INTEREST RATE MODELING AND A TIME SERIES MODEL FOR FUNCTIONAL DATA

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF STATISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Chung Kwan Pong
August 2010

This dissertation is online at: http://purl.stanford.edu/nv255rn6736

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Tse Lai, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Paul Switzer

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Guenther Walther

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

In finance, an interest rate derivative is a financial instrument where the underlying asset is an interest rate at which payments are made based on a notional amount. A common approach to price interest rate derivatives is through the use of interest rate models. However, a drawback with this approach is that calibration of interest rate models does not involve the interest rate being modeled. Hence, calibrated models may not be good representations of interest rates and may not produce reliable derivative prices.

To deal with the issue, we propose a time series modeling approach to analyze interest rates, specifically, the zero-coupon yield curves. In this approach, yield curves are modeled as functional data and we introduce models that are based on the well-known autoregressive model in time series analysis. The objective of this approach is to understand the dependency of the yield curves on historical data and to predict future yield curves before they are observed.

The proposed models are illustrated and compared with the time series of US Treasury zero-coupon yield curves. We explore how individual models perform during different times in an economic cycle. We also propose a way to predict future caplet prices by combining yield curve prediction using functional time series models and historical implied volatilities of caplets. The time series approach that we propose are shown to work well against existing models such as the Hull-White model.

Acknowledgements

I would like to express my sincere appreciation to my advisor, Professor Tze Leung Lai, for his immense support throughout the course of my PhD career. He has given me insightful comments and advices on my research. I thank him for his patience and encouragement. I would not be able to complete my thesis without the guidance from him.

I would like to thank Professor Paul Switzer for giving valuable suggestions on my research and future direction. I also thank my oral examination committee members, Professors Guenther Walther, Yinyu Ye, and Antoine Toussaint, who provided constructive feedback on my research.

I am grateful to many friends at Stanford. Specifically, I must thank Ling Chen and Kevin Sun for discussing research problems with me. I also thank Zhen Wei, Nicholas Johnson, Zongming Ma, Baiyu Zhou, Shaojie Deng, Wai Wai Liu, and Feng Zhang for their time with me.

On a more personal note, my girlfriend Yi Fang Chen has given me strength and motivation. I thank her for her patience and understanding even during hard times of this study. I would also like to thank my parents for their support and encouragement. Their confidence in me has inspired me to push limits and reach my full potential. I am grateful to have such a wonderful family that continues to support me in my future endeavors.

Contents

List of Tables

List of Figures

Interest Rate Modeling and a Time Series Model for Functional Data

Chung Kwan Pong

August 24, 2010

Chapter 1

Introduction

1.1 Motivation

In finance, an interest rate derivative is a financial instrument where the underlying asset is an interest rate at which payments are made based on a notional amount. The price of an interest rate derivative depends on the level of the interest rate and its expected change in the future. To price an interest rate derivative, a common approach is to define the future evolution of the interest rates using an interest rate model. There are three main types of interest rate models. A short rate model describes the short rate; an HJM model describes the instantaneous forward rate; and a market model describes the forward rate. These interest rate models are based on some parameters which are solved by a process called calibration. Calibration makes sure that the interest rate models produce prices that are close to the market prices of some interest rate derivatives. These model paramters are then used to price other interest rate derivatives.

A drawback of this approach is that the interest rate that is modeled by the interest rate model is never used in the calibration procedure. Hence, the calibrated model may not give a very good representation of the dynamics of the interest rate, and hence the prices it produces may be unreliable. In this thesis, we propose a time series approach to analyze the time series of zero-coupon yield curves. It aims to understand the dependence structure of the yield curve on past data and to make accurate forecasts of future yield curves. In the model, yield curves are treated as functional

data, where each curve is regarded as a functional observation.

In Chapter 4, we illustrate the use of this modeling approach by looking at how well it can predict the yield curve on future periods. We then propose to use these predictions in combination with the implied volaitlities of caplets to predict the prices of future caplets. Not only is this time series approach of predicting caplet prices shown to work well against pricing by calibrating interest rate models, but treating yield curves as functional data is seen to have an advantage because the forecasts of the model are also functional. When calculating caplet prices, points on the yield curve at maturities where no observations are recorded are needed.

1.2 Functional Data

A set of data can be considered functional if observations have common underlying functional structure such as curves or surfaces. For example, the weights of an individual recorded at different times is a functional observation because it can be considered as a function of the relationship between weight and time. If the weight information is recorded for more than one individual, it is considered a functional data set. A functional time series is a set of functional data where the observations are taken at successive times. In Chapters 3 and 4, the time series of yield curves is treated as a functional time series. We assume that the functional observations are spaced at uniform time intervals.

Normally, functional data are collected in discrete form rather than in functional form because it is not possible to record and store an infinite number of points. In order to analyze them in functional form, these data are first converted into functions via means such as spline or kernel smoothing. The time series of yield curves is treated in the same way. Section 2.2 discusses some parametric and nonparametric methods that estimate the zero-coupon yield curve from the prices of a number of bonds. This also makes it possible to convert the zero-coupon yield curve to the forward rate curve, which involves taking derivatives to the curve.

Analyzing data in functional form is not a new idea. Ramsay and Silverman (2005) describe several statistical methods that can be used to analyze functional data such as functional principal component analysis, functional canonical correlation analysis, and functional linear regression.

1.3 Outline of the Thesis

This thesis is organized as follows. Chapter 2 provides some background information about interest rate markets. In particular, it describes some popular interest rate derivatives such as swaps, caps, floors, and swaptions, and interest rate models that include short rate models, HJM models, and market models. Chapter 3 provides functional time series models that are used to model the time series of yield curves. The functional autoregressive model and the pointwise autoregressive functional model are presented. They are solved by first assuming the functional coefficients are linear combinations of basis functions and then using least squares method. Chapter 4 provides empirical studies that illustrate how the functional time series models can be applied to real world data. The first study compares the prediction accuracy of the one-day ahead out-of-sample forecast of several time series models. The second study shows how the results of the first study can be used to predict future caplet prices. Chapter 5 provides a summary of the main ideas and results of the thesis.

Chapter 2

Interest Rate Markets

This chapter provides some financial background of the interest rate markets. Section 2.1 provides some basic concepts and definitions of popular interest rate derivatives including bonds, LIBOR, forward rates, interest rate swaps, caps, floors, and swaptions. Section 2.2 describes methods that can be used to estimate the zero-coupon yield curve (as known as the zero curve) from given default-free bond prices. Section 2.3 describes several types of models in the literature that model the future evolution of interest rates. Short rate models describe the short rate; HJM models describe the instantaneous forward rate; the LIBOR market model describes the forward LIBOR. Section 2.4 discusses how the paramters in the interest rate models can be estimated.

2.1 Elements of Interest Rate Markets

2.1.1 Bond Market

In finance, a zero-coupon bond with face value 1 is a debt security that pays 1 unit of currency at the maturity date. A coupon bond is like a zero-coupon bond but with interest payments at specified times before the maturity. In the United States, these debt securities issued by the US Treasury are divided into 3 categories. Treasury bills (T-bills) are zero-coupon bonds with maturity less than 1 year. Treasury notes (T-notes) are semiannual coupon bonds with maturity between 1 and 10 years. Treasury bonds (T-bonds) are semiannual coupon bonds with maturity longer than 10 years.

Denote the price at time t of a zero-coupon bond with face value 1 and maturity T by $P(t,T)$. It is clear that $P(T,T)=1$. For a coupon bond with face value 1 and coupon payments c_i at times $T_1 < T_2 < \cdots < T_n = T$, it can be realized as a sum of zero-coupon bonds and its price is given by

$$\sum_{i=1}^{n-1} c_i P(t,T_i) + (1+c_n)P(t,T_n). \tag{2.1}$$

Under continuous compounding, the yield-to-maturity (or simply yield) y of such a bond is defined to be the solution of the following equation

$$\text{bond price} = \sum_{i=1}^{n-1} c_i e^{y(T_i-t)} + (1+c_n)e^{-y(T_n-t)}.$$

The spot rate is the yield of a zero-coupon bond (under continuous compounding) and is given by

$$R(r,T) = -\frac{\log P(t,T)}{T-t}.$$

The relationship between interest rates and their maturities is called the term structure.

2.1.2 LIBOR and Forward Rates

LIBOR stands for London InterBank Offered Rate. It is a daily reference rate based on the interest rates that banks in the London wholesale money market charge for borrowing funds to each other. It is an annualized, simple interest rate that will be delivered at the end of a specified period. Denote LIBOR by $F(t,T)$, where t is the current time and T is the maturity. Then it is defined as

$$F(t,T) = \frac{1}{T-t}\left(\frac{1}{P(t,T)} - 1\right).$$

A forward rate agreement (FRA) is a contract today t for a loan between T_1 and T_2. It gives its holder a loan at time T_1, with a fixed simple interest rate for the period $T_2 - T_1$, to be paid at time T_2 in additional to the principal. The rate agreed in a FRA is called the forward rate and is

denoted by $F(t, T_1, T_2)$. Its value is defined as

$$F(t, T_1, T_2) = \frac{1}{T_2 - T_1} \left(\frac{P(t, T_1)}{P(t, T_2)} - 1 \right). \tag{2.2}$$

The instanteneous forward rate at time t with maturity T is denoted by $f(t, T)$ and is given by

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

From this definition, the zero-coupon bond price $P(t, T)$ can be expressed as

$$P(t, T) = \exp \left\{ -\int_t^T f(t, u) du \right\}. \tag{2.3}$$

The short rate $r(t)$ is defined as

$$r(t) = \lim_{T \to t} f(t, T) = f(t, t). \tag{2.4}$$

It represents the interest rate at which a loan is made for an infinitesimally short period of time from time t.

2.1.3 Interest Rate Swaps

An interest rate swap is a contract between two parties in which they agree to exchange one stream of cash flows based on fixed interest rate κ with another stream based on variable interest rate, which is usually taken to be LIBOR. Let $t \leq T_0 < T_1 < \cdots < T_n$, where t is the current time and $T_1, \ldots, T_n$ are times at which payments occur. T_n is called the maturity of the swap. Often, $T_i - T_{i-1} = \delta$ and δ can be 1, 1/2, or 1/4 year. If $t < T_0$, then the swap is called a forward swap. Let N be the notional of the swap. Then at each time T_i, $i = 1, \ldots, n$, one party pays the other party a fixed amount $N\delta\kappa$ and receives a floating amount $N\delta F(T_{i-1}, T_i)$. The fixed interest rate κ is called the swap rate. Its value at time t is denoted by $r^{swap}(t, T_0, T_n)$ and is defined as follows:

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

2.1.4 Caps and Floors

An interest rate caplet with reset date T and settlement date $T+\delta$ is a European call option on LIBOR, that pays its owner at $T+\delta$ if the rate exceeds the strike rate K in the amount of $\delta \max(0, F(T, T+\delta) - K)$.

An interest rate cap is a series of caplets. Let $T_0 < T_1 < \cdots < T_n$ be future dates and K be the strike rate. Then at each time T_i, $i = 1, \ldots, n$, the cap pays $(T_i - T_{i-1}) \max(0, F(T_{i-1}, T_i) - K)$. Interest rate caps are designed to provide insurance against the interest rate of a floating rate loan rising above a certain level K.

Floorlet and floor are European put option counterparts of caplet and cap, respectively. Using the same notations, a floorlet pays $\delta \max(0, K - F(T, T+\delta))$, and a floor is a series of floorlets and pays $\delta \max(0, K - F(T_{i-1}, T_i))$ at time T_i for $i = 1, \ldots, n$.

In the market, caplets and caps (as well as floorlets and floors) are quoted in terms of their implied volatilities. Their prices are obtained by plugging the implied volatilities into Black (1976) formula. Suppose the implied volatility at time t for the ith caplet is $\sigma_t^{\text{caplet}^i}$. Then its price at time t is given by

$$(T_i - T_{i-1}) P(t, T_i) \, \text{Black}(F(t, T_{i-1}, T_i), K, \sigma_t^{\text{caplet}^i} \sqrt{T_{i-1} - t}), \tag{2.5}$$

where

$$\text{Black}(L, K, \sigma) = L\Phi\left(\frac{\log(L/K)}{\sigma} + \frac{\sigma}{2}\right) - K\Phi\left(\frac{\log(L/K)}{\sigma} - \frac{\sigma}{2}\right), \tag{2.6}$$

and Φ is the standard normal cumulative distribution function. The relationship between the cap price and the cap implied volatility σ_t^{cap} is given by

$$\sum_{i=1}^{n} (T_i - T_{i-1}) P(t, T_i) \, \text{Black}\left(F(t, T_{i-1}, T_i), K, \sigma_t^{\text{cap}} \sqrt{T_{i-1} - t}\right).$$

2.1.5 Swaptions

A swaption with strike κ is an option that grants its owner the right to enter into a swap with fixed rate κ at the maturity T_0 of the swaption. Suppose the payments of the swap occur at times $T_1, \ldots, T_n$. The duration of the swap $T_n - T_0$ is called the tenor of the swaption. The payoff of the

swaption at maturity T_0 is given by

$$\delta \max\{0, r^{swap}(T_0, T_n) - \kappa\} \sum_{i=1}^{n} P(T_0, T_i).$$

2.2 Yield Curve Estimation

The term structure of interest rate can be described by the price of a zero-coupon bond with face value 1 versus its maturity. Alternatively, it can be described by the yield curve, which is the relationship between the yield of a zero-coupon bond and its maturity. The yield curve is usually, but not always, an increasing function of time t. Its shape and level reveal conditions in the economy and the financial markets, and so yield curves are monitored closely by economists and market practitioners.

To estimate the yield curve at current time 0, one takes a set of n reference default-free bonds such as the US Treasury bonds, which is seen in (2.1) to be dependent on the function $P(0, \cdot)$. A parametric or nonparametric model is assumed on $P(0, T)$ and least squares regression is used to estimate the parameters. That is, the following quantity is minimized over the parameters of the model: $\sum_{i=1}^{n}(B_i - \hat{B}_i)^2$, where B_i is the observed price of the ith bond, and $\hat{B}_i$ is model price of the ith bond.

A parametric model on $P(0, T)$ assumes certain form on it so that by varying the parameters, the model is able to reproduce the the shapes seen in historical yield curves: increasing, decreasing, flat, humped, and inverted. For example, Nelson and Siegel (1987) assume the following model on the instanteneous forward rate.

$$f(0, s) = \beta_0 + \beta_1 \exp\left\{-\frac{s}{\tau}\right\} + \beta_2 \frac{s}{\tau} \exp\left\{-\frac{s}{\tau}\right\},$$

which implies (by (2.3))

$$P(0, t) = \exp\left\{-\beta_0 t - (\beta_1 + \beta_2)\tau(1 - e^{-t/\tau}) + t\beta_2 e^{-t/\tau}\right\}.$$

Svensson (1994) generalized the Nelson-Siegel model by assuming the instantenous forward rate has

the following form:

$$f(0,s) = \beta_0 + \beta_1 \exp\left\{-\frac{s}{\tau_1}\right\} + \beta_2 \frac{s}{\tau_1} \exp\left\{-\frac{s}{\tau_1}\right\} + \beta_3 \frac{s}{\tau_2} \exp\left\{-\frac{s}{\tau_2}\right\},$$

which is capable of producing additional U and humped shapes for $P(0,t)$:

$$P(0,t) = \exp\left\{-\beta_0 t - (\beta_1+\beta_2)\tau_1(1-e^{-t/\tau_1}) + t\beta_2 e^{-t/\tau_1} - \beta_3\tau_2(1-e^{-t/\tau_2}) + t\beta_3 e^{-t/\tau_2}\right\}.$$

The nonparametric approach to estimate the yield curve is to express the function $P(0,T)$ using spline basis functions. The regression parameters can then be estimated using ordinarily least squares method. We will use this approach in Chapter 4 with a spline function called B-spline. Details of B-spline is given in Appendix A.

2.3 Stochastic Interest Rate Models

Interest rate models are mathematical models that describe the dynamics of interest rates in the future. These models are used to price interest rate derivatives that depend on future interest rates. The quantities that are modeled can be different in different interest rate models. This section gives a brief description of three kinds of interest rate models: short rate models, HJM framework, and LIBOR market model, where the quantities of interest are short rate, instantaneous forward rate, and forward LIBOR, respectively.

2.3.1 Short Rate Models

Short rate models are stochastic models that prescribe the dynamics of the short rate r_t, which was defined in (2.4). It can be shown that in the absence of arbitrage, under some technical conditions, the price of a zero-coupon bond is related to short rate by the following formula:

$$P(t,T) = E\left[\exp\left\{-\int_t^T r_s ds\right\}\middle| r_t = r\right].$$

Hence, future bond prices can be calculated if the dynamics of the short rate is specified. Interest rate models with different forms have been proposed in the literature and this section describes some of them. In the following models, W_t denotes a standard Brownian motion under a risk-neutral probability measure.

Vasicek Model

Under the Vasicek model (Vasicek, 1977), the short rate r_t is modeled as

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

The Vasicek model was the first one to capture the mean-reverting property of the interest rate. Mean reversion refers to the fact that interest rate flutuates around some level and this property is what sets interest rates apart from other financial prices such as stock prices. In the model, b is the long-term mean level of r_t, a is the speed of reversion, and σ is the volatility.

Under the Vasicek model, the price of a zero-coupon bond has the following closed-form expression:

$$\begin{aligned} P(t,T) &= \alpha(t,T)e^{-\beta(t,T)r_t}, \\ \alpha(t,T) &= \exp\left\{\left(b - \frac{\sigma^2}{2a^2}\right)[\beta(t,T) - (T-t)] - \frac{\sigma^2}{4a}\beta^2(t,T)\right\}, \\ \beta(t,T) &= \frac{1}{a}\left(1 - e^{-a(T-t)}\right). \end{aligned}$$

Cox-Ingersoll-Ross (CIR) Model

Cox, Ingersoll, and Ross (1985) modify the Vasicek model due to the fact that the short rate process r_t in the Vasicek model can become negative. The CIR model assumes the following dynamics for the short rate r_t:

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

This model is also a mean-reverting process. The square root ensures that r_t stays nonnegative.

Under the CIR model, the price of a zero-coupon bond has the following closed-form expression:

$$\begin{aligned} P(t,T) &= \alpha(t,T)e^{-\beta(t,T)r_t}, \\ \alpha(t,T) &= \left[\frac{2he^{(a+h)(T-t)/2}}{2h+(a+h)[e^{(T-t)h}-1]}\right]^{2ab/\sigma^2}, \\ \beta(t,T) &= \frac{2\left(e^{(T-t)h}-1\right)}{2h+(a+h)[e^{(T-t)h}-1]}, \end{aligned}$$

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

Hull-White Model

Vasicek and CIR models have only a few number of paramters and they usually can not reproduce the initial term structure exactly. A model that can fit the initial term structure perfectly is the Hull-White model, which was introduced by Hull and White (1990) and has a more general form than the Vasicek model:

$$dr_t = (b_t - ar_t)dt + \sigma dW_t. \tag{2.7}$$

Under this model, the price of a zero-coupon bond is given by

$$\begin{aligned} P(t,T) &= \alpha(t,T)e^{-\beta(t,T)r_t}, \\ \alpha(t,T) &= \frac{P(0,T)}{P(0,t)}\exp\left\{-\beta(t,T)\frac{\partial}{\partial t}\log P(0,t) - \frac{\sigma^2(1-e^{-2at})}{4a}\beta^2(t,T)\right\}, \\ \beta(t,T) &= \frac{1}{a}\left(1-e^{-a(T-t)}\right). \end{aligned}$$

Also, the price of a caplet at current time 0 with strike K on the LIBOR $F(T,T+\delta)$ is given by

$$(1+\delta K)P(0,T)\left[\frac{1}{1+\delta K}\Phi(-d_2) - \frac{P(0,T+\delta)}{P(0,T)}\Phi(-d_1)\right], \tag{2.8}$$

where

$$\begin{aligned} d_1 &= \frac{\log\left[\frac{P(0,T+\delta)}{P(0,T)}(1+\delta K)\right]}{\Sigma\sqrt{T}} + \frac{1}{2}\Sigma\sqrt{T}, \\ d_2 &= d_1 - \Sigma\sqrt{T}, \\ \Sigma &= \sigma\frac{1-e^{-a\delta}}{a}\sqrt{\frac{1-e^{-2aT}}{2aT}}. \end{aligned}$$

Some Other Short Rate Models

The Vasicek, CIR, and Hull-White models belong to a more general type of models called affine models. Affine models consist of the class of diffusion models for r_t which have the nice property that the zero-coupon bond price $P(t,T)$ has explicit solutions as $P(t,T) = e^{-A(t,T)-B(t,T)r_t}$. Duffie and Kan (1996) showed that a necessary and sufficient condition for a model to be affine is that it has the following form:

$$dr_t = (b_t + \beta_t r_t)dt + \sqrt{a_t + \alpha_t r_t}dW_t,$$

where b, β, a, α are deterministic functions of time. It can be shown that $A(t,T)$ and $B(t,T)$ satisfy the following system of two ordinary differential equations:

$$\begin{aligned} B'(t,T) &= -\beta_t B(t,T) + \frac{\alpha_t}{2}B^2(t,T) - 1, \\ A'(t,T) &= b_t B(t,T) + \frac{1}{2}a_t B^2(t,T), \end{aligned}$$

with terminal conditions $B(T,T) = 0$ and $A(T,T) = 0$. In order to solve for A and B, the strategy is to first solve for B and integrate it into A's differential equation to get A. However, in general there are no closed-form solutions for A and B. Numerical methods are needed to solve for these ordinary differential equations.

Some examples of affine models are given below:

- Ho-Lee model: $dr_t = \theta_t dt + \sigma dW_t$.
- Extended CIR model: $dr_t = (b_t - \beta r_t)dt + \sigma\sqrt{r_t}dW_t$.

2.3.2 Heath-Jarrow-Morton (HJM) Framework

The HJM framework was proposed by Heath, Jarrow, and Morton (1992). It is a general framework that models the dynamics the instantaneous forward rate $f(t,T)$. The motivation behind the development of the framework is that while short rate models which prescribe the dynamics of the instantaneous spot rate r_t is a natural way of modeling interest rates, these models have shortcomings. First, short rate models such as the Vasicek and CIR models have only a few number of paramters and they are not able to reproduce the initial yield curve exactly. Second, short rate models do not capture the full dynamics of the forward rate curve; it only captures the dynamics of a point on the curve.

In the HJM framework, the instantaneous forward rate $f(t,T)$ is modeled as follows by the following k-factor model:

$$df(t,T) = \sum_{i=1}^{k} \sigma_i(t,T)s_i(t,T)dt + \sum_{i=1}^{k} \sigma_i(t,T)dW_i(t), \tag{2.9}$$

where $(W_1(t), W_2(t), \ldots, W_k(t))^T$ is a k-dimensional Brownian motion. They showed that under an equivalent martingale measure, the condition of no-arbitrage implies that

$$s_i(t,T) = \int_t^T \sigma_i(t,u)du.$$

Since the drift term in (2.9) is a function of the volatility, once the volatility is given, the drift is constrained to be the given expression and no drift estimation is needed. However, models developed under the general HJM framework are usually non-Markovian, which makes it difficult to use tree methods when pricing interest rate derivatives because it leads to non-recombining trees. In these cases, Monte Carlo simulations are used instead (Brigo and Mercurio, 2006).

2.3.3 LIBOR Market Model

The LIBOR market model is an interest rate model that prescribes the dynamics of forward LIBOR rates. It was developed by Brace, Gatarek, and Musiela (1997) and is also known as the BGM model.

Section 2.1.4 describes the connection between the caplet price and its Black implied volatility through Black (1976) formula. In fact, this formula was an extension of the work of Black and Scholes (1973) on option pricing theory on equity options. It replaces the volatility of the spot price in the Black-Scholes formula with that of the forward price.

The Black formula for caplets assumes that the forward LIBOR follows a geometric Brownian motion with constant volatility. It is the formula used by the market to quote caps and floors. However, it can be shown that Gaussian HJM models, which are HJM models with deterministic volatility function $\sigma_i(t,T)$ in (2.9), are not able to reproduce the Black formula for caplets used by the market. The incompatibility on the caplet pricing formula leads to the development of the LIBOR market model.

Consider times $T_0 < T_1 < \cdots < T_n$, where T_i's are reset dates and T_{i+1}'s are settlement dates for $i = 0, 1, \ldots, n-1$. Denote $\mathbb{Q}^{T_i}$ to be the T_i-forward measure, which is the forward risk-neutral measure with respect to the numeraire $P(t,T_i)$. Let $F_i(t) = F(t,T_i,T_{i+1})$. The LIBOR market model models a set of n forward LIBOR $F_i(t)$, $i = 0, \ldots, n-1$ under $\mathbb{Q}^{T_i}$ as lognormal processes

$$\frac{dF_i(t)}{F_i(t)} = \mu_i(t) dW_t^{T_i},$$

where $\mu_i(t)$ is a deterministic function, and $W_t^{T_i}$ is a Brownian motion under $\mathbb{Q}^{T_i}$.

Under the LIBOR market model, the price of the ith caplet is given by

$$(T_i - T_{i-1}) P(t,T_i)[F(t,T_{i-1},T_i)\Phi(d_{1i}) - K\Phi(d_{2i})],$$

where

$$\begin{aligned} d_{1i} &= \frac{\log(F(t,T_{i-1},T_i)/K)}{\nu_i\sqrt{T_{i-1}-t}} + \frac{1}{2}\nu_i\sqrt{T_{i-1}-t}, \\ d_{2i} &= d_{1i} - \nu_i\sqrt{T_{i-1}-t}, \\ \nu_i^2 &= \frac{1}{T_{i-1}-t}\int_t^{T_{i-1}} \mu_{i-1}(u)du. \end{aligned} \tag{2.10}$$

Therefore, it is seen that this recovers the caplet pricing formula of Black model used in the market. For this reason, it is named with the term "market model".

2.4 Parameter Estimation in Interest Rate Models

Currently, the market practice in the financial industry to estimate the parameters of an interest rate model is to calibrate it to certain interest rate products which are called calibrating instruments. The goal of calibration is to be ensure that the model is able to produce prices that are close to the prices of the calibrating instruments. A standard way of doing this is to minimize the sum of squares between the observed prices and the model prices over the calibrating instruments. In Section 2.2, the estimation of model parameters of the parametric models is an example of calibrating to bond data.

Calibrating instruments are usually taken to be similar to the derivatives that are being valued. Hence, caps, floors, and swaptions, which are the most popular interest rate derivatives, are often taken as calibrating instruments.

To calibrate the LIBOR market model to caplet implied volatilities σ_i, the ν_i's in (2.10) is taken to be σ_i's. The $\mu_i(t)$ are assumed to have the form $\mu_i(t) = \mu(t, T_i)$. Define $\mu_{i,k}^2$ by

$$\mu_{i,k}^2 = \int_{T_{k-1}}^{T_k} \mu(u, T_i) du, \quad 1 \le k \le i,$$

and let T_0 be the current time t. Then (2.10) becomes $(T_{i-1} - t)\sigma_i^2 = \sum_{k=1}^{i-1} \mu_{i,k}^2$. Rebonato (2002) proposes the following form for $\mu(t, T) = \xi(t)\eta(T)\rho(T - t)$, where $\xi(t)$ is the purely time-dependent component, $\eta(T)$ is the purely forward-rate-specific component, and $\rho(T-t)$ is the time-homogeneous component. Rebonato (2002) and Brigo and Mercurio (2006) also recommend using parametric models for these three building blocks when estimating these functions.

An alternative way to fit the term structure models is to use multivariate time series of bond yields instead of the prices of financial instruments. For one-factor short-rate models, discrete-time approximations of the continuous-time stochastic models of r_t are used. Short-maturity Treasury bills are taken as proxies for r_t. Lai and Xing (2008, Section 9.7.3) illustrates this approach to estimate the parameters and to test the validity of different short-rate models using generalized method of moments by Chan et al. (1992). Aït-Sahalia (1996) and Stanton (1997) proposed nonparametric model to estimate the drift and volatility functions of $dr_t = \mu(r_t)dt + \sigma(r_t)dW_t$ through the use of its stationary distribution.

Chapter 3

A Functional Time Series Approach to Interest Rate Modeling

In the last chapter, we saw several types of interest rate models that prescribe the dynamics of the short rate, instantaneous forward rate, and the forword LIBOR. Each of the models can be seen as modeling the evolution of some time series. For example, in short rate models, r_t can be thought of as a continuous-time time series. The calibration procedure attempts to resemble the dynamics of the interest rate defined by the model and the dynamics of the interest rate observed in the market. It is done by looking for model paramters so that the model produces prices that are close to the market prices of certain calibrating instruments. In a sense, the calibrated model defines how the interest rate should evolve based on hisorical or present data. In fact, this is exactly what time series models attempt to do, namely, understanding the underlying structure which allows one to explore how the time series is generated, and also making forecasts of future points based on historical data before the actual data is measured.

The price of an interest rate derivative depends on the future evolution of interest rates, which can be defined by an interest rate model. However, the assumptions made by an interest rate

model might be unrealistic. Also, during the calibration procedure, the interest rate modeled by the interest rate model is never used. Hence, the model calibrated to financial instruments may not provide accurate dynamics of the interest rate, and as a result, the prices of interest rate derivatives it produces may be unrealiable. In addition, since there are so many interest rate models and different models may have different assumptions about interest rates. These assumptions need to be justified before pricing can be done.

While interest rate models use stochastic different equations to define the dynamics of interest rates, a statistical counterpart that can be used to analyze time series of interest rates is time series models. In this section, we propose a time series approach to analyze yield curve data. It has two advantages. First, interest rate data are actually used to analyze its own structure. Second, the model is more flexible in terms of assumptions on the interest rate. An interest rate model uses Brownian motion and sometimes requires mean-reverting property, which may not may not hold in reality; the proposed model has no such restriction. In these time series model, the yield curve data are treated as functional data. This is motivated by Carmona and Tehranchi (2006), who proposed an infinite-dimensional HJM model that replaces $\sum_{i=1}^{k}$ in (2.9) by $\sum_{i=1}^{\infty}$.

This chapter is organized as follows. Section 3.1 introduces the notion of functional data. In analyzing interest rate data, the data are treated as functional time series. Section 3.2 gives a brief review of time series models. Section 3.3 introduces a time series model that we will use to analyze interest rate data in functional form. The model is based on autoregressive model and is called the autoregressive functional time series model. Section 3.4 discusss two ways to approximate the coefficient functions. One is via discretizing the functions, and the other is via assuming that the coefficients are linear combinations of a common set of predefined basis functions. After the coefficients are estimated, Section 3.5 explains the procedure of forecasting future observations based on historical data. Section 3.6 describes an algorithm to select a subset of basis functions in order to obtain a better representation of the coefficient functions and achieve better prediction performance. Section 3.7 describes an extension of the autoregressive functional model which incorporates exogenous time series. Section 3.8 introduces the pointwise autoregressive functional model which is an alternative model for functional time series.

3.1 Yield Curves as Functional Data

The interest rate that we attempt to model is the yield curve of zero-coupon bonds. In reality, the US Treasury releases bills, notes, and bonds with some specific maturities such as $0.25, 0.5, 1, 2, \ldots, 10$ years. Hence, at each time, we are only able to obtain a set of points on the yield curves from the bonds. The time series of bond yields can be considered as a multivariate time series. Clearly, one way to model this multivariate time series is to apply some time series models such as the vector autoregressive moving-average model. However, suppose we consider the HJM model in Section 2.3.2 that models the instantaneous forward rate. Once the model parameters is determined by calibration, the future dynamics of the forward curve $f(t,T)$ is given. Then zero-coupon prices and their yields can be obtain by $P(t,T) = \exp\{-\int_t^T f(t,u)du\}$ and $R(t,T) = \frac{1}{T-t}\int_t^T f(t,u)du$. Transformation between bond prices and forward rates can be done because for each time t, the entire forward curve $f(t,T)$ is known. On the other hand, this kind of transformation is not possible with multivariate time series and multivariate models. This motivates us to first consider the time series of yield curves as a functional time series.

In Section 2.2, we describe parametric and nonparametric methods to estimate yield curves. These methods are very useful now in our analysis of the yield curves as functional time series because they transform the multivariate bond yield data into functional yield curve data. Also, in cases where the multivariate bond yield data have missing data, or when the bond yields are recorded at different maturities at different times, then multivariate time series models are not applicable, but the methods in Section 2.2 are still able to produce a functional time series for us to analyze. As a result, transformations between different rates which require derivatives or integrals are possible. In particular, we will use the nonparametric splines method to estimate yield curves in the empirical study of the US Treasury zero-coupon bond yield data in Chapter 4. Figure 3.1 shows some of the zero-coupon bond yield curves obtained by the nonparametric splines approach.

3.2 Review of Time Series Models

There are numerous time series models and they take on different forms and represent different stochastic processes. Three broad classes of time series models are autoregressive (AR), integrated

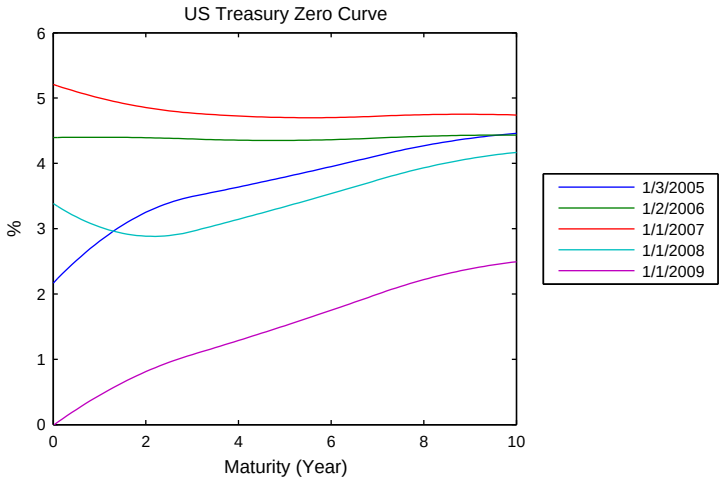

Figure 3.1: U.S. Treasury zero-coupon bond yield curve on the first trading day of each of years 2005-2009 obtained by the nonparametric splines approach.

(I), and moving average (MA) models. These models assume linear relationships between present information and past information. In particular, an AR model assumes a linear relationship between each observation and a number of past observations, an I model assumes that the kth difference of the time series is a white noise process, and an MA model assumes that every observation is a linear combination of previous (unobserved) white noise terms. These ideas can be combined to form autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models.

One can also incorporate external time series into any of the above time series models when the observed time series is driven by it. Hence, we could have models denoted by ARMAX or ARIMAX, where the "X" represents "exogenous".

The univariate time series models above can be generalized to deal with vector-valued time series data. A "V", which stands for "vector", is added to the front of the acroynms. For example, a multivariate ARMA model is called a vector autoregressive moving average model and is denoted by VARMA.

3.3 Autoregressive Functional Time Series Models

The autoregressive model is a basic and intuitive model that is used to analyze time series. The idea behind the model is that every observation is dependent on some past observations, and the dependency is linear. Since the model is linear, the problem is treated as a linear regression problem and it has a closed-form solution. Because of the usefulness of the autoregressive model, we will generalize its idea to the functional case. In this section, we give the autoregressive model for functional time series, which is based on the univariate and multivariate autoregressive models.

Given a univariate time series $x_1, x_2, \ldots$, the autoregressive model with order p is given by

$$x_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_{t-i} + \epsilon_t,$$

where ϵ_t are the error terms. In the model, every observation x_t is a linear combination of the past p observations x_{t-i}, $i = 1, \ldots, p$. When solving for the solutions $\hat{\alpha}_i$'s, the problem is regarded as a multiple linear regression, and the sum of the squared errors is minimized. It can be generalized to the multivariate case where we have a k-dimensional time series $\mathbf{x}_1, \mathbf{x}_2, \ldots$. The autoregressive model becomes

$$\mathbf{x}_t = \boldsymbol{\alpha}_0 + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t.$$

This model is called the vector autoregressive model with order p. $\boldsymbol{\alpha}_0$ is the intercept, which is now a k-dimensional vector. Each of the coefficients $\mathbf{A}_i$ is a $k \times k$ matrix. Hence, the model relates each dimension of the observation $\mathbf{x}_t$ with every dimension of the past p observations $\mathbf{x}_{t-i}$, $i = 1, \ldots, p$, linearly. When solving for least squares estimates, the problem is regarded as a multivariate linear regression, and the sum of the squared Euclidean norms of the error vectors $\boldsymbol{\epsilon}_t$ is minimized.

Consider a functional time series $x_t(\theta)$, where $t = 1, 2, \ldots$ denotes the time and we assume that the series has common domain $\theta \in \Theta$. Θ is assumed to be a closed interval for practical applications. Following the definitions of the autoregressive models given above, the autoregressive functional model for $x_t(\theta)$ with order p is given by

$$x_t(\theta) = \alpha_0(\theta) + \sum_{i=1}^{p} \int \alpha_i(\vartheta, \theta) x_{t-i}(\vartheta) d\vartheta + \epsilon_t(\theta), \tag{3.1}$$

where $\epsilon_t(\theta)$ is a sequence of uncorrelated random processes that satisfies

1. $E(\epsilon_t(\theta_0)) = 0$ for $\theta_0 \in \Theta$.
2. $E(\epsilon_t(\theta_1)\epsilon_t(\theta_2)) = \sigma(\theta_1, \theta_2)$ for $\theta_1, \theta_2 \in \Theta$ (covariance function).
3. $E(\epsilon_{t_1}(\theta_1)\epsilon_{t_2}(\theta_2)) = 0$ for $t_1 \neq t_2$ and $\theta_1, \theta_2 \in \Theta$.

In this model, each functional observation $x_t(\theta)$ is related to the p most recent past observations $x_{t-i}(\theta)$, $i = 1, \ldots, p$, in a linear fashion. The coefficients are the intercept function $\alpha_0(\theta)$ and the two-dimensional $\alpha_i(\vartheta, \theta)$'s; they are analogous to the $\boldsymbol{\alpha}_0$ and A_i's in the vector autoregressive model. Definite integrals, which are taken over Θ, are used in order for the linear relationship to make sense. It can also be viewed as a linear operator T_i, which is the integral transform $T_i x(\theta) = \int \alpha_i(\vartheta, \theta) x(\vartheta) d\vartheta$.

The assumptions made on the error functions $\epsilon_t(\theta)$ can be seen as extensions of the assumptions on the error vectors in the vector autoregressive model. We require that the errors have mean 0; their covariance does not depend on time; and errors at different times are uncorrelated. Also, we require that the covariance function σ to be positive definite, which means that for any sequences $\theta_1, \ldots, \theta_n \in \Theta$ and $\xi_1, \ldots, \xi_n \in \mathbb{C}$, the sum

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \sigma(\theta_i, \theta_j) \xi_i \bar{\xi_j}$$

is real-valued and nonnegative.

3.4 Estimation of the Solutions

Suppose we have observations $x_t(\theta)$ for $t = 1, \ldots, T$. Then model (3.1) can be written in the following form in matrix notation:

$$\mathbf{X}_0(\theta) = \alpha_0(\theta)\mathbf{1} + \sum_{i=1}^{p} \int \mathbf{X}_i(\vartheta) \alpha_i(\vartheta, \theta) d\vartheta + \boldsymbol{\epsilon}_t(\theta), \tag{3.2}$$

where

$$\mathbf{X}_i(\theta) = \begin{bmatrix} x_{T-i}(\theta) \\ x_{T-1-i}(\theta) \\ \vdots \\ x_{p+1-i}(\theta) \end{bmatrix}, \quad i = 0, 1, \ldots, p,$$

and $\mathbf{1}$ is a vector of 1's. In order to solve for the $\hat{\alpha}_i$'s, we minimize the sum of squares of the errors terms, integrated over Θ. Mathematically, the criterion C is

$$C = \int \left\| \mathbf{X}_0(\theta) - \alpha_0(\theta)\mathbf{1} - \sum_{i=1}^{p} \int \mathbf{X}_i(\vartheta)\alpha_i(\vartheta, \theta)d\vartheta \right\|^2 d\theta.$$

This is an infinite-dimensional problem because unknown functions are needed to be solved. In order to deal with the situation, we will impose restrictions to reduce the problem into a finite-dimensional one. This can be done in one of the two following ways: discretizing the functional data, or reducing the dimemsions using basis functions. However, subsequent parts of this thesis will be based on the basis functions approach because of the drawbacks of the first approach which will be discussed later.

3.4.1 Discretizing the Domain of Definition of the Functions

One way of approximating the solutions is to discretize the interval Θ into subintervals. For the sake of simplicity, we will assume the intervals are defined by equally-spaced points $\theta_1, \ldots, \theta_m$ with $\Delta\theta = \theta_{i+1} - \theta_i$. Discretizing model (3.2) gives

$$\mathbf{X}_0(\theta_k) = \alpha_0(\theta_k) + \sum_{i=1}^{p} \sum_{j=1}^{m} \mathbf{X}_i(\theta_j)\alpha_i(\theta_j, \theta_k)\Delta\theta + \boldsymbol{\epsilon}_t(\theta_k), \qquad k = 1, \ldots, m.$$

This corresponds to a multivariate linear regression and can be represented in matrix form as

$$\mathbf{Y} = \mathbf{ZA} + \mathbf{E},$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{X}_0(\theta_1) & \cdots & \mathbf{X}_0(\theta_m) \end{bmatrix},$$

$$\mathbf{Z} = \Delta\theta \begin{bmatrix} \mathbf{1} & \mathbf{X}_1(\theta_1) & \cdots & \mathbf{X}_1(\theta_m) & \mathbf{X}_2(\theta_1) & \cdots & \mathbf{X}_p(\theta_m) \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} \alpha_0(\theta_1) & \alpha_0(\theta_2) & \cdots & \alpha_0(\theta_m) \\ \alpha_1(\theta_1,\theta_1) & \alpha_1(\theta_1,\theta_2) & \cdots & \alpha_1(\theta_1,\theta_m) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1(\theta_m,\theta_1) & \alpha_1(\theta_m,\theta_2) & \cdots & \alpha_1(\theta_m,\theta_m) \\ \alpha_2(\theta_1,\theta_1) & \alpha_2(\theta_1,\theta_2) & \cdots & \alpha_2(\theta_1,\theta_m) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_p(\theta_m,\theta_1) & \alpha_p(\theta_m,\theta_2) & \cdots & \alpha_p(\theta_m,\theta_m) \end{bmatrix},$$

$$\mathbf{E} = \begin{bmatrix} \boldsymbol{\epsilon}_t(\theta_1) & \cdots & \boldsymbol{\epsilon}_t(\theta_m) \end{bmatrix}.$$

The discretized version of the criterion C is

$$\sum_{k=1}^{m} \left[\mathbf{X}_0(\theta_k) - \alpha_0(\theta_k) - \sum_{i=1}^{p} \sum_{j=1}^{m} \mathbf{X}_i(\theta_j)\alpha_i(\theta_j,\theta_k)\Delta\theta \right]^2 \Delta\theta.$$

The solution $\hat{\mathbf{A}}$ is the least squares solution which is given by

$$\hat{\mathbf{A}} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y}.$$

Since $\hat{\mathbf{A}}$ contains only the values of the coefficient functions at discrete points, to recover the functions $\hat{\alpha}_0(\theta)$ and $\hat{\alpha}_i(\vartheta,\theta)$, one could interpolate the values $\hat{\alpha}_0(\theta_k)$ and $\hat{\alpha}_i(\theta_j,\theta_k)$ in $\hat{\mathbf{A}}$.

3.4.2 Basis Functions Approach

A drawback of the previous approach is that if the functional data are sufficiently smooth, then one cannot discretize the functions into too many intervals; otherwise, it will cause the discretized time series to be nearly linearly dependent, which leads to inaccuracies in the inversion of the matrix $\mathbf{Z}^T\mathbf{Z}$. On the other hand, there also cannot be too few intervals that would result in low resolution of the data.

In this section we introduce the basis function approach. It is common in the literature to use

basis functions when dealing with functional data. For example, Cont and Fonseca (2002) use basis functions to solve for functional principal components of time series of implied volatility surfaces of European options, Ramsay and Silverman (2005) use basis functions to perform regression analysis of functional data.

The idea of the approach is to assume that the unknown functional coefficients α_i are finite linear combinations of some pre-defined basis functions such as splines. By making this assumption, the goal now is to solve for the coefficients of the basis functions. As we shall see, we have transformed an infinite-dimensional problem into a finite-dimensional problem.

Suppose the α_i have the following expansions in terms of basis functions:

$$\begin{gathered}\alpha_0(\theta) = \textstyle\sum_{l=1}^{L} a_{0l} h_l(\theta) = \mathbf{a}^T \mathbf{h}(\theta), \\ \alpha_i(\vartheta, \theta) = \textstyle\sum_{k=1}^{K} \sum_{l=1}^{L} a_{ikl} g_k(\vartheta) h_l(\theta) = \mathbf{g}^T(\vartheta) \mathbf{A}_i \mathbf{h}(\theta), \quad i = 1, \ldots, p,\end{gathered} \tag{3.3}$$

where $\mathbf{A}_i$ is a matrix with kl-th entry given by a_{ikl}, $\mathbf{h}(\theta) = [h_1(\theta), \ldots, h_L(\theta)]^T$, $\mathbf{g}(\vartheta) = [g_1(\vartheta), \ldots, g_K(\vartheta)]^T$, and L and K are the number of basis functions in $\mathbf{h}(\theta)$ and $\mathbf{g}(\vartheta)$, respectively. From this definition, $\mathbf{h}(\theta)$ is the basis of the intercept function $\alpha_0(\theta)$. The basis of the two-dimensional functions $\alpha_i(\vartheta, \theta)$, $i = 1, \ldots, p$, is the tensor product of two sets of one-dimensional basis $\mathbf{g}(\vartheta)$ and $\mathbf{h}(\theta)$. The model now becomes

$$\begin{aligned}\mathbf{X}_0(\theta) &= \mathbf{a}^T \mathbf{h}(\theta) + \sum_{i=1}^{p} \int \mathbf{X}_i(\vartheta) \mathbf{g}^T(\vartheta) \mathbf{A}_i \mathbf{h}(\theta) d\vartheta + \boldsymbol{\epsilon}_t(\theta) \\ &= \mathbf{a}^T \mathbf{h}(\theta) + \sum_{i=1}^{p} \mathbf{G}_i \mathbf{A}_i \mathbf{h}(\theta) + \boldsymbol{\epsilon}_t(\theta) \\ &= \mathbf{G}\mathbf{A}\mathbf{h}(\theta) + \boldsymbol{\epsilon}_t(\theta),\end{aligned} \tag{3.4}$$

where $\mathbf{G}_i$, $\mathbf{G}$ and $\mathbf{A}$ are defined by

$$\mathbf{G}_i = \int \mathbf{X}_i(\vartheta) \mathbf{g}^T(\vartheta) d\vartheta,$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{1} & \mathbf{G}_1 & \cdots & \mathbf{G}_p \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}^T \\ \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_p \end{bmatrix} = \begin{bmatrix} a_{01} & \cdots & a_{0L} \\ a_{111} & \cdots & a_{11L} \\ \vdots & \ddots & \vdots \\ a_{1K1} & \cdots & a_{1KL} \\ a_{211} & \cdots & a_{21L} \\ \vdots & \ddots & \vdots \\ a_{pK1} & \cdots & a_{pKL} \end{bmatrix}.$$

The minimizing criterion C can now be written as

$$C = \int \|\boldsymbol{\epsilon}_t(\theta)\|^2 \, d\theta = \int \|\mathbf{X}_0(\theta) - \mathbf{G}\mathbf{A}\mathbf{h}(\theta)\|^2 \, d\theta.$$

Minimization of this criterion is done by taking derivative to C with respect to $\mathbf{A}$ and setting it to zero. The resulting equation is called the normal equation and is given by

$$\mathbf{G}^T\mathbf{G}\hat{\mathbf{A}} \int \mathbf{h}(\theta)\mathbf{h}^T(\theta)d\theta = \mathbf{G}^T \int \mathbf{X}_0(\theta)\mathbf{h}^T(\theta)d\theta. \tag{3.5}$$

Hence, the solution $\hat{\mathbf{A}}$ is given by

$$\hat{\mathbf{A}} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T \int \mathbf{X}_0(\theta)\mathbf{h}^T(\theta)d\theta \left(\int \mathbf{h}(\theta)\mathbf{h}^T(\theta)d\theta \right)^{-1}$$

Equivalently, the above can be expressed in terms of Kronecker product.

$$\left[\int \mathbf{h}(\theta)\mathbf{h}^T(\theta)d\theta \otimes (\mathbf{G}^T\mathbf{G}) \right] \mathrm{vec}(\hat{\mathbf{A}}) = \mathrm{vec}\left(\mathbf{G}^T \int \mathbf{X}_0(\theta)\mathbf{h}^T(\theta)d\theta \right)$$

$$\mathrm{vec}(\hat{\mathbf{A}}) = \left[\int \mathbf{h}(\theta)\mathbf{h}^T(\theta)d\theta \otimes (\mathbf{G}^T\mathbf{G}) \right]^{-1} \mathrm{vec}\left(\mathbf{G}^T \int \mathbf{X}_0(\theta)\mathbf{h}^T(\theta)d\theta \right)$$

The estimate $\hat{\mathbf{A}}$ is composed of matrices $\hat{\mathbf{a}}^T$ and $\hat{\mathbf{A}}_i$, $i = 1, \ldots, p$, which are the estimates of $\mathbf{a}^T$ and $\mathbf{A}_i$. Then the estimates of the coefficient functions, $\hat{\alpha}_0(\theta)$ and $\hat{\alpha}_i(\vartheta, \theta)$, can be calculated by plugging $\hat{\mathbf{a}}$ and $\hat{\mathbf{A}}_i$ into equations (3.3).

3.5 Forecasting

Forecasting is an important aspect of time series analysis. Suppose that we have a functional time series $x_1(\theta), \ldots, x_T(\theta)$ and we are now at time T. The goal is to use the functional autoregressive model to predict the outcome of the time series at some time index $T + l$ in the future, based on historical data. The time index T is called the forecast origin, and the time index l is called the forecast horizon. Denote the l-step ahead forecast of the time series by $x_T^{(l)}$. Forecasting is done in a similar fashion as in the univariate and multivariate autoregressive models.

1-Step Ahead Forecast

From the functional autoregressive model (3.1) with order p, we have

$$x_{T+1}(\theta) = \alpha_0(\theta) + \int \alpha_1(\vartheta, \theta) x_T(\vartheta) d\vartheta + \cdots + \int \alpha_p(\vartheta, \theta) x_{T-p+1}(\vartheta) d\vartheta + \epsilon_{T+1}(\theta).$$

After obtaining the estimates of the coefficients $\hat{\alpha}_i$, the 1-step ahead forecast can be calculated as

$$\begin{aligned} x_T^{(1)}(\theta) &= \hat{\alpha}_0(\theta) + \int \hat{\alpha}_1(\vartheta, \theta) x_T(\vartheta) d\vartheta + \cdots + \int \hat{\alpha}_p(\vartheta, \theta) x_{T-p+1}(\vartheta) d\vartheta \\ &= \hat{\mathbf{a}}^T \mathbf{h}(\theta) + \int \mathbf{g}^T(\vartheta) \hat{\mathbf{A}}_1 \mathbf{h}(\theta) x_T(\vartheta) d\vartheta + \cdots + \int \mathbf{g}^T(\vartheta) \hat{\mathbf{A}}_p \mathbf{h}(\theta) x_{T-p+1}(\vartheta) d\vartheta \\ &= \tilde{\mathbf{G}} \hat{\mathbf{A}} \mathbf{h}(\theta), \end{aligned}$$

where

$$\tilde{\mathbf{G}} = \left[\begin{array}{cccc} 1 & \int x_T(\vartheta) \mathbf{g}^T(\vartheta) d\vartheta & \cdots & \int x_{T-p+1}(\vartheta) \mathbf{g}^T(\vartheta) d\vartheta \end{array} \right].$$

Multistep Ahead Forecast

Again from the functional autoregressive model (3.1), we have

$$x_{T+l}(\theta) = \alpha_0(\theta) + \int \alpha_1(\vartheta, \theta) x_{T+l-1}(\vartheta) d\vartheta + \cdots + \int \alpha_p(\vartheta, \theta) x_{T+l-p}(\vartheta) d\vartheta + \epsilon_{T+l}(\theta).$$

The idea is to replace the $x_{T+l-i}(\theta)$, $i = 1, \ldots, p$, on the right hand side of the above equation by its forecast if it is an observation in the future. Mathematically, an l-step ahead forecast $x_T^{(l)}(\theta)$ for

$l \geq 2$ can be obtained as

$$\begin{aligned} x_T^{(l)}(\theta) &= \hat{\alpha}_0(\theta) + \int \hat{\alpha}_1(\vartheta,\theta) x_T^{(l-1)}(\vartheta) d\vartheta + \cdots + \int \hat{\alpha}(\vartheta,\theta) x_T^{(l-p)}(\vartheta) d\vartheta \\ &= \bar{\mathbf{G}}\hat{\mathbf{A}}\mathbf{h}(\theta), \end{aligned}$$

where

$$\bar{\mathbf{G}} = \left[\begin{array}{cccc} 1 & \int x_T^{(l-1)}(\vartheta)\mathbf{g}^T(\vartheta)d\vartheta & \cdots & \int x_T^{(l-p)}(\vartheta)\mathbf{g}^T(\vartheta)d\vartheta \end{array} \right],$$

and $x_T^{(i)}(\theta)$ is taken to be $x_{T+i}(\theta)$ if $i \leq 0$. The above formula provides a recursive relation for multistep ahead forecast. The l-step ahead forecast can be computed by repeatedly applying the 1-step ahead forecast procedure to obtain the forecasts $x_T^{(1)}(\theta), x_T^{(2)}(\theta), \ldots, x_T^{(l)}(\theta)$ in order.

3.6 Basis Selection

In Section 3.4, we approximate the solutions of the functional autoregressive model (3.1) by assuming the coefficient functions are linear combinations of basis functions (3.3). This reduces the model to (3.4), which can be treated as a multivariate regression problem. L 1-dimensional basis functions are used for $\alpha_0(\theta)$, and KL basis functions are used for each of $\alpha_i(\vartheta,\theta)$, $i = 1,\ldots,p$, where K is the number of basis functions of $\mathbf{g}$, L is the number of basis functions of $\mathbf{h}$, and p is the order of the model. Hence, the total number of basis functions used to describe the α_i's is equal to $KLp + L$. It is also the number of entries of $\mathbf{A}$. We see that this number increases with K, L, and p. When numerous basis functions are used, it is possible that some are less important and including them might lead to overfitting of the model.

One way to overcome the problem is that instead of assuming that α_0 is a linear combination of the h_l's and each α_i, $i = 1,\ldots,p$, is a linear combination of $g_k h_l$'s, we assume α_0 is a linear combination of a subset of all h_l's and each α_i, $i = 1,\ldots,p$, is a linear combination of a subset of all $g_k h_l$'s. Different α_i's will then be represented by different sets of basis functions. In this section we will discuss how the basis function representations of the α_i's are chosen.

We start from the matrix representation (3.4): $\mathbf{X}_0(\theta) = \mathbf{G}\mathbf{A}\mathbf{h}(\theta) + \boldsymbol{\epsilon}_t(\theta)$. In this representation, each entry of $\mathbf{A}$ represents the coefficient of a basis function. Suppose that we restrict an entry of $\mathbf{A}$

to be zero. This is equivalent to dropping the basis function corresponding to that entry. Hence, if we are given the set of basis functions (which is a subset of the original set of basis functions) that is used to explain each α_i, the solution can be obtained by performing a constrained minimization of the criterion C, with the constraints being the entries of $\mathbf{A}$ corresponding to the unused basis functions are 0. Before we discuss how the basis functions are chosen, it is necessary to understand how to solve for $\mathbf{A}$ when such constraints exist.

Consider a slightly more general problem where we have constraints of the form $\mathbf{u}_i^T A \mathbf{v}_i = \gamma_i$, with $\mathbf{u}_i$ being $(1 + pK_1)$-vectors, $\mathbf{v}_i$ being K_2-vectors, for $i = 1, \ldots, I$. In particular, if $\gamma_i = 0$, $\mathbf{u}_i = \mathbf{e}_k$, and $\mathbf{v}_i = \mathbf{e}_l$, where $\mathbf{e}_j$ is a vector with a 1 at the jth entry and 0 elsewhere, then the constraint $\mathbf{e}_k^T \mathbf{A} \mathbf{e}_l = 0$ corresponds to the kl-entry of $\mathbf{A}$ is 0. We are interested in $\hat{\mathbf{A}}^*$ which is the solution of the following optimization problem:

$$\hat{\mathbf{A}}^* = \underset{\substack{\mathbf{A} \\ \mathbf{u}_i^T \mathbf{A} \mathbf{v}_i = \gamma_i, 1 \le i \le I}}{\arg\min} \int \|(\mathbf{X}_0(\theta) - \mathbf{G}\mathbf{A}\mathbf{h}(\theta)\|^2 \, d\theta.$$

The derivative of the criterion with respect to $\mathbf{A}$ is given by

$$\begin{aligned} & \frac{d}{d\mathbf{A}} \int (\mathbf{X}_0(\theta) - \mathbf{G}\mathbf{A}\mathbf{h}(\theta))^T (\mathbf{X}_0(\theta) - \mathbf{G}\mathbf{A}\mathbf{h}(\theta)) d\theta \\ = & \frac{d}{d\mathbf{A}} \int -2\mathbf{X}_0^T(\theta)\mathbf{G}\mathbf{A}\mathbf{h}(\theta) + \mathbf{h}^T(\theta)\mathbf{A}^T\mathbf{G}^T\mathbf{G}\mathbf{A}\mathbf{h}(\theta) d\theta \\ = & -2\mathbf{G}^T \int \mathbf{X}_0(\theta)\mathbf{h}^T(\theta) d\theta + 2\mathbf{G}^T\mathbf{G}\mathbf{A} \int \mathbf{h}(\theta)\mathbf{h}^T(\theta) d\theta. \end{aligned}$$

Then the normal equations for the above optimization are given by

$$\mathbf{G}^T\mathbf{G}\hat{\mathbf{A}}^* \int \mathbf{h}(\theta)\mathbf{h}^T(\theta) d\theta + \sum_{j=1}^{I} \mathbf{u}_j \lambda_j \mathbf{v}_j^T = \mathbf{G}^T \int \mathbf{X}_0(\theta)\mathbf{h}^T(\theta) d\theta, \tag{3.6}$$

$$\mathbf{u}_i^T \hat{\mathbf{A}}^* \mathbf{v}_i = \gamma_i, \quad i = 1, \ldots, I, \tag{3.7}$$

where λ_j are the Lagrange multipliers. Equation (3.6) can be expressed in terms of the solution $\hat{\mathbf{A}}$

of the unconstrained model in the following way:

$$\hat{\mathbf{A}}^* = \hat{\mathbf{A}} - (\mathbf{G}^T\mathbf{G})^{-1}\sum_{j=1}^{I}\mathbf{u}_j\lambda_j\mathbf{v}_j^T\left(\int \mathbf{h}(\theta)\mathbf{h}^T(\theta)d\theta\right)^{-1}.$$

Plugging it into equation (3.7), we obtain

$$\sum_{j=1}^{I}\lambda_j\cdot\left[\mathbf{u}_i^T(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{u}_j\mathbf{v}_j^T\left(\int \mathbf{h}(\theta)\mathbf{h}^T(\theta)d\theta\right)^{-1}\mathbf{v}_i\right] = \mathbf{u}_i^T\hat{\mathbf{A}}\mathbf{v}_i - \gamma_i,$$

for $i = 1, \ldots, I$. This is a system of linear equations with I equations and I unknowns. Assuming that the λ_j can be solved, the solution $\hat{\mathbf{A}}^*$ of the constrained optimization is given by

$$\hat{\mathbf{A}}^* = \hat{\mathbf{A}} - \sum_{j=1}^{I}\lambda_j\cdot\left[(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{u}_j\mathbf{v}_j^T\left(\int \mathbf{h}(\theta)\mathbf{h}^T(\theta)d\theta\right)^{-1}\right].$$

We can check that this solution indeed satisfies the constraints by observing that

$$\begin{aligned}\mathbf{u}_i^T\hat{\mathbf{A}}^*\mathbf{v}_i &= \mathbf{u}_i^T\hat{\mathbf{A}}\mathbf{v}_i - \sum_{j=1}^{I}\lambda_j\cdot\left[\mathbf{u}_i^T(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{u}_j\mathbf{v}_j^T\left(\int \mathbf{h}(\theta)\mathbf{h}^T(\theta)d\theta\right)^{-1}\mathbf{v}_i\right]\\ &= \mathbf{u}_i^T\hat{\mathbf{A}}\mathbf{v}_i - (\mathbf{u}_i^T\hat{\mathbf{A}}\mathbf{v}_i - \gamma_i)\\ &= \gamma_i.\end{aligned}$$

The goal now is to decide which basis functions to use for each α_i. We will do it in two stages. Let N be the number of elements in $\mathbf{A}$. As mentioned before, $N = KLp + L$ is the total number of basis functions used to describe the α_i since each entry $\mathbf{A}$ is the coefficient of a basis function. Firstly, for each $1 \leq n \leq N$, we pick the n basis functions that best fit the data and constrain the coefficients that correspond to unchosen basis functions to have zero values in $\mathbf{A}$. These n basis functions are chosen in a forward stepwise fashion. Secondly, we choose the optimal number of basis functions according to prediction errors through a learning technique.

We will now describe our iterative algorithm to select n basis functions for each $1 \leq n \leq N$. Let $\hat{\mathbf{A}}_n^*$ be the solution calculated using our algorithm for each n. Since n basis functions are selected for $\hat{\mathbf{A}}_n^*$, this matrix has n non-zero entries and the other $N - n$ entries are restricted to be zero. In

particular, $\hat{\mathbf{A}}_0^*$ is the zero matrix that can be viewed as the solution with constraints that all entries are zero, or that none basis functions are chosen. The algorithm is given as follows.

1. For $n = 0, 1, \ldots, N-1$, do the following two steps.
2. Start from $\hat{\mathbf{A}}_n^*$. For each of its zero entries, say A_{ij}, remove the constraint that $A_{ij} = 0$ and calculate the solution $\hat{\mathbf{A}}_{(n+1)ij}^*$.
3. Let $\hat{\mathbf{A}}_{n+1}^*$ be the $\hat{\mathbf{A}}_{(n+1)ij}^*$ that minimizes the integral of the squared residual function. That is,
$$\hat{\mathbf{A}}_{n+1}^* = \underset{\hat{\mathbf{A}}_{(n+1)ij}^*}{\arg\min} \int \left\| \mathbf{X}_0(\theta) - \mathbf{G}\hat{\mathbf{A}}_{(n+1)ij}^* \mathbf{h}(\theta) \right\|^2 d\theta.$$

The idea behind the above algorithm is that at each step, we look for the basis function which, if included, would result in the smallest fitting error. The algorithm is done in a forward stepwise fashion: the choices of the basis functions for all $1 \leq n \leq N$ are found in one pass. Theoretically, one can of course search for the subset of n basis functions that results in the smallest fitting error by looking at all $N!/(n!(N-n)!)$ combinations. However, in functional autoregression $N(= KLp + L)$ is generally not a small number that calculating the solutions of all combinations is not easy.

It now remains to choose the total number of basis functions n to be used in the model. This number will be chosen by looking at a measure of the prediction performance when different numbers of basis functions are used. The algorithm depends on two parameters: r and δ. r is the number of (historical) periods in which we would assess the prediction performance. δ is a small positive constant which will be described below. Recall that T is the number of time periods of the data (or the index of the last period). The algorithm is given below.

1. Set $n = 0$.
2. For each time $t = T, T-1, \ldots, T-r+1$, do steps 3 and 4.
3. Use a moving window (or data from the beginning) of historical data relative to time t (not including itself) to get a 1-step ahead forecast $x_{t-1,n}^{(1)}(\theta)$ of $x_t(\theta)$ using the functional autoregressive model with n basis functions chosen using the method described above.

4. Evaluate the out-of-sample squared perdiction error

$$\rho_n = \sum_{t=T-r+1}^{T} \int \left(x_t(\theta) - x_{t-1,n}^{(1)}(\theta) \right)^2 d\theta.$$

5. If $\rho_n > \rho_{n-1} + \delta$ or $\rho_n > \rho_{n-1}(1+\delta)$, where δ is a small positive number, set the optimal number of basis functions to be $n-1$; otherwise, increment n by 1 and repeat steps 2 to 4.

In the above algorithm, our measure of the prediction error is the 1-step ahead prediction error of the model. This is similar to cross-validation in that the data is divided into training and test sets; the model is applied to the training set and is assessed on the test set. If $r = 1$, then the model is applied on the data in all but the last period; a 1-step ahead perdiction is obtained and is compared with the data in the last period. The sequence $\rho_1, \ldots, \rho_N$ is expected to be first decreasing, and then after some point, it will either increase or fluctuate. One can strategically choose the optimal n by looking at the sequence ρ_n, and we have provided two ways in point 5 above. The idea behind this algorithm comes from the intuition that the optimal number of basis functions used for data $x_1(\theta), \ldots, x_T(\theta)$ and for data $x_1(\theta), \ldots, x_{T-1}(\theta)$ should be similar. Therefore, if we are able to get the optimal number of basis functions for the latter set of data, we can use that as an estimate of the optimal number of basis functions for the former set of data.

3.7 Autoregressive Functional Exogenous Model (ARFX)

In this section we introduce the autoregressive functional exogenous model (FARX). This model relates every observation not only with past observations of the same series (autoregressive), but also with present and past observations of another time series (exogenous series). An application of the model is given in the next chapter, where we try to model the time series of yield curves, with the federal funds rate as the exogenous time series.

For simplicity, we will only incorporate the present value of the exogenous time series into the model. It is straightforward to generalize it to include past values of the exogenous time series. The way that we include the exogenous time series into the model depends on whether it is univariate, multivariate, or functional. In all three cases, the exogenous time series is included into the model

in a linear fashion.

3.7.1 Models

Univariate Exogenous Variables

Let y_t be a univariate time series. The autoregressive model with exogenous variable y_t is given by

$$x_t(\theta) = \alpha_0(\theta) + \sum_{i=1}^{p} \int \alpha_i(\vartheta, \theta) x_{t-i}(\vartheta) d\vartheta + \beta_1(\theta) y_t + \epsilon_t(\theta).$$

Multivariate Exogenous Variables

Let $\mathbf{y}_t$ be a m-dimensional multivariate time series. The autoregressive model with exogenous variable $\mathbf{y}_t$ is given by

$$x_t(\theta) = \alpha_0(\theta) + \sum_{i=1}^{p} \int \alpha_i(\vartheta, \theta) x_{t-i}(\vartheta) d\vartheta + \sum_{i=1}^{m} \beta_i(\theta) y_{it} + \epsilon_t(\theta). \tag{3.8}$$

Functional Exogenous Variables

Let $y_t(\theta)$ be a functional time series. The autoregressive model with exogenous variable $y_t(\theta)$ is given by

$$x_t(\theta) = \alpha_0(\theta) + \sum_{i=1}^{p} \int \alpha_i(\vartheta, \theta) x_{t-i}(\vartheta) d\vartheta + \int \beta(\vartheta, \theta) y_t(\vartheta) d\vartheta + \epsilon_t(\theta). \tag{3.9}$$

3.7.2 Solving for the Solutions

In all of the above cases, the solutions can be solved using basis functions approach similar to the one we saw in the last chapter. We assume the same basis function expansions for $\alpha_0(\theta)$ and $\alpha_i(\vartheta, \theta)$, $i = 1, \ldots, p$, and the same definitions for $\mathbf{A}$, $\mathbf{X}$, and $\mathbf{G}$. Since the univariate case is just a specific case of the multivariate, we will only solve the cases where the exogenous time series is multivariate or functional.

Solutions for the Multivariate Case

Suppose each $\mathbf{y}_t$ is m-dimensional and denote $\mathbf{y}_t = [y_{1t}, \ldots, y_{mt}]^T$. Define $\mathbf{Y}_i$ to be the following vector:

$$\mathbf{Y}_i = \begin{bmatrix} y_{i,T} \\ y_{i,T-1} \\ \vdots \\ y_{i,p+1} \end{bmatrix},$$

and assume $\beta_i(\theta) = \mathbf{b}_i^T \mathbf{h}(\theta)$ has the same basis function expansions as $\alpha_0(\theta)$. Then model (3.8) can be written as

$$\begin{aligned} \mathbf{X}_0(\theta) &= \mathbf{GAh}(\theta) + \sum_{i=1}^{m} \mathbf{Y}_i \mathbf{b}_i^T \mathbf{h}(\theta) + \boldsymbol{\epsilon}_t \\ &= \mathbf{F}_1 \mathbf{C}_1 \mathbf{h}(\theta) + \boldsymbol{\epsilon}_t, \end{aligned}$$

where

$$\mathbf{F}_1 = \begin{bmatrix} \mathbf{G} & \mathbf{Y}_1 & \cdots & \mathbf{Y}_m \end{bmatrix}, \quad \mathbf{C}_1 = \begin{bmatrix} \mathbf{A} \\ \mathbf{b}_1^T \\ \vdots \\ \mathbf{b}_m^T \end{bmatrix}.$$

Hence, the model is reduced into the same form that we had for the autoregressive functional model in (3.4). It could then be solved by the same approach. The solution is give by

$$\hat{\mathbf{C}}_1 = \left(\mathbf{F}_1^T \mathbf{F}_1\right)^{-1} \int \mathbf{X}_t(\theta) \mathbf{h}^T(\theta) d\theta \left(\int \mathbf{h}(\theta) \mathbf{h}^T(\theta) d\theta \right)^{-1}.$$

Solutions for the Functional Case

Define $\mathbf{Y}_0(\theta)$ by

$$\mathbf{Y}_0(\theta) = \begin{bmatrix} y_T(\theta) \\ y_{T-1}(\theta) \\ \vdots \\ y_{p+1}(\theta) \end{bmatrix},$$

and assume $\beta(\vartheta, \theta) = \mathbf{g}^T(\vartheta)\mathbf{B}\mathbf{h}(\theta)$ has similar basis function expansions as α_i for $i \geq 1$. Then model (3.9) can be written as

$$\begin{aligned} \mathbf{X}_0(\theta) &= \mathbf{G}\mathbf{A}\mathbf{h}(\theta) + \int \mathbf{Y}_0(\vartheta)\mathbf{g}^T(\vartheta)d\vartheta\mathbf{B}\mathbf{h}(\theta) + \boldsymbol{\epsilon}_t \\ &= \mathbf{F}_2\mathbf{C}_2\mathbf{h}(\theta), \end{aligned}$$

where

$$\mathbf{F}_2 = \begin{bmatrix} \mathbf{G} & \int \mathbf{Y}_0(\vartheta)\mathbf{g}^T(\vartheta)d\vartheta \end{bmatrix}, \quad \mathbf{C}_2 = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}.$$

Again, $\hat{\mathbf{C}}_2$ is obtained in a similar way:

$$\hat{\mathbf{C}}_2 = \left(\mathbf{F}_2^T\mathbf{F}_2\right)^{-1} \int \mathbf{X}_0(\theta)\mathbf{h}^T(\theta)d\theta \left(\int \mathbf{h}(\theta)\mathbf{h}^T(\theta)d\theta\right)^{-1}.$$

3.7.3 Forecasting

The forecasting of the time series at a time in the future for the functional autoregressive exogenous model is just the same as the forecasting for the functional autoregressive model. The only difference is that we must assume that future values of the exogenous time series is known when doing predictions. Of course, we can also apply a time series model on the exogenous time series independently to obtain estimates of its future values, and then use them to get forecasts of the time series $x_t(\theta)$. Alternatively, we can model the functional exogenous model so that it uses only past values of the exogenous time series. Mathematically, the 1-step ahead forecast is given by

$$x_T^{(1)}(\theta) = \tilde{\mathbf{F}}\hat{\mathbf{C}}\mathbf{h}(\theta),$$

where in the case when the exogenous time series is univariate or multivariate,

$$\tilde{\mathbf{F}} = \begin{bmatrix} \tilde{\mathbf{G}} & y_{1T} & \cdots & y_{mT} \end{bmatrix}, \quad \hat{\mathbf{C}} = \hat{\mathbf{C}}_1,$$

and in the case when the exogenous time series is functional,

$$\tilde{\mathbf{F}} = [\ \tilde{\mathbf{G}} \quad \int y_T(\vartheta)\mathbf{g}^T(\vartheta)d\vartheta\ ], \quad \hat{\mathbf{C}} = \hat{\mathbf{C}}_2.$$

$\tilde{\mathbf{G}}$ in both cases is the same as before. Multistep ahead forecasts are obtained in the same way by repeatedly applying the 1-step ahead forecast procedure.

3.8 Pointwise Autoregressive Functional Model (PARF)

While the functional autoregressive model can be considered to be a more general version of the vector autoregressive model, in this section we develop an alternate time series model for functional time series $x_t(\theta)$. The model is again a linear relationship between each observation and past observations, but the linearity is expressed in a different way that for $\theta_0 \in \Theta$, $x_t(\theta_0)$ is related only to $x_{t-j}(\theta_0)$ for a number of positive j. This model is referred to as the pointwise autoregressive functional model (PARF).

3.8.1 Model

A pointwise autoregressive model for functional data with order p is given by

$$x_t(\theta) = \alpha_0(\theta) + \sum_{s=1}^{p} \alpha_s(\theta)x_{t-s}(\theta) + \epsilon_t(\theta), \tag{3.10}$$

where $\epsilon_t(\theta)$ is a sequence of uncorrelated random processes that satisfies the following conditions:

1. $E(\epsilon_t(\theta_0)) = 0$ for every $\theta \in \Theta$.
2. $E(\epsilon_t(\theta_1)\epsilon_t(\theta_2)) = \sigma(\theta_1, \theta_2)$ for $\theta_1, \theta_2 \in \Theta$ (covariance function).
3. $E(\epsilon_{t_1}(\theta_1)\epsilon_{t_2}(\theta_2)) = 0$ for $t_1 \neq t_2$ and $\theta_1, \theta_2 \in \Theta$.

One can think of the model as having a univariate autoregressive model for every $\theta_0 \in \Theta$. Since there is no interaction between different values of θ, the coefficients α_i, $i = 1, \ldots, p$, are functions of one paraemters in our case, as opposed to functions of two parameters in the autoregressive functional model. Also, integrals are not involved. In fact, this pointwise model is a special case

of the functional autoregressive model. Let the coefficients of the the two models be $\alpha_i^{FAR}(\vartheta,\theta)$ and $\alpha_i^{PFAR}(\theta)$. If $\alpha_i^{FAR}(\vartheta,\theta)=\alpha_i^{PFAR}(\theta)\delta(\vartheta-\theta)$, where $\delta(\cdot)$ is the Dirac delta function with the property that $\int_{-\infty}^{\infty} f(x)\delta(x-x_0)dx=f(x_0)$ for any continuous function f on $\mathbb{R}$, then then the FAR becomes a PFAR since

$$\int \alpha_i^{FAR}(\vartheta,\theta)x_{t-i}(\vartheta)d\vartheta=\alpha_i^{PFAR}(\theta)\int x_{t-i}(\vartheta)\delta(\vartheta-\theta)d\vartheta=\alpha_i^{PFAR}(\theta)x_{t-i}(\theta).$$

While the Dirac delta can be considered to be a measure which places a point mass at 0, it is not a true function. The solutions of this model can not be solved using the same procedure as in the autoregressive functional model. Hence, it is worth to look at this model separately.

3.8.2 Solving for the Solutions

Given functional observations $x_1(\theta),\ldots,x_T(\theta)$, model (3.10) can be expressed in matrix form

$$\mathbf{y}(\theta)=\mathbf{X}(\theta)\boldsymbol{\alpha}(\theta)+\boldsymbol{\epsilon}(\theta),$$

where

$$\mathbf{y}(\theta)=\begin{bmatrix} x_T(\theta)\\ x_{T-1}(\theta)\\ \vdots\\ x_{p+1}(\theta)\end{bmatrix},\mathbf{X}(\theta)=\begin{bmatrix} 1 & x_{T-1}(\theta) & x_{T-2}(\theta) & \cdots & x_{T-p}(\theta)\\ 1 & x_{T-2}(\theta) & x_{T-3}(\theta) & \cdots & x_{T-p-1}(\theta)\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & x_p(\theta) & x_{p-1}(\theta) & \cdots & x_1(\theta)\end{bmatrix},\boldsymbol{\epsilon}(\theta)=\begin{bmatrix}\epsilon_T(\theta)\\ \epsilon_{T-1}(\theta)\\ \vdots\\ \epsilon_{p+1}(\theta)\end{bmatrix}.$$

The goal is to minimize the sum of the squares of the error functions, integrated over Θ:

$$C=\int \|\mathbf{y}(\theta)-\mathbf{X}(\theta)\boldsymbol{\alpha}(\theta)\|^2\,d\theta.$$

One way of solving the solutions of this model is to discretize it into points $\theta_1<\theta_2<\cdots<\theta_n$ and then treat it as a univariate autoregressive model for each θ_i to solve for the coefficients. The functional coefficients can be recovered by applying any convenient methods such as splines approximation or interpolation. However, a drawback of this approach is that n, which is the

number of points we have discretized the interval into, is the number of equations we have to solve. Hence, using larger values of n means that more computational efforts are needed. Also, while we expect that the contribution of $x_{t-s}(\theta)$ on $x_t(\theta)$ does not change a lot for a small change in θ, or that the coefficients $\alpha_s(\theta)$ are smooth, it may not always be the case if discretization is used. An alternative way of estimating the solutions is to use the basis functions approach. This approach is less computational intensive, and it will ensure that the coefficients are smooth.

Suppose the α_s have the form

$$\alpha_s(\theta) = \sum_{i=1}^{N_s} a_{si} h_{si}(\theta) = \mathbf{h}_s^T(\theta)\mathbf{a}_s,$$

and define $\mathbf{a}$ and $\mathbf{H}$ by

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_p \end{bmatrix}, \qquad \mathbf{H} = \begin{bmatrix} \mathbf{h}_0^T & 0 & \cdots & 0 \\ 0 & \mathbf{h}_1^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{h}_p^T \end{bmatrix}.$$

Then $\mathbf{y}(\theta)$ can be written as

$$\mathbf{y}(\theta) = \mathbf{X}(\theta)\mathbf{H}(\theta)\mathbf{a} + \boldsymbol{\epsilon}(\theta),$$

and the solution is obtained by taking derivative to C with respect to $\mathbf{a}$ and then set it to zero. It is given by

$$\hat{\mathbf{a}} = \left[\int \mathbf{H}^T(\theta)\mathbf{X}^T(\theta)\mathbf{X}(\theta)\mathbf{H}(\theta)d\theta\right]^{-1} \left[\int \mathbf{H}^T(\theta)\mathbf{X}^T(\theta)\mathbf{y}(\theta)d\theta\right].$$

3.8.3 Forecasting

Following the same idea as the autoregressive models introduced earlier, the 1-step ahead forcast $x_T^{(1)}(\theta)$ is given by

$$\begin{aligned} x_T^{(1)}(\theta) &= \hat{\alpha}_0(\theta) + \sum_{s=1}^{p} \hat{\alpha}_s(\theta) x_{T+1-s}(\theta) \\ &= \mathbf{x}^T(\theta)\hat{\boldsymbol{\alpha}}(\theta) = \mathbf{x}^T(\theta)\mathbf{H}(\theta)\hat{\mathbf{a}}, \end{aligned}$$

where $\mathbf{x}^T(\theta)$ is given by

$$\mathbf{x}^T(\theta) = \left[\begin{array}{cccc} 1 & x_T(\theta) & x_{T-1}(\theta) & \cdots & x_{T+1-p}(\theta) \end{array} \right].$$

The l-step ahead forcast $x_T^{(l)}(\theta)$ is obtained by repeatedly applying the 1-step ahead forecast procedure above.

Chapter 4

Empirical Study

4.1 Yield Curve Data

The time series of yield curves that we will analyze consists of daily U.S. Treasury zero-coupon yield curves for each of the 1227 trading days between 6/28/2004 to 3/10/2009 with maturities 0.25, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 years. Every curve gives the relationship between the yields of zero-coupon bonds and their maturities. The zero-coupon yield curves, which are obtained directly from Bloomberg, are bootstrapped from the coupon yield curves. Bootstrapping is an iterative process that solves for the prices of zero-coupon bonds one-by-one from coupon bonds. This procedure is needed because in the market, bonds with long maturities (for example, longer than 1 year) generally have coupon payments.

Applying the nonparametric splines method described in Section 2.2, we transform the yield curves from multivariate to functional form. The splines are taken to be B-spline basis functions of order 3 with two internal knots $\{3, 7\}$ years and boundary knots at 0 and 10 years.

When a time series model is used to analyze a time series, it is often required that the time series is weakly stationary. Such a time series has mean and autocorrelation that do not vary with respect to time. However, many financial time series such as stock prices or interest rates tend to be nonstationary (which is called unit-root nonstationary). One method that is used to convert a nonstationary time series to a stationary one is by differencing. We apply the same idea to our data.

That is, we consider a first difference to the time series of yield curves in the time dimension. The augmented Dickey-Fuller test, which is a statistical test for a unit-root in a time series sample, with lag 2 is applied to the differenced yield curves at maturities 0.25, 0.5, 1, ..., 10 years. All 12 tests are rejected at 1% level. This suggests that no unit-root is present in each of the 12 time series.

Figure 4.1 shows the yield curve data and Figure 4.2 shows the differenced yield curve data. The yield curve starts out from a low level in mid 2004; it increases to a high level in around 2006 before it decreases again afterwards. The zero-coupon yield curve level is affected by the federal funds rate, which is the interest rate that banks charge for each other for overnight loans of reserve balances. The level of the federal funds rate reflects the economic condition. In the past, the Federal Reserve responded to potential slow-down during recessions by lowering the target federal funds rate to stimulate economic growth, and for opposite reasons, it raised the target federal funds rate during boom to slow down economic growth. There is a close relationship between the federal funds rate and the yield curve. The low level of the yield curve in around mid-2004 is the consequence of the global economic recession between 2000-2003, during which the target federal funds rate was reduced from 6.5% to 1%. The economy recovers from the recession afterwards and the target federal funds rate was gradually increased to 5.25% in 2007, which is where the yield curve reaches its highest level. Due to the financial crisis in 2007-2010 which was triggered by the subprime mortgage crisis, the target federal funds rate was reduced to $0 - 0.25\%$ in late 2008; this is where yield curve level dropped again.

Apart from the general level of the yield curve across all maturities, yields with different maturities do show different characteristics. Yields with longer maturities tend to be more stable than yields with shorter maturities.

4.2 Yield Curve Prediction Performance

In this section, we look at the prediction performance of the functional time series models compared with the performance of some benchmark models on the differenced yield curve data. To do this, the data is divided into 39 disjoint moving windows. Each window consists of 30 trading days of functional data. For each of the moving windows, every time series model is fitted to the data to obtain estimates of the coefficients. Using these estimates, one-day ahead forecasts $x_t^{(1)}(\theta)$ are

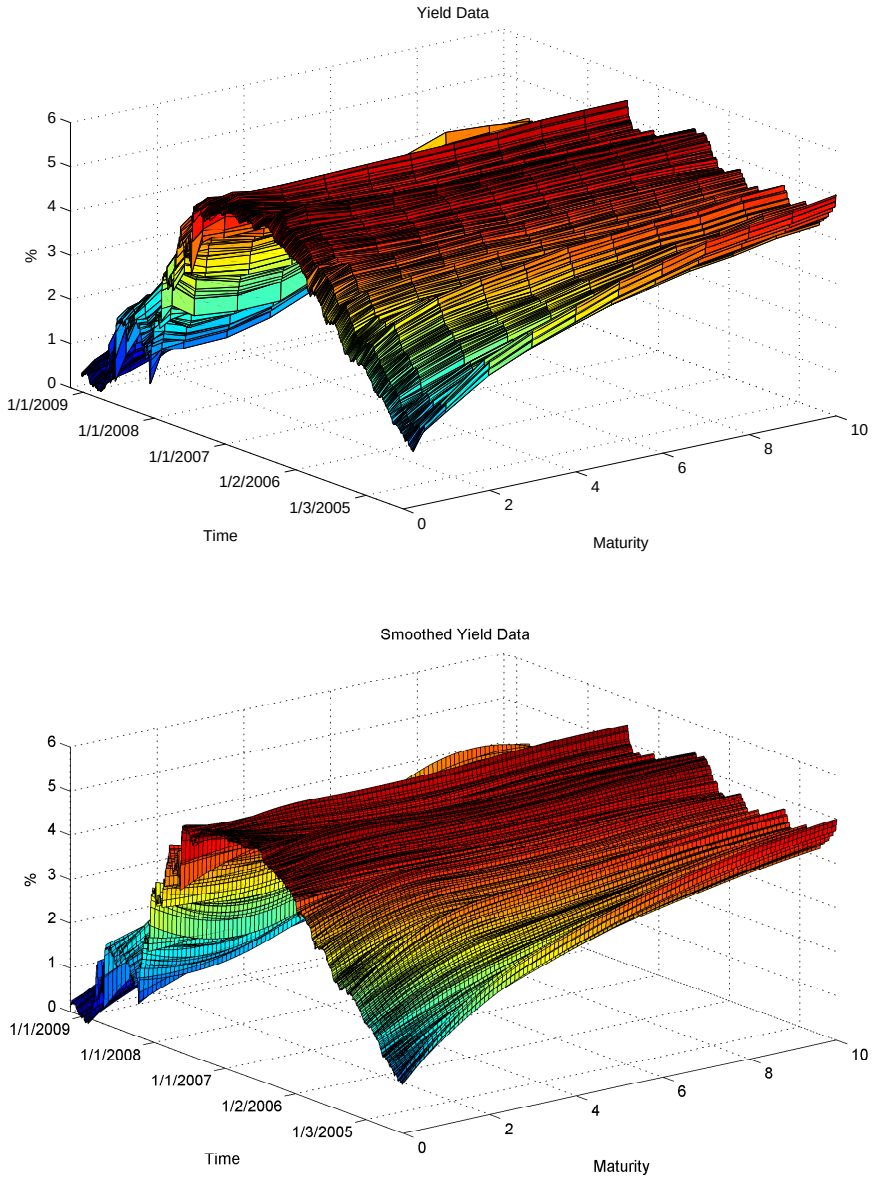

Figure 4.1: U.S. Treasury zero-coupon bond yield curves from Jun 28, 2004 to Mar 10, 2009. The top graph is the original data and the bottom graph is obtained by applying nonparametric splines method on the original data.

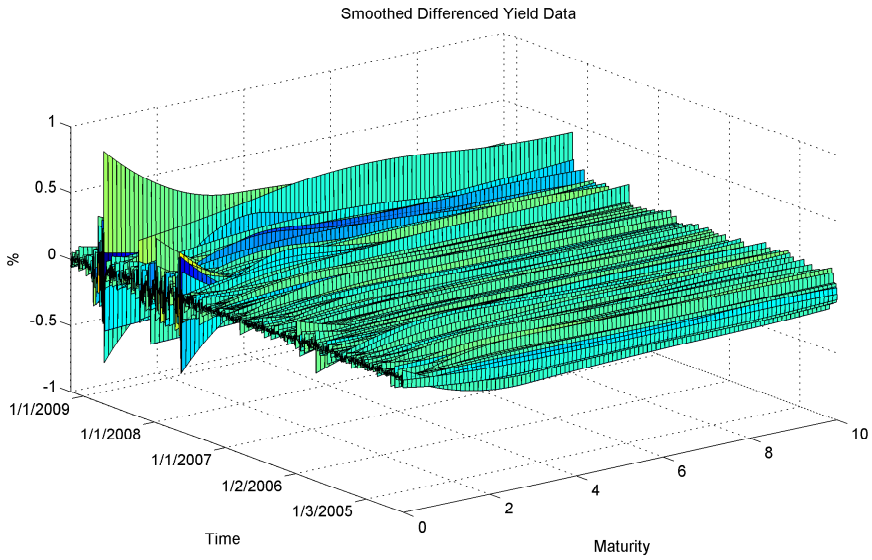

Figure 4.2: Differenced U.S. Treasury zero-coupon bond yield curves from Jun 28, 2004 to Mar 10, 2009. Obtained by differencing the smoothed yield data of Figure 4.1.

calculated and are compared with the actual observation $x_{t+1}(\theta)$ on the day after the end of the window. Instead of calculating the error as the integral of the absolute difference between $x_t^{(1)}(\theta)$ and $x_{t+1}(\theta)$ integrated over the interval $[0, 10]$ (which is a natural choice to compare two differenced yield curves over all maturities), we will use a discretized version of this error because some of the benchmark models are univariate or multivariate models.

$$error = \sum_{i=1}^{12} \left| x_{t+1}(\theta_i) - x_t^{(1)}(\theta_i) \right|, \tag{4.1}$$

where $(\theta_1, \ldots, \theta_{12}) = (0.25, 0.5, 1, 2, \ldots, 10)$ years. That is, the sum is taken over the maturities where the data are originally available. Since these maturities span the interval $[0, 10]$ quite evenly, using this discretized error over the integrated error should not have significant impact on the relative performance of the time series models.

The prediction performance of the following 7 models are compared:

- "Naive": A prediction approach that takes the current observation as the 1-day ahead forecast.
- ARF: Autoregressive functional model of order 2.

- ARF (B.S.): Autoregressive functional model of order 2 with basis selection. The number of basis functions is determined separately for each moving window using the algorithm described in Section 3.6 with $r = 1$.
- ARFX: Autoregressive functional exogenous model of order 2. The exogenous time series is the univariate time series of differenced effective federal funds rate.
- PARF (B): Pointwise autoregressive functional model of order 2. Basis functions approach is used to solve for the coefficients.
- PARF (D): Pointwise autoregressive functional model of order 2. Discretization approach is used to solve for the coefficients.
- VAR: Vector autoregressive model of order 2.

The reason that the orders are all chosen to be 2 is to keep the number of parameters low. In particular, a VAR(p) model applied on a 12-dimensional data has 300 parameters.

4.2.1 Benchmark Models

Errors of one-day ahead forecasts for the models in each of the 39 moving windows are reported in Table 4.1. Among the 7 time series models, the "naive", PARF (D), and VAR models serve as benchmark models because they are existing models applied on the data set. The "naive" model is the most basic among all because it does not require any sophiscated model or parameter estimation method. The idea behind this approach is that in order to predict the observation on the next day, the most basic guess is to use the observation on the present day. One can of course take a weighted average of the observations in the past few days as the forecast, which would possibly do a better job in predicting the next observation. However, determining the length of the period and the weight function may not be easy. The "naive" model would work well if changes in the differenced yield curves between consecutive days are not significant, and it would not work so well otherwise. Alternatively, by looking at the errors produced by this model, one can see when the changes in the differenced yield curves are large. From Table 4.1, it can be seen clearly that the data can be divided into two periods according to the volatility of the differenced yield curves. From window 1

to window 25 (July 2004 to July 2007), the errors are mostly in the order of 10^{-1}. Occasional spikes exist but it drops to normal level fairly quickly. However, its error in window 26 jumps to 2.0610 and this level of error occur quite frequently in subsequent moving windows. High volatility in yield curve movement can occur in unstable economic or political enviroment in which the economy faces more uncertainty. In fact, as mentioned before, the latter period corresponds to the financial crisis in 2007-2010 which was triggered by the subprime mortgage crisis.

Models perform differently in windows 1-25, which we will refer to as the stable economy period, and in windows 26-39, which we will refer to as the unstable economy period. Errors tend to be larger in unstable period then in stable period. Table 4.2 gives a summary of the prediction errors in windows 1-25 and Table 4.3 gives a summary of the prediction errors in windows 26-39.

The VAR model is another benchmark model that is considered in this study. In the literature, VAR has been a popular model for interest rate data. For example, Tsay (2002) applied a VARMA(2,1) to a bivariate time series of US monthly interest rate with maturities 1 and 3 years from April 1953 to January 2001. They observed a unidirectional linear relationship from the 3-year rate to the 1-year rate. The application of VAR to interest rate data in our study is different from Tsay's work in nature because we are concerned with the prediction performance. We applied VAR(2) model to the original 12-dimensional multivariate time series of bond yields. Although one would not expect it to perform badly, its results shown in Table 4.1 are quite disappointing. It produces the largest errors among the models in more then three-fourth of the times and its error is usually about at least 5 to 6 times the errors of the next largest error. It is rather surprising it does not do as good as the "naive" prediction. However, its bad prediction performance does not imply that there is no autoregressive structure in the difference yield curve data because there are other autoregressive models that do better. One explanation is that it has too many parameters in its model and bad prediction performance is the result of overparametrization. It is mentioned earlier that the VAR(2) model has 300 parameters. Another possible explanation is due to high dimensionality of the data. Since every functional observation of differenced yield curve is smooth, the univariate time series of data at two adjacent maturities (for example, 0.25 and 0.5 years) could be nearly identical. Since VAR is solved by applying least squares method to a linear regression model, if the regression matrix is given by X, then having two or more dimensions of the time series

to be nearly linearly dependent would mean that the matrix $X^T X$ is nearly singular. This could result in inaccuracy in its inversion. Hence, the solutions of the VAR model are affected. This could be a reason why VAR performs badly against other models.

The solutions of PARF (D) are obtained by discretizing the interval [0,10] into the 12 maturities 0.25, 0.5, 1, ..., 10 and then applying a univariate autoregressive model of order 2 to the time series at each maturity. It has the simplicity of not having to consider the covariance structure of different dimensions of the multivariate time series of differenced bond yields which saves a lot of parameters from models such as the VAR model. However, its performance is still not as good as the "naive" model. This is more apparent in unstable economy period, where the error of PARF (D) is often 2 to 3 times the error of the "naive" model. This suggests that treating the multivariate bond yield data as 12 univariate time series may not be appropriate, especially during times of unstable economy.

4.2.2 Autoregressive Functional Models

We start by discussing PARF (B), which is the same model as PARF (D), but it uses basis functions approach instead of discretization approach to solve for model parameters. Although the PARF model (3.10) is stated so that yields of different maturities have no interaction with each other, when basis functions approach is used, the time series of the differenced yield curves at each maturity is influenced by those at neighboring maturities. This is because the basis functions are assumed to have certain forms and can not move completely free across different maturities when they are being solved. It is seen from Table 4.1 that the prediction performance of PARF (B) is better than that of the benchmark models by substantial amounts in terms of the summary statistics. In particular, it is interesting to see that all of its errors are very close to half of the errors of PARF (D) in same windows. This suggests that correlations between yields with different maturities do exist and one should include them in modeling the yield data.

The fact that correlations between yields with different maturities exist can be seen from the differenced yield data shown in Figure 4.2. Particular patterns are observed. For example, many of the differenced yield curves do not cross the zero level. That is, either the whole differenced yield curve is positive or the whole curve is negative, which corresponds to a positive or negative change to the whole yield curve. This type of change is common and could be due to news, economic events

or policies such as change in the expected inflation rate that affect the market in both short and long runs. In fact, many of the differenced yield curves, especially during stable economic period, are close to constant across all maturities, not only having the same sign. These correspond to parallel shifts in the yield curve.

The prediction performance of PARF (B) depends on which period it is predicting. During stable economic period, it is the clear winner among all models in terms of all five summary statistics. Its average error is approximately two-third of the next largest average error. However, during unstable economic period, its performance is similar to that of the "naive" model, and it does not perform as good as other autoregressive functional models.

In general, ARF does have a better prediction performance than the benchmark models, but similar to other models, its performance varies depending on which period it predicts. During stable economic period, it is seen from Table 4.2 that although ARF gives a slightly better overall prediction performance than the "naive" model, no evidence suppports that the former consistently outperforms the latter in all moving windows. However, the unstable economic period is where it stands out from the benchmark models. During this period, it produces small errors quite consistently with error standard deviation 0.3970, which is the smallest among all models, and maximum error 1.8801, which is also the smallest. Although it might not be the best model in terms of prediction during the unstable period, its performance is acceptable given the high volatility of the differenced yield curves.

The ARF model is meant to be the primitive model where more sophisticated methods, ARF (B.S) and ARFX, are based on. We did not expect it to be the best performing model, but it is enough to see it does no worse than the benchmark models and better during some times. ARF (B.S.) is the same model as ARF but it aims to improve prediction performance by choosing only a subset of the basis functions used for ARF in a forward stepwise fashion. We can see that during stable economic period, selecting basis using the algorithm provided in Section 3.6 improves prediction accuracy by a notable amount over ARF in many moving windows. However, it is still slightly behind PARF (B) if we consider the mean, 0.3359 vs 0.2614, and the standard deviation, 0.2371 vs 0.2038. During unstable economic period, unfortunately, ARF (B.S.) does not carry any prediction improvements to ARF. It produces some large prediction errors and the maximum error is 3.8953

compared to ARF's 1.8801. One explanation to this phenomenon is to recall that the idea behind chooing the number of basis functions comes from the intuition that the optimal number of basis functions used for data $x_1(\theta), \ldots, x_{T-1}(\theta)$ and for data $x_1(\theta), \ldots, x_T(\theta)$ should be similar. However, this could be violated during times of high volatility.

The ARFX model is an ARF model with an exogenous time series which is the differenced effective federal funds rate. As mentioned previously, the federal funds rate is the interest rate at which banks lend federal balance at the Federal Reserve to each other overnight. Since banks can negotiate with each other about the interest rate when loans are made, the actual interest rate used might be different in different transactions, and the weighted average of this rate across all transactions is called the effective federal funds rate. The motivation behind incorporating this exogenous time series into the model is that it is known to have significant impact on short-term yields. This is because banks usually use borrowed funds to hold inventory of notes and bonds, which they plan to re-sell to customers. Hence, in general, banks are unwilling to hold these securities if the securities yield less than the federal funds rate. Besides the federal funds rate, of course there are other factors such as news, economic events, or investors' psychology that affect the level of the yield curve. However, these factors are hard to quantify. Relatively speaking, the federal funds rate is given as numbers and is easily obtained. That is why it is used in this study.

It can be seen from Table 4.1 that using federal funds rate as an exogenous variable does not always give better prediction performance than ARF. During stable economic period, it performs worse than ARF by a slight amount. While its errors behave similarly to that of the "naive" model (Table 4.2), one can argue that ARFX is worse than the "naive" model because ARFX has a larger error standard deviation. In contrast, the federal funds rate becomes more useful in the unstable economic period. In this period, FARX can be considered as the best model because it produces the smallest error mean 0.7445 and median 0.6596 among all models. Also, it is able to produce small errors consistently due to its small standard deviation 0.4221 (second smallest).

To conclude this section, we compared prediction performance of 7 models by looking at 1-day ahead prediction error. we noted that models perform differently during stable economic period and during unstable economic period. Models that perform well during stable period does not mean that it performs well in unstable period, and vice versa.

The stable economic period is from July 2004 to July 2007. This period is characterized by the differenced yield curves having small magnitude across all maturities and small volatility between days. In terms of summary statistics in Table 4.2, PARF (B) is the best model. This suggests that during the stable period, the market can be treated as an autonomous system, where exogenous shocks rarely exist and do not have much impact on the yield curve. The time series of yield curves can be explained well by an autoregressive functional model which models each curve only using the curves in the past few days. Also, it does not require a complicated model in this period because models such as ARF with more paramters tend to overfit the data and result in higher prediction error.

This unstable economic period is from July 2007 to February 2009, which corresponds to the financial crisis triggered by the subprime mortgage crisis. The differenced yield curves in this period tend to have larger magnitude across maturities and larger volatility between days. The curves often have irregular shapes. In terms of summary statistics in Table 4.3, the best performing model is ARFX, where the exogenous time series is the univariate time series of differenced effective federal funds rates. This suggests that this period can not be treated as an autonomous system. The market is affected by exogenous factors and one has to include them when modeling the yield curve data. Also, models with more paramters such as ARF and ARFX tend to perform better, possibly because a more complicated covariance structure exists in the data.

Table 4.1: One-day ahead prediction error of the yield curve of several models. Error is calculated as the sum of the absolute difference between the forecast and the observed yield curves over maturities 0.25, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. ARF (B.S.) is the ARF model with basis selection. PARF (B) and PARF (D) correspond to the PARF model where the coefficients are approximated by the basis function or discretization approaches, respectively.

	"Naive"	ARF	ARF (B.S.)	ARFX	PARF (B)	PARF (D)	VAR
1	1.0120	0.2280	0.2275	0.2951	0.2209	0.4964	2.7454

Continued on next page

Table 4.1 – continued from previous page

		"Naive"	ARF	ARF (B.S.)	ARFX	PARF (B)	PARF (D)	VAR
	2	0.8660	0.7073	0.7073	0.7082	0.6064	1.2119	1.3043
	3	0.2800	0.3218	0.1214	0.2410	0.1626	0.3278	2.0547
Jan 05	4	0.4538	0.4350	0.2300	0.4530	0.2001	0.3997	1.0421
	5	0.6601	0.1865	0.1958	0.1942	0.1586	0.3209	1.0503
	6	0.7065	0.5517	0.5620	0.4799	0.5267	1.0473	7.4892
	7	0.1287	0.1337	0.0372	0.1087	0.0748	0.1535	0.7603
	8	0.1850	0.3766	0.1846	0.3725	0.1918	0.3887	0.9123
	9	0.3530	0.5049	0.3712	0.5400	0.1140	0.1734	0.4808
	10	0.2723	0.4220	0.4774	0.5175	0.4915	0.9903	0.3887
	11	0.6240	0.4987	0.6195	0.4784	0.5097	1.0009	0.2947
	12	0.2677	1.1373	0.6671	1.4314	0.0653	0.1432	1.3569
Jan 06	13	0.5560	0.7282	0.7251	0.8302	0.2448	0.4817	0.5828
	14	0.3610	0.1585	0.1642	0.1519	0.1997	0.3784	0.4882
	15	1.2810	1.0328	0.8606	1.1043	0.4792	0.9598	1.3600
	16	0.3960	0.4189	0.3593	0.3108	0.2118	0.4114	1.6564
	17	0.1591	0.1081	0.0725	0.1532	0.0855	0.1712	1.1491
	18	0.1730	0.2562	0.1704	0.2033	0.1171	0.2373	0.8787
	19	0.6430	0.0495	0.1946	0.1302	0.2573	0.5075	0.5543
	20	0.1535	0.2289	0.2403	0.6929	0.2231	0.4445	1.2531
	21	0.4610	0.4457	0.4537	0.3881	0.1072	0.2148	1.2690
Jan 07	22	0.1779	0.0762	0.0497	0.0953	0.0368	0.0690	0.6924
	23	0.1099	0.1979	0.0802	0.1465	0.2070	0.4216	0.8116
	24	0.3089	0.2302	0.2729	0.2234	0.1699	0.3273	0.8913
	25	0.3890	0.3607	0.3529	0.3171	0.8734	1.7857	1.9551
	26	2.0610	0.6500	0.5090	0.6629	0.9339	1.8178	1.5501
	27	0.6557	0.3609	0.4787	0.4268	0.6195	1.3543	1.6064

Continued on next page

Table 4.1 – continued from previous page

		"Naive"	ARF	ARF (B.S.)	ARFX	PARF (B)	PARF (D)	VAR
	28	0.2463	0.3108	0.3052	0.3023	0.2933	0.5554	2.2618
	29	0.6439	0.3672	0.7041	0.3483	1.4399	2.9565	3.1354
Jan 08	30	0.3843	0.6324	0.3728	0.6563	0.2460	0.4857	5.9915
	31	2.4315	0.7281	0.6666	0.6492	1.0083	2.2129	2.5678
	32	1.4193	0.3180	0.2654	0.2579	0.2750	0.5654	3.4211
	33	0.8022	0.7572	0.2715	0.7572	0.2383	0.5327	4.1148
	34	0.8560	1.1265	0.8904	1.1314	0.8641	1.7216	2.2567
	35	0.3675	0.6863	0.5151	0.7534	0.5810	1.2159	0.3984
	36	4.6860	1.0035	3.8953	0.9821	4.1799	8.3194	2.0886
	37	1.8760	0.9723	0.7686	0.5857	0.7430	1.4918	4.7551
	38	0.4395	0.9816	0.7323	1.0290	0.9826	2.0282	1.6626
Jan 09	39	0.8567	1.7679	1.4673	1.8801	1.1487	2.2865	2.7443
Minimum		0.1099	0.0495	0.0372	0.0953	0.0368	0.0690	0.2947
Mean		0.7360	0.5246	0.5190	0.5382	0.5151	1.0412	1.8455
Median		0.4538	0.4220	0.3712	0.4530	0.2460	0.5075	1.3569
Maximum		4.6860	1.7679	3.8953	1.8801	4.1799	8.3194	7.4892
SD		0.8422	0.3669	0.6279	0.3910	0.6954	1.3934	1.5619

Table 4.2: A Summary of windows 1-25 (stable economic period) of Table 4.1.

	"Naive"	ARF	ARF (B.S.)	ARFX	PARF (B)	PARF (D)	VAR
Minimum	0.1099	0.0495	0.0372	0.0953	0.0368	0.0690	0.2947
Mean	0.4391	0.3918	0.3359	0.4227	0.2614	0.5226	1.3369
Median	0.3610	0.3607	0.2403	0.3171	0.2001	0.3997	1.0421
Maximum	1.2810	1.1373	0.8606	1.4314	0.8734	1.7857	7.4892
SD	0.2969	0.2767	0.2371	0.3268	0.2038	0.4130	1.4036

Table 4.3: A Summary of windows 26-39 (unstable economic period) of Table 4.1.

	"Naive"	ARF	ARF (B.S.)	ARFX	PARF (B)	PARF (D)	VAR
Minimum	0.2463	0.3108	0.2654	0.2579	0.2383	0.4857	0.3984
Mean	1.2661	0.7616	0.8459	0.7445	0.9681	1.9674	2.7539
Median	0.8291	0.7072	0.5908	0.6596	0.8035	1.6067	2.4148
Maximum	4.6860	1.7679	3.8953	1.8801	4.1799	8.3194	5.9915
SD	1.1994	0.3970	0.9323	0.4221	0.9960	1.9798	1.4518

4.3 Principal Component Analysis on the Yield Curves

It is documented in the literature that when principal component analysis is applied to time series of differenced yield curves, then the first few principal components represent certain characteristics of the yield curves:

1. The first principal component is called the parallel shift component. The factor loadings are roughly constant over different maturities. As its name suggests, this corresponds to a parallel shift in the yield curve.

2. The second principal component is called the tilt component. The factor loadings have a monotone change over the maturities, and loadings with short-maturity and long-maturity have opposite signs.

Table 4.4 gives the loadings of the first 2 principal components when principal component analysis is applied to 3 periods of the differenced yield curve data: 5/26/2006-9/1/2006, 8/15/2008-11/21/2008, 11/21/2008-2/27/2008.

During the period 5/26/2006-9/1/2006, when the economy is stable, the principal components conform to the characteristics of the parallel shift, and tilt components mentioned above. However, the principal components deviates from what is observed in the literature during the unstable economic period. The period 8/15/2008-11/21/2008 corresponds to the financial crisis when Lehman Brothers declared bankruptcy and Merrill Lynch was acquired by Bank of America. It is seen that the loadings of the first principal component tend to have a trend going towards 0. In the subsequent period 11/21/2008-2/27/2009, the first principal component has a trend going away from 0, and the second principal component does not have a monotone trend.

In addition, the variation explained by the first principal component in the three periods are quite different. This number is 92% for the stable period, and is 74% and 87% in order for the two unstable periods. A low percentage suggests that the differenced yield curves tend to have irregular shapes in that period and so it is not easy to capture these shapes using one principal component. The cumulative variation explained by the first two principal components is 97% for the stable period, and is 92% and 92% for the two unstable periods.

These observations reinforce our observation in Section 4.2 that the differenced yield curves are more volatile during the unstable period. High volatility in yield curves could be due to external shocks, which makes it harder to perform predictions based only on historical yield curves.

Table 4.4: Loadings of the first 3 principal components when PCA is applied to the differenced yield curves during 3 periods.

	5/26/2006-9/1/2006		8/15/2008-11/21/2008		11/21/2008-2/27/2009	
Maturity (Year)	PC1	PC2	PC1	PC2	PC1	PC2
0.25	-0.077	0.820	-0.359	-0.621	-0.006	-0.046
0.5	-0.161	0.415	-0.284	-0.512	0.025	-0.048
1	-0.282	0.148	-0.274	-0.176	0.046	-0.210
2	-0.321	0.143	-0.302	0.019	0.165	-0.547
3	-0.331	0.051	-0.312	0.057	0.237	-0.527
4	-0.333	-0.030	-0.298	0.127	0.307	-0.227
5	-0.326	-0.037	-0.298	0.155	0.314	-0.246
6	-0.323	-0.083	-0.296	0.217	0.354	-0.033
7	-0.320	-0.122	-0.294	0.282	0.395	0.183
8	-0.301	-0.156	-0.262	0.227	0.382	0.227
9	-0.292	-0.178	-0.234	0.240	0.384	0.300
10	-0.277	-0.173	-0.227	0.206	0.378	0.293

4.4 Caplet Data

In Section 2.1.4, it is mentioned that a caplet is a European call option on forward LIBOR and a cap is a collection of caplets with different maturities. Caps and caplets provide insurance against LIBOR rising above certain level. These interest rate derivatives are traded actively in the market. Between this section and Section 4.6, we propose to use the results of the yield curve prediction to obtain one-day ahead prediction of caplet prices.

The data consist of the implied volatilities of daily cap data with strike 4% and maturities 1, 2, 3, 4, 5 years for every trading day from 6/28/2004 to 3/10/2009. The caps are based on 3-month

LIBOR. That is, the 1-year cap has 3 caplets maturing in 6, 9, and 12 months, the 2-year cap has 4 additional caplets maturing in 1.25, 1.5, 1.75, and 2 years, and so on. The data are obtained from Bloomberg.

It is mentioned briefly in Section 4.1 that zero-coupon bond yields can be obtained from coupon bond yields via bootstrapping. Using the same procedure, one can derive caplet prices from cap prices. The idea of this method is that to calculate the price of a caplet, one takes the difference between two caps which differ in only this caplet. However, in our data, caps with adjacent maturities differ by 4 caplets. In order to solve for caplet prices, we assume that caps with maturities $0.5, 0.75, 1, \ldots, 4.75, 5$ years exist (in addition to 1, 2, 3, 4, 5). The implied volatilities of caps with maturities 0.5 and 0.75 year are assumed to be the same as that with maturity 1 year, and the implied volatilities of other caps are interpolated linearly between caps with nearest-integer maturities. For each day, there are 19 caplets with different maturities and they are plotted in Figure 4.3.

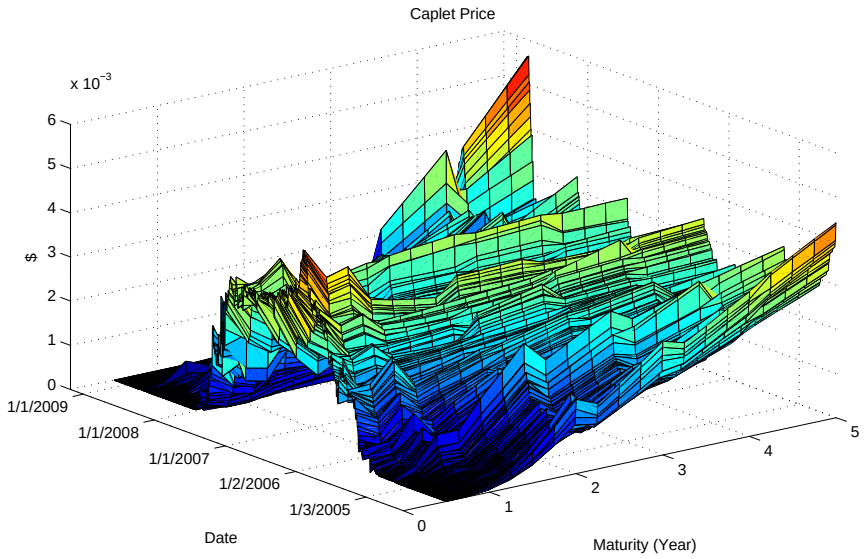

Figure 4.3: Daily caplet price data with strike 4% and maturities less than or equal to 5 years from Jun 28, 2004 to Mar 10, 2009.

4.5 Caplet Pricing Formula

Since different interest rate models define the dynamics differently, the caplet price formula inferred by them is also different. In this section we investigate how accurate these pricing formulas can produce market caplet prices based on historical interest rates. If the assumptions of these models are accurate descriptions of the market, then we expect that models can produce caplet prices that are close to the prices observed in the market.

The error measure is taken to be the root mean square error between the observed caplet price $C_{i,t+1}$ at time $t+1$ and the caplet price inferred by the model $C_{i,t}^{(1)}$ based on historical interest rate data at or before time t, where $i \in \{1, \ldots, 19\}$ denotes the ith caplet. The RMSE is defined as

$$\sqrt{\frac{1}{19}\sum_{i=1}^{19}\left(C_{i,t+1} - C_{i,t}^{(1)}\right)^2}.$$

4.5.1 Methodology

Formula (2.5) gives the price of a caplet on $F(T, T+\delta)$ under Black model:

$$\delta P(t, T+\delta)\,\text{Black}(F(t, T, T+\delta), K, \sigma_t^{\text{caplet}}\sqrt{T-t}),$$

where the function "Black" is defined by (2.6). In this model, the implied volatility σ_t^{caplet} represents the volatility of the forward LIBOR $F(t, T, T+\delta)$. A standard way of estimating the volatility is to take the standard deviation of the returns of a historical series $F(t-1, T, T+\delta), F(t-2, T, T+\delta), \ldots, F(t-l, T, T+\delta)$, multiplied by $\sqrt{252}$ to convert it to a time horizon of 1 year. l is the width of each moving window and is equal to 30 in our study. The forward LIBOR $F(t-j, T, T+\delta)$, $j > 0$, may not be observed but can be obtained by interpolating between other forward LIBOR's at the same time. Once the implied volatility is estimated, it is plugged into the caplet price formula above as the caplet price under Black model.

Formula (2.8) gives the price of a caplet under Hull-White model:

$$(1+\delta K)P(0,T)\left[\frac{1}{1+\delta K}\Phi(-d_2) - \frac{P(0,T+\delta)}{P(0,T)}\Phi(-d_1)\right].$$

This formula depends on model parameters a and σ, which can be estimated by matching the variance of the historical yields to the variance $\left(\frac{1-e^{-aT}}{aT}\right)^2 \frac{\sigma^2}{2a}$ under the Hull-White model. Model parameters are estimated by minimizing

$$\sum_{T\in\{0.25,0.5,1,\ldots,10\}} \left[\left(\frac{1-e^{-aT}}{aT}\right)^2 \frac{\sigma^2}{2a} - v_T\right]^2,$$

where v_T is the sample variance of historical yields with maturity T. Estimated parameters are plugged into the caplet price formula above as the caplet price under Hull-White model.

4.5.2 Results

The first two columns of Table 4.5 give the RMSE of the two pricing formulas. It is seen that the error that Hull-White model produces is larger than the error of Black model in most moving windows. However, Although both models use the fluctuations in interest rate to calculate the caplet price (Black uses standard deviation of LIBOR, and Hull-White uses variance of zero-coupon bond yields), Hull-White model is not affected as hard as Black model during the unstable period. In terms of the change in mean RMSE from stable to unstable period, Black model increases by 6 times, but Hull-White model increases by less than 2 times. Compared to the prediction errors produced by calibrating the models to caplet data (which is discussed in the next section), the errors produced by these pricing formulas are considered to be unacceptably large. An explanation of its performance is that the model assumptions may not be accurate descriptions of the market. For example, the Black model assumes that the forward rate is a geometric Brownian motion with constant volatility. If this assumption does not hold, then the standard deviation of returns of forward rates might not be a good estimate of the implied volatility.

4.6 Caplet Price Prediction

4.6.1 Methodology

Although the pricing formulas described in the last section provide an idea of what the caplet prices should be based on historical interest rates, the approach used by the market to price financial

derivatives is to calibrate an interest rate model to financial instruments similar in nature. In this section, we compare several methods that use this approach to predict caplet prices, some of which involve applications of yield curve prediction done in Section 4.2. The error measure is taken to be the same RMSE defined in the last section.

Predicting Yield Curve

Consider the price of a caplet (2.5) under the Black model. Suppose the strike rate and the time to maturity are given. Then under this model, the price of such caplet at time $t+1$ depends on the forward rate $F(t+1, T_{i-1}, T_i)$ and the implied volatility $\sigma_{t+1}^{\text{caplet}^i}$. A prediction of the price of this caplet can be obtained by predicting the levels of the forward rate and the implied volaility at time $t+1$.

In Section 4.2, we apply functional time series models on moving windows of yield curves and obtain 1-day ahead predictions. These predictions can be used in combination with formula (2.2), which expresses the forward rate in terms of zero-coupon bond prices, to predict the forward rate. The implied volatility $\sigma_{t+1}^{\text{caplet}^i}$ is estimated using the implied volatility $\sigma_t^{\text{caplet}^i}$ of the present day.

We compare the caplet prediction performance of 5 out of the 7 models that are considered in Section 4.2: "naive", ARF, ARF (B.S.), ARFX, and PARF (B). These 5 models are capable of producing yield curve forecasts in functional form, which can be used to calculate forward rates of the form $F(0, T, T+0.25)$.

A Reference

To assess the caplet price prediction performance of the 5 time series models mentioned above, we need to know the portion of the RMSE due to using the present implied volatilities as estimates of the implied volatilities of the next day. It can be obtained by using the term structure of the next day (assuming it is known) and calculate the caplet prices using the present implied volatilities. It is referred to as the "reference" in later discussions. We are interested in how close the RMSE produced by other time series methods can get to the RMSE of the reference, although there is no guarantee that the latter must be smaller than the former.

Hull-White Model

In addition to the above models, we also consider the performance of prediction using the Hull-White model (2.7). For every moving window, we calibrate the model to all the caplets in the window by minimizing the sum of squared difference between the observed caplet prices and model caplet prices (2.8) over the parameters σ and a. Then the 1-day ahead caplet prediction is calculated using (2.8) with the estimated values $\hat{\sigma}$ and $\hat{a}$.

4.6.2 Prediction Performance

RMSE of the prediction methods in each of the 39 moving windows is reported in Table 4.5. Tables 4.6 and 4.7 give summaries of the RMSE in the stable and unstable periods, respectively.

It can be seen from Table 4.5 that better yield curve prediction tends to result in better caplet prediction since the model that has the best overall yield curve prediction performance in the stable/unstable period also has the best overall caplet prediction performance in the same period. During stable economic period, PARF (B) can be considered as the best model because all 4 summary statistics in this period are the smallest among the 5 time series models. In particular, its mean is 0.7158 while others are more than 0.9. It offers a great reduction of mean RMSE from 0.9184 of the "naive" model towards 0.5836 of the reference. During the unstable economic period, the ARFX model with differenced effective federal funds rate as the exogenous variable has the best overall performance. 3 out of its 4 statistics shown in Table 4.7 are the smallest among the 5 time series models, with its "maximum" statistic 3.0899 only slightly above that of PARF (B) which is 3.0035.

The fact that better overall yield curve prediction results in better overall caplet prediction is reasonable, but there are times when a better yield curve prediction does not result in a better caplet prediction. For example, consider window 33. It is seen in Table 4.1 that the yield curve prediction error of ARF (B.S.), 0.2715, is much smaller than that of ARF, 0.7572. However, from Table 4.5, the caplet RMSE of ARF (B.S.), 0.7485, is larger than the RMSE of ARF, 0.6740. This phenomenon could be due to one of the following three reasons. First, the transformation from zero-coupon bond prices to caplet prices is not linear. In particular, formula (2.2) is used to convert bond prices to forward LIBOR. Second, the caplet implied volatilities on the next day are not known and are taken

as the implied volatilities on the present day. This creates an uncertainty in our prediction. During days where implied volatilities have large fluctuations, a better estimate of the forward LIBOR can actually cause a worse estimate of the caplet prices. Third, the yield curve prediction error (4.1) is an aggregate measure of the difference between the observed yield curve and the predicted yield curve over 12 maturities. It is possible that variations exist in the difference at different maturities. Hence, a yield curve prediction with a large prediction error might give a more accurate prediction for the caplets. However, having a more accurate prediction of the yield curve is definitely advantageous because, as we mentioned earlier, it tends to give a more accurate prediction of caplets.

Tables 4.6 and 4.7 reveal that the overall level of RMSE increases from the stable period to the unstable period. The median increases by about 0.5 quite uniformly across all models including the reference. This suggests that the implied volatilities of caplets have larger fluctuations in unstable period and this causes larger errors in predicting caplet prices. In contrast, the mean increases differently for different models, ranging between 0.4 to 1. This is because some time series models have larger variations when predicting the yield curve. Since a large yield curve prediction error can induce a large caplet price prediction error, some models also have larger variations when predicting caplet prices.

The Hull-White model is provided as an alternative way of predicting caplet prices. The summary statistics show that although it is on a par with prediction using time series models, a careful examination reveals that it is slightly behind the autoregressive models in both stable and unstable economic period. Hence, analyzing yield curves using functional time series models is indeed useful in predicting caplet prices.

Table 4.5: RMSE of one-day ahead prediction of caplets with strike 4% and maturities less than or equal to 5 years. "ARF (B.S.)" refers to ARF basis selection; and "PARF (B)" refers to PARF solved using basis functions approach.

$10^4\times$		Pricing Formula		Prediction						
		Black	Hull-White	Black						Hull-White
				Reference	"Naive"	ARF	ARF (B.S.)	ARFX	PARF (B)	
	1	2.3350	28.9978	0.5562	1.0118	0.5188	0.5190	0.4307	0.4625	0.5463
	2	3.4460	27.9088	0.3475	0.5975	0.4641	0.4641	0.4639	0.3343	0.8617
	3	2.3938	11.3342	1.0028	0.9551	1.0166	0.9064	0.9885	0.8526	0.7174
Jan 05	4	2.5101	10.4723	0.2513	0.7199	0.5374	0.4206	0.5684	0.2981	1.5948
	5	2.6849	9.4529	0.5546	0.9243	0.5386	0.5548	0.5580	0.4976	1.6129
	6	1.6540	16.2610	2.7523	3.0299	2.9248	3.0384	2.8785	2.8189	2.2984
	7	4.5309	14.6157	0.0860	0.1831	0.1767	0.0758	0.1486	0.0931	0.6259
	8	3.6883	13.6873	0.5057	0.7236	0.8557	0.7485	0.8507	0.6843	0.8177
	9	3.2243	10.2446	0.2661	0.5603	0.7561	0.7010	0.8183	0.2567	1.1185
	10	2.4878	9.7610	0.4188	0.5553	0.7265	0.7219	0.8590	0.7762	1.0079
	11	4.9846	8.2357	0.6200	0.8699	0.9016	0.9195	0.8584	0.8844	1.2277
	12	1.5449	18.8678	0.2817	0.3138	2.3945	2.7465	2.8159	0.3824	1.2000
Jan 06	13	2.0080	7.7798	0.7329	1.1089	1.4069	1.4069	1.5982	0.8501	1.1034
	14	1.6109	21.4210	0.5929	0.8618	0.6563	0.6512	0.6637	0.6634	0.6729
	15	1.3139	26.3624	0.2881	2.4908	2.0230	1.5229	2.1317	0.8957	0.9587
	16	0.8709	4.8208	0.5155	0.7504	0.8422	0.8358	0.7321	0.5804	0.8004
	17	2.0341	4.9176	0.3958	0.4418	0.4313	0.4238	0.4641	0.4462	0.4407

Continued on next page

Table 4.5 – continued from previous page

$10^4\times$		Pricing Formula		Prediction						
		Black	Hull-White	Black						Hull-White
				Reference	"Naive"	ARF	ARF (B.S.)	ARFX	PARF (B)	
	18	3.8221	22.0713	0.2149	0.3936	0.4325	0.7042	0.3659	0.3484	0.6162
	19	1.8452	24.0130	0.5915	1.1823	0.6501	0.6703	0.7020	0.7076	1.0510
	20	2.3137	19.7510	0.7862	0.8063	0.8359	0.8468	1.4904	0.8431	0.9829
	21	1.9701	20.7983	0.5241	0.9927	0.9003	1.0047	0.8301	0.5892	0.7811
Jan 07	22	1.7683	21.0229	0.0578	0.3450	0.1571	0.1996	0.1926	0.1118	0.4496
	23	1.8517	19.7205	0.4545	0.6804	0.5414	0.5495	0.4925	0.6931	1.0592
	24	3.1714	20.4072	1.1184	1.4974	1.2669	1.4482	1.2644	1.1432	1.2098
	25	2.5813	4.2084	0.6756	0.9636	0.8582	1.0592	0.7771	1.6811	1.0914
	26	3.4675	8.5310	0.2464	3.5209	1.1379	1.9702	1.1305	1.6514	1.7977
	27	12.4116	9.8998	1.6753	2.0505	1.7747	1.7544	1.7963	1.7398	1.3896
	28	4.4496	10.1313	0.9381	1.1844	1.2539	1.3525	1.2559	1.3644	0.5176
	29	12.7665	9.2834	0.7771	1.0368	0.9901	0.8525	0.9651	1.6549	1.9044
Jan 08	30	8.1866	33.6798	0.7908	0.7608	0.9545	0.8353	0.9561	0.8449	0.9383
	31	17.5314	45.4247	2.1685	2.5020	3.2286	3.2705	3.0899	2.9481	1.5650
	32	29.7376	41.4822	0.4739	1.2274	0.6869	0.3914	0.6985	0.4298	2.1128
	33	9.7228	28.9897	0.1765	0.6724	0.6740	0.7485	0.6740	0.3678	1.1199
	34	13.5217	12.9377	1.2865	1.4651	1.4875	1.5941	1.4856	1.5679	0.3588
	35	11.7530	10.7957	0.8585	1.1029	0.8536	0.7835	0.9416	0.8128	1.5471
	36	16.5983	11.1634	1.7828	4.4126	2.4652	4.2009	1.6674	3.0035	3.2690
	37	28.3697	51.1576	2.1349	4.1343	2.9746	2.7313	2.5290	2.9011	2.6266
	38	16.7127	70.3869	1.1912	1.6937	0.3730	0.8995	0.3339	0.7309	1.9378
Jan 09	39	25.1129	57.0542	1.2995	0.9633	1.7264	1.2639	1.9079	1.0521	1.8615

Continued on next page

Table 4.5 – continued from previous page

$10^4\times$	Pricing Formula		Prediction						
	Black	Hull-White	Black						Hull-White
			Reference	"Naive"	ARF	ARF (B.S.)	ARFX	PARF (B)	
Minimum	0.8709	4.2084	0.0578	0.1831	0.1571	0.0758	0.1486	0.0931	0.3588
Mean	6.9997	20.4628	0.7793	1.2740	1.1127	1.1741	1.1122	0.9991	1.2255
Median	3.2243	16.2610	0.5915	0.9633	0.8557	0.8468	0.8584	0.7762	1.0914
Maximum	29.7376	70.3869	2.7523	4.4126	3.2286	4.2009	3.0899	3.0035	3.2690

Table 4.6: A summary of caplet prediction RMSE in windows 1-25 (stable economic period) of Table 4.5.

$10^4\times$	Pricing Formula		Prediction						
	Black	Hull-White	Black						Hull-White
			Reference	"Naive"	ARF	ARF (B.S.)	ARFX	PARF (B)	
Minimum	0.8709	4.2084	0.0578	0.1831	0.1571	0.0758	0.1486	0.0931	0.4407
Mean	2.5059	15.8853	0.5836	0.9184	0.9125	0.9256	0.9577	0.7158	0.9939
Median	2.3350	16.2610	0.5155	0.8063	0.7561	0.7219	0.7771	0.6634	0.9829
Maximum	4.9846	28.9978	2.7523	3.0299	2.9248	3.0384	2.8785	2.8189	2.2984

Table 4.7: A summary of caplet prediction RMSE in windows 26-39 (unstable economic period) of Table 4.5.

$10^4\times$	Pricing Formula		Prediction						
	Black	Hull-White	Black						Hull-White
			Reference	"Naive"	ARF	ARF (B.S.)	ARFX	PARF (B)	
Minimum	3.4675	8.5310	0.1765	0.6724	0.3730	0.3914	0.3339	0.3678	0.3588
Mean	15.0244	28.6370	1.1286	1.9091	1.4701	1.6177	1.3880	1.5050	1.6390
Median	13.1441	20.9637	1.0647	1.3463	1.1959	1.3082	1.1932	1.4662	1.6814
Maximum	29.7376	70.3869	2.1685	4.4126	3.2286	4.2009	3.0899	3.0035	3.2690

Chapter 5

Summary

This thesis proposes a functional time series approach to analyzing time series of interest rate data, and in particular, time series of yield curves. The motivation of the research is that the common market practice of using an interest rate model to price interest rate derivatives does not involve the interest rate being modeled. This is because normally interest rate models are calibrated to financial instruments that are close in nature to the derivatives being priced. However, if the interest rate being modeled is not used during calibration, then the calibrated model may not give a very good representation of the dynamics of the interest rate, and hence the prices that it produces may be unreliable.

A functional time series approach is therefore suggested to analyze the time series of yield curves. Yield curves are treated as functional data in the model because it allows derivatives/integrals to be taken, which are needed in the tranformation between the yield curve and the forward curve. Also, in cases where the multivariate bond yield data have missing data, or when the bond yields are recorded at different maturities at different times, then multivariate models are not applicable.

Two functional time series models are proposed: the autoregressive functional model (ARF) and the pointwise autoregressive model (PARF). ARF can be considered as a generalization of the vector autoregressive model (VAR) to the functional case, and PARF can be considered as having a univariate autoregressive model at each maturity of the yield curve time series. The autoregressive functional exogenous (ARFX) model is also proposed which incorporates an exogenous

time series into ARF. In all of these models, the coefficient functions are approximated by finite linearly combinations of basis functions, and the basis function coefficients are solved by least squares method.

Empirical studies are carried out to illustrate how the proposed models can be applied to real world data. In the first study, yield curve prediction performance of functional time series models as well as several benchmark models are investigated. It is noted that the data can be divided into two periods based on the behavior of the data. The first period is from July 2004 to July 2007 and is referred to as the stable economic period, and the second period is from July 2007 to February 2009 and is referred to as the unstable economic period. While an apparent difference between the two periods is that yield curves are more volatile in the unstable period, it is concluded that the stable period can be treated as an autonomous system because yield curves are explained well by PARF, and the unstable period is affected by external factors because the yield curves are explained well by ARFX, where the exogenous variable is the federal funds rate.

In the second empirical study, we consider a time series of caplet prices with fixed strike and different maturities. We propose a way to predict future caplet prices by combining yield curve prediction using functional time series models and historical implied volatilities of caplets. It is seen that models which have better yield curve prediction tend to have better caplet price prediction. The best time series model in each of the two economic periods is shown to be superior to the Hull-White model in caplet price prediction in the same period.

Appendix A

Splines

In this section, we review the spline method that is used for data smoothing or interpolation for one-dimensional data. A spline is a piecewise polynomial function over a domain partition on an interval $[a, b]$. Denote a spline function by $f : [a, b] \to \mathbb{R}$. We partition the interval $[a, b]$ into $K + 1$ disjoint subintervals $[t_i, t_{i+1}]$, $i = 0, 1, \ldots, K$, where $a = t_0 \leq t_1 \leq \cdots \leq t_K \leq t_{K+1} = b$. On each of these $K + 1$ subintervals, we define a polynomial q_i so that

$$f(x) = \begin{cases} q_0(x), & t_0 \leq t < t_1 \\ q_1(x), & t_1 \leq t < t_2 \\ \vdots & \\ q_K(x), & t_K \leq t < t_{K+1} \end{cases}$$

The points t_i are called knots. Since each subinterval is represented by a separate polynomial, f is typical not analytic at the knots, which means the kth derivative of f is not continuous for some non-negative integer k. Smoothness constraints are sometimes imposed to ensure that a spline has continuous derivatives at its knots up to certain order. We say that a spline is of order M if it is a piecewise polynomial of order M (degree at most $M - 1$), and is $M - 2$ times continuously differentiable. In particular, a spline of order 1 is a step function; a spline of order 2 is a continuous linear piecewise function.

A commonly used spline is the cubic spline. It is a spline of order 4 (cubic piecewise polynomials)

and with 2-times continuous derivatives at the knots. Mathematically, q_i's are cubic polynomials

$$q_i(t) = a_{i0} + a_{i1}t + a_{i2}t^2 + a_{i3}t^3$$

that satisfy

$$\begin{aligned} q_{i-1}(t_i) &= q_i(t_i), \quad i = 1, \ldots, K, \\ q_{i-1}^{'}(t_i) &= q_i^{'}(t_i), \quad i = 1, \ldots, K, \\ q_{i-1}^{''}(t_i) &= q_i^{''}(t_i), \quad i = 1, \ldots, K. \end{aligned}$$

In general, an order-M spline can be represented by the following set of basis functions:

$$\begin{aligned} B_j(x) &= x^{j-1}, \quad j = 1, \ldots, M, \\ B_{M+l} &= (x - t_l)_+^{M-1}, \quad l = 1, \ldots, K, \end{aligned}$$

where s_+ denotes the positive part.

A natural cubic spline is another commonly used spline. It is a cubic spline with additional constraints that the second derivatives of the function are 0 at the two boundaries. Namely,

$$q_0^{''}(a) = 0, \quad q_K^{''}(b) = 0.$$

The constraints attempt to regulate the behavior of the spline near the boundaries by setting the boundaries to be inflection points, so that the function is linear beyond the boundary knots. A natural cubic spline with K knots is represented by K basis functions:

$$B_1(x) = 1, \quad , B_2(x) = x, \quad , B_{k+2}(x) = d_k(x) - d_{K-1}(x),$$

where

$$d_k(x) = \frac{(x - t_k)_+^3 - (x - t_K)_+^3}{t_K - t_k}.$$

A.1 B-splines

B-splines, which stand for basis splines, provide an alternative set of spline basis functions. It has the property that every spline function of given degree and knot placements can be represented as a linear combination of B-spline basis functions of that same degree and knot placements, while B-spline has an additional property that each of its basis functions has minimal support. A B-spline is defined as follows. Suppose we have a knot sequence $\tau_1 < \tau_2 < \cdots < \tau_n$ and two boundary knots τ_0 and τ_{n+1}, where $\tau_0 < \tau_1$ and $\tau_n < \tau_{n+1}$. We first define an augmented knot sequence t which satisfy

- $t_1 \leq t_2 \leq \cdots \leq t_K \leq \tau_0$.
- $t_{j+K} = \tau_j, \quad j = 1, \ldots, n$.
- $\tau_{n+1} \leq t_{n+K+1} \leq t_{n+K+2} \leq \cdots \leq t_{n+2K}$.

While the values of the augmented knots beyond the boundary knots are not defined strictly, one normally would take them to equal to the boundary knots τ_0 and τ_{n+1}. Denote the ith B-spline basis function of order p by $B_{i,p}(t)$. Then these basis functions are defined by the following recursive relations

$$B_{i,0}(t) = \begin{cases} 1 & \text{if } t_i \leq t < t_{i+1} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, n + 2K - 1$, and

$$B_{i,p}(t) = \frac{t - t_i}{t_{i+p-1} - t_i} B_{i,p-1}(t) + \frac{t_{i+p} - t}{t_{i+p} - t_{i+1}} B_{i+1,p-1}(t),$$

for $i = 1, \ldots, n + 2K - p$. In order to avoid division by 0, we let $B_{i,1} = 0$ if $t_i = t_{i+1}$. This implies $B_{i,p} = 0$ if $t_i = t_{i+1} = \cdots = t_{i+p}$. From the recursive relations, we see that for a knots sequence with n knots, a B-spline with order p has $n + p$ basis functions. Each basis function is a polynomial of degree p and is always non-negative. In particular, a cubic B-spline refers to the B-spline with order 3. $B_{i,p}$ has local support only on the interval $[t_i, t_{i+p+1})$. On any interval $[t_i, t_{i+1})$, at most $p + 1$ basis functions of degree p are non-zero, and they are $B_{i-p,p}(t), B_{i-p+1,p}(t), \ldots, B_{i,p}(t)$. At a knot with t_j which is repeated r times, the basis function $B_{i,p}(t)$ is $p - r$ times continuously differentiable.

The local support property of B-spline basis functions has important computational implications. To illustrate it using an example in this thesis, suppose we need to calculate the matrix whose ij-entries are $\int B_{i,p}(t)B_{j,p}(t)dt$. Then this matrix will be a band matrix, whose non-zero entries are confined to a diagonal band. Since we know exactly where this matrix is zero (or non-zero), calculations at these entries can be omitted.

Figure A.1 shows the order 1 and 2 B-spline basis functions on the interval [0,10] with knot placements at $1, 2, \ldots, 9$.

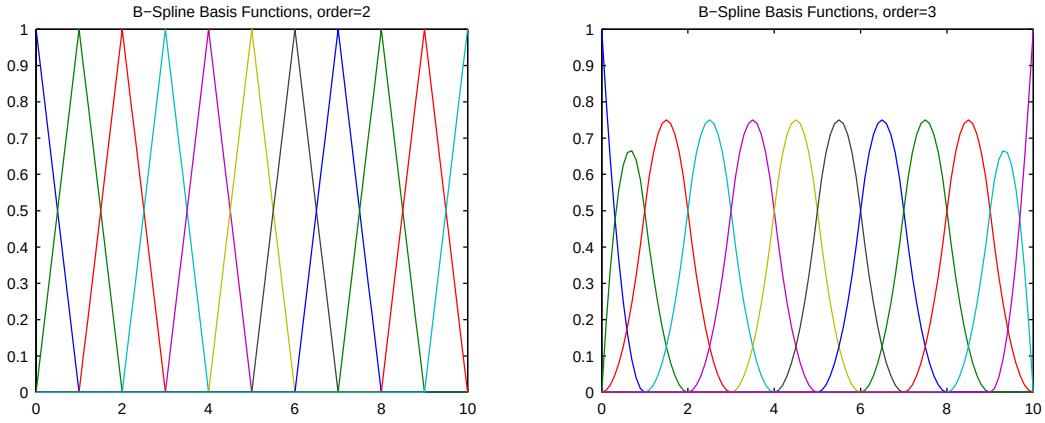

Figure A.1: B-spline basis functions on the interval [0,10] with knots $1, 2, \ldots, 9$. Left figure shows linear (order=2) basis functions; right figure shows quadratic (order=3) basis functions.

A.2 Spline Smoothing

Given data (x_i, y_i), $i = 1, \ldots, n$, and spline basis functions $B_j(x)$, $j = 1, \ldots, J$, we can smooth the data into the function $f(x) = \sum_j \alpha_j B_j(x_i)$ by minimizing the criterion

$$\sum_{i=1}^{n} (y_i - f(x_i))^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{J} \alpha_j B_j(x_i) \right)^2 .$$

A.3 Spline Smoothing in Higher Dimensions

We have reviewed a few one-dimensional spline functions that can be used to smooth data (x_i, y_i), where each x_i is a scalar, into a one-dimensional function. Suppose now each x_i is a two-dimensional vector, and we are again interested in an estimate of the data in a functional form. The one-dimensional splines that we have reviewed above can be extended to this case easily. Assume the domain of f is $[a_1, b_1] \times [a_2, b_2]$. Suppose we have two sets of one dimensional splines basis functions $g_j : [a_1, b_1] \to \mathbb{R}$ and $h_l : [a_2, b_2] \to \mathbb{R}$, where $j = 1, \ldots, J$ and $l = 1, \ldots, L$. Then we can construct two-dimensional basis functions by taking the tensor product of g_j and h_l. The new basis are

$$f_{jl}(u, v) = g_j(u)h_l(v), \quad j = 1, \ldots, J, \quad l = 1, \ldots, L.$$

There are a total of JL basis functions and f can be represented as

$$f(u, v) = \sum_{j=1}^{J} \sum_{l=1}^{L} a_{jl} f_{jl}(u, v) = \sum_{j=1}^{J} \sum_{l=1}^{L} a_{jl} g_j(u) h_l(v).$$

References

Yacine Aït-Sahalia. Nonparametric pricing of interest rate derivative securities. *Econometrica*, 64 (3):527–560, 1996.

Fischer Black. The pricing of commodity contracts. *Journal of Financial Economics*, 3(1-2):167–179, 1976.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

Alan Brace, Dariusz Gatarek, and Marek Musiela. The market model of interest rate dynamics. *Mathematical Finance*, 7(2):127–155, 1997.

Damiano Brigo and Fabio Mercurio. *Interest Rate Models - Theory and Practice*. Springer, 2006.

Rene Carmona and Michael Tehranchi. *Interest Rate Models: An Infinite Dimensional Stochastic Analysis Perspective*. Springer, 2006.

K.C. Chan, G. Andrew Karolyi, Francis Longstaff, and Anthony B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, 47(3):1209–1227, 1992.

Rama Cont and José Fonseca. Dynamics of implied volatility surfaces. *Quantitative Finance*, 2: 45–60, 2002.

John E. Cox, Jonathan Ingersoll, and Stephen Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.

Darrell Duffie and Rui Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6(4): 379–406, 1996.

David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60(1):77–105, 1992.

John Hull and Alan White. Pricing interest-rate-derivative securities. *University of Toronto*, 3(4): 573–592, 1990.

Tze Leung Lai and Haipeng Xing. *Statistical Models and Methods for Financial Markets*. Springer, 2008.

Charles Nelson and Andrew Siegel. Parsimonious modeling of yield curves. *Journal of Business*, 60 (4):473–489, 1987.

J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, second edition, 2005.

Riccardo Rebonato. *Modern Pricing of Interest-Rate Derivatives: The LIBOR Market Model and Beyond*. Princeton University Press, 2002.

Richard Stanton. A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance*, 52(5):1973–2002, 1997.

Lars E.O. Svensson. Esitmating and interpreting forward interest rates: Sweden 1992 - 1994. *National Bureau of Economic Research Working Paper*, 1994.

Ruey S. Tsay. *Analysis of Financial Time Series*. Wiley, 2002.

Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.