesting observations and properties of the resulting distribution are outlined in Section 7.3.4.

7.3.1 Clarification of Spatio Parameter

First of all, let us illustrate the role of β in distribution (7.61).

In Section 7.1.1, we have introduced the major difference between traditional entropies and the Tsallis entropy. Traditional entropies are always applied to quantify the energy used only between short-range interactions in a extensive system. The corresponding maximum entropy distribution thus reflects the decor relation between the variables in the system. The Tsallis entropy, however, can also be used to measure the energy between long-term interactions in a non-extensive system. Its corresponding maximum entropy distribution thus characterizes the strong correlation between the variables in the system. Therefore, by maximizing the Shannon entropy — a kind of traditional entropies — subject to the expectation constraint

on $f(X)$, the resulting ordinary exponential distribution (7.61) would *locally* stand only, namely, for a small range of values of the random variable X .

Such a local density function is parameterized by the spatial parameter β , which can be regarded as a constant rate in a short term. Alternatively, if here the random variable is a measure of time, for instance, in our term structure model where X denotes the time left to maturity, then we say that distribution (7.61) is available only for a small time scale. To avoid confusion on the physical nature of X , we give a unified name to the parameter β : the *spatio parameter*.

The role of the spatio parameter β can be further illustrated in the calibration method proposed by Brody and Hughston [19]. Assume we are given a set of bond prices $\{P_{0T_k}\}$ ($k = 1, \dots, n$) expressed by

$$\int_{T_k}^{\infty} \rho(x) dx = P_{0T_k}, \quad (7.62)$$

and the price of a perpetual annuity, which is expressed as the mean value w.r.t the term structure density by

$$\int_0^{\infty} x\rho(x) dx = \zeta. \quad (7.63)$$

Then the term structure density function that maximizes the Shannon entropy proves to be a piecewise exponential function in the form of

$$\rho(x) = \frac{1}{Z(\lambda, \mu)} \exp\left(-\lambda x - \sum_{k=1}^n \mu_k I_{T_k}(x)\right), \quad (7.64)$$

where $Z(\lambda, \mu) = \int_0^{\infty} \exp(-\lambda x - \sum_{k=1}^n \mu_k I_{T_k}(x)) dx$ is the normalization factor; λ and $\{\mu_k\}$ are determined by, respectively, the annuity constraint (7.63) and the bond price constraint (7.62). $I_{T_k}(x)$ is the step function defined by (7.31). Comparing (7.64) with (7.61), we find that the spatio parameter β in this case linearly grows as the random variable evolves from a short-time scale to a long-term scale.

This follows as a consequence of the fact that the step function controls the actual number of parameters $\{\mu_k\}$ working on the density function.

7.3.2 Basic Idea of Superstatistics

Inspired by the example illustrated in the previous subsection, we have a more general idea, assuming that β follows a particular distribution such that in the long-time run the whole term structure is described by a superposition of various Boltzmann exponential factors (7.61) with different β , or in short, a *superstatistics*.

Mathematically, the density function given a value of β in local is a conditional probability density function $\rho(x|\beta)$ given by the ordinary Boltzmann factor:

$$\rho(x|\beta) = \frac{1}{Z(\beta)} \exp(-\beta f(x)). \quad (7.65)$$

In the long-term run, the whole system is characterized by an average of local Boltzmann exponential factors by integrating over β . Thus the joint probability $\rho(x, \beta)$ is given by

$$\rho(x, \beta) = \rho(x|\beta)h(\beta), \quad (7.66)$$

where $h(\beta)$ denotes the distribution of β . Finally, the marginal probability density function $\rho(x)$ is obtained via

$$\begin{aligned} \rho(x) &= \int_{-\infty}^{\infty} \rho(x|\beta)h(\beta) d\beta \\ &= \int_{-\infty}^{\infty} h(\beta) \frac{1}{Z(\beta)} \exp(-\beta f(x)) d\beta. \end{aligned} \quad (7.67)$$

The superstatistics approach has been applied in many physical systems, such as Lagrangian turbulence [6, 10, 11], Eulerian turbulence [8], and defect turbulence [33]. In these turbulence systems, the velocity of a Brownian particle is assumed to follow a Langevin equation with the volatility coefficient σ . Thus the stationary probability density of the velocity is Gaussian distributed with mean 0 and variance β^{-1} ,

where $\beta \sim \sigma^{-2}$ and can be physically interpreted as the inverse temperature of ordinary statistical mechanics. With a further assumption on β that β follows a χ^2 -distribution, the corresponding superstatistics for the velocity is given by the Tsallis statistics (7.22). Based on the power-law distribution, experimentally measured non-Gaussian stationary distributions are successfully reproduced [9]. Following this line of argument, we see a link between the maximum Tsallis entropy distribution and the supersatistics.

7.3.3 Initial Calibration Algorithm

Now we attempt to design an initial calibration algorithm by use of the superstatistics method presented above.

At the first step, we should specify the conditional probability density function. In the short-term scale, it is plausible to introduce a continuously compounded flat term structure, for which a discount bond is priced by

$$P_0(x|\beta) = e^{-\beta x}, \quad (7.68)$$

where the local parameter β acts as a constant continuously compounded rate in a short term. For the term structure density $\rho(x) = -\frac{\partial P_0(x)}{\partial x}$, we obtain $\rho_0(x|\beta) = \beta e^{-\beta x}$, or

$$Z(\beta) = \beta, \quad (7.69)$$

$$f(x) = x, \quad (7.70)$$

if we rewrite the density function in the form of (7.65). It is immediately verified that the parameter β has dual roles: 1) the local short rate; and 2) the reciprocal of the mean of density function $\rho_0(x|\beta)$. Furthermore, it is known that the expectation of the term structure density is exactly the perpetual annuity so that β finds its third role.

The second step in the superstatistics approach is to find an approximate distribution for the spatio parameter β . Given the data observed from bond markets, for instance, x -year spot rates, we can easily derive the real distribution of β . Here in our theoretical research, we assume that β is χ^2 -distributed with degree n , i.e.

$$h(\beta) = \frac{1}{\Gamma(\frac{n}{2})} \left(\frac{n}{2\beta_0}\right)^{\frac{n}{2}} \beta^{\frac{n}{2}-1} \exp\left(-\frac{n\beta}{2\beta_0}\right), \quad (7.71)$$

where $\Gamma(\cdot)$ is a gamma function and hence the χ^2 -distribution is also called the *gamma distribution*. In fact, the form (7.71) is usually used as the distribution of the sum of n squared identified independently Gaussian distributed random variables $\{Y_i\}$ ($i = 1, 2, \dots, n$) with mean zero. We write

$$\beta := \sum_{i=1}^n Y_i^2, \quad (7.72)$$

and its expectation is calculated to be

$$\bar{\beta} = n\langle Y_i^2 \rangle = \int_0^\infty \beta h(\beta) d\beta = \beta_0. \quad (7.73)$$

Here the integration interval starts from zero because we always assume the interest rate is positive.

Before proceeding further, we should clarify the reason for the choice of χ^2 -distribution. Recall in Section 4.2, we have introduced a Laplace transform expression for the initial term structure density:

$$\rho_0(T) = \int_0^\infty \beta e^{-T\beta} \phi(\beta) d\beta \quad (7.74)$$

where $\phi(\beta)$ is the inverse Laplace transform of $\rho_0(T)$. Accordingly, for the discount function we have

$$P_{0T} = \int_0^\infty \phi(\beta) e^{-\beta T} d\beta. \quad (7.75)$$

The essential idea behind (7.74) or (7.75) is the same as superstatistics — they both treat the whole term structure as a weighted superposition of local continuously

compounded flat term structures if $\phi(\beta)$ here is interpreted as a distribution for the short rate β . Specifically, consider the standard gamma distribution with parameters M and λ

$$\phi(\beta) = \frac{1}{\Gamma(M)} \lambda^M \beta^{M-1} \exp(-\lambda\beta). \quad (7.76)$$

Substitution of (7.76) into (7.75) gives the formula for the discount bond:

$$P_{0T} = \frac{1}{\left(1 + \frac{T}{\lambda}\right)^M}. \quad (7.77)$$

Comparing the gamma distributions given in (7.71) and (7.76), we find $\lambda = \frac{M}{\beta_0}$ and $M = \frac{n}{2}$. As a consequence, the discount bond with a short rate process which follows the χ^2 -distribution (7.71) is priced by

$$P_{0T} = \frac{1}{\left(1 + \frac{\beta_0 T}{M}\right)^M}. \quad (7.78)$$

We see that with a χ^2 -distributed short rate β , the discount bond is calibrated to be compounded at the frequency $M = \frac{n}{2}$ over its life, with a constant annualized interest rate $\beta_0 = \bar{\beta}$. Thus the χ^2 -distribution assumption makes sense in our calibration.

Now let us turn back to the notion of superstatistics and proceed to the final step for determining the marginal distribution. Substitute Eqn. (7.69) for $Z(\beta)$, Eqn. (7.70) for $f(x)$, and Eqn. (7.71) for $h(\beta)$ into (7.67). A short calculation gives rise to

$$\rho(x) = \frac{2\Gamma\left(\frac{n+2}{2}\right) \beta_0}{\Gamma\left(\frac{n}{2}\right) n} \frac{1}{\left[1 + 2x \frac{\beta_0}{n}\right]^{\frac{n+2}{2}}}. \quad (7.79)$$

Note that the term $\frac{2\Gamma\left(\frac{n+2}{2}\right)}{\Gamma\left(\frac{n}{2}\right) n}$ equals one for any integer n . Hence for the term structure density $\rho(x)$ with a χ^2 -distributed short rate process β , we obtain a power-law distribution in the form of

$$\rho(x) \sim \frac{1}{\left[1 + \bar{\beta}(\bar{q} - 1)x\right]^{\frac{1}{\bar{q}-1}}}, \quad (7.80)$$

provided the following identifications are made:

$$\frac{1}{\bar{q} - 1} = \frac{n + 2}{2} \iff \bar{q} = 1 + \frac{2}{n + 2} \tag{7.81}$$

$$\bar{\beta}(\bar{q} - 1) = \frac{2\beta_0}{n} \iff \bar{\beta} = \frac{\beta_0}{2 - \bar{q}}. \tag{7.82}$$

Sum up matters so far. We regard the yields of n differently-maturing bonds P_{0T_k} ($k = 1, 2, \dots, n$) as given information and assume there exists a perpetual annuity. If we further suppose that the yields are χ^2 -distributed, then the calibrated term structure density is given in the form of (7.80). We summarize our discussion in the following proposition.

Proposition 7.4. *Given a set of yields $\{R_k\}$ ($k = 1, 2, \dots, n$) to different maturities $\{T_k\}$ and the existence of a perpetual annuity, if we assume the yields are χ^2 -distributed with degree n , then the term structure density function is given by*

$$\rho(x) = \frac{C}{\left[1 + \bar{\beta}(\bar{q} - 1)x\right]^{\frac{1}{\bar{q}-1}}}, \tag{7.83}$$

where C is the initial short rate; the values of \bar{q} and n are related by (7.81); and $\bar{\beta}$ is determined via

$$C = \bar{\beta}(2 - \bar{q}). \tag{7.84}$$

The price of the perpetual annuity is inferred via

$$\zeta = \frac{1}{\beta_0} \frac{n}{n - 2}, \tag{7.85}$$

where β_0 is determined by (7.82). The corresponding discount function P_{0x} is given by

$$P_{0x} = \frac{1}{\left[1 + \bar{\beta}(\bar{q} - 1)x\right]^{\frac{2-\bar{q}}{\bar{q}-1}}}. \tag{7.86}$$

Proof. It follows from our previous discussion that the term structure density is (7.80), and can be rewritten as

$$\rho(x) = \frac{C}{(1+ax)^b}, \quad (7.87)$$

where we write $a \triangleq \tilde{\beta}(\tilde{q} - 1)$ and $b \triangleq \frac{1}{\tilde{q}-1}$ to simplify the notations. C functions as the normalization factor. Moreover, $C = \rho(0)$, indicating that C is the initial short rate.

Note that $b - 1 = \frac{2-\tilde{q}}{\tilde{q}-1} = \frac{n}{2}$ by (7.81). Thus Eqn. (7.87) is integrable over $[0, \infty]$ provided that $1 < \tilde{q} < 2$. A simple calculation gives rise to

$$\int_0^\infty \frac{C}{(1+ax)^b} dx = \frac{C}{a(b-1)}. \quad (7.88)$$

Thus the relation (7.84) is proved as a consequence of the normalization condition and expressions for a and b .

Besides, since the price of the perpetual annuity can be expressed as the expectation of the tenor variable:

$$E[X] = \zeta, \quad (7.89)$$

it follows from the law of total expectation that

$$E[X] = E_{h(\beta)} [E_{\rho(x)}[X|\beta]] = \int_0^\infty \frac{1}{\beta} h(\beta) d\beta = \frac{1}{\beta_0} \frac{n}{n-2}. \quad (7.90)$$

Thus we have proved (7.85).

Finally, the discount bond is valued by use of the fact that

$$P_{0x} = \int_x^\infty \rho(u) du \quad (7.91)$$

$$\begin{aligned} &= \int_x^\infty \frac{C}{(1+au)^b} du \\ &= \frac{C}{a(b-1)} \frac{1}{(1+ax)^{b-1}}, \end{aligned} \quad (7.92)$$

for $x \in [0, \infty)$. Substituting the expressions of a , b , and C into (7.92), we obtain (7.86). \square

7.3.4 Observations and Properties

We observe many interesting properties for the resulting distribution.

First, equations (7.81) and (7.82) imply that $1 < \bar{q} < 2$, which precisely coincides with the range given in Section 7.1.2 and thus guarantees the power-law property.

Second, with the definition $N = \frac{1}{\bar{q}-1}$ or equivalently $N = \frac{2}{n+2}$, the term structure density (7.83) is given in terms of $\bar{\beta}$ and N as

$$\rho(x) = \frac{1}{\left[1 + \frac{\bar{\beta}x}{N}\right]^N}, \quad (7.93)$$

where N , just as we have pointed out in the entropic method, still counts the compounding frequency of the observed bonds. In the limit $N \rightarrow \infty$, the calibrated term structure converges to the flat term structure on a continuously compounded basis. Moreover, as a consequence of (7.81) and (7.82), the mean value β_0 of the χ^2 -distributed short rate tends to the long-term rate $\bar{\beta}$ as $N \rightarrow \infty$.

Furthermore, let us study the annuity pricing formula (7.85). We see that the price of the perpetual annuity does not equal, but differs a little from, the reciprocal of the mean value β_0 of β , although given any fixed value for β the corresponding annuity price equals $\frac{1}{\beta}$. If we are given continuously compounded bonds, namely, in the limit case $N \rightarrow \infty$ or $n \rightarrow \infty$, the annuity will be valued precisely at $\frac{1}{\beta_0}$.

Moreover, let us study the expectation of the random variable X . For an individual discount bond, the tenor variable X is actually the duration of this bond, measuring the average time one takes to earn back as much as it has invested on the bond. Consequently, $E[X]$ calculates the mean of duration of a portfolio of bonds that mature differently and accordingly measures the average time one waits before recouping its investment on this bond portfolio.

Finally, a few remarks about the difference between the superstatistics approach

and the entropic approach are outlined here. These two methods are supported by different principles, although they produce the same power-law distribution when the only source of information is the value of a perpetual annuity. The difference can be justified by use of the *principle of minimum relative information* [47], an equivalent principle of maximum entropy. The *relative information entropy* is a non-commutative measure of the difference between two probability distributions p and q defined by

$$D(p||q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} dx, \quad (7.94)$$

where p typically represents the “true” distribution of observations, while q represents a model or approximation of p . Given a prior distribution q , the principle of minimum relative information suggests to choose a posterior distribution p under certain constraints which is as hard to discriminate from the original one q as possible. Accordingly the information gained in $D(p||q)$ should be as small as possible. Following this line of argument, in our calibration if we assume the prior distribution as

$$q(x) = \beta \exp(-\beta x)$$

just as what we do in the superstatistics method, and further impose a constraint on the mean value such that

$$\int_{-\infty}^{\infty} xp(x) dx = \frac{1}{b},$$

then the minimum relative information distribution is given by

$$p(x) = b \exp(-bx).$$

It shows that given an exponentially distributed prior, the posterior distribution by use of the entropic approach is still exponentially distributed. Whereas the posterior distribution by use of the superstatistics approach is power-law distributed.

7.4 Summary and Discussion

In this chapter we have designed two calibration algorithms for the initial term structure density, one based on the Tsallis entropy maximization and the other based on superstatistics.

Tsallis entropic approach

Based on maximizing the Tsallis entropy, we determine the initial term structure density that is consistent with multiple bond price data and the value of a perpetual annuity. The idea is to treat the Tsallis entropy as a functional of the term structure density and express the known data as constraints on the density function. With the calculus of variations and Lagrange multipliers, we obtain a piecewise power-law distribution for the initial density.

The initial distribution is parameterized by $1 - q$, where q is the entropy index, a physical measurement originally used in thermodynamics. In the power-law distributed term structure density, the power-law exponent N is defined via $\frac{1}{N} = 1 - q$. We prove that N is nothing new but the compounding frequency of the observed bonds. When we are given the prices of continuously compounded bonds, $N \rightarrow \infty$ or equivalently $q \rightarrow 1$. At this time, the calibrated density function reduces to the piecewise exponential form. In this sense, the entropy index q essentially measures the departure of the current term structure from flatness on a continuously compounded basis.

More observations from the comparison between the initial calibration algorithms proposed by Brody-Hughston and us could be found in Table 7.1 in Section 7.2.4.

Superstatistics approach

Inspired by the concept of superstatistics, we initially suppose that the term structure in a short term is flat associated with a constant continuously compounded rate

β , and further assume that the rate follows a χ^2 -distribution. Therefore, the whole term structure could be regarded as a superposition of local flat structures, and proves to follow the same power-law distribution as the entropic method indicates if the only source of information available is the existence of a perpetual annuity.

The power-law exponent N in the initial term structure density is defined via $\frac{1}{N} = \frac{2}{n+2}$, where n is the degree of the χ^2 -distribution for the local short rate. We prove that N accounts for the compounding frequency of the underlying bonds. Many interesting properties of the current model are observed when we are given the prices of continuously compounded bonds, namely, $N \rightarrow \infty$. First, the calibrated initial term structure becomes flatter on a continuously compounded basis as N increases. Second, in the limit $N \rightarrow \infty$ the mean value β_0 of the χ^2 -distributed local short rate tends to the long-term rate. Third, for an arbitrary N the perpetual annuity price does not equal but differs a little from the reciprocal of the mean value β_0 of short rate. In the limit $N \rightarrow \infty$, however, the annuity is valued precisely at $\frac{1}{\beta_0}$.

Finally, we should point out one future study direction of the superstatistics approach. Assuming that we are given more information, for example, the prices of multiple bonds for different maturities, we want to see if the calibrated density function under the framework of superstatistics is still power-law distributed.

□ End of chapter.



Chapter 8

Implementation of the Proposed Model

Based on the theoretical study presented in the previous chapters, we implement the proposed model with initial data in the US swap market for 15 Feb, 2007. First, we analyze in Section 8.1 the raw data and the bootstrapped zero rates. Preliminary computation based on the US term structure data in swap market gives rise to forward swap rates and LIBOR forward rates. Besides, given market quotes for implied cap volatilities, we immediately obtain the Black formula cap prices on 3-month US LIBOR rates. Next, in Section 8.2 we calibrate the initial term structure by maximizing, respectively, the Shannon entropy and the Tsallis entropy. Finally, we implement the proposed model in Section 8.3 and obtain the evolutions of short rates and bond prices over a long term. To test our model improvements over traditional models, we also run the simulation with the Hull-White model. A comparison of these two no-arbitrage models is presented in Table 8.7 in Section 8.3.3.

8.1 Data Description

To implement the proposed model we use two data sets, kindly supplied by a large international bank, for 15 Feb, 2007: US term structure data in swap market to

determine the underlying term structure of forward swap rates and LIBOR forward rates, and US derivatives data on implied volatilities of caps.

8.1.1 US Term Structure Data

The term structure data consist of the raw data observed from the US swap market and the zero rates bootstrapped from them.

The raw data with different maturities up to 30 years are given in Table 8.1. The raw data includes spot LIBOR rates in money market, Eurodollar futures, and swap rates of different maturities. In the US market, spot LIBOR rates are typically used to define short-term LIBOR zero rates. Eurodollar futures are then used for maturities up to two years. For longer maturities, swap rates are applied.

The corresponding zero rates, bootstrapped from the raw data and computed using continuously compounding, are provided in Table 8.2. Given these data, we plot the initial yield curve in Figure 8.1 together with the corresponding discount bond prices. Here the discount bond P_{0T} with yield y is valued by

$$P_{0T} = e^{-yT}. \quad (8.1)$$

The blue circles on the initial zero curve represent the data provided in Table 8.2. Other zero rates are computed by linear interpolation at time nodes $0 < T_1 < T_2 < \dots < T_n = T = 30$ years, where $T_i = 3i$ months.

We observe a partially inverted yield curve in Figure 8.1. First, at a very short horizon there is a positive sloping segment, represented by the first four circles. This indicates a growing economy in the subsequent half a year, namely until Aug, 2007. This is because a positive slope always reflects investors' expectation for the economy to grow in the future and for a rising inflation associated with this growth. With this expectation, the central bank will tighten monetary policy by raising short

Cash Rates		Eurodollar Futures		Swap Rates	
Tenor	Rate(%)	Expiry	Price	Tenor	Rate(%)
2-Day	5.2400	Mar 07	94.640	2-Year	5.215
1-Week	5.3044	Jun 07	94.665	3-Year	5.145
1-Month	5.3200	Sep 07	94.765	4-Year	5.130
2-Month	5.3450	Dec 07	94.900	5-Year	5.140
3-Month	5.3600	Mar 08	95.010	7-Year	5.170
		Jun 08	95.065	8-Year	5.190
		Sep 08	95.105	9-Year	5.210
				10-Year	5.230
				12-Year	5.270
				15-Year	5.320
				20-Year	5.360
				30-Year	5.370

Table 8.1: Raw data in the US swap market for 15 Feb, 2007

Zero Rates			
Tenor	Rate(%)	Tenor	Rate(%)
1-Day	5.3108	4-Year	5.0653
1-Month	5.3695	5-Year	5.0768
3-Month	5.3931	7-Year	5.1108
6-Month	5.3853	10-Year	5.1804
1-Year	5.3167	15-Year	5.2923
2-Year	5.1505	20-Year	5.3426
3-Year	5.0802	30-Year	5.3414

Table 8.2: US zero rates for 15 Feb, 2007

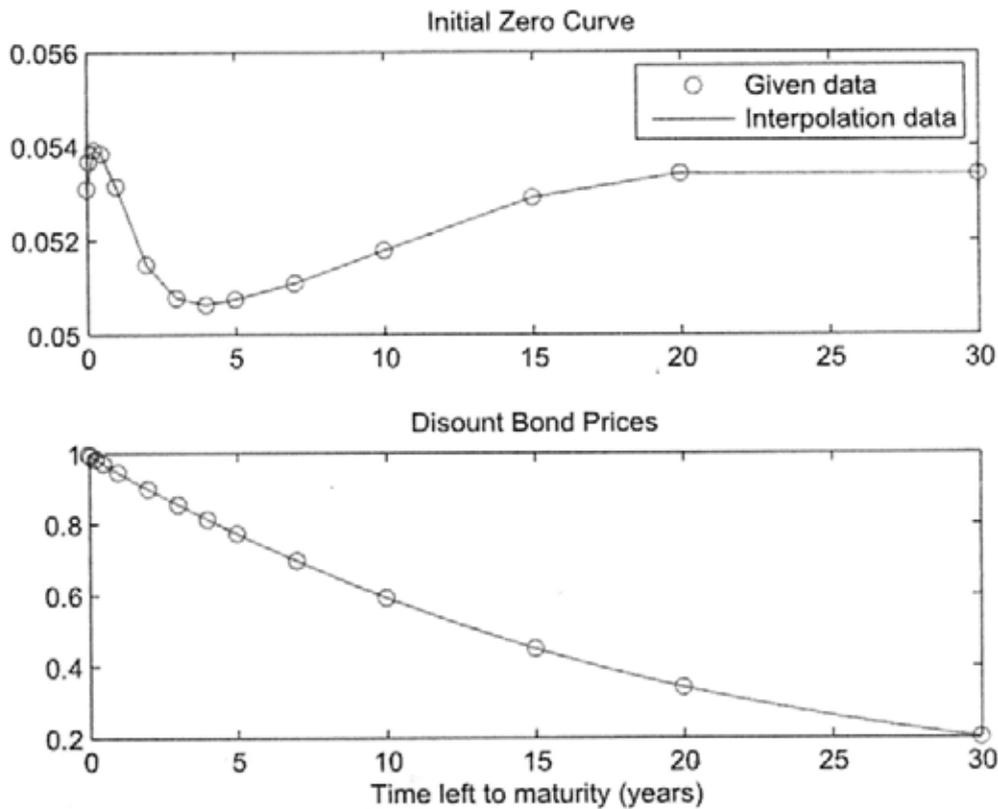


Figure 8.1: Initial Term Structure

term interest rates in the future to slow economic growth and dampen inflationary pressure. Next, starting from the fourth circle, the yield curve is downward sloping and reaches its lowest point at the eighth circle. This inverted segment indicates an expectation of interest rate cuts or even an economic decline in the subsequent four years since 2007, particularly bottomed out in 2011. This is because an inverted yield curve always occurs when long-term investors believe the economy will slow or even decline in the future. With this predication, the central bank will loosen monetary policy by lowering rates to stimulate the economy. Finally, since the ninth circle, the slope of the yield curve becomes positive again, indicating the anticipation of a slow economic recovery between 2011 and 2012 and a full recovery afterwards.

There are two turning points on the initial yield curve – Aug, 2007 (the fourth circle) and Feb, 2011 (the eighth circle) – indicating, respectively, rate cuts and hikes. Recall the real economic environment since 2007. We have witnessed the financial crisis breaking out in the third quarter of 2008 and observed certain indications at the beginning of 2010 of the economic recovery. It remains to be seen if the recession has really bottomed out and a full economic recovery will come in 2012 as expected.

Moreover, the second graph in Figure 8.1 represents the bond price as a monotone decreasing function of the time left to maturity. If the discount bond matures today, its value equals the face value one; otherwise its value reduces as maturity lengthens.

The term structure data can be used to evaluate forward swap rates and LIBOR forward rates [26, 27]. Consider a forward swap with principle \$1, where two parties agree to exchange at dates $\{T_{i+1}, \dots, T_{i+j}\}$ the floating LIBOR rates $\{L(T_i), \dots, L(T_{i+j-1})\}$ for a fixed rate. The forward swap rate is the fixed rate that gives this contract zero initial value. The value of the forward swap rate at the present time 0 with maturity T_k is given by

$$Swap = \frac{P_{0T_1} - P_{0T_k}}{\sum_{i=2}^k (T_i - T_{i-1}) P_{0T_i}}. \quad (8.2)$$

The LIBOR forward rate as seen at 0 for a period $[T_{i-1}, T_i]$ is calculated by

$$F_0(T_{i-1}, T_i) = \frac{P_{0T_{i-1}} - P_{0T_i}}{(T_i - T_{i-1}) P_{0T_i}}. \quad (8.3)$$

Using the discount bond prices calculated via (8.1), we obtain the forward swap rates maturing from 1 year to 30 years and the LIBOR forward rates for various time intervals $[T_{i-1}, T_i]$ ($T_i = 3i$ months). We display the values of swap rates in Table 8.3, and plot the behavior of swap rates and LIBOR rates in Figure 8.2.

Forward Swap Rates			
Tenor	Rate(%)	Tenor	Rate(%)
1-Year	5.3537	7-Year	5.1333
2-Year	5.1649	10-Year	5.1956
3-Year	5.0973	15-Year	5.2860
4-Year	5.0862	20-Year	5.3256
5-Year	5.0994	30-Year	5.3344

Table 8.3: Forward swap rates

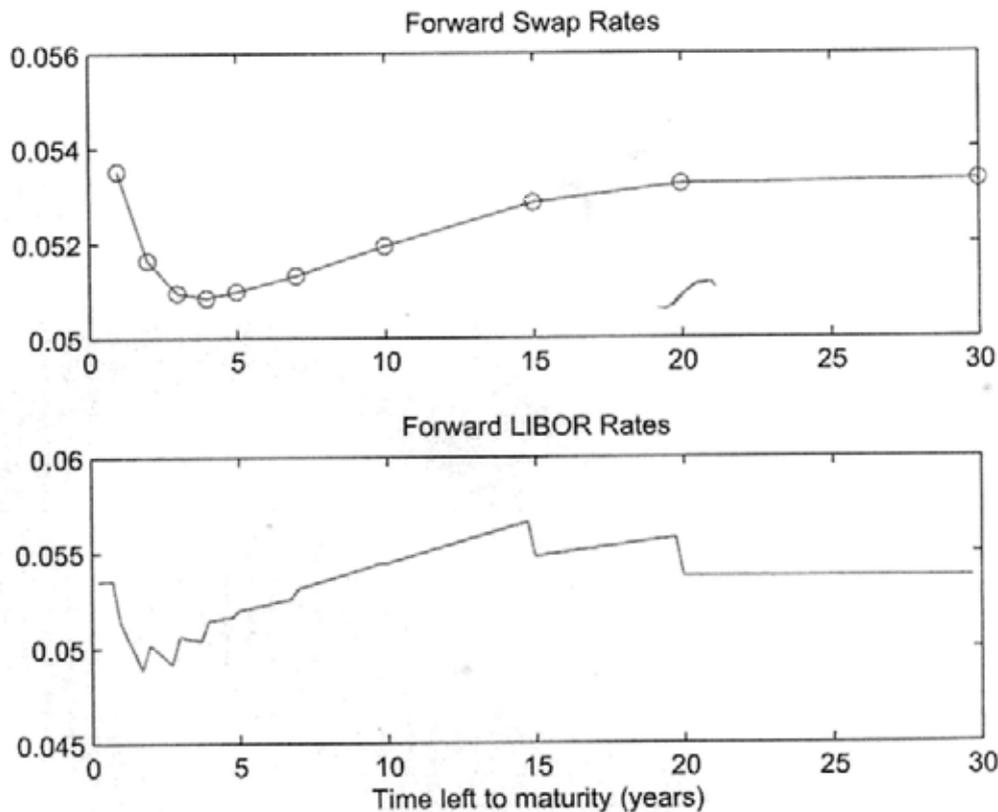


Figure 8.2: Forward swap rates and LIBOR forward rates

In the first graph of Figure 8.2, we observe that the swap rate curve exhibits an analogous shape as the initial yield curve shows in Figure 8.1. The inverted segment during the first three years reflects investors' prediction for the economy to decline in the future until 2010. The flat curve between the third and the fifth data points

implies an economic adjustment during 2010 and 2012. Afterwards, the positive sloping curve indicates an economic recovery starting from 2012. In the second graph of Figure 8.2, we see that the forward LIBOR rate increases stepwise as maturity lengthens, although for very short (less than 2 years) and long (after 20 years) maturities there are some exceptions. The stepwise trend is due to the fact that the forward rate is a piecewise function depending on the time interval it relies in. The flat trend after 20 years is due to the incomplete information on yields for that period — we are given only data on time nodes $T = 20$ and $T = 30$ but without any other information in between.

8.1.2 US Derivatives Data

The derivatives data we use here are daily quotes for the implied Black volatilities of at-the-money-forward (ATMF) US caps. The cap under consideration is a portfolio of call options on the 3-month LIBOR forward rate. Each option of the cap is known as caplet, which is observed at time T_i with the payoff occurring at time T_{i+1} . The strike price of each ATMF cap is taken as the corresponding forward swap rate with quarterly compounding.

The market convention to quote a cap price is to quote the implied volatility which sets the Black model price equal to the market price. Actually there are two types of implied volatilities that put into the Black formula. One type is called spot volatility, namely, to use a different volatility for each caplet. The other type is called flat volatility, namely, to use the same volatility for all the caplets constituting any individual cap. The data given on the left side in Table 8.4 are the flat volatilities quoted in the market for caps maturing from 1 year to 30 years.

There is a one-to-one mapping between the flat volatility and the present value

Flat Volatilities		Cap Prices	
Tenor	Rate(%)	Tenor	Price(%)
1-Year	13.92	1-Year	0.1576
2-Year	15.39	2-Year	0.5261
3-Year	15.88	3-Year	0.9854
4-Year	15.66	4-Year	1.4710
5-Year	15.30	5-Year	1.9646
7-Year	14.58	7-Year	2.9619
10-Year	13.94	10-Year	4.2699
15-Year	12.71	15-Year	6.4842
20-Year	12.07	20-Year	8.3716
30-Year	11.48	30-Year	11.1256

Table 8.4: Flat cap volatilities on 3-month US LIBOR for 15 Feb, 2007

of a cap for a certain maturity. Specifically, for maturity T_k the cap is priced by

$$Cap = \sum_{i=2}^k P_{0T_i}(T_i - T_{i-1}) (F_i \mathcal{N}(d_1^i) - R \mathcal{N}(d_2^i)), \quad (8.4)$$

where $F_i = F_0(T_{i-1}, T_i)$ is the LIBOR forward rate calculated via (8.3) and R denotes the forward swap rate given in Table (8.3) for maturity T_k . $\mathcal{N}(\cdot)$ denotes the standard normal distribution and

$$\begin{aligned} d_1^i &= \frac{1}{\sigma \sqrt{T_{i-1}}} \left(\log \left(\frac{F_i}{R} \right) + \frac{1}{2} \sigma^2 T_{i-1} \right) \\ d_2^i &= d_1^i - \sigma \sqrt{T_{i-1}}, \end{aligned} \quad (8.5)$$

where σ denotes the flat volatility for maturity T_k . The cap prices for different maturities, each corresponding to the implied volatility given on the left side in Table 8.4, are also listed in the same table on the right side, to make the equivalence between caps and implied volatilities more prominent..

We plot the implied flat volatilities and their corresponding cap prices in Figure 8.3. It demonstrates a typical ‘‘hump’’ pattern for the flat volatility curve. The peak of the hump appears at the third point, namely, in the year of 2010. A possible explanation is given as follows [29]. Typically, it is the central bank that controls

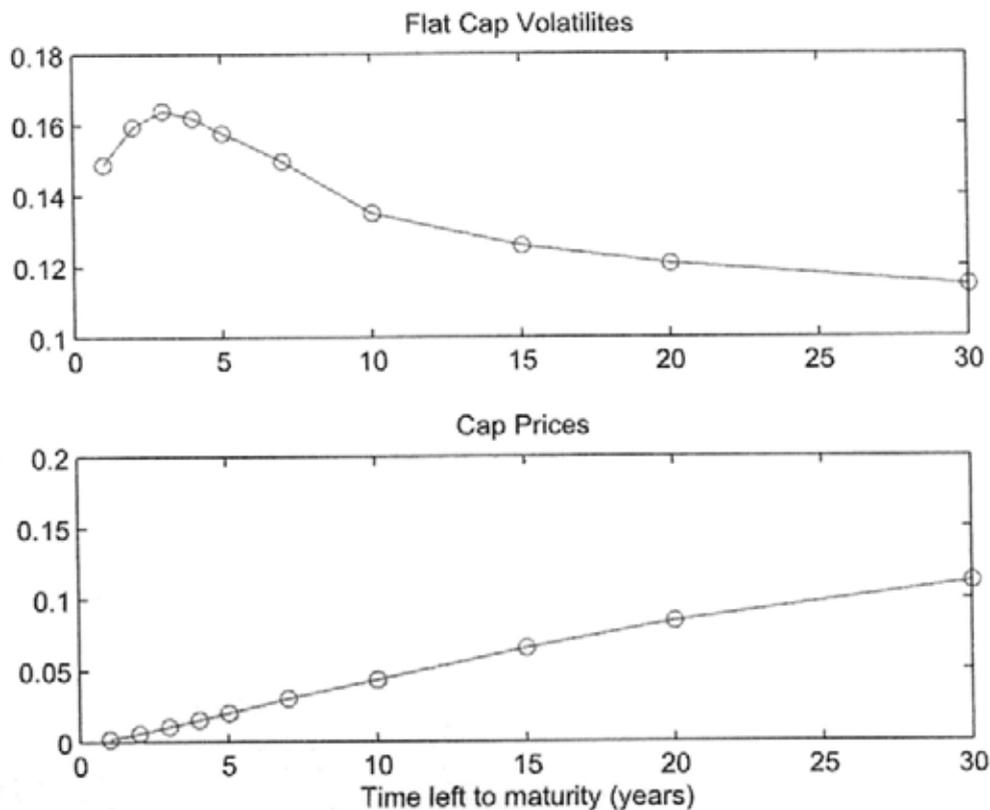


Figure 8.3: Cap prices and implied flat volatilities

rates at the short end of the zero curve and the traders determine the 2- and 3-year rates. For maturities beyond 3 years, the volatilities tend to decline because of the mean version of interest rates.

8.2 Initial Calibration

The first step to implement a term structure model is initial calibration. In Chapter 7 we have designed an iterative algorithm, which is based on the Tsallis entropy maximization, to determine the initial term structure density in terms of the initial short rate and the specified bond prices with different maturities. An alternative algorithm is proposed by Bordy and Hughston [19] to maximize the Shannon en-

tropy. In this section, given the short rate and bond prices $\{P_{0T_k}\}$ (indicated by the continuously compounded zero rates given in Table 8.2) with maturities $\{T_k\} = 1, 2, 3, 4, 5, 7, 10, 15, 20, \text{ and } 30$ years, we will use these two calibration algorithms to calibrate the initial term structure for 15 Feb, 2007.

This section is organized as follows:

1. Calibrate the initial term structure density in Section 8.2.1 by maximizing the Shannon entropy. The results are demonstrated in Figure 8.4–Figure 8.6.
2. Calibrate the initial term structure density in Section 8.2.2 by maximizing the Tsallis entropy. The results are demonstrated in Figure 8.7–Figure 8.9.
3. Inspect and verify the connection between these two calibration algorithms in Section 8.2.3. The convergence of the maximum Tsallis entropy distribution as the compounding frequency increases is demonstrated in Figure 8.10.
4. The key experimental results are displayed in Table 8.5.

8.2.1 Initial Calibration Based on Shannon Entropy Maximization

The initial calibration is performed by the following steps:

1. Deduce the maximum entropy term structure density $\rho(x) = \sum_{k=0}^n I_{T_k T_{k+1}}(x) r_k e^{-Rx}$ and the calibrated bond price $P_{0x} = P_{0T_k} - \frac{r_k}{R}(e^{-RT_k} - e^{-Rx})$ by following the steps indicated in Proposition 8 in [19] (proposed by Brody and Hughston). In the simulation, we assume $x = 3i$, $i = 0, 1, 2, \dots, 1200$ months and take the maximum maturity $x = 300$ years as an approximation of the infinity future.
2. Plot the initial term structure density function in Figure 8.4.

3. Display the values of $\{r_k\}$ in Table 8.5 and plot r_k as a function of T_k in Figure 8.5.
4. Plot the calibrated bond prices with different maturities up to 30 years in Figure 8.6, together with the given data in the same plot.
5. Test the algorithm accuracy:
 - (a) Test if $\int_0^\infty \rho(x) dx = 1$;
 - (b) Test if $\rho_t(0) = r_t$;
 - (c) Test if the calibrated bond prices coincide with the given data, by calculating the standard error

$$SEE = \frac{\sum (PP_{cal}(x) - PP(x))^2}{NN}, \quad (8.6)$$

where PP_{cal} denotes the calibrated data and PP denotes the given data. NN counts the number of bond price data. In our simulation, $NN = 1201$.

All the test values are displayed in Table 8.5.

Figure 8.4 demonstrates the initial term structure density with maturity horizon. The decreasing trend in general is due to its financial role as interest return on the discount bond. However, different from the smooth trend indicated by the bond price curve in Figure 8.1, the density curve decreases stepwise. This is attributed to its piecewise nature, namely, over each time interval $[T_k, T_{k+1})$ divided by the given maturities, the density function decreases exponentially.

Figure 8.5 displays the parameter r_k of the resulting maximum entropy distribution, which reflects the information indicated by the values of available bonds over

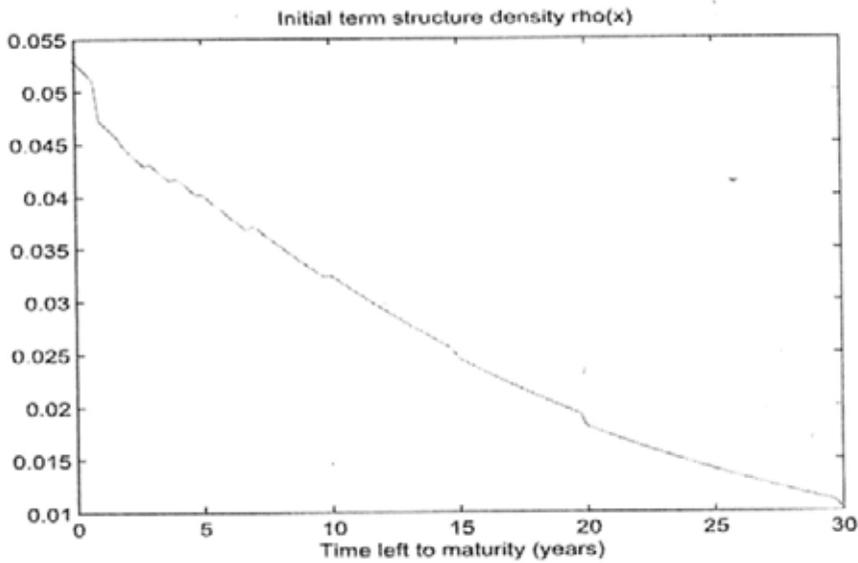


Figure 8.4: Initial term structure density calibrated by maximizing the Shannon entropy

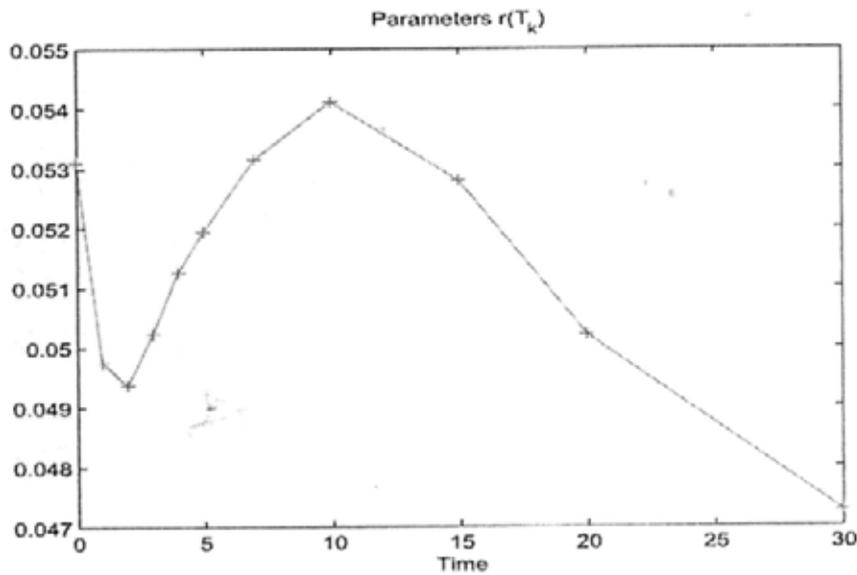


Figure 8.5: Parameters $r(T_k)$ in the maximum Shannon entropy distribution

$[T_k, T_{k+1})$. At the short end (over $[0, 10]$), the curve exhibits an analogous shape as the zero curve shows in Figure 8.1. However, after maturity 10 years, the curve is falling sharply since the bond price for long maturities tends to be vanishing.

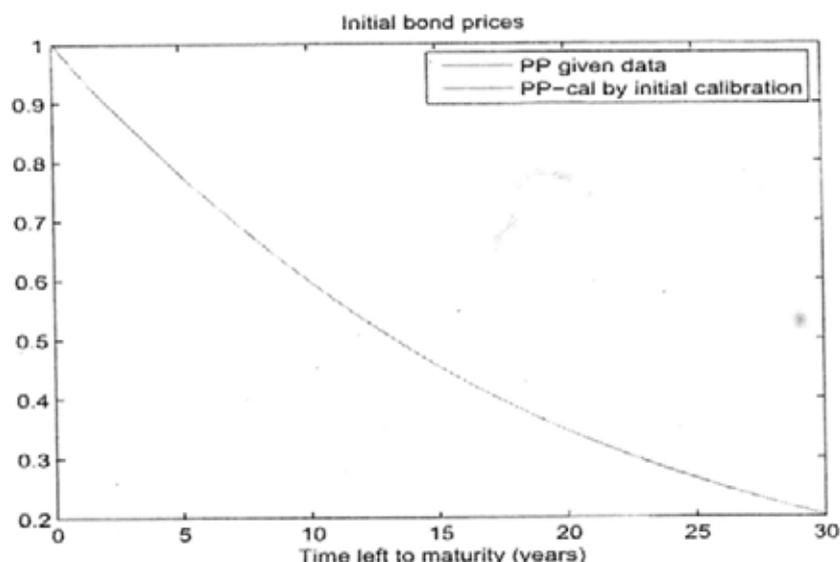


Figure 8.6: Calibrated bond prices together with the given data

Figure 8.6 demonstrates the initial calibrated values of bonds for different maturities, together with the given data. We find a perfect match in between, which shows that the calibrated density is a good candidate for the initial term structure.

Table 8.5 displays the computed values of the involved parameters and results for the tests carried out during the calibration procedure. The experimental values obtained from the Shannon entropic method are recorded in the third column. There $N \rightarrow \infty$ means that the bonds we consider in this experiment are continuously compounded. Pay particular attention to the last three rows. It shows that the normalization condition is “almost” satisfied. The slight difference from the theoretical value one is unavoidable since a density function with an infinity maturity is out of reach in simulation. Furthermore, the zero difference between $\rho(0)$ and r_0 verifies that $\rho_t(0)$ actually represents the short rate at time t . Finally, the standard error, defined via Eqn. (8.6), between the calibrated bond prices and the given data proves to be negligible, which further confirms our observation in Figure 8.6.

8.2.2 Initial Calibration Based on Tsallis Entropy Maximization

The initial calibration is performed by the following steps:

1. Deduce the maximum entropy term structure density $\rho(x) = \sum_{k=0}^n I_{T_k T_{k+1}}(x) \frac{C}{\left[1 + \frac{\beta_k}{N} + \frac{\beta}{N}x\right]^N}$ and the calibrated bond price $P_{0x} = P_{0T_k} - \frac{CN}{R(N-1)} \left[\frac{1}{\left(1 + \frac{\beta_k}{N} + \frac{\beta}{N}T_k\right)^{N-1}} - \frac{1}{\left(1 + \frac{\beta_k}{N} + \frac{\beta}{N}x\right)^{N-1}} \right]$ by following the steps indicated in Proposition 7.2 in Chapter 7. Here N denotes the compounding frequency of discount bonds. In our simulation, we assume the bonds are compounded monthly, i.e. $N = 12$.
2. Plot the initial term structure density function in Figure 8.7.
3. Display the values of $r(T_k) \triangleq Ce^{-\beta_k}$ in Table 8.5 and plot $r(T_k)$ as a function of T_k in Figure 8.8.
4. Plot the calibrated bond prices with different maturities up to 30 years in Figure 8.9, together with the given data in the same plot.
5. Test the algorithm accuracy:
 - (a) Test if $\int_0^\infty \rho(x) dx = 1$;
 - (b) Test if $\rho_t(0) = r_t$;
 - (c) Test if the calibrated bond prices coincide with the given data, by calculating the standard error (8.6).

All the test values are displayed in Table 8.5.

Figure 8.7 shows the initial term structure density with maturity horizon. This figure looks quite similar to Figure 8.4, which displays the maximum Shannon entropy distribution. In the next subsection, we will demonstrate how the Tsallis

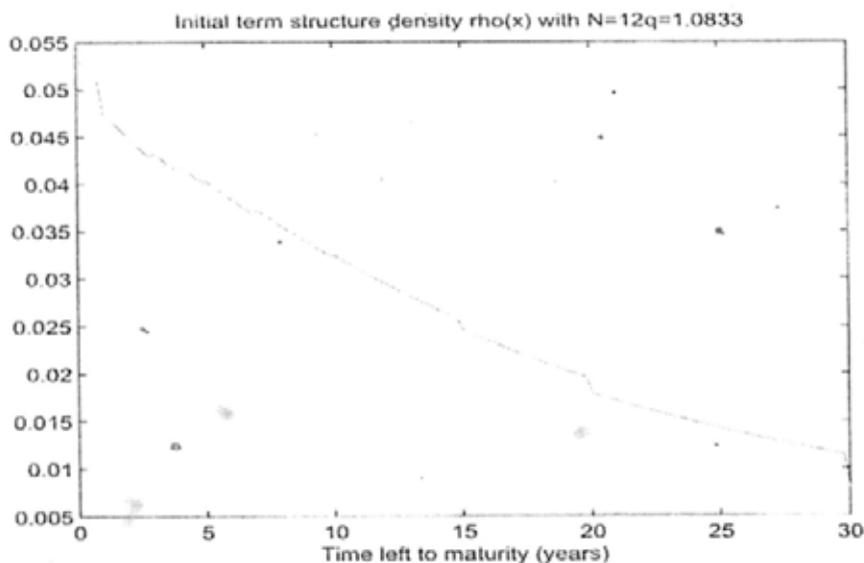


Figure 8.7: Initial term structure density calibrated by maximizing the Tsallis entropy

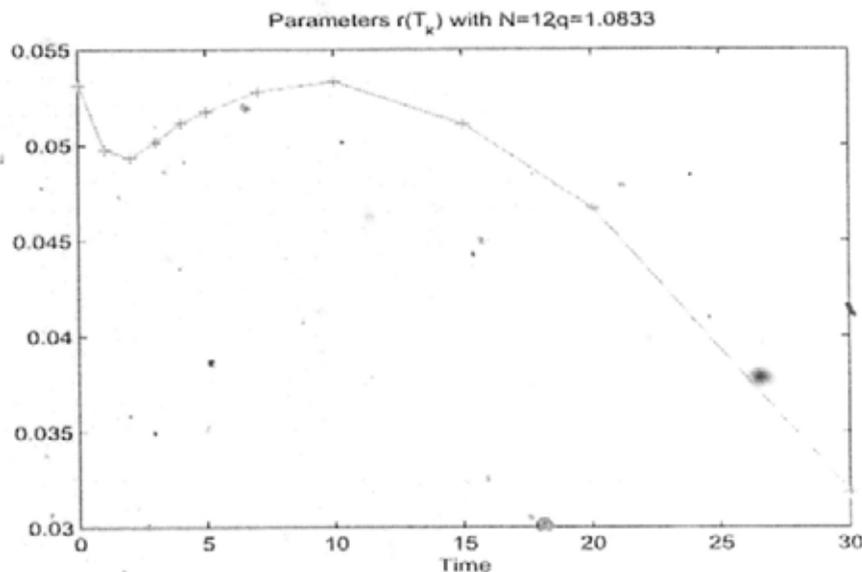


Figure 8.8: Parameters $r(T_k)$ in the maximum Tsallis entropy distribution

distribution approaches the Shannon distribution as the compounding frequency increases.

Figure 8.8 displays $r(T_k)$, which is defined via $r(T_k) = Ce^{-\beta_k T_k}$. Such a definition is made to compare the parameter β_k in the Tsallis case with r_k in the Shannon

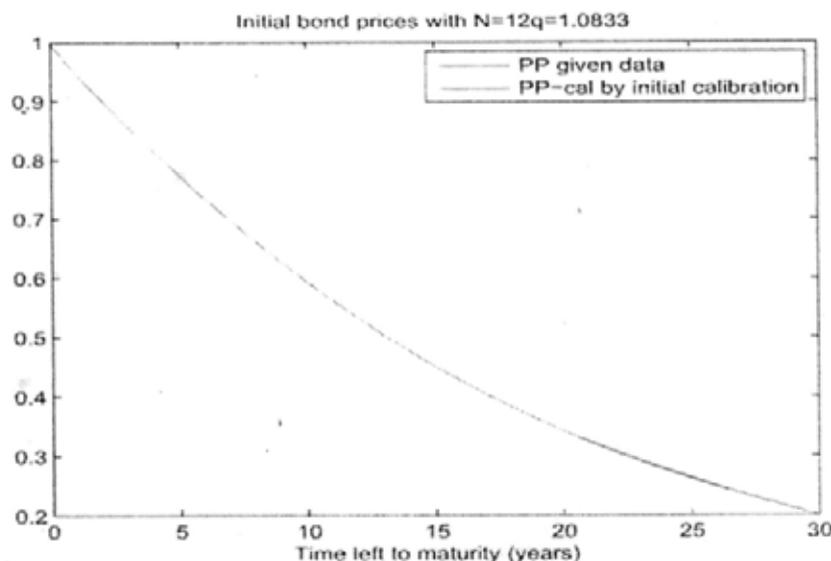


Figure 8.9: Calibrated bond prices with the given data

case such that we can clearly observe how the Tsallis distribution converges to the Shannon distribution as the compounding frequency N increases.

Figure 8.9 demonstrates the initial calibrated values of bonds for different maturities, together with the given data. We observe a perfect match in between, which shows that the calibrated density is a good candidate for the initial term structure.

We record the Tsallis experimental results in the last column in Table 8.5. There $N = 12$ indicates that the bonds we consider in this experiment are compounded monthly. The parameter values $r(T_k) \triangleq Ce^{-\beta k}$ are listed in the middle rows. The last three rows display the test results for the algorithm accuracy. First, the density nature of $\rho(x)$ is confirmed by showing that the normalization condition is “almost” satisfied. Furthermore, the zero difference between $\rho(0)$ and r_0 verifies that $\rho_t(0)$ actually represents the short rate at time t . Finally, the standard error between the calibrated bond prices and the given data proves to be negligible, which further confirms our observation in Figure 8.9.

8.2.3 Connection Between the Two Calibration Algorithms

In Chapter 7 we have proved that the piecewise power-law distribution (derived by maximizing the Tsallis entropy) provides the most general model for the initial term structure density because the power-law exponent N functions as the compounding frequency of bonds. For example, if we are given the prices of monthly compounded bonds, then $N = 12$. Specifically, when the bonds are continuously compounded, i.e. $N \rightarrow \infty$, the power-law distribution would reduce to the ordinary exponential form. To demonstrate this theoretical finding, we perform experiments with compounding frequency of, respectively, 1, 2, and 6 months, i.e. $N = 12, 6$, and 2. The calibrated initial densities with different compounding frequencies are illustrated in Figure 8.10. The key experimental results are displayed in Table 8.5.

In Figure 8.10 the curve for $N \rightarrow \infty$ is calibrated by maximizing the Shannon entropy and corresponds to the continuously compounded bonds. The curves for $N = 2, 6$, and 12 are calibrated by maximizing the Tsallis entropy and correspond to the bonds compounded, respectively, semiannually, bimonthly, and monthly. We observe that for short maturities less than 10 years it is hard to distinguish among the calibrated term structure densities with different compounding frequencies. For long maturities larger than 20 years, the distinction is evident. The semiannually compounded curve is farthest away from the continuously compounded curve. As the compounding frequency N increases, the density curve approaches the continuously compounded curve closer.

Inspecting the data record in Table 8.5, we find that for each maturity T_k the value of $r(T_k)$ (excluding $r(0)$, which is set to be the initial short rate in all cases) in the Tsallis case becomes larger as N increases, although they are always smaller than the value obtained in the Shannon case. Looking at the test results, we observe

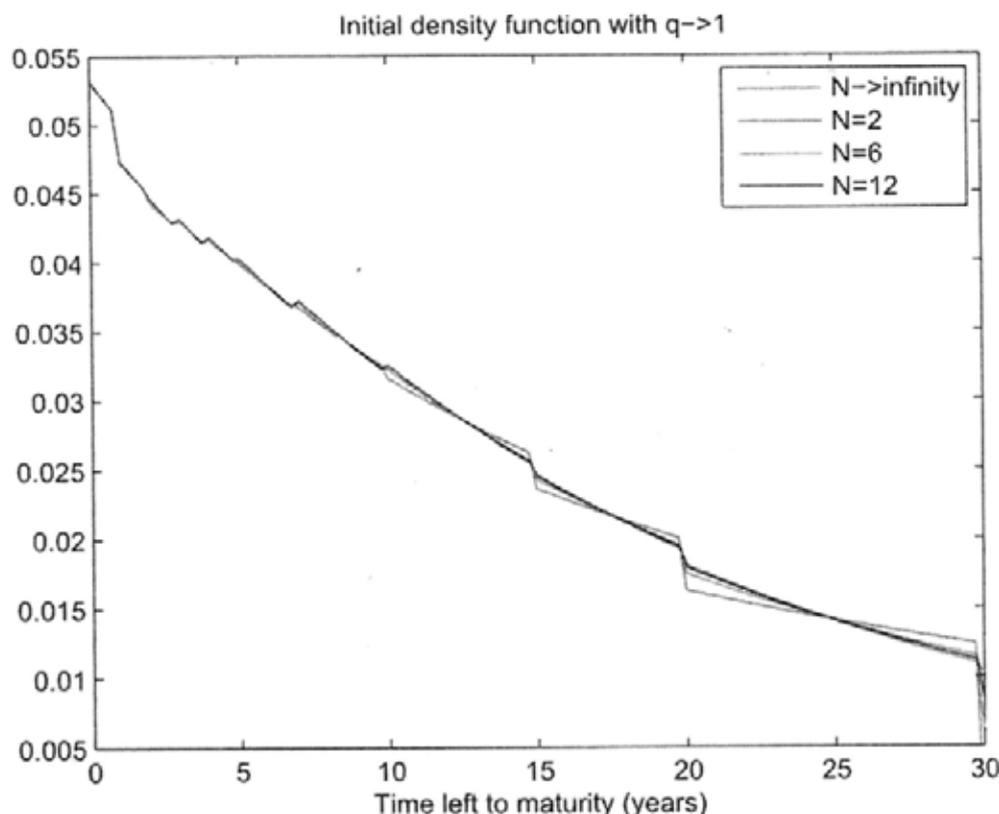


Figure 8.10: Initial term structure density with $N \rightarrow \infty$

that the density integral with $N = 2$ differs a little from the theoretical value one, whereas the integral with $N = 12$ is quite close to one. Moreover, we also find that the standard error between the calibrated bond prices and the given data is reducing as the compounding frequency N increases. Specifically, the smallest error occurs when $N \rightarrow \infty$. All these phenomena verify the relationship between these two calibration algorithms, that the maximum Tsallis entropy distribution converges to the maximum Shannon entropy distribution as the compounding frequency N tends to infinity.

Note that we cannot conclude that the Shannon entropy is a better candidate than the Tsallis entropy for our initial calibration, even though in the Shannon

$\rho(x) =$		Shannon	Tsallis		
		$\sum_{k=0}^n I_{T_k}^{T_{k+1}}(x) r_k e^{-Rx}$	$\sum_{k=0}^n I_{T_k}^{T_{k+1}}(x) \frac{C}{[1 + \frac{\beta_k}{N} + \frac{R}{N}x]^N}$		
Parameters (%)		$N \rightarrow \infty$	$N = 2$	$N = 6$	$N = 12$
Tenor	R	5.0927	5.1359	5.1071	5.0998
0	$r(0)$	5.3108	5.3108	5.3108	5.3108
1	$r(1)$	4.9758	4.9534	4.9685	4.9722
2	$r(2)$	4.9376	4.8919	4.9227	4.9302
3	$r(3)$	5.0243	4.9608	5.0036	5.0140
4	$r(4)$	5.1277	5.0446	5.1007	5.1143
5	$r(5)$	5.1947	5.0622	5.1519	5.1734
7	$r(7)$	5.3166	5.0752	5.2389	5.2781
10	$r(10)$	5.4129	4.8982	5.2478	5.3312
15	$r(15)$	5.2806	4.2136	4.9355	5.1096
20	$r(20)$	5.0219	2.9413	4.3145	4.6681
30	$r(30)$	4.7265	0.0000	1.7520	3.1771
Test					
$\int_0^\infty \rho(x) dx$		0.9998	0.9694	0.9970	0.9991
$\rho(0) - r(0)$		0	0	0	0
$SEE (\times 10^{-4})$		2.2656	15.417	9.3252	5.8213

Table 8.5: Experimental results in the initial calibration

case the density integral is closer to one and the standard error is smaller. This is because the original zero rates given in this experiment are computed using continuously compounding. If the zero rates are calculated using other compounding frequency, for example, semiannually compounding, then the maximum Tsallis entropy distribution with $N = 2$ will be the best candidate. In conclusion, we should select the proper value for N according to the real information on the given rates.

8.3 Model Implementation in the Risk-Neutral World

Given the calibrated initial term structure density, we come to implement the proposed term structure model using initial data in the US swap market for 15 Feb,

2007. In order to test our model improvements over traditional models, we also run the simulation with the Hull-White model. This section is organized as follows:

1. Implement the proposed model in Section 8.3.1 by using the Monte Carlo simulation.
2. Implement the Hull-White model in Section 8.3.2 by constructing the Black-Karasinski tree.
3. Compare these two models in Section 8.3.3 in various major characteristics.

8.3.1 Implementation of the Proposed Model

In this section, we will show how to use the proposed model to predict the future bond prices. Without loss of generality, we start in the risk-neutral world, where the market risk premium is set to be zero. Considering the dynamical equations for $\rho_t(x)$ and r_t (see Proposition 4.1):

$$\begin{aligned} d\rho_t(x) &= \left(r_t \rho_t(x) + \frac{\partial \rho_t(x)}{\partial x} \right) dt + \rho_t(x) \sigma_t(x) dW_t^*, \\ dr_t &= \left(r_t^2 + \frac{\partial \rho_t(x)}{\partial x} \Big|_{x=0} \right) dt + r_t \sigma_t(x) dW_t^*, \end{aligned}$$

we implement the proposed model by the following steps:

1. Calibrate the initial term structure $\rho_0(x)$ by maximizing the Shannon entropy.
 - (a) In this experiment, we choose the Shannon entropy, rather than the Tsallis entropy, as our calibration basis because the original zero rates given in Table 8.2 are computed using continuously compounding.
 - (b) Set $X_{\max} = 300$ years to approximate the infinity maturity and $\Delta x = \frac{3}{12}$ to indicate that the ATMF caps under consideration are options on the 3-month LIBOR forward rate.

2. Trace the evolutions of the density function $\rho_t(x)$ and the short rate r_t .
 - (a) Set $T_{\max} = 30$ years since the initial zero rates are provided with maturities up to 30 years. Set $\Delta t = \frac{3}{12}$ so that we can predict the yield curves on 120 time nodes. Divide each time interval $[t, t + \Delta t]$ into 90 steps so that the evolutionary trajectory is actually daily updated. Pick 1000 paths in Monte Carlo for each simulation step so that there are totally 90×1000 trials for each modeled yield curve at time node $T_i = 3i$ months.
 - (b) Rewrite $\sigma_t(x) = \nu_t(x) - \bar{\nu}_t$ on account of the zero-mean constraint on σ_t . Here $\nu_t(x)$ can be freely specified. For simplicity we assume $\nu(x) = ae^{-bx}$. Although $\nu_t(x)$ is independent of t , the volatility $\sigma_t(x)$ is still updated timely since the expectation $\bar{\nu}_t$ w.r.t ρ_t varies as the term structure density evolves. To begin with, we use $a = 0.1$, $b = 0.15$.
 - (c) Make a 3-dimensional plot of the term structure density $\rho_t(x)$ in Figure 8.11.
 - (d) Plot $\rho_t(0)$ and r_t in the same plot in Figure 8.12 to verify the relationship $\rho_t(0) = r_t$.
 - (e) Verify the normalization condition by calculating the density integral over $[0, \infty)$. The integral results at a set of selective time slots are listed in Table 8.6.
3. Compute the prices of interest rate derivatives.
 - (a) Compute the bond prices $B_t(x)$ at different time t with various maturities $t + x$. Make a 3-dimensional plot of $B_t(x)$ in Figure 8.13.
 - (b) Compute the yields of $B_t(x)$ at different time t with various maturities $t + x$ and make a 3-dimensional plot of it in Figure 8.14.

- (c) Compute the LIBOR forward rates at different time t for time interval $[t + x_{i-1}, t + x_i]$ and make a 3-dimensional plot of it in Figure 8.15.
- (d) Compute the instantaneous forward rates at different time t for a future time $t + x$ and make a 3-dimensional plot of it in Figure 8.16.
4. Auxiliary tests for the initial term structure.
- (a) Calculate the initial values of bonds in three different ways: first by integrating the initial density function, second by interpolating the market data, and third by calibrating with given yields. Plot them in the same plot in Figure 8.17.
- (b) Plot the model computed initial yields together with the market data (already given in Table 8.2) in the same plot in Figure 8.18.
- (c) Implement the finite difference scheme for the initial prices of caps by use of the model computed short rates and forward rates. Plot the resulting cap prices and the Black formula prices (already given Table 8.4) in the same plot in Figure 8.19.
- (d) Compute the flat volatilities implied by the model computed cap prices. Plot them together with the market implied volatilities (already given in Table 8.4) in the same plot in Figure 8.20.
- (e) Compute the sum of squared error (SSE. Also called least squares) between the implied volatilities originally provided and obtained from the model:

$$SSE = \sum_x (\hat{\sigma}_c(x) - \sigma_c(x))^2, \quad (8.7)$$

where $\hat{\sigma}_c$ denotes the model implied volatilities and σ_c denotes the market data given in Table 8.4.

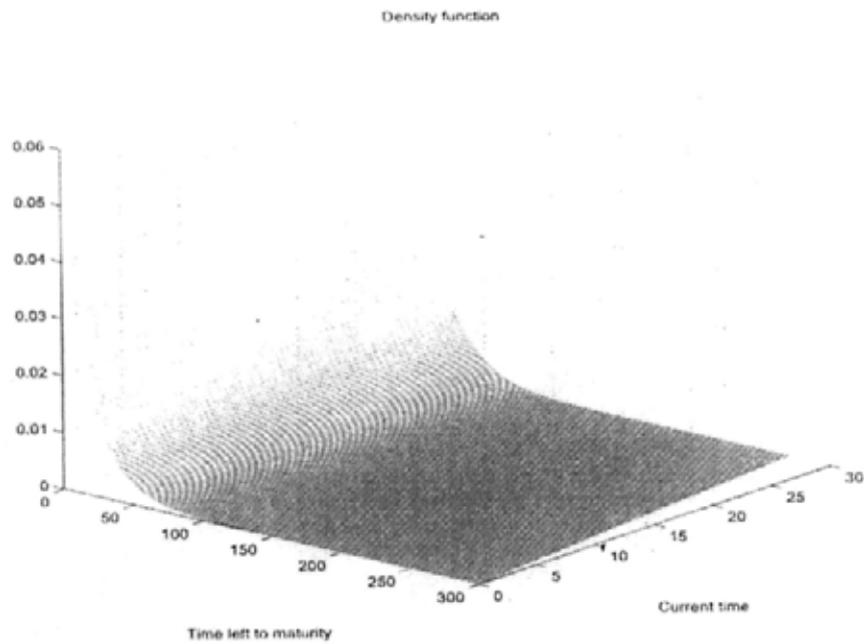
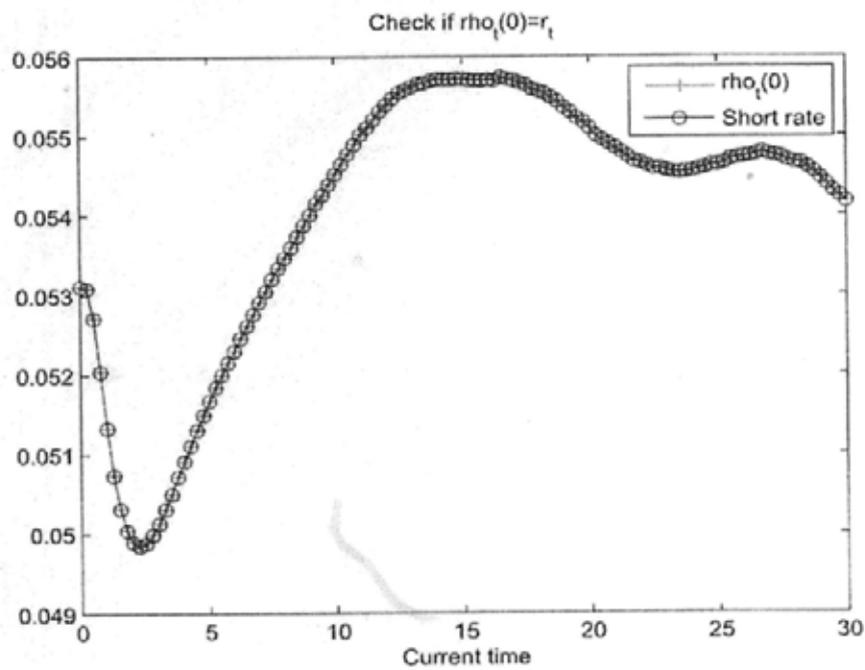


Figure 8.11: Term structure density evolution

Figure 8.12: Test for $\rho_t(0) = r_t$

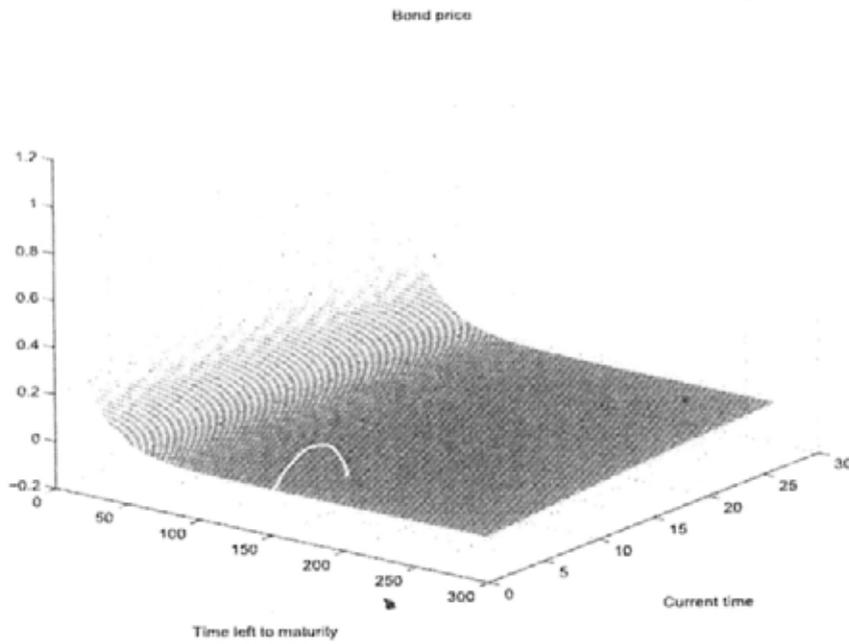


Figure 8.13: Bond price evolution

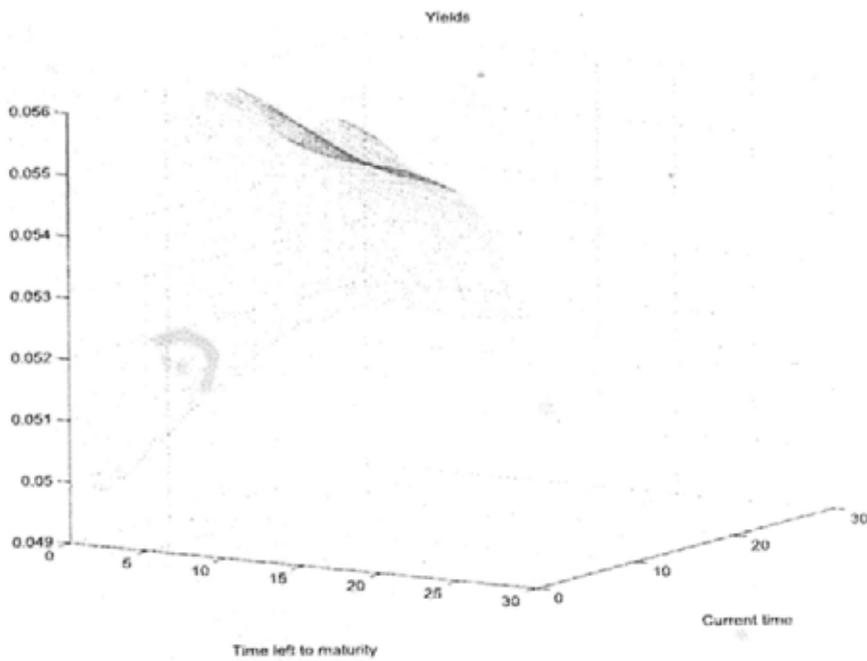


Figure 8.14: Yield evolution

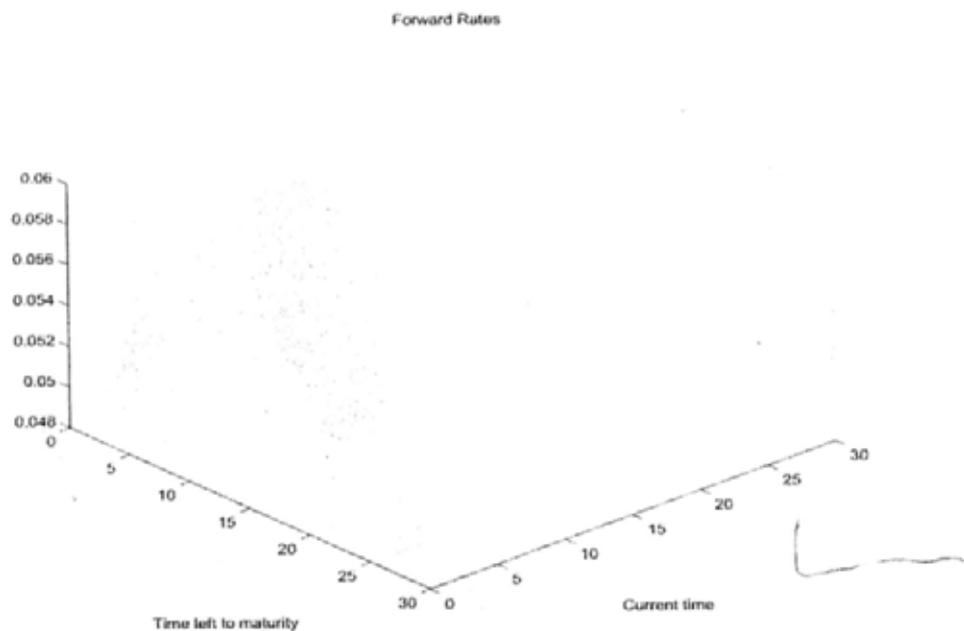


Figure 8.15: LIBOR forward rate evolution

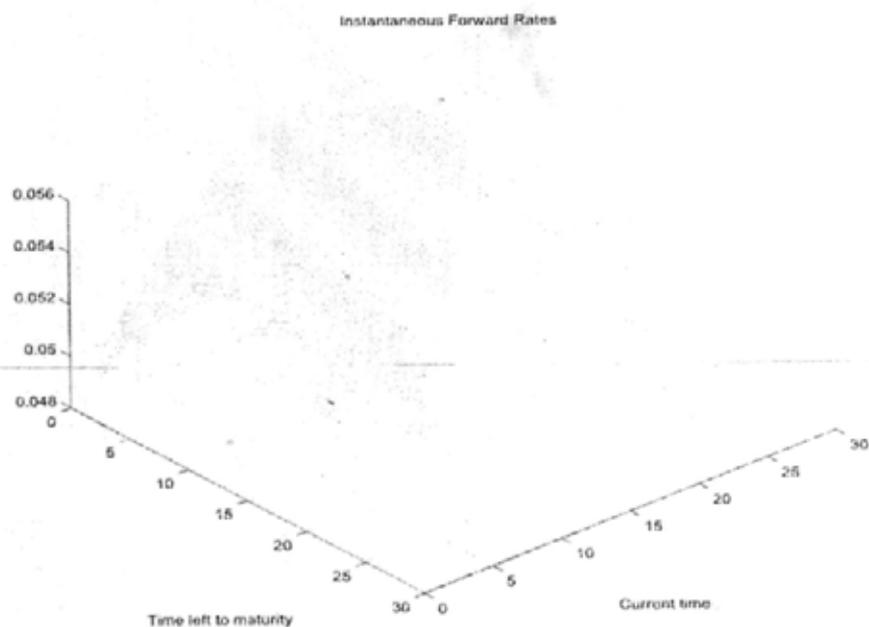


Figure 8.16: Instantaneous forward rate evolution

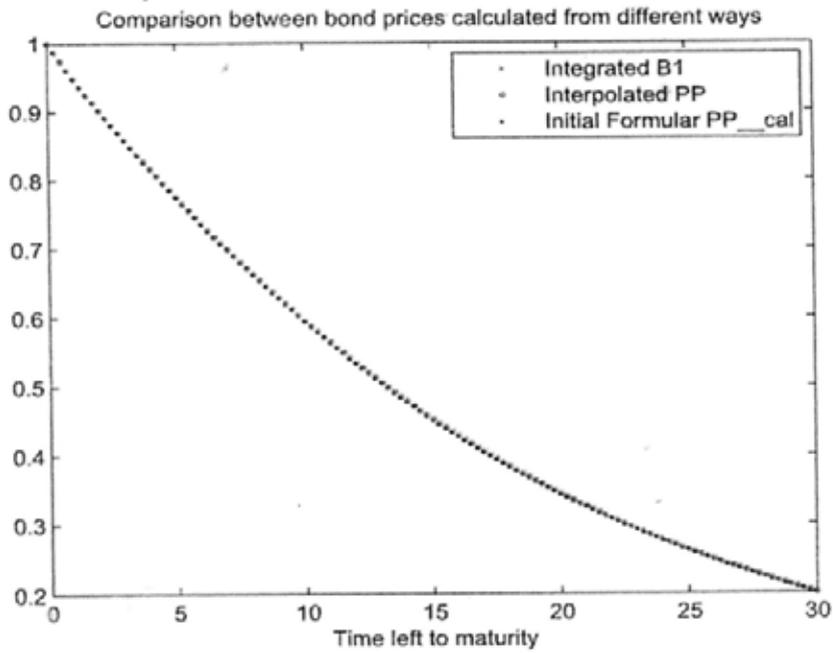


Figure 8.17: Initial bond price comparison

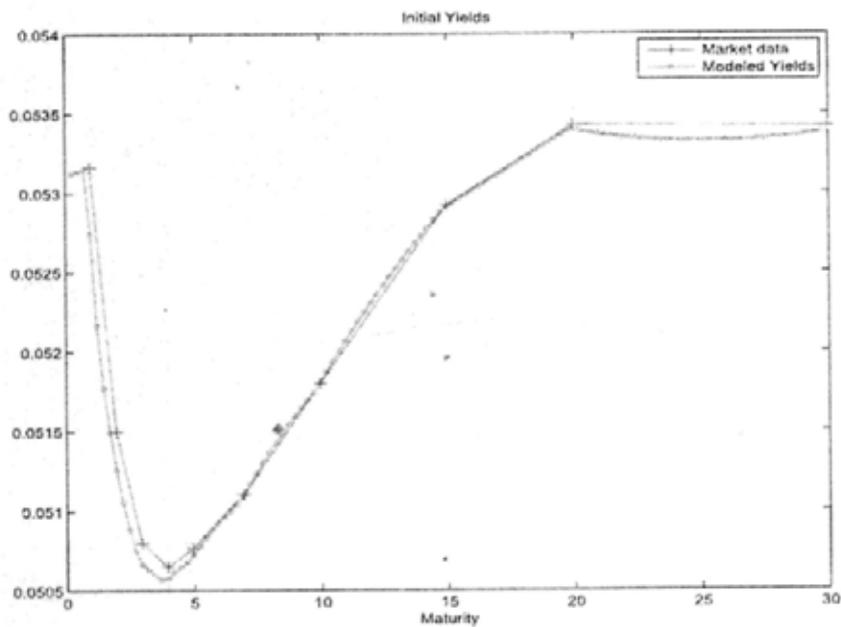


Figure 8.18: Initial yield comparison

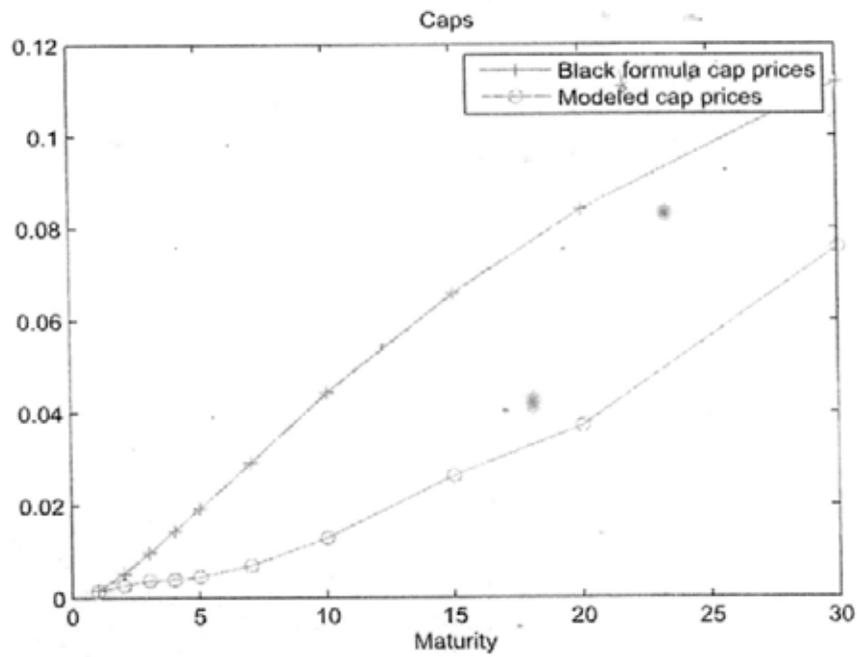


Figure 8.19: Cap price comparison

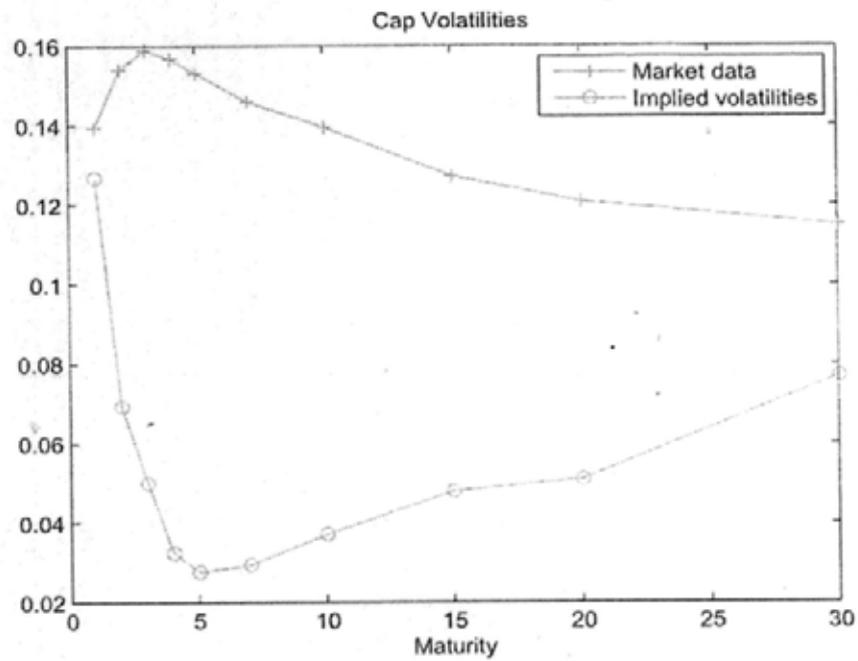


Figure 8.20: Implied volatility comparison

Test of Normalization Condition			
t	$\int_0^{\infty} \rho_t(x) dx$	t	$\int_0^{\infty} \rho_t(x) dx$
1-Year	1.000	7-Year	1.003
2-Year	1.000	10-Year	1.003
3-Year	1.001	15-Year	1.007
4-Year	1.001	20-Year	1.011
5-Year	1.002	30-Year	1.023

Table 8.6: Normalization test

Figures 8.11, 8.12 and Table 8.6 demonstrate the results obtained from Step 2.

The term structure density surface in Figure 8.11 displays the evolution of the density curve for $0 \leq t < 30$ years. At any time t , the curve gives rise to density values with different maturities up to 300 years and is shown to decrease stepwise as maturity lengthens, just as the initial term structure density curve behaves.

Figure 8.12 demonstrates the evolution of the short rate. We observe three things here. First, it verifies the relation $\rho_t(0) = r_t$, namely, that the short rate can be incorporated into the term structure density. Second, we observe mean reversion for the short rate curve. Specifically, the short rate values within 30 years are bounded between the maximum value 0.05573 and the minimum value 0.04985. Therefore, we conclude that the fluctuation of the short rate is bounded within a reasonable range.

Table 8.6 displays the sum of discrete density function values at various selective time, all of which prove to approach the unit theoretical value. Thus we have verified the normalization condition of $\rho_t(x)$.

Figures 8.13-8.16 demonstrate the results obtained from Step 3, for the evolutions of the bond price, the bond yield, the forward rate, and the instantaneous forward rate, respectively. They are all interest rate derivatives so that we can immediately model their processes once the term structure density process or the short rate

process is depicted.

The bond price surface in Figure 8.13 displays the evolution of the bond price curve for $0 \leq t \leq 30$ years. At any time t , the curve gives bond prices with different maturities up to 300 years and is shown to be a decreasing function w.r.t maturity.

Since Figure 8.14, we only display the corresponding values within maturity 30 years since the values for longer maturities make no sense. Considering an individual bond, at time t we would predict its value only with maturities up to $30 - t$ rather than 300 years. That is why we observe upper triangular shapes for the yield surface and the (instantaneous) forward rate surface. We find that the yield curve in Figure 8.14 at an arbitrary time $t < 30$ years still exhibits a partially inverted shape as the initial yield curve shows. The (instantaneous) forward rate curve in Figures 8.15 (Figure 8.16) still increases stepwise just as the initial curve behaves, although the curve is becoming more smooth as time involves.

Figures 8.17-8.20 demonstrate the results obtained from Step 4, exhibiting perfect match between the given market data, for instance, the given initial bond prices, yields, or cap values, and the corresponding model computed values.

In Figure 8.17 we find a perfect fit among the bond prices calculated in three ways: first by integrating the density function, second by interpolating the market data, and third by calibrating with given yields.

In Figure 8.18 we compare the given yields with those model computed yields. A satisfactory match is observed, although there exists a slight difference for short maturities (less than 4 years) and long maturities (more than 20 years). The large difference at the long end is probably caused by the incomplete information on yields after 20 years.

In Figure 8.19 and 8.20 we compare the model computed cap prices or implied cap volatilities with the market data given in Table 8.4. In both figures we find the model computed data do not fit the market data quite well. The smallest difference occurs at the initial time, indicating that our initial calibration is successful. However, since maturity 5 years, the gap is widening significantly and then tends to narrow as maturity approaches 30 years. Calculated results show that the sum of squared error between the model implied volatilities and the market data is as large as 9.344%. This indicates that there will be an error between our predicted values for the future bond prices or interest rates and the market data since in the current finite difference scheme we have used the predicted values. However, we have to point out that the error for bond price or interest rate should be much smaller than the error for cap price since a %1 error in the price of the underlying security may lead to a 25% error in an option price [29]. Such an error between model computed values and market data is natural and expectable since the volatility structure used in our experiment is not carefully chosen.

8.3.2 Implementation of the Hull-White Model

In this section, we will implement the Hull-White model in the risk-neutral world by constructing the Black-Karasinski trinomial tree [31]. Consider the dynamical equation for the short rate process r_t :

$$d \ln r(t) = (\theta(t) - a \ln r(t))dt + \sigma dW_t, \quad (8.8)$$

where $\theta(t) = \dot{F}_{0t} + aF_{0t} + \frac{\sigma^2}{2a}(1 - e^{-2at})$, and F_{0t} denotes the forward interest rate seen at the initial time 0 for time t . To begin with, we assume the constants $a = 0.15$ and $\sigma = 0.25$ in Eqn. (8.8). The model is implemented by the following steps:

1. Construct the Black-Karasinski tree with $\Delta t = \frac{3}{12}$, which is calibrated against

the initial yields given in Table 8.2.

- (a) Display the Black-Karasiniski tree of short rates in Figure 8.21 for the first 13 time nodes, namely, within 3 years.
 - (b) Make a 3-dimensional plot of the short rates r_t in the tree. Show the plot in Figure 8.22.
2. Compute the price of a discount bond maturing in 30 years at each node of the tree.
- (a) Display the Black-Karasiniski tree of bond prices in Figure 8.23 for the first 13 time nodes, namely, within 3 years.
 - (b) Make a 3-dimensional plot of the bond prices $B_t(x)$ in the tree. Show the plot in Figure 8.24.
 - (c) Compare the model computed bond prices $P_{0,30}$ with the market data by calculating the standard error (8.6) in between.
3. Auxiliary tests for the initial term structure.
- (a) Implement the finite difference scheme for the initial prices of caps by use of the short rates in the trinomial tree. Plot the resulting cap prices and the Black formula prices (already given Table 8.4) in the same plot in Figure 8.25.
 - (b) Compute the flat volatilities implied by the model computed cap prices. Plot them together with the market implied volatilities (already given in Table 8.4) in the same plot in Figure 8.26.
 - (c) Compute the SSE, defined by Eqn. (8.7), between the implied volatilities originally provided and obtained from the model.

4. Calibrate the parameters a and σ by minimizing the SSE between the model implied cap volatilities and the given market data.
 - (a) Redo steps 1 and 3 with different values of the parameters a and σ . Search the optimal pair (a, σ) such that the SSE defined via Eqn. (8.7) reaches the minimum.
 - (b) Plot the model computed cap prices after calibration in Figure 8.27, together with the market data.
 - (c) Plot the model implied cap volatilities after calibration in Figure 8.28, together with the market data.

Node/year	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75	3
1	0.1568	0.1557	0.1547	0.1538	0.1469	0.1458	0.1408	0.1378	0.1411	0.1397	0.1383	0.1370	0.1407
2	0.1263	0.1254	0.1246	0.1239	0.1183	0.1158	0.1134	0.1110	0.1136	0.1125	0.1114	0.1103	0.1133
3	0.1017	0.1010	0.1003	0.0998	0.0953	0.0932	0.0913	0.0894	0.0915	0.0906	0.0897	0.0888	0.0913
4	0.0819	0.0813	0.0808	0.0804	0.0767	0.0753	0.0735	0.0720	0.0737	0.0729	0.0722	0.0715	0.0735
5	0.0660	0.0655	0.0651	0.0647	0.0618	0.0605	0.0592	0.0580	0.0595	0.0587	0.0582	0.0576	0.0592
6	0.0531	0.0527	0.0524	0.0521	0.0498	0.0487	0.0477	0.0467	0.0478	0.0475	0.0468	0.0464	0.0477
7	0.0428	0.0425	0.0422	0.0420	0.0401	0.0392	0.0384	0.0376	0.0385	0.0383	0.0377	0.0374	0.0384
8	0.0345	0.0342	0.0340	0.0338	0.0323	0.0316	0.0309	0.0303	0.0310	0.0307	0.0304	0.0301	0.0309
9	0.0277	0.0275	0.0274	0.0272	0.0260	0.0254	0.0249	0.0244	0.0250	0.0247	0.0245	0.0242	0.0249
10	0.0223	0.0222	0.0220	0.0219	0.0209	0.0205	0.0201	0.0196	0.0201	0.0199	0.0197	0.0195	0.0201
11	0.0180	0.0179	0.0178	0.0177	0.0169	0.0165	0.0162	0.0159	0.0162	0.0160	0.0159	0.0157	0.0161

Figure 8.21: Black-Karasinski short rate tree

Figures 8.21 and 8.22 demonstrate the results obtained from Step 1, by constructing the Black-Karasinski tree. The Black-Karasinski tree is a trinomial interest rate tree. At each time node, the short rate can increase (move up), maintain (keep unchanged), or decrease (fall down) with specified probabilities. Here we use forward induction to construct the interest rate tree and display partial values of the short rate in Figure 8.21 as an illustration. The nodes labeled with small numbers

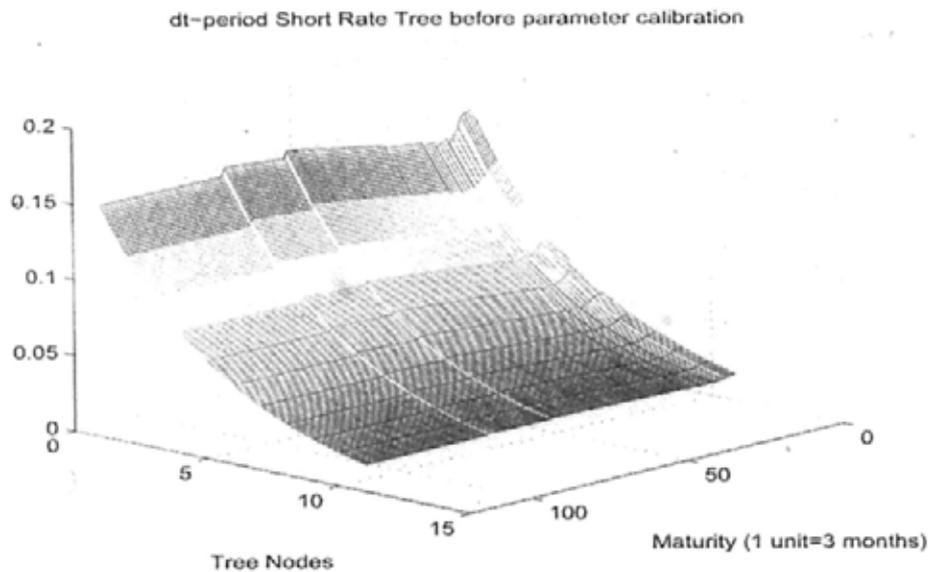


Figure 8.22: dt Period short rates before parameter calibration

indicate the growing direction of interest rates, whereas the large number nodes indicate the declining direction. We find that the resulting interest rates are bounded between the maximum value 0.159 and the minimum 0.0157. The short rate surface in Figure 8.22 displays the evolution of the short rate curve for $0 \leq t \leq 30$ years.

Node/year	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75	3
1	0.1244	0.1263	0.1283	0.1302	0.1322	0.1340	0.1357	0.1374	0.1389	0.1406	0.1422	0.1439	0.1454
2	0.1404	0.1425	0.1446	0.1468	0.1490	0.1510	0.1529	0.1547	0.1565	0.1584	0.1603	0.1621	0.1639
3	0.1562	0.1585	0.1608	0.1631	0.1655	0.1676	0.1698	0.1718	0.1739	0.1760	0.1781	0.1802	0.1822
4	0.1717	0.1742	0.1766	0.1791	0.1816	0.1840	0.1863	0.1886	0.1909	0.1932	0.1956	0.1979	0.2002
5	0.1868	0.1894	0.1921	0.1947	0.1974	0.2000	0.2025	0.2050	0.2075	0.2100	0.2126	0.2152	0.2177
6	0.2014	0.2042	0.2070	0.2098	0.2127	0.2154	0.2181	0.2208	0.2235	0.2263	0.2291	0.2319	0.2347
7	0.2155	0.2184	0.2214	0.2244	0.2274	0.2303	0.2332	0.2361	0.2390	0.2420	0.2450	0.2481	0.2511
8	0.2291	0.2321	0.2352	0.2384	0.2415	0.2447	0.2478	0.2509	0.2540	0.2572	0.2604	0.2636	0.2669
9	0.2420	0.2453	0.2485	0.2518	0.2551	0.2584	0.2617	0.2650	0.2683	0.2717	0.2751	0.2785	0.2820
10	0.2545	0.2578	0.2612	0.2647	0.2681	0.2716	0.2750	0.2785	0.2820	0.2856	0.2892	0.2928	0.2965
11	0.2664	0.2699	0.2734	0.2770	0.2806	0.2842	0.2878	0.2915	0.2951	0.2989	0.3026	0.3065	0.3103

Figure 8.23: Black-Karasinski bond price tree

Figures 8.23 and 8.24 demonstrate the results obtained from Step 2, specifically displaying the bond prices in the Black-Karasinski tree and plotting them in a 3-

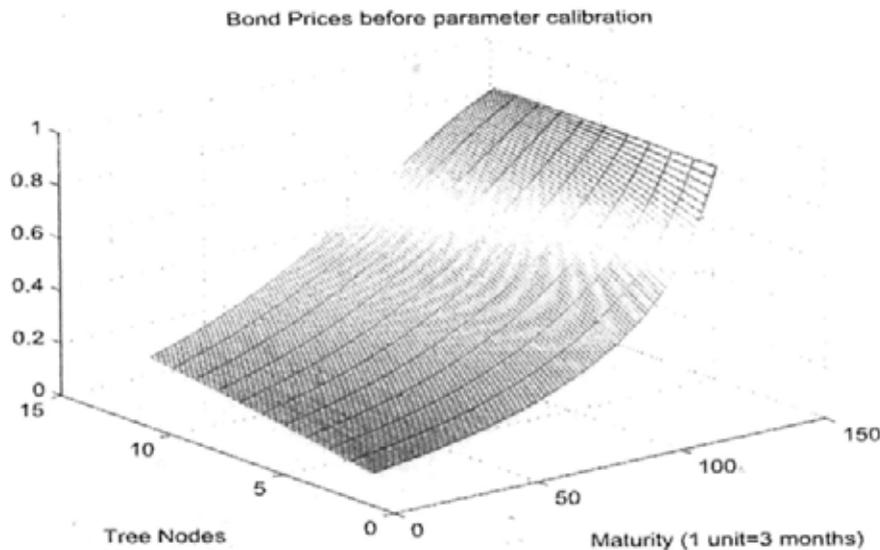


Figure 8.24: Discount bond prices before parameter calibration

dimensional graph. In Figure 8.23, the point at the intersection of the middle node of the tree (node 6) and the initial time (Year 0) gives the initial price $P_{0,30}$ of the discount bond maturing at the end of 30 years. Comparing the model computed data with the market data indicated by the zero rates in Table 8.2, we calculate the absolute error to be 7.1524×10^{-6} (relative error is 3.5511×10^{-5}), a quite small difference in between. In Figure 8.24 we find that the bond prices at all maturing nodes (at time 120 months as shown in the figure) tend to 1. The initial bond prices, especially those at node 1 with high interest rates, tend to be vanishing since the time there left to maturity is quite long.

Figures 8.25 and 8.26 demonstrate the results obtained from Step 3. In these two figure we compare the model computed cap prices or implied cap volatilities before parameter calibration with the market data given in Table 8.4. In either figure we observe a large difference in between. Moreover, the SSE between the model implied volatilities and the market data proves to be as large as 2.175%.

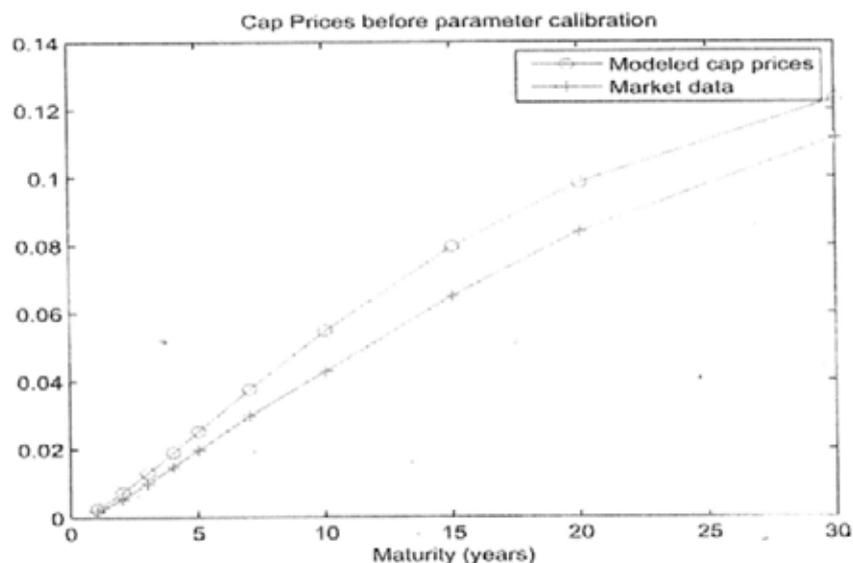


Figure 8.25: Cap prices before parameter calibration

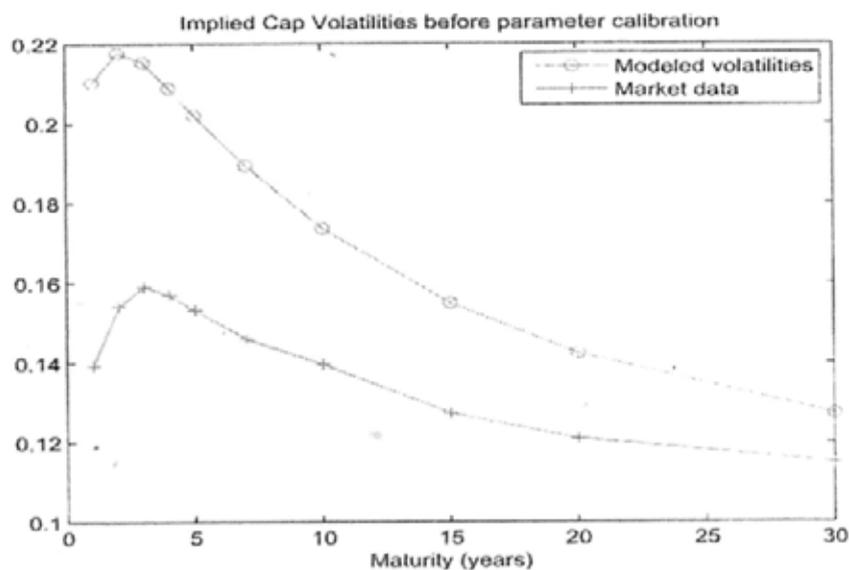


Figure 8.26: Implied cap volatilities before parameter calibration

In order to minimize the difference between the model implied volatilities and the market data (equivalently to minimize the SSE), we use `fminsearch` function in Matlab and deduce the optimal values for a and σ :

$$a = 0.06906, \quad \sigma = 0.17199.$$

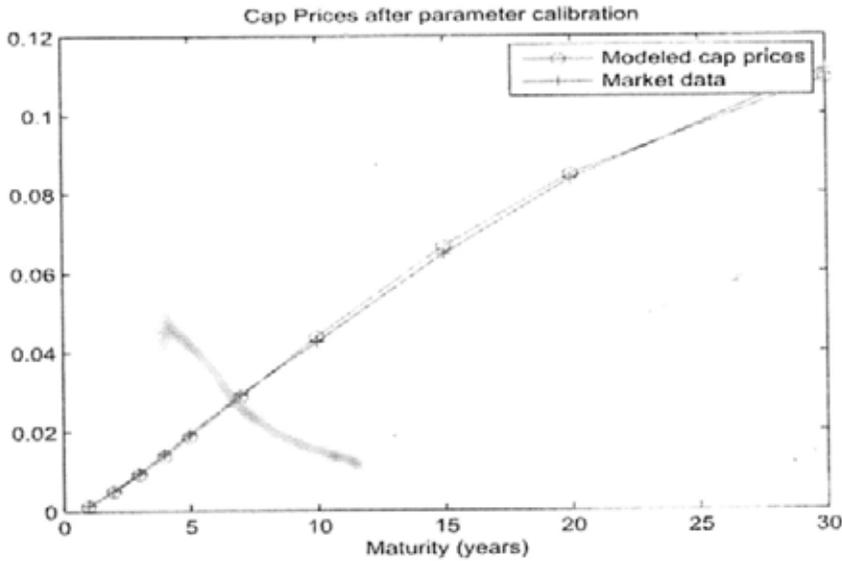


Figure 8.27: Cap prices after parameter calibration

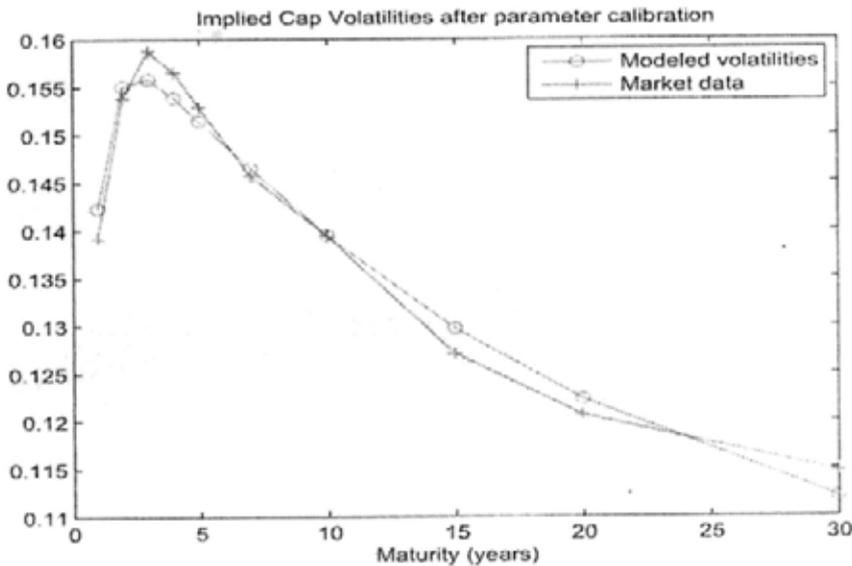


Figure 8.28: Implied cap volatilities after parameter calibration

Then we find the SSE immediately diminishes to be 5×10^{-5} . Figure 8.27 and 8.28 demonstrate the comparison after the parameter calibration. The model implied volatilities, as shown in Figure 8.28, fit the market data perfectly for most time, except the slight difference around 5 years and 15 years.

8.3.3 Comparison and Discussion

Table 8.7 unfolds a comparison between the proposed model and the Hull-White model in 13 major characteristics. We will conduct a discussion in the below on the advantages of each model. For easier reading, the item number before each detailed statement corresponds to the number in Table 8.7.

A. These two models have the following similarities:

1. Both are no-arbitrage models.
2. Both automatically fit the initial term structure.
4. Both exhibit mean reversion for the modeled short rate evolution.

B. The newly proposed model owns the following advantages over the Hull-White model:

3. The new model needs to specify fewer external processes. It is mainly controlled by two structures only: the initial term structure and the volatility structure. However, the Hull-White model is also dominated by the dynamics of the short rate.
5. The new model guarantees interest rate positivity, whereas the Hull-White model fails.
8. The short rates resulting from the new model evolve with less fluctuation than those from the Hull-White model.
9. Compared with the Hull-White model, the new model leads to less relative error between the model computed bond prices and the market data.

C. The Hull-White model owns the following advantages over the newly proposed model:

	Hull-White Model	New Model	Comments
1. Model Type	No-arbitrage model	No-arbitrage model	
2. Initial Term Structure	Input	Input	
3. Degrees of Freedom	$r_t, P_0(t), \sigma$	$P_0(t), \sigma$	
4. Mean Reversion	Yes	Yes	
5. Positive Definite	No	Yes	
6. Numerical Scheme	Trinomial tree	Monte Carlo	
7. Short Rates	Δt -rates	Instantaneous rates	
8. Short Rate Range	[0.0092, 0.2838] (after cal)	[0.04985, 0.05573]	less fluctuation for new model
9. Relative Error of Bond	6.73×10^{-5} (after cal)	0.13×10^{-5}	less error for new model
10. SSE of Cap Volatility	2.175% (before cal); 0.005% (after cal)	9.344%	less SEE for HW model
11. Parameter Calibration	para cal required to min the SSE of implied cap vol		
12. Dependence on σ	σ affects the accuracy; but not availability	Inappropriate vol structure even leads to an "explosion" and unavailability	more sensitive for new model
13. Time Needed	Less than 3 min.	1h 32min.	time-consuming for new model

Table 8.7: Comparison of the proposed model and the Hull-White model

- 6, 7, 13. The most noteworthy advantage of the Hull-White model lies in the implementation speed. Since the Hull-White model is manipulated by constructing a trinomial interest rate tree, the model implementation takes only a few minutes. Even when the parameter calibration is performed, the model implementation takes no longer than 3 minutes. On the other hand, the manipulation of the proposed model proves to be quite time-consuming — around 1 hour and 30 minutes for each run. This is because here we apply the Monte Carlo simulation. The more trials we use in the simulation, the longer it takes to implement the model.
10. After the parameter calibration, the model implied cap volatilities in the Hull-White model fit the market data perfectly.
- 11-12. Comparing the results obtained before and after the parameter calibration in the Hull-White model, we find that the optimal volatility could improve the model accuracy. Whereas in the proposed model an arbitrarily specified volatility would lead to unpalatable results. After trying several forms for the volatility structure, we find that an inappropriate volatility may even result in an “explosion” during the implementation and leads the model to fail to work. In this sense, the new model has stronger dependence on the volatility structure.

Above discussion shows that each model has its own merits and limitations. In order to explore more advantages of the proposed model and its availability in practice, we need to further refine our designs for the external processes. For example, we should choose a more proper form for the volatility structure to precisely reflect the volatility of market. Besides, a term structure model is most often implemented in a non-risk-neutral world, where the market risk premium should be taken up.

However, in the financial industry the value of the risk premium is generally not assigned by practitioners but rather calibrated with the market information. That is another research topic and deserves further study.

Chapter 9

Conclusions and Future Works

9.1 Conclusions

Following the initial study of Brody and Hughston on applying information geometry to interest rate modeling, we have proposed a novel term structure model and investigated its application in the US swap market. Different from the traditional term structure models that impose assumptions on either bonds or rates, the newly proposed model is characterized by the evolution of a density function which is obtained from the derivative of the discount function with respect to the time left till maturity. We have proved that such a density function can be interpreted as interest return on the discount bond.

The introduction of the term structure density turns the problem of term structure analysis into a problem of statistical study on probability distributions. In our research we have tackled two major problems:

1. **Dynamical Problem** — we have characterized the evolutionary trajectory of the term structure by developing the density dynamics;
2. **Distance Problem** — we have depicted the distance evolution for a pair of yield curves.

To fix the Dynamical Problem, we have followed three steps: calibrated the initial term structure density, specified the market risk premium, and chosen a volatility structure.

First, we have introduced two initial calibration methods, one by maximizing the Tsallis entropy and the other by the notion of superstatistics. With the entropic method, the explicit form of the term structure density function proves to be a piecewise power-law (Pareto) function, parameterized by $1 - q$. The entropy index q here, which is a well-known physics quantity, now finds its financial interpretation as the measure of departure of the current term structure from flatness on a continuously compounded basis. Our empirical experiments in the US swap market have fully demonstrated this observation. Moreover, we have proved that the piecewise power-law distribution provides the most general model for the initial term structure density since the power-law exponent acts as the compounding frequency of the observed bonds. With the superstatistics method, we have initially supposed that the term structure in a short term is flat associated with a constant continuously compounded rate, and further assumed that the rate follows a χ^2 -distribution. As a consequence, the whole term structure is regarded as a superposition of local flat structures, and proves to follow the same power-law distribution as the entropic method indicates if the only source of information available is the existence of a perpetual annuity.

Next, given the calibrated initial term structure density, we have developed the term structure dynamics in the risk-neutral world and proved that the evolution is fully and only determined by the initial density and the volatility structure. By use of a density volatility that possesses zero mean, we have deduced a concise martingale representation for the bond pricing formula.

Finally, as an illustration of the importance of volatility structure in term structure analysis, we have redesigned the HJM volatilities for interest rate positivity under the framework of the current model. Besides, for convenience of simulation, we have also developed the dynamics of the HJM bond volatility and ours, both under the current framework.

In the detailed study of the Distance Problem, after mapping yield curves to density functions on a Hilbert space, we have used the notion of information geometry to define a metric on the Hilbert space. Thus the difference between a pair of yield curves is measured by the Bhattacharyya spherical distance between the term structure densities. We have proved that two yield curves with large initial Bhattacharyya spherical distance in the risk-neutral world would diverge from each other with a significant probability.

Finally, we have implemented the proposed model with initial data in the US swap market for 15 Feb, 2007. To test our model improvements over traditional models, we have also run the simulation with the Hull-White model and compared these two no-arbitrage models in various major characteristics. It shows that each model has its own advantages and disadvantages. For example, the proposed model would not perfectly fit the market quotes for implied cap volatilities if the volatility structure is not carefully chosen.

9.2 Future Works

The followings are possible directions of future research.

1. In the theoretical study of the model, we attempt to
 - (a) impose a constraint on the initial condition of term structure such that the trajectories associated with two yield curves will converge as time

involves.

Our research on the Distance Problem is preliminary and deserves further study. In the present research, the yield curves are simply assumed to be flat at the beginning and their evolutions are governed by the same volatility structure and market risk premium. We have only pointed out the conditions under which the yield curves tend to diverge rather than converge. However, our aim is to bound the initial distance such that the subsequent evolving yield curves would eventually converge as time passes. If such a bound exists, we can conclude that the evolution of the yield curve would be indifferent from the initial calibration error.

- (b) explore the connections with various existing term structure models, especially the HJM model and the LIBOR model.

The HJM model focuses on the evolution of the instantaneous forward rate and the LIBOR model on the forward rate. Given the equivalence between the term structure density and bond price or interest rate, we could link the traditional term structure models together and fix their drawbacks, just as what we have done in Chapter 5 to the HJM model volatility for interest rate positivity under the framework of the proposed model.

- (c) use the Tsallis entropic method to calibrate the risk-neutral distribution for the terminal price of an asset.

Brody and Buckley [20] have explored an application of entropy maximization in the field of asset pricing. Consider a situation where we are given the initial value S_0 of an asset, together with the present value C_0 of an option on this asset with strike price K_0 and maturity T . To avoid

arbitrage opportunities, the following equations must hold:

$$\int_0^{\infty} x\rho(x) dx = S_0e^{rT}, \quad (9.1)$$

$$\int_0^{\infty} (x - K_0)^+ \rho(x) dx = C_0e^{rT}. \quad (9.2)$$

Here x denotes the terminal price of the underlying asset, r is the risk-free constant rate, and $\rho(x)$ is the risk-neutral probability density function for x . Our aim is to evaluate another option C_K on the same underlying asset but with a different strike price K :

$$\int_0^{\infty} (x - K)^+ \rho(x) dx = C_Ke^{rT}. \quad (9.3)$$

Brody and Buckley [20] have tackled this problem by calibrating the risk-neutral density $\rho(x)$ based on the Rényi entropy maximization. However, as pointed out in Chapter 7, the Rényi entropy is unstable in the sense that a small deformation of the distribution would lead to a great change of the corresponding entropy value. Therefore, the Rényi entropy cannot reproduce experimentally observable quantities. On the contrary, the Tsallis entropy is stable and this inspires us to calibrate the risk-neutral density $\rho(x)$ by maximizing the Tsallis entropy. Study in this direction is in progress and deserves further experimental support.

- (d) extend our study to coupon-bearing bonds.
2. In the practical study of the model, we will focus on the identification of external processes such as the volatility structure and the market risk premium.
- (a) Implement our proposed model in non-risk-neutral world.

In this thesis, our implementation of the proposed model is conducted only in the risk-neutral world for simplicity, where the market risk pre-

mium is immaterial. In practice, however, the risk premium process should be modeled and calibrated with available market data.

- (b) Update the volatility structure according to the rule we propose in Proposition 5.3.

In our model implementation, we have used a time-independent volatility structure for simplicity. However, the volatility structure should promptly reflect the updated market information. Proposition 5.3 provides a way to update the volatility structure timely and precisely based on the current term structure in real markets. It is therefore desirable to employ the time-varying volatility structure in practice.

- (c) Depict the distance evolution in numerical experiments.

Although our theoretical study on the distance evolution issue is concerned with simple term structures — the initial yield curves are assumed to be flat and their subsequent evolutions are governed by the same volatility structure and market risk premium, the distance evolution for complex term structures can be illustrated in numerical experiments. The experiments may provide inspiration in return for our further theoretical study.

Bibliography

- [1] <http://en.wikipedia.org/wiki/History-of-entropy>.
- [2] <http://www-history.mcs.st-and.ac.uk/Biographies/Clausius.html>.
- [3] Tsallis Statistics, Statistical Mechanics for Non-extensive Systems and Long-Range Interactions. Technical report, Center for the Study of Complex Systems, 2007. <http://www.cscs.umich.edu/~crshalizi/notabene/tsallis.html>.
- [4] A. Rényi. On Measures of Information and Entropy. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 547–561, 1960.
- [5] A. Rényi. *Probability Theory*. North-Hollan, Amsterdam, 1970.
- [6] A. Reynolds. Superstatistical Mechanics of Tracer-Particle Motions in Turbulence. *Physical Review Letters*, 91:084503, 2003.
- [7] A.K. Rajagopal and S. Abe. Implications of Form Invariance to the Structure of Nonextensive Entropies. *Physical Review Letters*, 83(9):1711–1714, 1999.
- [8] B. Castaing, Y. Gagne, and E.J. Hopfinger. Velocity probability density functions of high Reynolds number turbulence. *Physica D: Nonlinear Phenomena*, 46(2):177–200, 1990.

- [9] C. Beck. Dynamical foundations of nonextensive statistical mechanics. *Physical Review Letters*, 87:180601, 2001.
- [10] C. Beck. Lagrangian acceleration statistics in turbulent flows. *Europhysics Letters*, 64(2):151–157, 2003.
- [11] C. Beck. Superstatistics: Recent developments and applications. In C. Beck, G. Benedek, A. Rapisarda, and C. Tsallis, editor, *Complexity, Metastability and Nonextensivity*. World Scientific, 2005. arXiv:cond-mat/0502306.
- [12] C. Beck and E.G.D. Cohen. Superstatistics. *Physica A: Statistical Mechanics and its Applications*, 322:267–275, 2003.
- [13] C. Beck and F. Schlögl. *Thermodynamics of Chaotic Systems: An Introduction*. Cambridge University Press, Cambridge, 1993.
- [14] C. Tsallis. Possible Generalization of Boltzmann-Gibbs Statistics. *Journal of Statistical Physics*, 52:479–487, 1988.
- [15] D. Health, R. Jarrow, and A. Morton. Bond Pricing and the Term Structure of Interest Rate: A New Methodology for Contingent Claims Valuation. *Econometrica*, 60(1):77–105, 1992.
- [16] D. Mindell, J. Segal, S. Gerovitch. From Communications Engineering to Communications Science: Cybernetics and Information Theory in the United States, France, and the Soviet Union. In *Science and Ideology: A Comparative History*, pages 66–95. Routledge, London, 2003.
- [17] D.C. Brody and L.P. Hughston. Applications of Information Geometry to Interest Rate Theory. In P. Sollich *et al.*, editor, *Disordered and Complex Systems*, volume 553, pages 281–287. AIP Conference Proceedings, 2001.

- [18] D.C. Brody and L.P. Hughston. Interest rates and information geometry. In *Proceedings of the Royal Society of London, Series A*, volume 457, pages 1343–1363, 2001.
- [19] D.C. Brody and L.P. Hughston. Entropy and information in the interest rate term structure. *Quantitative Finance*, 2:70–80, 2002.
- [20] D.C. Brody, I.R.C. Buckley, and I.C. Constantinou. Option price calibration from Rényi entropy. *Physics Letters A*, 366:298–307, 2007.
- [21] E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.
- [22] E.T. Jaynes. Information Theory and Statistical Mechanics. *Physical Review*, 106(4):620–630, 1957.
- [23] E.T. Jaynes. Information Theory and Statistical Mechanics II. *Physical Review*, 108(2):171–190, 1957.
- [24] E.T. Jaynes. Gibbs vs Boltzmann Entropies. *American Journal of Physics*, 33(5):391–398, 1965.
- [25] F.C. Klebaner. *Introduction to Stochastic Calculus With Applications*. Imperial College Press, London, 1998.
- [26] F.D. Jong, J. Driessen, and A. Pelsser. LIBOR and Swap Market Models for the Pricing of Interest Rate Derivatives: An Empirical Analysis. Technical Report 2000-35, Center for Economic Research, 2000.

- [27] F.D. Jong, J. Driessen, and A. Pelsser. On the Information in the Interest Rate Term Structure and Option Prices. *Review of Derivatives Research*, 7:99–127, 2004.
- [28] J. Havrda and F. Charvát. Quantification Method of Classification Processes. *Kybernetika*, 3:30–35, 1967.
- [29] J. Hull. *Options, Futures, and Other Derivatives*. Prentice Hall, 5th edition, 2003.
- [30] J. Hull and A. White. Pricing Interest-Rate-Derivative Securities. *Review of Financial Studies*, 3(4):573–592, 1990.
- [31] J. Hull and A. White. Numerical Procedures for Implementing Term Structure Models I: Single-Factor Models. *Journal of Derivatives*, 2(1):7–16, 1994.
- [32] J.C. Cox, J.E. Ingersoll, and S.A. Ross. A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–407, 1985.
- [33] K.E. Daniels, C. Beck and E. Bodenschatz. Defect turbulence and generalized statistical mechanics. *Physica D: Nonlinear Phenomena*, 193:208–217, 2004.
- [34] L. Boltzmann. *Lectures on Gas Theory*. University of California Press, Berkeley, 1964.
- [35] M. Musiela. Stochastic PDEs and term structure models. *Journées Internationales de Finance, IGR-AFFI*, La Baule, 1993.
- [36] N. Wiener. *Cybernetics: or Control and Communication in the Animal and the Machine*. The MIT Press, MA, 1948.

- [37] O.A. Vasicek. An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5(2):177–188, 1977.
- [38] P.C. Mahalanobis. On tests and measures on group divergence. *Journal of the Asiatic Society of Bengal*, 26:541–588, 1930.
- [39] P.C. Mahalanobis. On the generalised distance in statistics. In *Proceedings of the National Institute of Sciences of India, Series A*, volume 2, pages 49–55, 1936.
- [40] P.T. Landsberg and V. Vedral. Distributions and channel capacities in generalized statistical mechanics. *Physics Letters A*, 247(3):211–217, 1998.
- [41] R. Clausius. On the Application of the Mechanical Theory of Heat to the Steam-Engine. In *The Mechanical Theory of Heat – with Its Applications to the Steam-Engine and to Physical Properties of Bodies*. John van Voorst, London, 1865.
- [42] R. Rendleman and B. Batter. The Pricing of Options on Debt Securities. *Journal of Financial and Quantitative Analysis*, 15(1):11–24, 1980.
- [43] R.S. Ingarden. Information geometry in functional spaces of classical and quantum finite statistical systems. *International Journal of Engineering Science*, 19:1609–1633, 1981.
- [44] S. Abe. Stability of Tsallis Entropy and Instabilities of Rényi and Normalized Tsallis Entropies: A basis for q -exponential distributions. *Physical Review E*, 66:046143, 2002.
- [45] S. Abe and A.K. Rajagopal. Microcanonical foundation for systems with power-law distributions. *Journal of Physics A*, 33:8733–8738, 2000.

-
- [46] S. Abe and Y. Okamoto. *Nonextensive Statistical Mechanics and Its Applications*, volume 560 of *Lecture Notes in Physics*. Springer-Verlag, Heidelberg, 2001.
- [47] S. Kullback and R.A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [48] Shun-ichi Amari. *Differential-geometrical methods in statistics*. Springer, 1990.
- [49] Shun-ichi Amari. *Methods of information geometry*. American Mathematical Society, 2000.
- [50] Sueli I. R. Costa, Sandra A. Santos, João E. Strapasson. Fisher Information Matrix and Hyperbolic Geometry. In M.J. Dinneen, editor, *Proceeding of IEEE ISOC ITW2005 on Coding and Complexity*, pages 34–36, 2005.
- [51] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.
- [52] T.S.Y. Ho and S.B. Lee. Term Structure Movements and Pricing Interest Rate Contingent Claims. *Journal of Finance*, 41(5):1011–1029, 1986.