THE FLORIDA STATE UNIVERSITY

COLLEGE OF SOCIAL SCIENCES AND PUBLIC POLICY

STRIPPING THE YIELD CURVE WITH MAXIMALLY SMOOTH FORWARD CURVES

By

YANIV JERASSY-ETZION

A Dissertation submitted to the
Department of Economics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy.

Degree Awarded:
Spring Semester, 2010

The members of the committee approve the dissertation of Yaniv Jerassy-Etzion defended on January 7, 2010.

Paul M. Beaumont
Professor Directing Dissertation

Gershon Tenenbaum
University Representative

David Rasmussen
Committee Member

Stefan Norrbin
Committee Member

Approved:

Bruce Benson, Chair, Department of Economics

David Rasmussen, Dean, College of Social Sciences and Public Policy

The Graduate School has verified and approved the above-named committee members.

To Shoshanit, Yael and Noa.

ACKNOWLEDGEMENTS

This work owes to a long line of dedicated individuals who have helped me conduct and finish this research. First in line, and high above the rest, is my advisor, mentor, and friend, Dr. Paul Beaumont. Paul's wisdom, energy, resourcefulness and above all dedication to the cause, have been a major thrust that enabled me to finish this project. I am forever indebted to him and his wife for the long evening hours and many weekends dedicated to helping me accomplishing this goal. Many thanks are due to my PhD committee members for their service and their contribution to this project. To the many friends at the Economics department at Florida State University, faculty, staff and peers, all of which have contributed to the culmination of this study. A special thank you to my friends Gershon Tenenbaum and Avi Burstein for logistical and moral support, and to my extended family who kept me motivated along the journey. Finally to my lovely wife Shoshanit, who never stopped believing in my ability, for her wisdom, patience, and endless love.

— Yaniv

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Continuous discount functions and forward rate curves are needed for nearly all asset pricing applications. Unfortunately, forward curves are not directly observable so they must be constructed from existing fixed-income security prices. In this paper I present two algorithms to construct maximally smooth forward rate and discount curves from the term structure of on-the-run U.S. treasury bills and bonds. I use on-the-run treasuries to get the most recent and liquid prices available. The maximum smoothness criterion produces more accurate prices for derivatives such as swaps and ensures that no artificial arbitrage will be introduced when using the constructed forward curve for pricing out-of-sample securities. When coupon bonds are included among the securities it is necessary to both strip the coupon payments and interpolate the spot curve. To be consistent, these steps must be done simultaneously but this complication usually leads to highly nonlinear algorithms.

The first method I describe uses an iterated, piecewise, quartic polynomial interpolation (IPQPI) of the forward curve that only requires the solution of linear equations while maintaining minimal pricing errors and maximum smoothness of the interpolated curves.The second method uses a genetic programming (GP) algorithm that searches over the space of differentiable functions for maximally smooth forward curves with minimal pricing errors.

I find that the IPQPI method performs better than the GP and other algorithms commonly used in industry and academics.

CHAPTER 1

INTRODUCTION

1.1 Summary of this Research Project

This research extends the model of Adams and van Deventer (1994), (hereafter AvD) and modified by Lim and Xiao(2002), by introducing a procedure that simultaneously strips and smoothes the yield curve iteratively by bootstrapping. The new method, the Iterated Piecewise Quartic Polynomial Interpolation (referred to as the *IPQPI* method) is based on constructing a series of piecewise maximally smooth polynomials which make up the forward curve and integrating the coupon stripping procedure into the curve construction process. The IPQPI method is then tested and compared to some of the prevailing methods currently in use. The IPQPI method appears to perform better than the alternatives based on pricing ability, smoothness, and the practicality of its implementation.

The IPQPI method has some drawbacks that were well documented in past research for all polynomial based methods, however. This leads me to explore an alternative method to constructing the yield curve which is based on Genetic Programming. This method is based on evolutionary principles and is independent of the researcher's notion about what the specific function might be. I briefly describe the guiding principals of the genetic programming and give a detailed description of the experiment I construct. Genetic programming's ability to produce a forward curve that prices well and is sufficiently smooth is shown to be limited in the context of the platform I used and the specification I provided to the system to perform its intended task.

In the remainder of this chapter I provide a broad overview of the research of yield and forward curve construction and then I define some important concepts and establishes a consistent set of notational conventions. The theoretical and empirical literature related to term structure models is then reviewed in chapter two . In chapter three I take a closer look at the AvD method, and then describe my IPQPI extension. the final chapter is dedicated to Genetic Programming. I review Genetic Programming and its main building blocks at the beginning of the chapter and then and I describe my experiment design and its outcome.

1.2 Overview of the Research

The Term Structure of Interest Rates is a function that defines the relationship between yield-to-maturity and the time-to-maturity on fixed income securities

Asset pricing in general, and the valuation of fixed income securities in particular, depend upon an accurate estimation of the yield curve, especially when the cash flow that these securities generate does not occur at the same point in time as the cash flow stemming from the securities that are used to estimate the term structure itself. Cochrane (2001) noted that prices of all assets are a function of two arguments: the expected value of the payoff, and the stochastic discount function. The fundamental asset pricing equation is:

$$P_t = E[m_t(\tau) \cdot X_{t+\tau}]$$

where P_t is the time t price of the asset, $m_t(\tau)$ is the stochastic discount function at time t for a time $t + \tau$ payoff, and $X_{t+\tau}$ is the cash payoff at time $t + \tau$. In the case of a zero coupon bond we write (where the payoff is \$1):

$$\delta_t(\tau) = E[m_t(\tau)].$$

Knowing the discount function, $\delta_t(\tau)$ for all maturities τ, one can discount any cash flow and price any asset. However, since $\delta_t(\tau)$ is not directly observable, it must be estimated. Past research has shown that deriving the discount rate from the forward rate produces a discount function which has the desired properties and accuracy sufficient for pricing assets. Modeling the forward curve directly forces the researcher to recover the zero coupon yield curve from the observable yield on existing securities and then use some kind of an interpolation method to determine the curve between the relatively few available observations.

Diebold and Li (2006) indicated that past research with regards to the construction of the yield curve and the estimation of the discount function, have diverged into theoretical and empirical branches. The theoretical methods (or the *Equilibrium Approach*) typically characterize the behavior of the short term rates, which is determined by a set of parameters, which in turn can be estimated from real data. Such are, among others, the seminal papers by Vasicek (1977) and Cox et al. (1985). This approach is useful when one is trying to synthesize the dynamics that are at play in determining the behavior of the curve under different economic scenarios, but they are not developed enough at this point to help practitioners with asset pricing.

Empirical methods (*No Arbitrage Approach*) for term structure construction are independent of a specific economic theory. They merely aim to represent the term structure at a point in time based on available information, mainly for asset pricing purposes. The construction of the term structure is subject to several statistical and mathematical constraints, all of which are concerned with the exclusion of arbitrage possibility in the pricing process. To that end, the major requirements from the constructed curve is continuity and smoothness. These properties will ensure that all assets are priced logically and that each distinct discount value is yielding a distinct asset price, assuming an identical cash flow. The method proposed in this paper belongs to this category of research.

Fabozzi (2000) describes the two approaches to valuing a treasury bond, or determining the discount rate δ_t. The traditional approach is concerned with discounting each cash flow (coupons and principal) to it's present value and adding these values to arrive at the value of the bond today. In this process, each coupon, and the principal payments are discounted to the present time by a single, unified, discount rate δ_t . It is noted as follows:

$$P_{t_i} = \sum_{t=1}^{n} \delta_t \cdot Z_i$$

where Z_i is the time $i >= t$ cash flow.

However, from a theoretical standpoint, this approach is flawed, since it views all cash flows from the bond as identical, thereby discounting all at the same rate. Theoretically, each cash flow from the bond is independent of the other flows. Therefore each of these payments should be discounted independently by the proper discount rate that corresponds to that cash flow's maturity. An intuitively appealing reason, among others, is that the flows that are scheduled for the near future are more certain to happen than the more distant ones. This justifies a higher discount rate. The no arbitrage approach accomplishes this by viewing each payment as a zero coupon "mini" bond, discounting each independently by the proper discount rate for its maturity. The value of all of these mini-bonds combined is the value of the bond. This is noted as follows:

$$P_t = \sum_{t=1}^{n} \delta_t(i) \cdot Z_i$$

This approach is superior to the traditional one, in that it does not allow an agent to buy the bond, and then sell the cash flows from the coupons for a profit.

Viewing each coupon as a separate bond is not just an academic exercise to bond valuation It is also conducted in practice. The practice of selling the cash flows from the individual coupons separate from the principal is called Stripping. There is an active market for stripped securities, and they are referred to as *treasury strips*. This forces practitioners to determine the theoretical rate on a zero coupon security, for all maturities, hence construct the estimated zero yield curve, in order to assist with the pricing of these securities.

The complete term structure of spot interest (or the yield on zero coupon bonds) rates is not readily observable and must be estimated. This is because the term structure is constructed from a limited number of treasuries. In the case of the U.S., The *on-the-run treasuries*, the most recently issued and most liquid of all treasuries, are the ones used to construct the curve. Some of the reasons *off-the-run treasuries* are not used are that they are not very liquid, and hence their prices are biased by a liquidity premium, that will introduce a bias to the zero rates they will produce. The on-the-run treasuries are issued in a few maturities - which gives rise to the interpolation issue which was eluded to above. Out of these few observations only the ones that mature in one year or less are true discount bonds. Hence a major part of the estimating the entire yield curve is actually concerned with extracting the spot rates (zero coupon rates), as well as discount factors, from the coupon payments of the longer-maturity bonds. This necessitates a solution to the following problems:

I. Researchers must first recover the yield on a zero coupon curve from a limited number of observations of yields on coupon bonds, by interpolating the curve between the available observations using one of the many methods available (see Hagan and West (2006) for an extensive survey of available methods), while preserving the desired smoothness and continuity of the curve.

II. The data contains securities with different attributes such as different levels of liquidity, coupon or non coupon bonds, and securities with different tax treatments, all of which have a bearing on their prices. Researchers must account for all of these factors and arrive at the true zero coupon rate for each maturity.

A.V-D proposed a method for fitting the yield curve by defining a maximum smoothness criteria for the forward curve. A necessary condition for the A.V-D method to work is the existence of a zero coupon curve.

The method I present is designed to combine the interpolation and smoothing features of the method proposed by A.V-D (and augmented by Lim and Xiao (2002)) with a zero coupon stripping process. The result is a simultaneous approach to stripping and smoothing which produces an elegant tool for practitioners. The importance of combining the two processes was noted by Hagan and West (2006). In a standard linear bootstrapping method that is independent of the stripping process, a linear search for the first "unknown" node (the terminal bond cash flow) is conducted. It is preformed at maturity of the bond one is trying to price by scanning the yields at that maturity, until the correct yield which prices the bond correctly is discovered. Interpolation then takes place between the last known node and the one that was just now estimated. The coupons are then priced of this linear approximation, and will inevitably be biased. By combing the stripping and the smoothing process, the initial estimate for any bond x is being revised, along with the pricing of its coupons, every time I add a bond, $x+1$, $x+2$...,$x+m$ etc'. This ensures that the curve is not 'rigid' in the sense that the initial estimate of the yield on bond x is allowed to accommodate the pricing characteristics of all the other bonds and coupons that follow it on the curve.

A second contribution is made by introducing a Genetic Programming (GP) approach to fit the forward curve. Genetic Programming is a branch of research in computational science that uses evolutionary principles to search for an optimal solution by searching over functional spaces. The interesting feature that GP adds to our topic of research is that we will attempt to fit the yield curve without any preconceived notions about the functional form beyond those restrictions implied by financial theory. In particular, the GP does not impose the restriction that the forward curve should be of the piece-wise polynomial class of functions, as was widely assumed in past research. The GP experiment failed to produce the desired outcome and I discuss possible reasons for that at the end of Chapter 4.

Additional motivation for yield curve modeling comes from the fact that other than pricing assets, the term structure of interest rates has additional important applications (which are outside of the focus of this research). For example, it is of major importance in macroeconomics, particularly in the field of Monetary Policy (see Svensson (1995), Dahlquist and Svensson (1996), Shiller (1991), where it has been shown that the spread between long and short term rates is useful in predicting future interest rates, and forward rates. This, in turn, helps practitioners predict income and inflation (see Estrella and Mishkin (1998)).

1.3 Definitions and Notation

A *zero coupon bond* (also referred to as a "discount bond"), is a financial instrument that promises to pay a single payment at a specified date in the future. A zero coupon bond does not make any intermediate payments between the time of purchase,and the time of maturity. U.S. government obligation with maturities of less than 12 months are issued in the form of a discount bond.

A *coupon bond* is a bond that makes intermediate interest payments at specified dates during the time the bond is outstanding. At maturity, the last interest payment, as well as the face value (principal) of the bond are due. Coupon payments are made every six months, but the rate is always quoted on an annual basis. U.S government obligations that are issued with more than 12 months to maturity are issued in this format. The *coupon yield* is the simplest measure of yield and is determined at the time the bond is issued. It is simply the rate of the coupon, expressed as percentage of the face value of the bond.

The measure of return that takes into account the coupon yield as well as changes in the market price of the bond is called the *current yield*. The current yield is defined as the coupon payment divided by the current price of the bond.

The *yield to maturity* (denoted y) measures the total return one earns on a bond, assuming it is held to maturity. In addition to the factors accounted for by the current yield, yield to maturity also accounts for the reinvestment, at the same rate of return that the bond has, of cash flows received via the coupon payments.

The discount bond market, which we describe below, is characterized by the following assumptions, as in Adams (2001):

Continuous Trading: Trading occurs continuously from the current time to some distant future time such that all maturities of all instruments fall between now and the trading horizon.

Frictionless Markets: Transactions costs are absent, and there are no tax implications to any trade. No restrictions on trade, and goods are infinitely divisible.

Competitive Markets: Any agent can buy and sell without affecting market prices.

Efficient Market Information is available to all agents simultaneously, and full use of it is possible at all times.

Complete Markets: Any desired cash flow can be constructed from a suitable financing strategy based on a portfolio of discount bonds.

No Arbitrage: The price of a portfolio is the sum of constituent parts.

Profit Maximization: All agents are rational and prefer more to less.

Now we establish some notational conventions that will we use throughout the paper, and explain the mathematical relationships between the different curves that play a role in our problem

Let the time t price of a T-period zero coupon bond, with a face value of \$1 be

$$\delta_t(T) = \frac{1}{(1 + y/n)^n}$$

where n is the number of compounding periods until maturity. For pure discount bonds, the price is $\delta_t(T) \in (0, 1]$ which we refer to as the discount rate at time t of a \$1 zero coupon bond paying off in period $t + T$.

When we have continuous-time compounding we have

$$\delta_t(T) = \exp\left(-y_t(T)\, T\right).$$

For a coupon bond we let $P_t(T)$ be the period t price of a bond that matures at time $t+T$. If the bond has a face value of \$1 and pays a coupon of C_i at dates t_i, $t \leq t_1 < \cdots < t_n$ (where $t_n \leq T$), then the yield to maturity, or *internal rate of return*, is the rate y which solves the equation

$$P_t(T) = 1 \cdot \exp(-y\, T) + \sum_{i=1}^{n} C_i \, \exp(-y\, t_i).$$

Note that the above expression assumes continuous compounding. If we assume a fixed coupon C that is quoted in annual terms but paid semi-annually for $n > 1$ payments with simple compounding, we solve for the yield to maturity from

$$P_t = \sum_{i=1}^{n} \frac{C/2}{(1+y/2)^{(i-1-\nu/\epsilon)}} + \frac{1}{(1+y/2)^{(n-1+\nu/\epsilon)}}$$

where ϵ is the number of days between the previous coupon payment and the next coupon payment, and $\nu \leq \epsilon$ is the number of days from the purchase date (settlement date) to the next coupon payment. This is essentially places a restriction on the yield to reflect the accurate maturities of the cash flows.

A *Holding Period Return* on the bond is the rate of return on the bond over some period of time that is less than the time to maturity and is defined as

$$H_t(n, T) = \frac{\delta_{t+n}(T-n)}{\delta_t(T)}$$

where n is the holding period, T is the maturity date of the zero coupon bond, and t and $t+n < t+T$, are the buying and selling dates, respectively. The complication is that $\delta_{t+n}(T-n)$ is not known at time t.

The *Forward Rate* is the rate at which one can contract today, to borrow or lend funds tomorrow. The forward rate represents the consensus expected value trading agents hold with respect to the spot rate to be realized at some point in the future.

Let $F_t(n, T)$ be the time t price of a forward contract to deliver a $T-n$ maturity zero-coupon bond at date $t+n$ which matures at date $t+T$. The forward rate can be decomposed into a series of spot rates. The arbitrage free condition dictates that the following two investment plans yield an identical rate of return (Assume a T period investment horizon): (1) Invest funds today for n periods (at today's spot rate, $\delta + n$), then reinvest again next period for an additional $T-n$ periods (at today's implied forward rate, $F_t(n,t)$), and (2) Invest for T periods, at today's spot rate for a T period investment horizon at today's spot rate, $\delta_t(T)$. So in order to avoid arbitrage, it must be the case that

$$F_t(n, T) = \frac{\delta_t(T)}{\delta_t(n)}.$$

Define the continuously compounded yield to maturity of the forward contract (commonly referred to as the *forward rate*) to be $f_t(n, T)$ and compute it from

$$F_t(n, T) = \exp\left(-f_t(n, T)\,(T-n)\right).$$

Then, from the definition of the zero coupon yields, we get the forward rate as follows:

$$f_t(n,T) = \frac{1}{T-n}\left(y_t(T)\,T - y_t(n)\,n\right).$$

In the limit, as $n \to T$, we get the *instantaneous forward rate* at time t for maturity T as

$$f_t(T) = \frac{d}{d\,T}\left(y_t(T)\,T\right) = y_t(T) + T\,\frac{d}{d\,T} y_t(T).$$

Thus, the forward rate is the spot rate plus the slope of the spot curve times its maturity. The forward curve is above the spot curve when spot rates are increasing with maturity (the normal yield curve) and below the spot curve when spot rates are decreasing with maturity (an inverted yield curve). Note that the forward curve intersects the spot curve at the maximum or the minimum of the spot curve to the extent a minima or a maxima exist. If the yield curve is flat then the forward rate equals the spot rate for all maturities.

Integrating the forward curve gives

$$\int_0^T f_t(\tau)\,d\tau = \int_0^T d(\tau\, y_t(\tau)) = T\,y_t(T)$$

so that

$$y_t(T) = \frac{1}{T}\int_0^T f_t(\tau)\,d\tau.$$

In this sense, the spot rate is the average of the forward rates.

We note that,

$$\delta_t(T) = \exp\left(-\int_0^T f_t(\tau)\,d\tau\right) = e^{-y\cdot T}$$

so $\delta(T)$ is smooth with respect to $f(\cdot)$ and $y(\cdot)$, and

$$f_t(T) = -\frac{\frac{d}{dT}\delta_t(T)}{\delta_t(T)}$$

so the forward rate is the percentage change in the discount rate.

To summarize, the term structure of interest rates can therefore be represented by the discount rate, the instantaneous forward rate, or the yield to maturity, as follows:

$$\delta_t(T) = \exp(-y_t(T)\,T) = \exp\left(-\int_0^T f_t(\tau)\,d\tau\right) \tag{1.1}$$

$$f_t(T) = y_t(T) + T\,\frac{d}{dT} y_t(T) = -\frac{\frac{d}{dT}\delta_t(T)}{\delta_t(T)} \tag{1.2}$$

$$y_t(T) = \frac{1}{T}\int_0^T f_t(\tau)\,d\tau = -\frac{1}{T}\ln\delta_t(T). \tag{1.3}$$

Finally, we note that the spot rate and the forward rate start at the same point, $y_t(0) = f_t(0)$.

CHAPTER 2

OVERVIEW OF THE LITERATURE

2.1 Theoretical Foundations of the Term Structure

There were four early economic theories with regards to the term structure of interest rates. The first is the *expectation hypothesis* (EH) (see Lutz (1940), among others), which relies on the assumption that bond prices are derived from the expectations agents have with regards to future spot rates. The premise is that the holding period return is equal to repeated investment in a series of short term IOUs, which collectively equal the holding period for the long term bond. Alternatively it can be thought of as if the expected return for a fixed unit if time is equal for bonds of all maturities.

The second is the *liquidity preference hypothesis* (see Hicks (1946)) (LP), which is an extension of the EH approach, but it places a higher degree of importance on the risk preference of the agent. Risk averse agents will demand higher forward rates compared with actual expected spot rates. This will increase with time to maturity, since the time to maturity is the main source of risk.

The *market segmentation hypothesis*(see Culbertson (1957)) (SH), also attempted to explain the premium placed on long-maturity bonds. It postulates that agents have risk preference that is highly segmented, causing bonds of different maturities to trade in disconnected markets. It follows that the supply and demand for bonds in one segment does not affect the price of bonds in the other segment. The result is that different yields will exist for different maturities, and these yields will behave differently under similar market conditions.

The *preferred habitat theory* (see Modigliani and Sutch (1966)) (PH) is essentially a blend of both SH and LP. They adopt the principle of equality between expected returns, modified by risk premiums on longer maturities securities, which is the basic premise of the LP, but they also blend SH and stress the principle of habitats in which lenders and borrowers are choosing long term securities for various reasons, such as long investment horizon etc'. Modigliani and Sutch are also incorporating the views of Keynes (1936) and Duesenberry (1958) with regards to the stochastic process that govern the evolution of interest rates. Keynes postulated that there is some Long Term Rate that is the common experience of all agents in the economy, and that the spot rate gravitates towards. Duesenberry argued that interest rates are exploratory in nature and do not gravitate towards a long term "acceptable" level. Modigliani and Sutch's model accommodate both views.

Roll (1971) set up one of the early attempts to test the theories above. He constructed a mean variance model and a condition for market efficiency to relate ex-ante expectations and ex post data observations. The underlining assumption was that rational expectations should not differ from the ex-post results in a systematic manner, hence it is a testable hypothesis to see if rational expectations are consistently correct ex post or not. Roll counted four risk elements that are a major determinants of interest rates: (1) Default; (2) Liquidity; (3) Inflation Uncertainty; and (4) incompatibility between bond maturity and investors horizon. He then adds an additional one which is termed Portfolio Interaction. In essence it is a combination of the rational expectations hypothesis which links forward rates and expected future spot rates with the Sharpe (1964) Capital Asset Pricing Model. Roll derives a relation between the portfolio risk of bonds and the term structure Liquidity Premium, which is the main premise of the Liquidity Preference Theory.

2.2 Equilibrium Approach

Beginning with Vasicek (1977), a number of partial Equilibrium models have attempted to model the behavior of the term structure of interest rates. He and others such as Brennan and Schwartz (1979), John C. Cox and Ross (1981), as well as Langetieg (1980) have all made ad hoc assumptions with regards to the stochastic process that govern the evolution of interest rates. A characterization of the term structure was then deduced from these assumptions in the context of an efficient market. These models yielded very specific formats for the stochastic process, and did not conform well to the observed data. The yield curve one would actually observe in the marketplace demonstrated more varied shapes than the ones proposed by these models. Hence practitioners could not rely on these early models to price securities.

Cox et al. (1985) (CIR) were the first to formulate a model for the term structure of interest rates in the context of a general equilibrium. They allow consumption decisions, technology driven production, as well as multiple classes of investment vehicles and risk preferences to enter as determinants of the curve. This approach essentially combines the PH, as well as EH, and LP in the following sense: they allow the agents a choice of what type of bonds to buy (which is in essence SH), when to consume and when to save (PH), and the choice of different investment vehicles (LP). Their model is one in which identical individuals are seeking to maximize their lifetime utility, and the production possibilities are determined by the state of the technology which changes randomly over time (A). The agent is to maximize utility, $U(\cdot)$:

$$\int_{t}^{t'} U(C(s), A(s), s)ds$$

subject to a standard budget constraint. The supply in the economy is a derivative of the state of the technology with evolves over time. Each agent is choosing (1) optimal consumption C^*, where $C(s)$ is consumption at time $t \leq s \leq t^{'}$; (2) optimal proportion of wealth to be invested in each of the production processes; and (3) optimal proportion of wealth to be invested in contingent claim. Any remaining wealth will be lent out, or borrowed to satisfy the budget constraint. In equilibrium, the interest rate and the expected rate of return on contingent claims are such that all real capital is invested in production. The source of variation in the economy is stemming from only one factor - the state of the technology, A , which follows a random process. In equilibrium, under a steady state A, interest rates are gravitating towards a long-term value θ. This behavior exhibits four important properties: (1) no negative rates (2) zero rates can only go up (3) when rates increase, the corresponding variance increases as well, and (4) there is a steady state distribution for the rates.

CIR noted the price of a bond at time T as P and write the first order condition to the maximization problem as

$$1/2\sigma^2 rP_{rr} + k(\theta - r)P_r + P_t - \lambda rPr - rP = 0$$

where the subscripts denote partial derivatives, r is the short-term interest rate, and the term λr is the covariance of changes in interest rates with the percentage of the portion of optimally-invested wealth. Note that factor k determines the speed of the gravitation of the rates towards their steady state θ.

The term $1/2\sigma^2 rP_{rr} + k(\theta - r)P_r + P_t$ represents the expected price change of the bond, and is the result of Ito's formula. It follows that the expected rate of return on the bond is $r + (\lambda r P_r/P)$, or, the instantaneous return premium on the bond is proportional to its interest elasticity.

The solution to the problem is an efficient allocation between consumption and investment, which include an equilibrium price for the bond

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

where A is the equilibrium amount allocable to investment which is being discounted by B (the yield to maturity), which in turn is a function of risk aversion, as well as time to maturity. This expression is roughly equivalent to $\delta_t(T) = \exp[-(T-t)y_t(T)]$ in our notation.

The absence of arbitrage in equilibrium suggests that the bond price is also equal to the risk free rate plus a risk premium only. The dynamics these prices are following are denoted by:

$$dP = r[1 - \lambda B(t,T)]Pdt - B(t,T)P\sigma\sqrt{r}dz_t$$

CIR note that as at the limit $(T \to \infty)$,the yield to maturity will converge to a steady-state-dependant condition:

$$\frac{2\kappa\theta}{\gamma + \kappa + \lambda}$$

which implies

$$f_{t,T\to\infty}(T) \to y_t(T)$$

so that the forward rate converges to the spot rate in the limit. This, in turn, implies that the derivative of the yield curve must converge to zero as $T \longrightarrow \infty$. Finally, when the time of the purchase of the bond approaches the maturity of the bond $(t \to T)$, the yield to maturity will approach the spot rate for all parameter values.

The CIR model shared similar short-comings with the partial equilibrium models which have preceded it: it was able to explain movements in the level of the term structure as well as some of the changes in slope but did not work well to explain curvature, or inverted curve. Subsequent studies that have followed in the footsteps of CIR have generalized the CIR framework in an effort to achieve a better fit to the data. One such improvement was made by Longstaff (1990) who introduced a non-linear relationship between technology and production. Longstaff and Schwartz (1992) added a second layer of dynamics in the form of volatility in short term interest rates. This enabled the estimated term structure to be more flexible in its shape, taking into account not only the level of interest rates, but the volatility of them as well.

Hull and White (1990) also extended the work of Vasicek (1977) and CIR. They presented extensions for the two above models. Their contribution is showing that the stochastic process that governs short term rates can be deduced from the term structure of interest rates, and the term structure of spot or forward interest rate volatilities. Knowing the nature of this stochastic process enables either model to price contingent claims. The conjecture is that the market's expectations with regards to future rates involves time-dependent parameters, as well as state variables and the spot rate itself. To incorporate this, the CIR model is modified to include $\theta(t)$, which is a time-dependent drift, in the following manner. For Vasicek's model :

$$dr = [\theta(t) + a(t)(b - r)]dt + \sigma(t)dz$$

here θ is a time dependent drift and a, b, and σ are all positive constants, r is the short term rate, and dz is a Wiener process. For the CIR model the equivalent expression is

$$dr = [\theta(t) + a(t)(b - r)]dt + \sigma(t)\sqrt{r}\,dz$$

Hull and White report that the extended Vasicek's model analytically derives parameters and short term rates, as well as European bond options, which is an appealing practical tool.

Jordan and Mansi (2003) estimated the term structure from a sample that contained on-the-run treasuries only. This alleviated some of the issues associated with different attributes of securities such as different levels of liquidity, coupon or non coupon bonds, and securities with different tax treatments, but exacerbated the interpolation problem as there are only 8 observed data points across the spectrum of maturities, so the gaps are wider. They compare a number of methods and they show that continuous time models perform better than discrete ones, and that there are two sources of errors in the price prediction: (1) interpolation errors, which is related to the method used by the modeler to estimate the curve; and (2) random pricing errors. The random pricing error is small (about 7%) in relation to the total pricing error.

Dai and Singleton (2000) augmented the work of CIR, Vasicek and others, by defining the "affine" model for the Term Structure. Their aim was to formally characterize and point out the differences and similarities of the different specifications of the affine models used by their predecessors. The defined an"admissability test" for affine models based on their price predictability. Those models that pass the test, were then classified into separate "families" based on their factors. Each of these families can be characterized by a "maximal" model, that is the generalization of all members of that family. The idea is to find that "maximal" model and compare it to earlier models in order to find an over-identifying factors in the earlier models. In the their model the price of a zero coupon bond $\delta_t(T)$is:

$$\delta_t(T) = E_t^Q[e^{-\int_t r_s ds}]$$

where E_t^Q is the risk neutral expectation, and r is the short term rate. The N-factor affine term structure model is constructed assuming the short rate, $r(t)$ is a an affine function of a vector of unobserved state variables $Y(t) = (Y_1(t), Y_2(t)....Y_N(t))$.

The instantaneous short rate is defined as

$$r(t) = \delta_0 + \Sigma_{i=1}^{N} \delta_i Y_i(t) = \delta_0 + \delta_y^{'} Y(t)$$

and $Y(t)$ follows the vector process

$$dY(t) = \tilde{\kappa}(\tilde{\theta} - Y(t))dt + \sum \sqrt{S(t) d\tilde{W}(t)}. \tag{2.1}$$

where last term is the "affine" diffusion" and $\tilde{W}(t)$ is an N dimensional independent standard Brownian motion under the risk neutral measure, $\tilde{\kappa}$, as well as $\sum$ are $N \times N$ matrices, and $S(t)$ is a diagonal matrix with the i_{th} element defined as

$$[S(t)]_{ii} = \alpha_i + \beta_i^{'} Y(t) \tag{2.2}$$

where both the drift in (2.1) and the conditional variances in (2.2) are affine in $Y(t)$.

This family of models accommodates time varying means and volatilities of the state variables through affine specifications of the risk neutral drift volatility coefficients. Another nice property is that they produce a closed—form expression for the price of a zero-coupon bond, which simplifies implementation. Dai and Singleton have shown that the models that have preceded them imposed undue restrictions, and essentially over-identified the "true" (maximal) model. They have attempted to fit the model to the data, without much success, but note that the reason for it lies with their formulation of the risk premium—which they say may not be linear, as they have modeled it. Cochrane (2001) indicated that these models all maintain multiple factors as determinants of bond prices, not just short term rates, as most earlier models. The general form is a linear one, in which the log price of the bond is determined by a number of state variables. Other bond prices are also part of the state variables. Short term rates will be forecasted by lagged values of both short and long term rates. Campbell (1997) listed some of the advantages of the affine class models: first, log bond yields inherit the conditional normality assumed for the underlying state variables. Second, because log bond yields are a linear functions of the state variables, one can re-normalize the model so the yields themselves are the state variables. Some of the disadvantages are: (1) affine models are limiting the way in which rate volatility is able to change with the level of rates; and (2) the covariance matrix of bond returns of affine model has a rank of K since all K bonds are linearly related. Therefor, one must add an error term to the model.

The equilibrium approach, however, fell short of the most basic requirement posed by practitioners - it was not able to price assets. The central role that the yield curve have in pricing contingent claims, and the growing popularity of these instruments in financial markets, as well as the deficiencies that the theoretical models described above exhibited, gave rise to a new developments in yield curve research. An empirical estimation of the yield curve, one that is accurate enough to price assets in the marketplace, began to develop along side the equilibrium models. The method was to fit a function to an observed set of data, while satisfying several basic assumptions that stem from the theory. We call this branch of research the no-arbitrage approach.

2.3 No Arbitrage Approach

One of the early building blocks of the no-arbitrage approach was McCulloch (1971). He developed a technique of fitting a smooth curve (discount function) to observations of prices of bonds with varying maturities using polynomial splines. The discount function, is then utilized in derivation of security prices, forward rates, and the yield curve.

The discount function is defined as the present price (value) of \$1 payable in T periods from now. It is continuously differentiable, and monotonly decreasing. The set up is as follows:

$$p = 100\,\delta_t(T) + C\int_0^T \delta_t(T)dt$$

where p is the price of the bond, and C is the coupon rate. $\delta_t(T)$ is the discount function for the T time maturity and it takes the form of

$$\delta_t(T) = 1 + \Sigma_{j=1}^{k} a_i h_i(T) \tag{2.3}$$

where h is continuously differential between 0 and k, and δ is 1 at the present time.

McCulloch notes that the actual form of h is unknown and is subject to debate but any form must satisfy $h(0) = 0$, and h is continuously differentiable. He proposes the use of polynomial spline in the form of:

$$h_j(t) = m^j \; j = 1, 2, \ldots, k.$$

The theoretical motivation he offers is that: (1) a polynomial does not depend on the distribution of T, and (2) it does not provide a greater capability for providing resolution for the value of T where T is more likely to occur. Since investors are much more concerned with small deviations in the very short term (short maturities), but are less so for very long maturities securities, he can use a relatively low order polynomial to fit the observations more accurately close to the origin of the curve which is appealing to practitioners.

The discount function $\delta(T)$ is an exponentially decaying function where the rate of decay is actually the instantaneous forward rate, noted as:

$$f_t(T) = -\delta_t^{'}(T)/\delta_t(T).$$

Exponentiate both sides and the instantaneous forward rate can be stated as

$$f_t(T) = \lim_{h \to 0} \left[\frac{\delta_t(T)/\delta_t(T+h) - 1}{h} \right]$$

Combining with 2.3 one can see that:

$$\hat{f}_t(T) = \frac{-\sum_j \hat{a}_j f_j'(T)}{1 + \sum_j \hat{a}_j h_j(T)}$$

The mean forward rate is the average of the instantaneous forward rate $(f(m))$:

$$f_t(T_1, T_2) = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} T^2 \cdot T_1(T) dT$$

This gives the rate of decay of the discount rate $\delta(m)$ The yield curve is the average of the rate of decay or

$$y_t(T) = \frac{1}{T} \int_o^T f_t(x) dx$$

McCulloch (1975) is extending his model by examining the tax-adjusted yield curve. But a greater contribution of this paper for our purposes is the introduction of cubic splines. Since an r degree spline is a piecewise function with an r degree polynomial which has $(r-1)$ continuous derivatives, the r_{th} derivative is a step function. As noted above, in the short end of the yield curve one can find many observations for note maturities, much more so than for long maturities. This justifies the use of cubic splines, which provide greater flexibility since the parameters in one interval are not affected by observations in another. The intervals of the cubic spline must be spaced out correctly to include roughly equal number of observation in each adjacent knot. Another motivation for the use of cubic spline is the shape of the forward curve it produces, which is smoother than the one produced by the quadratic one. McCulloch reports that the cubic spline performed better in fitting the long term as well, exhibiting a "credible" 3.5 % or 4.5 % premium for long maturities over shorter ones. This is was an improvement over the polynomial spline that showed unreasonable long term premium that was between -9 % and +57%.

Vasicek and Fong (1982) (VF) proposed using exponential splines to fit calculated spot rate observations. Their model was:

$$P_{t,k} + A_{t,k} = \delta_t(T_k) + \sum_{j=1}^{L_k} C_k \delta_t(T_k - j + 1) + Q_k - W_k + \epsilon_k, \; k = 1, \ldots, n \tag{2.4}$$

where $\delta_t(T)$ is the discount function - the value, at time t, of one unit of payment due at time T. n is the number of bonds used to estimate the term structure, T_k is the time to maturity of the k-th bond, C_k is the semi annual coupon rate of the k-th bond, expressed as a fraction of the bond value, and P_k is the price of the k-th bond, expressed as a fraction of the par value. Q_k is the is the price discount attributed to the effect of taxes. W_k is the price discount due to call features, and ϵ_k is a residual error with a mean of 0. We also have $A_k = C_k(L_k - T_k)$ is the accrued interest portion on the k-th bond,where $L_k = T_k + 1$ being the number of coupons to be received.

Polynomial splines, VF argued, are not well suited to fit this function since they have different curvatures than exponentials, which makes them inferior in trying to fit the yield curve, and although some modifications can be made to "make them close to an exponential curve by choosing a sufficiently large number of knot points, the local fit is not good". Specifically, their critique of McCulloch is based on the fact that the derivatives of the log of the discount function are not stable under his construct, especially when it was combined with estimated tax rates, which increased its sensitivity to errors in the tax rate estimation. Also, polynomials cannot be forced to asymptote for long maturities, as observed in the data. It is also argued that this is the reason McCulloch's curve is not stable for long maturities. Working with the log of δ_t was not possible since the transformation is not linear. VF instead applied the transformation directly to the argument of the function δ_t. This transformation allows the fitting of the model to be done with linear combinations of splines:

$$t = -\frac{1}{\alpha}\ln(1-x), 0 \leq x < 1.$$

They then approximate δ_t by a new function , $G(x)$:

$$\delta_t = \delta(-\frac{1}{\alpha}\ln(1-x)) \equiv G(x)$$

and $G(x) \sim (1-x)^{\gamma/\alpha}$ where α is now the limiting value for the forward rate, and can be fitted to the data along with the other parameters.

If $g_i(x)$ is the base of a polynomial spline space, then any spline in that space is a linear combination of the base, hence if $G(x)$ is in the space, we have:

$$G(x) = \sum_{i=1}^{m} \beta_i g_i(x)$$

for x between 0 and 1.

$G(x)$ is now a decreasing function defined over the interval $0 \leq x < 1$, with $G(0) = 1$ and $G(1) = 0$ The model is now linear in $G(x)$ and can be fitted with polynomial splines. The model can be re-written and the term structure can be estimated using the following format:

$$P_k + A_k = \sum_{i=1}^{m} \beta_i g_i(x) + \sum_{j=1}^{L_k} C_k g_i(X_{jk}) - q\frac{C_k}{P_k}\left(\frac{dP}{dY}\right) - wI_k + \epsilon_k.$$

where $W_k = \frac{dP}{dY}_k I_k$, Now call the left hand side U_k, and define

$$Z_{ki} = g_i(X_{k1}) + \sum_{j=1}^{L_k} C_k g_i(X_{kj})\, i = 1, 2, \ldots, m.$$

One gets:

$$Z_{k,m+1} = -\frac{C_k}{P_k}\left(\frac{dP}{dY}\right)$$

and $Z_{k,m+2} = -I_k$ for $k = 1, 2, ...n$.

Now define β as

$$\hat{\beta} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}U$$

were $U = (U_k), Z = (Z_{ki})$, and Ω is the residual covariance matrix from the term structure model in (2.4). The sum of squares here is a function of α only and is defined as: $S(\alpha) = U'\Omega^{-1}U - \hat{\beta}'z'\Omega^{-1}U$ The value of α that minimizes this sum of squares is the one to use to produce β, and when fitting the curve. Estimating the parameters, the model reduces to :

$$\hat{\delta}_t = \sum_{i=1}^{m} \hat{\beta}_i g_i (1 - exp^{-\hat{\alpha}t}) \tag{2.5}$$

for $t \geq 0$.

If one rewrites (2.5) in terms of t, then any interval between knot points $G(x)$ is a cubic polynomial, and $\delta(t)$ takes the form:

$$\delta_t = a_0 + a_1 e^{-\alpha t} + a_2 e^{-2\alpha t} + a_3 e^{-3\alpha t}.$$

$\delta(t)$ is then continuous in the first and second derivatives at the knot points. VF named this family of models "Third Order Exponential Splines".

The major contribution VF see in using exponential splines is that they increase the stability of the projected forward curve, and improve the asymptotic properties of the forward curve for long maturities so that the curve is more robust in the longer maturities.

Shea (1985) reviewed VF's work and made the following observations:

1. Their model is indeed linear in the β s but not so in the α s. It is therefore inescapable to use nonlinear routines in estimating it.

2. VF asserted that present value functions are exponential decays, but they did not enforce that condition. This caused these functions to deviate from that pattern.

3. The Discount function transformation G(x) was almost never linear when estimated using polynomial spline approximations. This caused great variations in the short term maturity space and contributed to the instability of the model.

4. VF were correct to point out that polynomials have a difficulty in modeling exponential functions, but neglected to state that this does not extend to local polynomial approximation.

Shea summarizes that the using exponential splines in the way VF did is not more likely to produce stable forward curves than in using ordinary polynomial splines, as long as the assumed exponential form of the discount function is modeled in only a piecewise fashion. Splines approximations are joint, but local, approximations of a function. Hence one can expect that "in a piecewise polynomial function, each polynomial piece of which is estimated over a not too wide domain, might very well mimic a piecewise exponential". It follows that there is no material difference between polynomial and exponential splines, and the fact the exponential splines require additional manipulation of the data, as well as a higher degree of computational burden, Shea proposes to give preference to polynomials .

Nelson and Siegel (1987) have proposed a second order differential equation based model to capture the determinants of the yield curve shape. Their model is motivated by the expectation hypothesis as well as no arbitrage arguments. They construct the following

$$f(T) = \beta_0 + \beta_1 exp(-T/\tau) + \beta_2[(T/\tau)exp(-T/\tau)] \tag{2.6}$$

where $f(T)$ is the instantaneous forward rate to maturity T, τ is a constant, the βs are determined by the initial state of the yield curve, and T is the time to maturity.

The reason for this construction is that their survey of past research showed that virtually all yield curve research showed the term structure function to be monotone, humped or occasionally S shaped. The construction above is a ready format that yields such functions: $\beta_1 exp(-m/\tau)$ is a monotonically decreasing factor of the time to maturity (if β_1 is positive), and the humped shape is achieved by the last factor, $\beta_2[(T/\tau)\exp(-T/\tau)]$. When time to maturity approaches infinity, the forward rate is approaching the constant β_0, and as we approach the time to maturity the rate approaches $\beta_0 + \beta_1$. The claim is that if spot rates are generated by a differential equation, then the forward rates being forecasted, will be the solutions to these equations.

If (2.6) is integrated between 0 and m and then divided by T one gets

$$y(T) = \beta_0 + (\beta_1 + \beta_2)[1 - \exp(-T/\tau)]/(T/\tau) - \beta_2 \exp(-T/\tau)$$

where β_0 is the long-term to maturity effect, β_1 is the effect of the short term maturities, and β_2 is medium, and all the βs are linear, given τ. They test their model on US Treasuries and find a high correlation (.96) between the present value of a long term bond implied by the fitted curves generated by their model, and the actual reported price of the bonds.

Gimeno and Nave (2006) report that nine out of the thirteen central banks that report their estimation methods to the Bank of International Settlements, use a function developed by Nelson and Siegel (1987) and augmented by Svensson (1995).

Jordan and Mansi (2003) have compared five methods for curve construction using on the run treasuries only. They show that the exponential function based Nelson Siegel method, and the four-parameter model of Mansi and Phillips (2001) outperform several other methods in pricing ability. However, as Redlemen (2004) pointed, these methods still lack in pricing accuracy, since one cannot reconstruct the par value of the bond by using the present value of the cash flows modeled.

Svensson (1995) showed how central bankers can use the Nelson and Siegel model (albeit extended, and a bit more flexible) to predict the forward rate curve. This is important since central banks are including the forward rate curve as one of the forward looking indicators when devising policy. Svensson makes the argument that the forward rate curve can be thought of as a time-path for short term rates. It follows that a logical interpretation is to divide the expectations into short, medium, and long term rates. This is very useful to central bankers, much more so than looking at spot short rates which are essentially the average of future short rates. Svensson claims that yield error minimization in the context of monetary policy is more relevant than price error minimization. Trying to minimize pricing errors can result in large yield errors, since prices are very sensitive to yield errors in the short term.

Svensson's innovation was to add a fourth term to the Nelson and Siegel construct. The new, modified Nelson and Siegel model, was specified as

$$\begin{aligned} y(T) =& \beta_0 + \beta_1 \exp(-T/\tau) + \beta_2[(T/\tau_1)\exp(-T/\tau_1)] \\ &+ \beta_3[(T/\tau 2)\exp(-T/\tau_2)] \end{aligned}$$

This formulation imposes a horizontal asymptote on the term structure. This is necessary since the expectations today as to what interest rate will be 20 years from now, should not be materially different than the expectation of the rates 30 years from now. This asymptote stabilizes the long term rates, an issue that has plagued most previous models. Since the forward rate plots a time path for the evolution of spot rates, it is easier to interpret in the context of monetary policy in terms of short, medium and long term rates. This is most useful for policy making. Svensson documents that this extended Nelson and Siegel model is able to fit the data better than the original one.

Fama and Bliss (1987) have constructed the yield curve via estimated forward rates at the observed maturities. They rely on CIR and others that have documented that interest rates are slow mean reverting. This fact is the driving force behind the forward rate's ability to predict future spot rates. This is true mostly for extended forecast horizons. They also documented that the expected premiums on maturities between 1 and 5 years vary through time, and can have negative values. This is in contrast to the intuitively appealing liquidity preference theory.

Diebold and Li (2006) (DL) extended the Nelson and Siegel approach to forecast the yield curve out of sample. They interpret the betas as factors of level slope and curvature , instead of them being related to maturities horizons. This can be done since, unlike factor analysis, the Nelson-Siegel approach imposes structure on the factor loadings:

$$y_t(\tau) = \beta_{1t} + \beta_{2t}(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau}) + \beta_{3t}(\frac{1 - e^{\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau})$$

Nelson and Siegel's interpretation of β_1 is that it is a long term factor, since it does not decay to 0 in the limit. Alternatively DL claim that it can also be viewed as the LEVEL factor, since if β_1 rises, it raises the yield on all maturities, uniformly. β_2 which was viewed by Nelson and Siegel as the short term factor due to it starting at 1 and the exhibiting quick decay, is interpreted by DL as the slope of the curve. DL define that slope as the difference in yield between the 10 year bond and the 3 month bill, which is a good approximation of the slope, but a better one is the difference $y_\infty - y_0$ which is exactly β_2. They also note that an increase in β_2 will affect short term rates much more so than long terms, thereby changing the slope. DL define the curvature of the curve as being twice the two year yield minus the sum of the yields of the 3 month bill and the 10 yr bond. β_3 has hardly any loading on very short or very long maturities, and mostly affects maturities in the middle part of the curve, as Nelson and Siegel pointed out. Changing the middle of the curve, without changing the extremes, will affect the curvature of the curve.

DL list the following stylized facts that were documented by their predecessors, and to which the DL model is claimed to adhere:

1. The average yield curve is increasing and concave. In this context, average means average of the βs.

2. The yield curve assumes a variety of shapes over time.

3. Yield dynamics are persistent, but spread dynamics are much less so. In this context, β_1 is persistent, but β_2 is less so.

4. The short end of the yield curve is more volatile than the long end. Here, the short end is positively related to β_1 and β_2, where as the long end is only dependent on β_1.

5. Long term rates are more persistent than short term ones. In the DL model this is exhibited by β_1 being the most persistent factor, and it being related to the long term.

DL's aim was to build a model that would exhibit a behavior that resembles the facts that were documented in the past. To accomplish that, they fit the curve based on the following procedure:

1. Fix λ at a value that maximizes the loading on the curvature factor.

2. Estimate the βs using OLS on the monthly yield data and the corresponding residuals. Since the maturities are not equally spaced (more observations for the short term), a larger weight is placed on the space where more observations exist.

DL perform a horse race to compare the predictability power of different methods of yield curve construction. They report that for long term forecasting their model performs better than the competitors, and the count the advancements their model is making from the ones that preceded it:

1. It can be used to produce yield at maturities other than those observed in the data (which is appealing to individuals who are trying to price securities).

2. It guarantees a smooth yield curve and forward curve.

3. It guarantees a positive forward rate for all horizons.

4. It guarantees that the discount function will start at 1 and will approach 0 as maturity approaches infinity.

Ho and Lee (1986) priced interest rate contingent claims using an Interest Rate Movement Model. They take the complete yield curve as given and derive the stochastic process that causes it to change over time, in an arbitrage free manner, and is consistent with the equilibrium framework. The pricing of these contingent claims is done relative to an observed term structure, and is different from CIR in that the term structure is not endogenized.

Ho and Lee noted that their approach is an improvement over their predecessors in that all information contained in the observed term structure is utilized in the pricing process, and in pricing a straight (discount) bond, the theoretical price is assured to be determined by the observed term structure.

The model is assuming that:

1. The market is free of transaction costs and taxes.

2. The market clears at discrete points in time.

3. The bond market is complete in the sense that there are bonds trading for each maturity.

4. At each point in time, there are n possible states for the world.

The arbitrage free condition requires that if a portfolio of two risky bonds is assembled in such a way that the portfolio realizes a risk free rate over the next period, then the risk free rate must be the rate on a one period discount bond. Specifically in this model:

$$\pi h(T) + (1 - \pi)h \times (T) = 1$$

for n, $i > 1$, where π is some constant independent of time T (or, binomial probability of the state of the world), and h is the perturbation function that governs the stochastic process for the price of the bond (h* is the “other” state of the world in this binomial model). The price of the bond is therefore:

$$P_i^{(n)}(T) = [\pi P_{i+1}^{n+1}(T-1) + (1-\pi)P_i^{n+1}(T-1)]P_i^n(1)$$

This states that the price of the bond equals the expected bond value at the end of the period, discounted by the prevailing discount rate.

The discount function evolves from one state to another independently of the sequence of the states of the world. It only depends on the number of each kind of states. The path-independent condition is given by:

$$h^*(T) = \frac{\phi^T}{\pi + (1-\pi)\phi^T}$$

where ϕ is the spread between h and h^*. There is a direct relationship between the spread and the rate variability between the states of the world. This implies that the longer the maturity of the bond is, the difference in terms of its price is larger as this spread is larger. Hence ϕ affects the bond's volatility. The one period rate $y_t(1)$ is

$$y_t(1) = -\ln P_i^n(1) = \ln \left[\frac{P(n)}{P(n+1)} \right] + \ln(\pi\phi^{-n} + (1-\pi)) + i \ln \phi.$$

Hence, the one period rate is the forward rate plus a bias. The bias would be zero if there will be no uncertainty [$\phi = 0$]. This stochastic process depends only on the information in the initial term structure, and the probabilities of the states of the world. In order to price a contingent claim using this model one would have to:

1. estimate the discount function at the time of the pricing using one of the available methods, such as cubic splines as in McCulloch (1975),

2. estimate π (the constant) and ϕ (the uncertainty coefficient) for an input into the model.

Heath et al. (1992) (hereafter, HJM) gave a unified framework for the no arbitrage approach and used it to price contingent claims. They defined the no-arbitrage pricing theory as having two purposes:

1. price all zero coupon , default free, bonds of varying maturities from a finite number of state variables, and

2. price all contingent claims using the price of the zero coupon bonds.

In this approach, given an initial forward rate curve and a stochastic process that governs its motion, an arbitrage free model is developed, which can price contingent claims in a manner that is not dependent of the market price of risk. This approach did not require an "inversion of the term structure" to eliminate the market price of risk from contingent claims value. The inversion of the yield curve to remove the market price of risk is necessary since these models [Vasicek (1977), Langetieg (1980), Brennan and Schwartz (1979), among others] are all deriving the price of the zero coupon bond from a set of state variables. These state variables are in themselves priced to reflect the market price of risk, hence they introduce this bias into the price of the zero coupon bond. This inversion carries several issues with it:

(i) It is computationally problematic since bond price functions are highly non-linear, (ii) spot rates and the parameters in the bond pricing function are not independent of the market price of risk. It is therefore possible that an arbitrage opportunity exists if you introduce a parameterized form of the market price of risk as a function of the state variables .

HJM generalized the approach advocated by Ho and Lee (1986), by constructing a continuous time economy, and adding additional factors to it. Lee and Ho imposed the exogenous stochastic process on zero coupon bond prices while Heath et. al. imposed it on the forward rates themselves. A given initial forward curve is assumed, then a general continuous stochastic process is defined for the evolution of the curve across time. In contrast to the Cox et al. (1985) general equilibrium model, HJM's model is taking the stochastic process for forward rates as a given and prices contingent claims from it, while CIR is fixing a particular market price for risk and endogenously deriving the stochastic process for forward rates.

White (1993), (hereafter, HW) innovated by introducing some numerical procedures to improve the arbitrage free approach to the modeling of the term structure. Within the arbitrage free approach, they identified three methods of modeling:

1. The short rate, as in Hull and White (1990),

2. The behavior of instantaneous forward rate at all future times, as in HJM , and

3. The price of discount bonds, δ_t , as in Ho and Lee (1986). Under this approach, the model specifies the behavior of all bond prices, at all times.

HW model the short rate as a markov process. They use a single factor, arbitrage free model, that can be extended to include several factors. The markov, risk-neutral process is specified for the short rate in terms of the unknown function (at time t) $\theta(t)$, which is the single factor in this model. The next step is a procedure that enables the modeler to choose the correct $\theta(t)$, which is consistent with the initial, given, term structure. This is accomplished utilizing a trinomial tree of probabilities of the state of the world in the next period. This procedure is capable of duplicating, at each node of the tree, both the expected drift in the short rate, and its instantaneous standard deviation. HW show that both the CIR, Vasicek, and the Ho and Li models can be implemented using a time varying state variables in a single factor model.

AvD presented a new approach to yield curve estimation and smoothing. They defined the criterion for the "best" curve, as one that has maximum smoothness in the forward curve.

They claimed their method can be used to "fit yield curves with one explicit function that is both consistent with all observed points on the yield curve and provides the smoothest possible forward rate curve consistent with the chosen functional form." This maximum smoothness feature is appealing to practitioners who are mostly concerned with precluding arbitrage that may be the result of kinks in the yield curve. The AvD. method is different from that of McCulloch and VF in that they actually define the maximum smoothness criteria on the forward curve, whereas McCulloch, VF, and others have fitted the splines on the observations of the yield curve, then mitigated the fluctuations in the forward curve by reducing the degree of the polynomial. As Shea noted, these efforts were not enough to eliminate unreasonable volatility between observations on the forward curve. AvD proposed dealing with the more sensitive forward curve itself, and imposing the maximum smoothness criteria directly on it.

AvD show that their method outperforms other methods (cubic spline of bond prices and of bond yields, among others) as measured by smoothness and accuracy ("goodness of fit"). This maximum smoothness approach is also a useful criterion in comparing different smoothness degrees generated by different functional forms. We will examine the AvD method in detail in the following chapter, as it is directly related to what we propose to do.

Roger J-B Wets (2002) pointed out that the only input into the AvD framework is the zero coupon curve. This, they indicate, necessitated the use of some interpolation method (which will introduce an error, per Jordan and Mansi (2003)), as well as selection of the securities to be used, which may not be unique. The method we propose in this paper resolves these issues They proposed a new method for constructing the zero curve for maturities longer than one year, after reviewing the limitations of the standard, widely used bootstrapping techniques. Their proposed method is guided by two principles: The *complete portfolio* and *smooth curves*. The complete portfolio principle is motivated by the fact that the zero curve is affected by the selection of inputs in its construction.

The advocated method starts by selecting a level of smoothness desired ($z \in C^q$), where z is the zero curve, whose qth derivative is a continuous piecewise linear function. This requires fixing a finite number of parameters. The interval $(0, T)$ is then broken into N

intervals, and the parameters are individually tailored or each one of those. The curve is then fitted piecewise. If the resulting total curve fails an accuracy criteria, an adjustment is made by the increasing the number of intervals in the estimation spectrum. essentially, they construct the following price error minimization problem: Find $z \in C^{q,pl}([0,T], N)$ so that $\| s - z(m_1 : m_L) \|$ is minimized where *pl* stands for *piecewise linear*, and s and m are an array of spot rates and maturities, respectively.

Turan and Karagozoglu (2000) have tested the effects of the AvD smoothing technique on the pricing ability of eurodollar future options. They provided evidence that the technique significantly improves pricing accuracy, at least in the context of the model proposed by Black (1990).

Hagan and West (2006)(HW) have reviewed a number of interpolation and bootstrapping methods that are in use by practitioners. Of particular importance to us is their critique of the AvD method. HW show that under certain conditions, the AvD method can produce a negative forward curve. Specifically they claim that the high degree of smoothness required by the method is achieved at the cost of losing any stiffness of the forward curve. Namely, a relatively small changes in the yield curve may cause the forward curve to react with large movements. They also point out that the AvD method does not preclude a negative forward rate. They show that if you have a flattened yield curve that has a bump on it (say, around the 5 years maturity node), you have a forward rate that is negative at some maturities. We make a note of their remark, but in our opinion if we actually observe a spot rate that has a blimp in it, the forward curve *should* reflect this pricing anomaly by exhibiting its own out-of-normal values. Another important contribution is their observation that bootstrapping and interpolation should not be considered separately. Indeed, the information in both processes is completing each other. Our approach adopts this point of view and does implement stripping and smoothing procedures that are intertwined.

CHAPTER 3

ITERATIVE PIECEWISE QUARTIC POLYNOMIAL INTERPOLATION METHOD

AvD have indicated that past research was focused on structuring the yield curve in a way that aimed to produce a meaningful forward curve. These past methods have modeled the discount function, δ_t, hoping to produce a forward rate that is "well behaved" in the sense that it was positive, robust and smooth. These attempts were pointed out by Shea (1985), among others, to be flawed, as the forward curve still exhibited unreasonable behavior. To construct a better forward curve, AvD took a different approach - by directly modeling the forward curve, rather than indirectly through the discount or the yield (spot) curve. The advantage of directly modeling the forward curve is clear if one examines the relationships between the forward, yield and discount rates, which we have specified in the introduction. The spot, discount and forward curves have varying degrees of sensitivities, with the forward curve being the most sensitive to changes in both the discount and spot rates. Modeling the forward rate by constructing the discount or spot rates is similar to fixing a pair of eyeglasses while wearing gloves. It is cumbersome to construct a reasonable forward curve that way. Instead, if one models the forward curve directly and embeds all of the desired properties into it, then by construction the spot and discount rates will have the characteristics that the modeler requires.

3.1 Overview of the AvD method

AvD's method is designed to construct the forward curve, given a set of zero coupon bonds. The model I present is based on their work, and extends it to include coupon bonds by integrating the coupon stripping process into model. Since my model is an extension of theirs, it is useful to review their work in some detail here. AvD's approach was to define a maximum smoothness criteria for the forward rate, and search for the curve that satisfied it while constrained to fit the data. Their definition of this maximum smoothness is the minimum of the second derivative (curvature) of the forward curve. A forward curve of this kind will ensure that the yield curve function will be forced to change very slowly. This will exclude abrupt changes to the yield function, thereby excluding potential arbitrage.

Let $\delta(t_i),\ i = 1, \ldots, m$ be the price of a zero coupon par 1 bond that matures at time t_i Let $f(t)$ and $y(t)$ denote the instantaneous forward and spot curves, respectively. Recall that

$$\int_0^{t_i} f(s)ds = -\log \delta(t_i),\ i = 1, \ldots, m.$$

AvD solve for the smoothest possible forward curve among the class of piecewise polynomials that satisfies m pricing constraints, where their notion of smoothness is the total curvature of the forward curve. They also impose initial and terminal constraints on the forward curve so that the optimization problem becomes

$$\min_f \int_0^{t_m} \left(f^{''}(s)\right)^2 ds \tag{3.1}$$

subject to the constraints:

1. $\exp\left(-\int_0^{t_i} f(s)ds\right) = \delta(t_i)$ for $i = 1, 2, ..., m$, which will ensure that the curve accurately prices each bond.

2. $f(0) = y(0)$, which will ensure that the forward rate and the spot rate are equal at time 0.

3. $\lim_{t\to\infty} f^{'}(t) = 0$ which will ensure that the forward rate is flat at the asymptote and will converge to $y(t)$

The optimization problem is complicated by the fact that the first constraint is a function of $f^{''}$. To see this, rewrite the first constraint as

$$\int_0^{t_i} f(s)ds = -\log \delta(t_i) \tag{3.2}$$

and integrate the left hand side by parts twice to get

$$\int_0^{t_i} f(s)ds = t_i f(t_i) - \frac{1}{2} t_i^2 f'(t_i) + \frac{1}{2} \int_0^{t_i} s^2 f''(s)ds$$

Define

$$g(t) = f''(t)\, 0 < t \leq t$$

and

$$Q_i = t_i f(t_i) - \frac{1}{2} t_i^2 f'(t), \text{ for } i = 1, 2, \ldots, m$$

so the optimization problem can be rewritten as

$$\min_f Z[g] = \int_0^{t_m} g^2(s) ds$$

subject to

$$\frac{1}{2} \int_0^{t_i} s^2 u(t_i - s) g(s) ds = -\log \delta t_i - Q_i, \; i = 1, \ldots, m.$$

This is a variational calculus problem. Thus, if the function g^* is the solution to the minimization problem then we have

$$\frac{d}{d\varepsilon} Z[g^* + \varepsilon h]\Big|_{\varepsilon=0} = 0$$

for any continuous function h defined over $[0, t_m]$, such that $h(s) = g(s) - g^*(s)$, where g^* is optimal.

AvD find the solution to be

$$\frac{d}{d\varepsilon} Z[g + \varepsilon h]\Big|_{\varepsilon=0} = 2 \int_0^T \left[g(s) + \frac{1}{4} s^2 \sum_{i=1}^{m} \lambda_i u(t_i - s) \right] h(s) ds \tag{3.3}$$

where $u(t_i)$ is an indicator function defined as

$$u(t_i - t_{i-1}) = \begin{cases} 1 & \text{if } t_{i-1} \leq t \leq t_i \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

AvD claim that the spline functions are of the form:

$$f_i(t) = c_i t^4 + b_i t + a_i \text{ for } t_{i-1} < t \leq t_i \text{ and } i = 1, 2, ..., m. \tag{3.5}$$

Lim and Xiao (2002) (LX), however, have pointed out an error in (3.3) above. They noted that Q_i, is, by it's definition a function of $g(s)$

$$Q_i = t_i \int_0^{t_i} \left[u(t_i - s) \int_0^s g(v) dv \right] ds - \frac{1}{2} t_i^2 \int_0^{t_i} u(t_i - s) g(s) ds + r_0 t_i$$

yet AvD, have treated Q_i as a constant when deriving (3.3).

LX reformulate the minimization problem, with Q_i expressed as a function of $g(\cdot)$:

$$\min_{g(\cdot),\lambda_i} Z[g,\lambda] = \int_0^{t_m} g^2(s)ds + \int_0^{t_m} [\sum_{I=1}^{m} \lambda_i u(t_i - s)[t_i \int_0^s g(v)dv$$
$$- \frac{1}{2}t_i^2 u(t_i - s)g(s) + \frac{1}{2}s^2 g(s)]]ds$$
$$+ \sum_{i=1}^{m} \lambda_i (r_0 t_i + \ln \delta(t_i))$$

where λ_i is the Lagrange multiplier of the ith constraint.

They show that $g^*(t)$ (the optimal g) is a continuous function in second order polynomial form in each interval, of the form

$$g^*(t) = -\frac{1}{4}\sum_{j=i+1}^{m} \lambda_j (t_j - t)^2, \text{ when } t_i < t \le t_{i+1}, \text{and } i = 0, \ldots, m-1.$$

Integrating g^* twice yields

$$f_i(t) = a_i t^4 + b_i t^3 + c_i t^2 + d_i t + e_i$$

for $i = 1, ..., m$ and $t_{i-1} < t \le t_i$. Thus, the maximum smoothness forward curve is an unconstrained fourth order polynomial function in each segment of the curve (between observations).

The next step is to determine the coefficients a, b, c, d and e. I follow LX and transform the problem into a quadratic form. Integrate $f''(t)$ to get:

$$\int_{t_{i-1}}^{t_i} f''(t)dt = \int_{t_{i-1}}^{t_i} (12a_i t^2 + 6b_i t + 2c_i)^2 dt = \frac{144}{5}\Delta_i^5 a_i^2$$
$$+ 36\Delta_i^4 a_i b_i + 12\Delta^3 b_i^2 + 16\Delta_i^3 a_i c_i + 12\Delta_i^2 b_i c_i + 4\Delta_i^1 c_i^2$$
$$= X_i^T h_i X_i$$

where

$$X_i = \begin{pmatrix} a_i \\ b_i \\ c_i \\ d_i \\ e_i \end{pmatrix}, h = \begin{pmatrix} \frac{144}{5}\triangle_i^5 & 18\triangle_i^4 & 8\triangle_i^3 & 0 & 0 \\ 18\triangle_i^4 & 12\triangle_i^3 & 6\triangle_i^2 & 0 & 0 \\ 8\triangle_i^3 & 6\triangle_i^2 & 4\triangle_i^1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$\Delta_i^l = t_i^l - t_{i-1}^l, \text{for } l = 1, \ldots, 5.$$

The objective function is now of the form:

$$\min_X X'hX$$

where

$$X = \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_{m+1} \end{bmatrix}, \quad h = \begin{bmatrix} h_1 & & & & 0 \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & h_{m+1} \end{bmatrix}.$$

The important observation to make is that the objective function is quadratic so that the gradient is linear in the coefficients. The constraint functions are also linear functions of the coefficients, and that is shown next. Note that an $m+1$ polynomial was added here so as to impose the terminal condition, which will be discussed below.

The objective function is minimized subject to the following constraints:

1. Fitting the observed prices with minimal error

2. Continuity of the spline function at the nodes: This imposes the condition that at the nodes, where the splines are connected, the function will be smooth with no abrupt changes in the slope. Alternatively this can be noted as $f_i(t_i) = f_{i+1}(t_i)$.

3. Continuity of the first derivatives (slope) of the spline function at the nodes ($f_i^{'}(t_i) = f_{i+1}^{'}(t_i)$)

4. Continuity of the second derivative (curvature) of the spline function at the nodes ($f_i^{''}(t_i) = f_{i+1}^{''}(t_i)$)

5. Boundary conditions on the initial point of the forward curve and the terminal slope of the forward curve. The initial point is determined be interpolating backward, using the one week, and 30 day rates to arrive at the instantaneous maturity rate . The terminal condition is designed so that we will have a flat forward curve at the last observed maturity, and beyond, so that $f(t) = \textit{constant}$ for all $t > t_m$. This is done by adding a spline in the segment beyond the last observed maturity by essentially adding a maturity at infinity $t_{m+1} = \infty$.

I now show how to formally impose these constraints into the minimization problem:

1. Fitting the observed prices: The zero-coupon bond price given by (1.3) may be written as

$$
\begin{aligned}
-\ln\delta(t_j) &= \int_0^{t_j} f(t)\,dt \\
&= \sum_{i=1}^{j} \int_{t_{i-1}}^{t_i} f_i(t)\,dt \\
&= \sum_{i=1}^{j} \int_{t_{i-1}}^{t_i} \big(a_i t^4 + b_i t^3 + c_i t^2 + d_i t + e_i\big)\,dt \\
&= \sum_{i=1}^{j} \Big(\frac{1}{5}a_i t_i^5 + \frac{1}{4}b_i t_i^4 + \frac{1}{3}c_i t_i^3 \\
&\qquad + \frac{1}{2}d_i t_i^2 + e_i t_i\Big)
\end{aligned}
\tag{3.6}
$$

so that the log of the zero-coupon bond price is linear in the coefficients X.

Recalling the previous notation $\Delta_j^n = t_j^n - t_{j-1}^n$, the difference of the log prices of two consecutive zero coupon bonds can be written as

$$
\begin{aligned}
-\ln\left(\frac{\delta(t_j)}{\delta(t_{j-1})}\right) &= \frac{1}{5}\Delta_j^5 a_j + \frac{1}{4}\Delta_j^4 b_j + \frac{1}{3}\Delta_j^3 c_j \\
&\quad + \frac{1}{2}\Delta_j^2 d_j + \Delta_j e_j.
\end{aligned}
\tag{3.7}
$$

The full set of pricing constraints for all m bonds may then be written in matrix form as

$$\vec{A}_1 \vec{X} = \vec{B}_1 \tag{3.8}$$

where

$$\vec{A}_1 = \begin{pmatrix} Dt_1 & 0_{1\times5} & \cdots & 0_{1\times5} & 0_{1\times5} \\ 0_{1\times5} & Dt_2 & \cdots & 0_{1\times5} & 0_{1\times5} \\ \vdots & \vdots & \ddots & \vdots & \\ 0_{1\times5} & 0_{1\times5} & \cdots & Dt_m & 0_{1\times5} \end{pmatrix}_{m\times5(m+1)}, \tag{3.9}$$

$$Dt_j = \big(\tfrac{1}{5}\Delta_j^5,\ \ \tfrac{1}{4}\Delta_j^4,\ \ \tfrac{1}{3}\Delta_j^3,\ \ \tfrac{1}{2}\Delta_j^2,\ \ \Delta_j\big)_{1\times5}, \tag{3.10}$$

and

$$\vec{B}_1 = \begin{pmatrix} \ln\big(\delta(t_1)/\delta(t_0)\big) \\ \ln\big(\delta(t_2)/\delta(t_1)\big) \\ \vdots \\ \ln\big(\delta(t_m)/\delta(t_{m-1})\big) \end{pmatrix}_{m\times1}. \tag{3.11}$$

2. The second set of constraints are now imposed to ensure that the forward curve remains smooth as it transitions through node points in the piecewise polynomial approximation. To ensure continuity at the node points it is required that the forward rate at node t_i has the same value whether computed using the left-side polynomial or the right-side polynomial:

$$f_{i+1}(t_i) = f_i(t_i),\ i = 1, \ldots, m, \tag{3.12}$$

or

$$(a_{i+1} - a_i)t_i^4 + (b_{i+1} - b_i)t_i^3 + (c_{i+1} - c_i)t_i^2 + (d_{i+1} - d_i)t_i + (e_{i+1} - e_i) = 0,\ i = 1, \ldots, m. \tag{3.13}$$

Define $T4_i = (t_i^4, t_i^3, t_i^2, t_i, 1)_{1\times 5}$ and write all m of these constraints in matrix form as

$$\vec{A}_2 x = \vec{B}_2 \tag{3.14}$$

where

$$\vec{A}_2 = \begin{pmatrix} -T4_1 & T4_1 & 0_{1\times5} & 0_{1\times5} & \cdots & 0_{1\times5} & 0_{1\times5} \\ 0_{1\times5} & -T4_2 & T4_2 & 0_{1\times5} & \cdots & 0_{1\times5} & 0_{1\times5} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0_{1\times5} & 0_{1\times5} & 0_{1\times5} & 0_{1\times5} & \cdots & -T4_m & T4_m \end{pmatrix} \tag{3.15}$$

is an $m \times 5(m+1)$ matrix and $\vec{B}_2$ is a $m \times 1$ vector of zeros.

3. To impose differentiability at the nodes it is required that

$$f'_{i+1}(t_i) = f'_i(t_i),\ i = 1, \ldots, m \tag{3.16}$$

or

$$4(a_{i+1} - a_i)t_i^3 + 3(b_{i+1} - b_i)t_i^2 + 2(c_{i+1} - c_i)t_i + (d_{i+1} - d_i) = 0,\ i = 1, \ldots, m. \tag{3.17}$$

Define $T3_i = (4t_i^3, 3t_i^2, 2t_i, 1, 0)_{1\times 5}$ and write all m of these constraints in matrix form as

$$\vec{A}_3 \vec{X} = \vec{B}_3 0_{m\times 1} \tag{3.18}$$

where

$$\vec{A}_3 = \begin{pmatrix} -T3_1 & T3_1 & 0_{1\times5} & 0_{1\times5} & \cdots & 0_{1\times5} & 0_{1\times5} \\ 0_{1\times5} & -T3_2 & T3_2 & 0_{1\times5} & \cdots & 0_{1\times5} & 0_{1\times5} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{1\times5} & 0_{1\times5} & 0_{1\times5} & 0_{1\times5} & \cdots & -T3_m & T3_m \end{pmatrix} \tag{3.19}$$

is an $m \times 5(m+1)$ matrix and $\vec{B}_3$ is a $m \times 1$ vector of zeros.

4. To ensure that the first derivatives of the forward curve are smooth at the nodes the following is imposed

$$f''_{i+1}(t_i) = f''_i(t_i), \; i = 1, \ldots, m, \tag{3.20}$$

or

$$12(a_{i+1} - a_i)t_i^2 + 6(b_{i+1} - b_i)t_i + 2(c_{i+1} - c_i) = 0, \quad i = 1, \ldots, m. \tag{3.21}$$

Define $t2_i = (12t_i^2, 6t_i, 2, 1, 0, 0)_{1\times5}$ and write all m of these constraints in matrix form as

$$\vec{A}_4\vec{X} = \vec{B}_4 \tag{3.22}$$

where

$$\vec{A}_4 = \begin{pmatrix} -T2_1 & T2_1 & 0_{1\times5} & 0_{1\times5} & \cdots & 0_{1\times5} & 0_{1\times5} \\ 0_{1\times5} & -T2_2 & T2_2 & 0_{1\times5} & \cdots & 0_{1\times5} & 0_{1\times5} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{1\times5} & 0_{1\times5} & 0_{1\times5} & 0_{1\times5} & \cdots & -T2_m & T2_m \end{pmatrix} \tag{3.23}$$

is an $m \times 5(m+1)$ matrix and $\vec{B}_4$ is a $m \times 1$ vector of zeros.

5. To ensure the boundary condition $f(0) = y_0$, we simply impose $e_1 = y_0$. The terminal boundary condition $f'(t_m) = 0$ is more difficult to impose. LX use the condition $d_1 = 0$ which is clearly incorrect. The terminal condition is imposed by adding an additional $(m+1)^{st}$ segment to the piecewise polynomial with the coefficient restrictions $a_{m+1} = b_{m+1} = c_{m+1} = d_{m+1} = 0$ so that $f(t) = e_{m+1}$ for all $t > t_m$. The terminal height of the forward function is left unconstrained and the continuity and smoothness constraints described above will ensure a smooth transition to the zero slope of the forward curve at node t_m.

These five boundary conditions may be written in matrix notation as

$$\vec{A}_5\vec{X} = \vec{B}_5 \tag{3.24}$$

where

$$\vec{A}_5 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & 0 \end{pmatrix}_{5\times 5(m+1)} \tag{3.25}$$

and

$$\vec{B}_5 = \begin{pmatrix} y_0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3.26}$$

Stacking all of these linear constraints gives

$$\vec{A}\vec{X} = \vec{B} \tag{3.27}$$

where

$$\vec{A} = \begin{pmatrix} \vec{A}_1 \\ \vec{A}_2 \\ \vec{A}_3 \\ \vec{A}_4 \\ \vec{A}_5 \end{pmatrix}_{(4m+5)\times 5(m+1)} \quad \text{and} \quad \begin{pmatrix} \vec{B}_1 \\ \vec{B}_2 \\ \vec{B}_3 \\ \vec{B}_4 \\ \vec{B}_5 \end{pmatrix}_{(4m+5)\times 1}. \tag{3.28}$$

The procedure for interpolating the zero coupon portion of the yield curve is continuing on the footsteps of LX except that the terminal condition on the forward curve is handled differently, as detailed above. An additional segment is added to the piecewise spline function with some additional coefficient restrictions on the terminal spline. Since the matrix sizes differ to reflect these changes, the details for this step are provided next. First, it is shown how to construct the objective function as a quadratic expression. Next,the constraints are constructed as linear equations and lastly, the PQPI system is solved.

Recall that there are m securities with maturities $t_1, \ldots, t_m$ where m is the number of securities. The settlement date in this list is not considered, so t_1 is the maturity of the first *real* security—the one-week LIBOR in this example. Define $t_0 = 0$.

3.2 Solving the IPQPI System

The constrained optimization problem may now be written in matrix notation as

$$\min_{\vec{X},\vec{\lambda}} Z(\vec{X},\vec{\lambda}) = \vec{X}'\vec{H}\vec{X} + \vec{\lambda}'(\vec{A}\vec{X} - \vec{B}) \tag{3.29}$$

where $\vec{\lambda}$ is the $4m+5$ vector of Lagrange multipliers corresponding to the constraints.

The first-order conditions are

$$\frac{\partial}{\partial \vec{X}} Z(\vec{X}, \vec{\lambda}) = 2\vec{H}\vec{X} + \vec{A}'\vec{\lambda} = 0 \tag{3.30}$$

and

$$\frac{\partial}{\partial \vec{\lambda}} Z(\vec{X}, \vec{\lambda}) = \vec{A}\vec{H}\vec{X} - \vec{B} = 0, \tag{3.31}$$

or

$$\begin{pmatrix} 2\vec{H} & \vec{A}' \\ \vec{A} & 0 \end{pmatrix} \begin{pmatrix} \vec{X} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} 0 \\ \vec{B} \end{pmatrix} \tag{3.32}$$

from which an explicit solution can be found

$$\begin{pmatrix} \vec{X}^* \\ \vec{\lambda}^* \end{pmatrix} = \begin{pmatrix} 2\vec{H} & \vec{A}' \\ \vec{A} & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \vec{B} \end{pmatrix}. \tag{3.33}$$

3.3 Iterative Piecewise Quartic Polynomial Interpolation Algorithm

The LX/AvD method that was presented above, requires zero coupon yields at m maturities as inputs. These are not generally available since coupon bonds have to be stripped to arrive at their zero coupon yields. To that end, I extend the AvD method by introducing a simultaneous stripping and bootstrapping design. I integrate the stripping process into the AvD smoothing routine to create a unified tool one can use in practice. The combined method is dubbed *Iterative Piecewise Quartic Polynomial Interpolation* (IPQPI), and its algorithm is illustrated next. To aid in the illustration of the method, the reader may refer to Figure 3.1 below.

Initially, a smooth zero curve is fitted through the observations that are already in the form of zero coupons (the short term, up to one year, maturities). I call this the *known portion* of the zero curve so all that is required at this stage is implementing the standard LX/AvD approach, to fit the curve. In the example in the figure, this portion of the curve is the first year - from 0 to $y(1)$, where $y1$ is the last known node - in our case, the one year bill. The coupon-bond observations are then added sequentially, but the coupons associated with the bonds' yields must be stripped, so as to arrive at the zero-coupon yield of each bond. The addition of the coupon bonds is implemented as follows: First, I use linear bootstrap to get an initial guess of the zero rate for the very first coupon bond maturity (the 2 year bond in this example)which is $y(2)$ in figure 3.1). Next, I define the aggregate value of all coupons that mature along the *known portion* of the curve as $K = c1$. K is then subtracted from the observed price of the 2 year bond ($P(t_i)$, which is given in the data) to arrive at the aggregate value of all other coupons and principal that lay beyond the *known portion* of the curve. I refer to that value as D and in figure 3.1. Mathematically this is expressed as $D = c2 + c3 + c4$, where $c2$ through $c4$ are coupon payments. At this point, a gradient search is performed for the right spot rate for the principal on maturity date. At the conclusion of this gradient updating, the value of D satisfies $P(t_i) = K + D$, and the forward curve has been updated through the maturity of the first coupon bond. Note as a result of the updating of $y(2)$, all coupons up to that point have been repriced, as a result of the polynomial having to change in order for $y(2)$ to be priced just right. The coupons labeled $c1$ as well as $c2$, $c3$ and $c4$ were all repriced up from their original location.

This process is continued by stepping through each coupon bond with increasing maturity. The smoothed zero curve and forward curves must be re-computed at each iteration of the process, and the coupons will continue to be repriced as each gradient search is conducted as a result of the polynomials changing to accommodate the most recent bond. The procedure is quite computationally intensive, however, the benefit is that a smooth forward curve which produces zero curves and discount functions which price all bonds with high precision is produced. After the final forward curve is computed, pricing errors and other statistics are calculated for the purpose of evaluating the estimated curve.

I now proceed to describe the algorithm in a more detailed manner. The inputs to the algorithm are: (1) the settlement date; (2) the over night repo rate or the federal funds rate (the anchor for the short end of the yield curve); (3) the yield, coupon and maturity date of the bills and bonds along the yield curve at the settlement date. The outputs of the algorithm are the coefficient vector, X, for the piece-wise fourth-order polynomial spline function describing the forward curve. From these coefficients we are able to construct the forward curve, the zero curve, and the discount function as described above.

Let y be the yield to maturity, and DM be the number of days to maturity. I denote the *smoothed* zero curve as ZC.

1. Compute the maturity in years, $t(i)$, of each security, $i = 1, \ldots, m$.

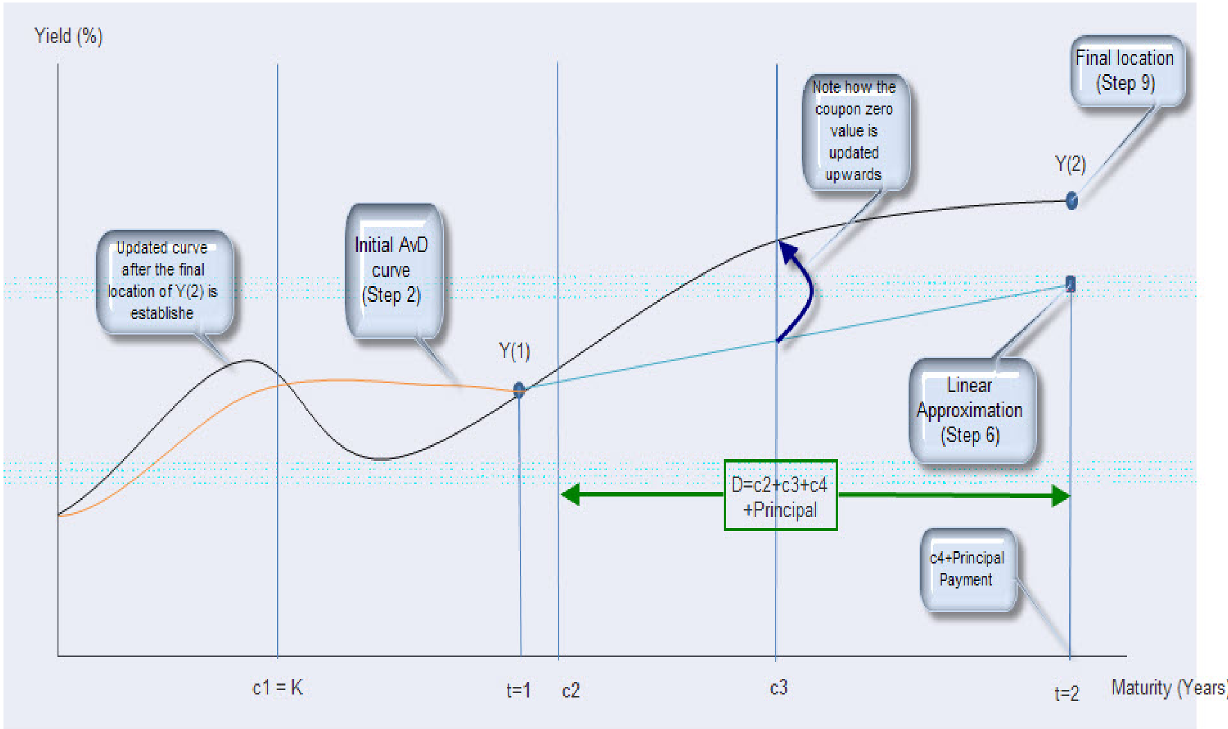

Figure 3.1: An illustration of the IPQPI method.

2. Use the LX/AvD method to fit a smooth forward curve through the known nodes for the zero coupon securities (bills). These securities are by definition already priced at the spot rate. This produces a coefficient vector X for the zero curve up through the longest maturity of the bills ($t(k) = 1$ in Figure 3.1).

3. We now turn our attention to the coupon bonds. First, we must determine the timing of the remaining cash flows form each coupon bond. We construct a vector that has three elements:

   (a) nc, the number of coupons remaining. In the case illustrated in Figure 3.1, these are coupons $c2$, $c3$ and $c4$.

   (b) w, the portion of the first coupon payment to which we are entitled (this is the number of days between the settlement date and the payment date of the next coupon.

(c) cd, number of days until each of the remaining coupon payments.

4. Recalling that coupons pay twice a year we set up the cash flow vector

$$Z_i = \begin{pmatrix} w \cdot coupon/2 \\ coupon/2 \\ \vdots \\ coupon/2 \\ coupon/2 + face\ value \end{pmatrix}$$

5. Compute the dirty price for the i^{th} bond as:

$$P = \sum_{j=1}^{nc} \frac{Z(j)}{(1+y)^{j-1+w}}$$

recall that nc is the number of coupons still outstanding, and y is the yield to maturity of the bond. A dirty price is a price that does not takes into account the portion of the first coupon that is payable to the previous bond holder. A so called 'clean' price does account for this proration.

6. Now we use the linear bootstrap method to get an initial guess for the spot rate of the i^{th} bond (the 2-year bond in this example). I show that as the square $y(2)$ in Figure 3.1

   (a) Let L be the number of coupon payments for the i^{th} bond that are made along the already known portion of the zero curve constructed in the previous step (that is $c1$ in Figure 3.1),and let $y(1)$ denote the maturity of the last known point along this zero curve.

   (b) Let $t(i)$ denote the maturity of the i^{th} bond—the bond we are currently pricing (the 2-year in our example) —and approximate the spot curve between maturities $t(k)$ and $t(i)$ using a linear function by *guessing* the endpoint spot rate (that is the square at $t = (2)$ in the case of the 2-year bond).

   (c) Define D to be that part of the i^{th} bond's price which is made up of coupons that mature after the last *known* date along the known zero curve ($y(1)$ in our example):

$$D = P - \sum_{j=1}^{L} Z(j) \exp[-t(j) \cdot y(j)]$$

(d) Use this linear approximated piece of the zero curve to price the quantity D above and adjust $y(t)$ until D is priced correctly. This provides an *initial* guess for the spot rate at $t(i)$.

7. Now add $y(n)$ to the "known" zero coupon rates. Using this new node we re-run the AvD to produce a new smooth forward curve up through maturity $t(n)$.

8. Price the i^{th} bond using the new forward curve:

$$\hat{P} = \sum_{j=1}^{nc} Z(j) \exp[-t(j)\, y(j)]$$

and compute the pricing error $P_{error} = |\hat{P} - P|$.

9. Iteratively adjust the terminal spot rate ($y(2)$ in my example) using the binary search algorithm until P_{error} is reduced to near zero. To determine which direction to adjust the curve, define $\tilde{y}(t_i)$ to be the estimated spot rate of the bond with maturity t_i. $\tilde{y}(t_i)$ is perturbed up and down by $\tilde{y}(t_i)/100$ along the gradient search $\tilde{y}(t_i)$ to get $\tilde{y}_u$ and $\tilde{y}_d$. From these perturbations we compute $\tilde{f}_u$ and $\tilde{f}_d$ and then $\tilde{P}_u$ and $\tilde{P}_d$. The derivative $d\tilde{t}(t_i)/d\tilde{y}(t_i)$ is then approximated using the centered difference method

$$\frac{d\tilde{P}(t_i)}{d\tilde{y}(t_i)} \approx \frac{\tilde{P}_u - \tilde{P}_d}{\tilde{y}_u - \tilde{y}_d} = \frac{\tilde{P}_u - \tilde{P}_d}{\tilde{y}(t_i)/50}. \tag{3.34}$$

We now update our estimate of the spot rate at t_i using

$$\tilde{\tilde{y}}(t_i) = \tilde{y}(t_i) + \frac{d\tilde{y}(t_i)}{d\tilde{P}(t_i)} \left(P(t_i) - \tilde{P}(t_i) \right). \tag{3.35}$$

This gradient updating of $\tilde{y}(t_i)$ is repeated until the estimated spot rate changes by less than 10^{-9}.

At the end of this process we adjusted the terminal point in the figure to the *circle* $y(2)$. Note that when we move the terminal point - the maturity of the bond, from the square to the circle at $y(2)$, the entire curve is adjusting and repricing the coupons all the way from the origin. It is clear that all the coupons $c1, c2, c3$ and $c4$ are being adjusted upward. Note also that the 6 month bill is still being priced correctly since that is a constraint we placed on the system.

10. Now we move on to the next coupon bond and return to step (3).

The algorithm is tedious but straight forward. The essential feature is that the forward curve is constructed sequentially one bond at a time. As each bond is added the entire curve must be recomputed so that all constraints are satisfied. By taking small steps, we reduce a multidimensional optimization problem into a sequence of univariate optimization problems.

3.4 Data and Results

I use yield curves from two different dates to test my method. The September 9th, 2009 data comes from that date's edition of the Wall Street Journal (WSJ), and represents a "normal", (upward sloping) yield curve. The August 31, 1981 data comes from the Treasury Bulletin (TB) for that month, and is chosen since it is an "inverted" yield curve (one in which the yield on the long term issues is lower than the yield on the short issues). Note that for 1981 data, the yield on the 180 day security is 17.16 percent, which is appreciably higher than the 14.74 percent on the 30 year bond.

Both the WSJ and the TB are providing the so-called "clean (bid) price", which is the price the purchaser is paying, net of any accrued interest that belongs to the seller. I translate these to "dirty prices" - the price which does not adjust for this interest, but rather reflects the gross price to be paid by the security's purchaser. I do not use this yield data, and it is presented here for the convenience of the reader only. The algorithm employed is computing the continuous-time, dirty price yield-to-maturity which is slightly different from the one presented here, which are clean price bond equivalent yields. This is also the reason why the curves shown below do not go through the center of the circled observations for the yields which come from the data.

The September 9th 2009 data shown in Table 3.1 is for the on-the-run treasuries for that date. It has 5 bills (one week, 30 days, 90 days, 180 days, and one year) and 6 coupon bonds (2 year, 3 year, 5 year, 7 year, 10 year, and 30 year), for a total of 11 securities, plus the zero maturity rate which is a product of a backward interpolation procedure that is aimed at anchoring the curve right at 0 maturity (since the first observation we have is the 7 day maturity), and is based on the slope between the one week, and the 30 day securities. The coupon column represents the annual coupon rate that is paid to the bond holder semiannually. Note that since the bills are in the form of a discounted note, they do not carry any coupon payments. The actual price column has the dirty price in dollar terms. The yield reported by the publications is the clean price, bid yield-to-maturity. From these

Table 3.1: Actual and estimated bond prices for the yield curve on date 9/9/2009.

Maturity	Coupon	Actual Price	Yield	Est. Price	Penny Error
9/9/2009	0	100.0000	0.0808	100.0000	0.0
9/17/2009	0	99.998160	0.08415	99.998160	0.0
10/10/2009	0	99.994580	0.09	99.994580	0.0
12/30/2009	0	99.969780	0.12982	99.969780	0.0
3/4/2010	0	99.899780	0.20806	99.899780	0.0
8/26/2010	0	99.626580	0.38978	99.626580	0.0
8/31/2011	1.0	100.18111	0.92	100.181347	-0.0237
8/15/2012	1.75	100.99389	1.4443	100.994116	-0.0226
8/31/2014	2.375	100.05905	2.3751	100.062716	-0.3666
8/31/2016	3.0	99.605836	3.0753	99.605594	0.0242
8/15/2019	3.625	101.43376	3.4828	101.427963	0.5797
8/15/2039	4.5	102.96196	4.3408	102.961960	0.0

11 observation the algorithm constructs the discount, spot and forward curves. The last two columns show my method's price estimates, and the pricing errors in pennies.

As shown in Table 3.1, the IPQPI method's computed forward curve prices the observed bonds quite accurately, with all errors of less than one penny. There are zero pricing errors for the zero coupon bonds. While one might expect this since the spot rates for these bonds are directly observed, recall that the forward curve is also influenced by the coupon payments from longer maturity coupon bonds intermingled amongst these bills. Thus, it is a nontrivial result to produce zero pricing errors for the bills when bonds are included in the set of securities. The largest error for the 2009 data set is .5797 cents on the ten year bond which amounts to a 0.0057% error. Note that for the 2009 data (Figure 3.2) the forward curve stays above the spot curve throughout the spectrum of maturities. This is since the spot rate is increasing as maturity rises. It is also apparent that the forward curve is much more volatile than the spot curve. This is the reason why modeling the forward curve rather than the spot or discount curves yields better results - the changes in spot or discount rates are simply hard to notice compared with the forward curve. If a smooth forward curve can be produced, then one is assured of producing smooth spot and discount curves. The forward curve produced by the IPQPI method never goes negative and is exhibiting stability at long maturities, as $y(T \to \infty) \to f(t_m)$

The 2009 curve is a classical upward sloping term structure that simple models such as

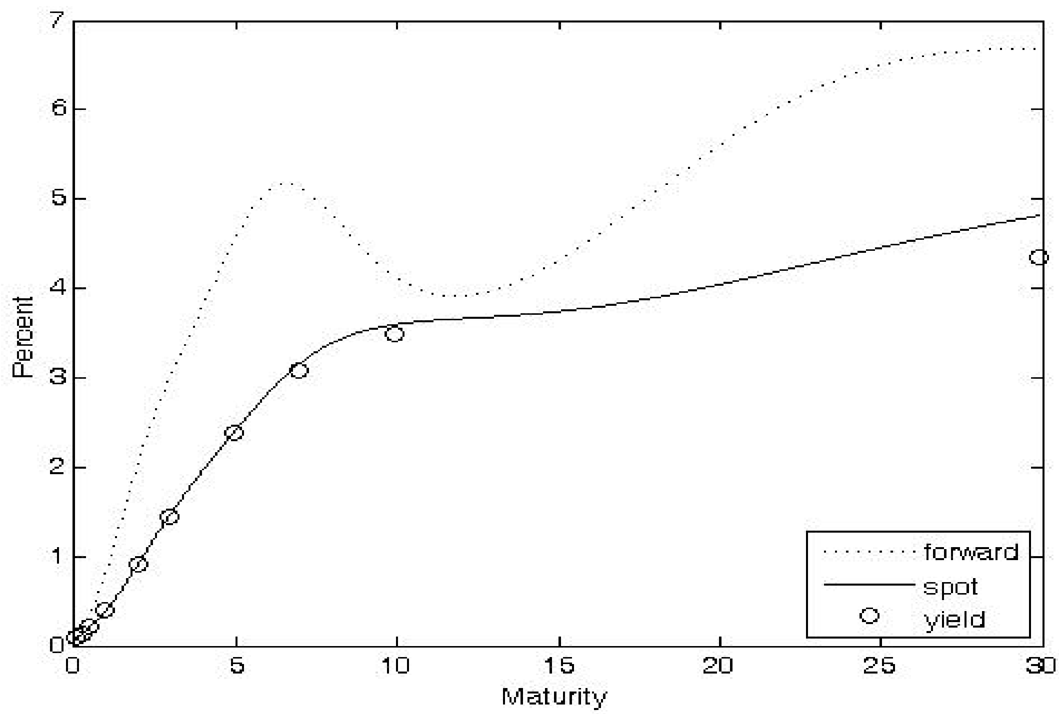

Figure 3.2: Forward and Spot rates for 9/9/2009.

Table 3.2: Actual and estimated bond prices for the yield curve on date 8/31/1981.

Maturity	Coupon	Actual Price	Yield	Est. Price	Abs. Error
8/31/1981	0	100.0000	14.37	100.0000	0.0
9/3/1981	0	99.8775	14.53	99.8775	0.0
9/10/1981	0	99.5867	14.91	99.5867	0.0
10/1/1981	0	98.7247	14.91	98.7247	0.0
11/27/1981	0	96.1989	16.28	96.1989	0.0
2/25/1982	0	92.2422	17.16	92.2422	0.0
8/12/1982	0	85.6506	16.90	85.6506	0.0
7/31/1983	15.875	99.806096	16.77	99.804069	0.2027
11/15/1984	16.0	103.75815	16.34	103.755268	0.2882
11/15/1986	13.875	97.009511	15.88	97.008831	0.0680
7/15/1988	14.0	94.663043	15.70	94.66145	0.1593
8/15/1991	14.875	97.896739	15.40	97.892565	0.4174
8/15/2001	13.375	90.519022	15.10	90.427431	9.1591
5/15/2011	13.875	98.009511	14.74	98.009511	0.0

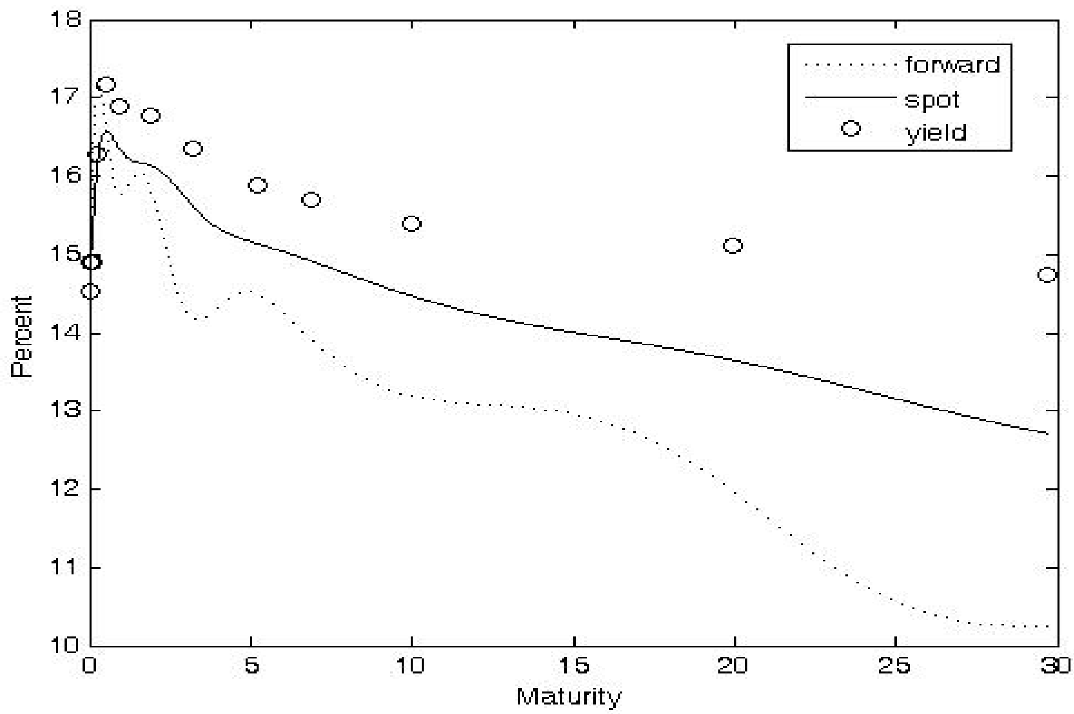

Figure 3.3: Forward and Spot rates for 8/31/1981.

Svensson's and Nelson-Siegel's can accommodate easily. A more challenging curve to fit is an inverted yield curve and, in particular, one that has considerable variation in the short term rates. A good example of such a term structure is the one from August 31, 1981.

Turning our attention to this more complex 1981 curve (Table 3.2) we note that most pricing errors are less than one penny, with the exception of the 20 year bond which carries a pricing error of 9.1 cents (0.101%). The 8/31/1981 term structure produces a maximum spot rate at six months maturity so the forward curve intersects the spot curve at that point. The spot and forward curves begin at the same point at the zero maturity date and both curves are very stable at long maturities. This latter property is non-trivial and often not observed in polynomial interpolation methods.

In addition, note that the IPQPI method presented is able to produce a forward curve that is flexible enough to accommodate the non trivial dynamics exhibited at the very short end of the 1981 curve. Figure 3.4 is a closeup look at the first year of maturities for 1981

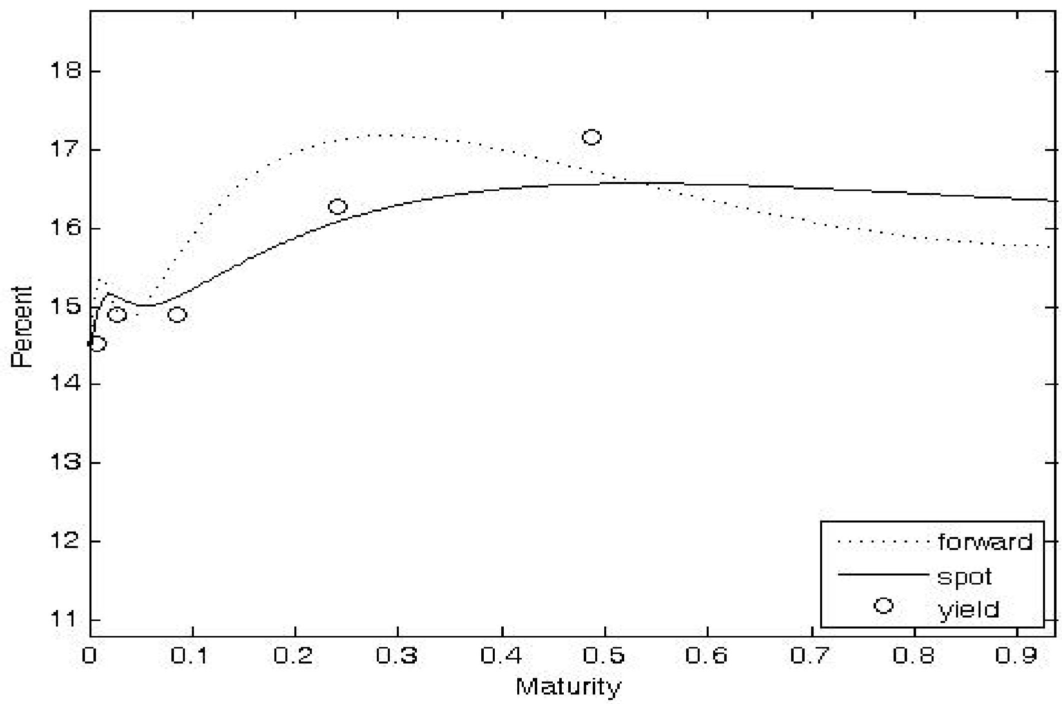

Figure 3.4: Forward and Spot rates for 8/31/1981 - Early Maturities.

curve. The six month (0.5 maturity in the figure) yield is the highest anywhere on the curve. The yield on the 30 day and 90 days are equal, and the produced polynomial is pricing all securities between these two at a lower yield than both of them - note that the forward curve intersects the spot curve at all inflection points, and is actually lower than then the sport rate immediately following the 30 day security.

3.5 Alternative methods and comparison

To test the IPQPI method and evaluate its performance, I compare it with several other prevalent methods:

LBLI: A simple linear bootstrap and linear interpolation method applied to the spot curve.

LBPQPI: A linear bootstrap method on the spot rates of the coupons with a piecewise quartic polynomial interpolation through these rates.

NS: The Nelson-Siegel method described in chapter two

SV: The Svensson extension of Nelson-Siegel described in chapter two

The linear bootstrap with linear interpolation (LBLI) method is simply a piecewise linear interpolation of the spot curve using a bootstrap method to compute the spot rates at each node. This is straightforward for zero coupon bonds when the spot rates at those maturities are directly observed. For bonds, we add one bond at a time and adjust the estimated spot rate up or down until that bond is correctly priced. Note that the linear segments determined in prior steps are not adjusted during this process. The resulting spot curve will be continuous but not differentiable at the node points. This method is admittedly a "straw man" but it is useful for illustrating some important features of the IPQPI method.

The LBPQPI method computes the spot rates at the maturities of the bonds in the same way as LBLI except, in the final step, we use a piecewise quartic polynomial interpolation of the computed spot rates to estimate the forward curve. This approach separates the interpolation and stripping steps and is the most straightforward extension of the various methods, including the Lim and Xiao (2002) method, applied to coupon bonds but that require spot rates as inputs.

The Nelson-Siegel (NS) and Svensson (SV) methods are described chapter two and, due to their widespread use, are the most serious contenders for the IPQPI method.

Evaluating the forward curve, is somewhat subjective because there is no unique solution to the curve stripping and interpolation problem. I chose to test the curves produced by each method on two key attributes: (1) smoothness, and (2) pricing ability.

Table 3.3 reports the pricing errors of the five algorithms for each of securities as well as some summary statistics to evaluate the methods. The first two columns give the maturities and actual prices of the on-the-run treasuries for the yield curve on September 9, 2009, and are replicated here from Table 3.1. The first six securities (including the settlement date in the first row) are zero coupon bills and the last six securities are the coupon bonds. The next columns show each method's pricing error in cents $\left[100 * \left(P(t_i) - \hat{P}(t_i)\right)\right]$ for each security. At the bottom of the table I report some useful statistics for each method. The "Ave Abs Error" row reports the average of the absolute pricing errors in pennies:

$$\text{Ave Abs Error} = \frac{1}{m}\sum_{i=1}^{m}\left|100\left(P(t_i) - \hat{P}(t_i)\right)\right|. \tag{3.36}$$

The Maximum Absolute Error is the largest pricing error on an individual security, in pennies. "Smoothness" reports the smoothness of the instantaneous forward curve as measured by the integral of the squared second derivatives

$$\text{Smoothness} = \left(\sqrt{\int_0^{t_m} \left(f''(t)\right)^2 \, dt}\right)^{-1} \tag{3.37}$$

$$\approx \left(\sqrt{\sum_{t=2}^{t_m - 1} \left(f(t+1) - 2f(t) + f(t-1)\right)^2}\right)^{-1} \tag{3.38}$$

where the second equation is the discrete approximation. The smoothness values in Tables 3.3, and 3.4 are computed with f measured in percentages rather than decimals. Taking the square root of the integral converts the units back to percents for easier interpretation. Taking the inverse of the measure means that the less "jerk" there is in the forward curve, the larger the Smoothness value will be. A straight line would have a smoothness score of infinity. There is a trade off between smoothness and pricing ability. For example, a linear forward curve would be very smooth but produce very large pricing errors. Alternatively, one could get zero pricing errors with a forward curve that allows discontinuities. Our goal is to find a curve that is as smooth as possible with minimal pricing errors.

Table 3.3 and associated Figures 3.5 and 3.6 shows the advantage of modeling the yield curve using the proposed IPQPI method. First,note that the LBLI method produces zero pricing errors for both the zero coupon bills and the coupon bearing bonds. This is what the LBLI algorithm is designed to do. The spot yields at each node are chosen to exactly price the security maturing at that node. Since there is no feedback from one node to previous nodes during the computations, zero pricing errors is an easy criteria to satisfy. The problem with this method is that, because the spot curve is piecewise linear, it produces discontinuities in the spot and forward curves at the nodes. The reason for the discontinuities is clear from equations (1.1). The very low smoothness statistic of 0.3747 reflects these discontinuities in the forward curve that are clearly visible in Figure 3.6. Even though all of the securities used to construct the curves are priced exactly, the discontinuities in the discount and forward curves would create large pricing errors in derivatives and other out-of-sample securities priced from the curves this method produces. The discontinuity introduces an imaginary arbitrage opportunity into the pricing of the securities that are on both sides of them. The

Table 3.3: Pricing errors and summary statistics of the five algorithms for the yield curve on 9/9/2009. The first column gives the maturity dates of the securities. The first six securities with maturities through 8/26/2010 are zero coupon bills. The pricing errors are reported in cents so an error of 7.0506 means that the bond was underpriced by \$0.070506. Max Abs Error is the largest error reported on an individual security, in cents. Ave Abs Error is the average absolute pricing error in cents. Smoothness is the inverse of the square root of the integral of the squared second derivative of the instantaneous forward curve.

Maturity	Actual Price	LBLI	LBPQPI	NS	SV	IPQPI
9 / 9/2009	100.000000	0.0000	0.0000	0.0000	0.0000	0.0000
9 /17/2009	99.998160	0.0000	0.0000	0.1347	-0.0627	0.0000
10/ 1/2009	99.994580	0.0000	0.0000	0.3272	-0.2129	0.0000
12/ 3/2009	99.969780	0.0000	0.0000	0.5388	-1.4698	0.0000
3/ 4/2010	99.899780	0.0000	0.0000	-0.3814	-4.1485	0.0000
8/26/2010	99.626580	0.0000	0.0000	-1.5419	-7.0121	0.0000
8/31/2011	100.181110	0.0000	-0.1384	3.3512	-4.1230	-0.0237
8/15/2012	100.993890	0.0000	-0.0146	-1.5876	7.0506	-0.0226
8/31/2014	100.059050	0.0000	0.2346	13.2677	23.3190	-0.3666
8/31/2016	99.605836	0.0000	1.0061	-42.5339	-39.4277	0.0242
8/15/2019	101.433760	0.0000	5.0049	32.8824	17.9175	0.5797
8/15/2039	102.961960	0.0000	-62.4890	-4.5924	-0.9220	0.0000
Ave Abs Error		0.0000	5.7302	8.4283	8.8055	0.0847
Max Abs Error		0.0000	62.4890	42.5339	39.4277	0.5797
Smoothness		0.3747	5212.6532	4518.4341	4403.6792	5187.5767

arbitrage is not reflected in the data, but it seems to exist to a trader that uses the LBLI method to price his trades.

The LBPQPI method applies a PQPI to the spot rates $\{y(t_1), \ldots, y(t_m)\}$ computed by LBLI rather than using the linear segments between the nodes of the spot curve in the LBLI method. As expected, this produces a very smooth forward curve. This process maintains the zero pricing errors for the zero coupon bills but introduces pricing errors for the coupon bonds. The thirty year bond has a large \$0.62 pricing error. This occurs because the linear segment between the ten year and thirty year node points used to compute the thirty year spot rate $y(t_{30})$ contains forty coupon payments for the thirty year bond. When PQPI is used to smooth the forward curve the value of these coupon payments can change dramatically and create large pricing errors. This method clearly illustrates the importance of simultaneously stripping the coupon bonds and interpolating the forward curve. Doing

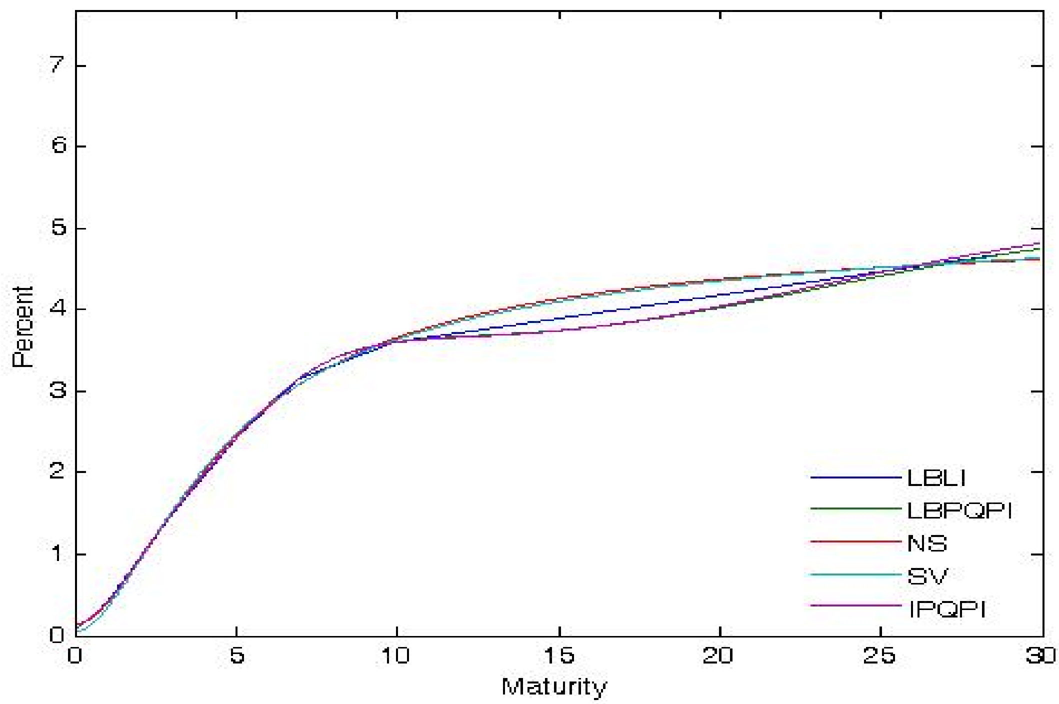

Figure 3.5: Spot rates for 9/9/2009 - all methods.

these steps sequentially introduces inconsistencies between the spot rates and the intervening forward and spot curves.

The very high smoothness score of the Nelson-Siegel method illustrates why this approach remains so popular. The primary disadvantage of the NS method is that the possible shapes of the spot and forward curves are quite limited and may produce substantial pricing errors. The average pricing error of 8.4 cents with maximum errors of 42 cents are too large for most applications. The Svensson extension of NS allows for more flexible curves and may improve pricing. In the 2009 data set presented here, the average pricing error is about the same as the NS method at about 8.8 cents, and a maximum error of 39.4.

The IPQPI method has low pricing errors, averaging at less than a tenth of a penny, while maintaining the smoothest possible forward curve among the class of piecewise polynomials. The largest pricing error is just over half a penny, and is observed on the ten year bond.

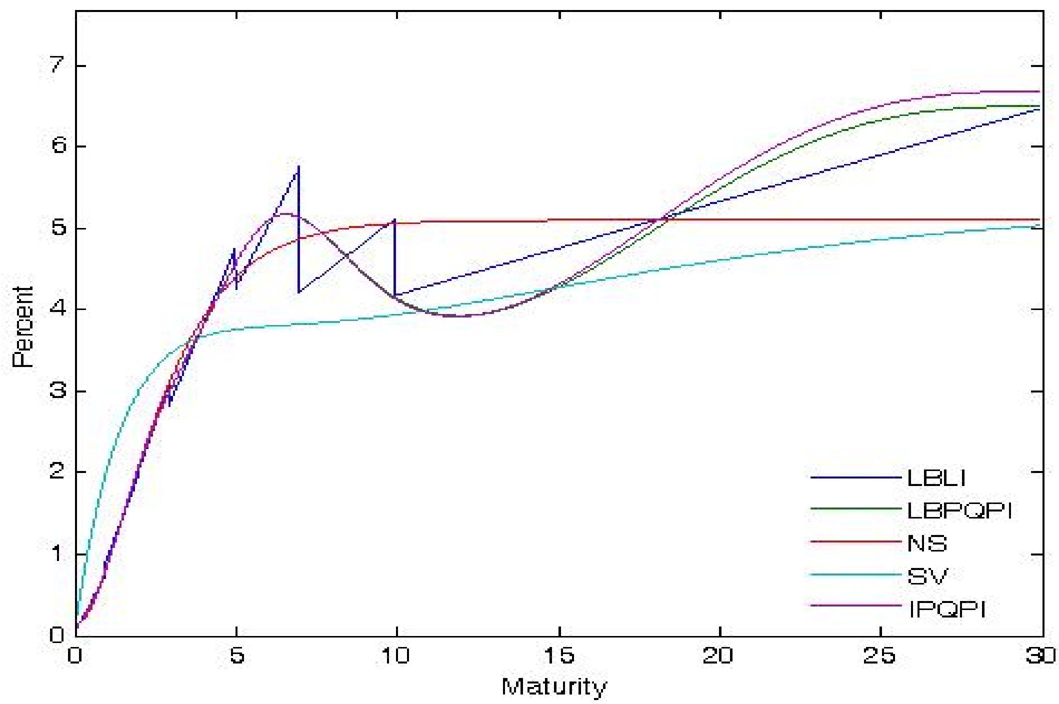

Figure 3.6: Forward curves for 9/9/2009 - all methods.

Table 3.4 and the related Figures 3.8 through 3.10 show the results for the 1981 data. Here, the inverted curve produces different results, but with the same basic interpretation. Worth noting is the low degree of smoothness of the IPQPI method, which is still the smoothest possible forward curve among the class of piecewise polynomials. The reason for this low smoothness is visible in Figures 3.8 and 3.10. The IPQPI method is able to accommodate for the dynamics of the short end of maturities and prices these short-maturities securities very well, albeit by paying a price by exhibiting a low degree of smoothness. The NS and the SV are not able to price securities at the short end of the curve, but preform better at the latest maturities. This is true especially for the SV method which is attributed to the additional term in the SV method (compared to the NS method) that allows more flexibility in the shapes of the forward and spot curves. All securities up to the 5 year one have similar pricing errors, but the 10 year security is priced better by SV compared to NS by a factor of 10, and both the 20 and 30 year are priced better by a factor

Table 3.4: Pricing errors and summary statistics of the five algorithms for the yield curve on 8/31/1981. The first column gives the maturity dates of the securities. The first six securities with maturities through 8/12/1982 are zero coupon bills. The pricing errors are reported in cents so an error of 7.0506 means that the bond was underpriced by $0.070506. Max Abs Error is the largest error reported on an individual security, in cents. Ave Abs Error is the average absolute pricing error in cents. Smoothness is the inverse of the square root of the integral of the squared second derivative of the instantaneous forward curve.

Maturity	Actual Price	LBLI	LBPQPI	NS	SV	IPQPI
8/31/1981	100.0000	0.0000	0.00000	0.0000	0.0000	0.0000
9/3/1981	99.8775	0.0000	0.00000	1.3552	1.5580	0.0000
9/10/1981	99.5867	0.0000	0.00000	3.9314	4.5971	0.0000
10/1/1981	98.7247	0.0000	0.00000	11.9557	13.9258	0.0000
11/27/1981	96.1989	0.0000	0.00000	9.6107	14.5283	0.0000
2/25/1982	92.2422	0.0000	0.00000	-6.7373	1.3031	0.0000
8/12/1982	85.6506	0.0000	0.00000	-5.9390	4.1361	0.0000
7/31/1983	99.806096	0.0000	-0.20870	-22.5051	-14.4496	0.2027
11/15/1984	103.75815	0.0000	0.71650	4.9817	2.5074	0.2882
11/15/1986	97.009511	0.0000	-2.61050	23.1559	11.0490	0.0680
7/15/1988	94.663043	0.0000	-2.60690	1.6472	-8.2545	0.1593
8/15/1991	97.896739	0.0000	-3.21920	-6.4815	0.6790	0.4174
8/15/2001	90.519022	0.0000	-14.5672	-25.2764	0.4120	9.1591
5/15/2011	98.009511	0.0000	-15.8360	19.3851	-0.1547	0.0000
Ave Abs Error		0.0000	2.8404	10.2116	5.5396	0.7353
Max Abs Error		0.0000	15.8360	25.2764	14.5283	9.1591
Smoothness		0.3043	4.2892	38490.9519	11763.2074	4.2892

of 100. This superior pricing ability comes at a price of a lesser degree of smoothness by a factor of 3. The largest pricing error exhibited by the IPQPI method is a 9.15 pennies on the 20 year bond. All other securities priced by the IPQPI method have an error that is less than one penny. Worth noting is the pricing error on the 30 year bond which is 0.

This chapter has described the Iterated Piecewise Quartic Polynomial Interpolation algorithm for simultaneously interpolating and stripping a yield curve consisting of coupon paying bonds. The algorithm is accurate, flexible and relies only upon the solution of linear systems of equations and produces maximally smooth forward curves.

The IPQPI method performs very well overall. Although the Nelson-Siegel algorithm produces smoother forward curves it does so at the cost of much larger pricing errors. IPQPI produces pricing errors an order of magnitude smaller than the Svensson algorithm while

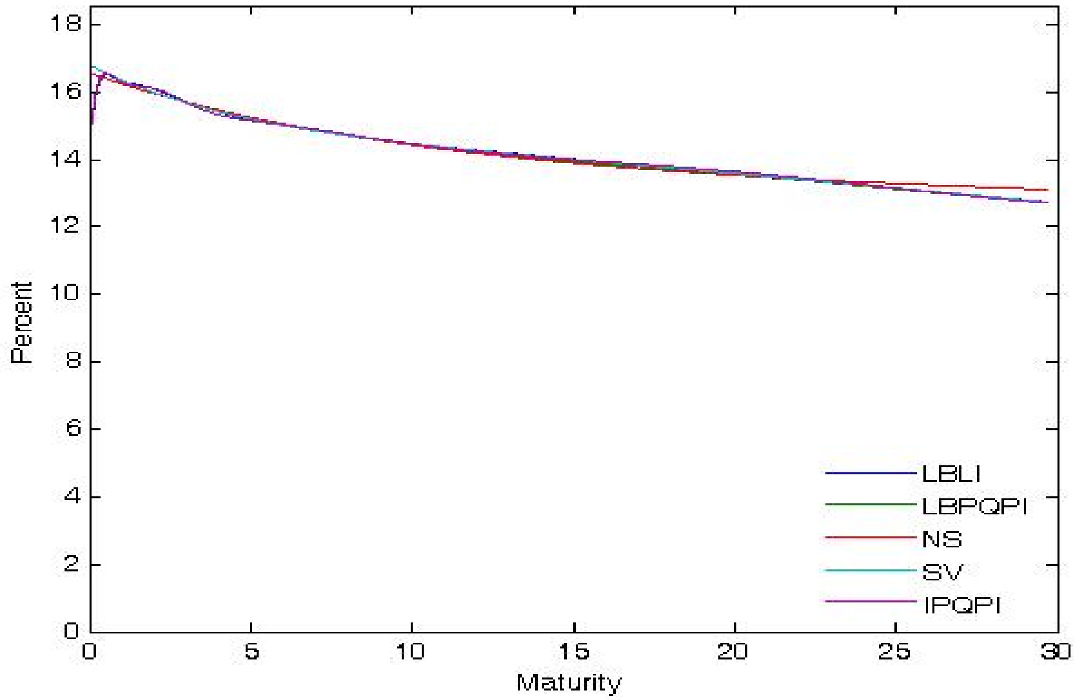

Figure 3.7: Spot rates for 8/31/1981 - all methods.

at the same time producing smooth forward curves. Computationally, IPQPI is very stable and an order of magnitude faster than the highly nonlinear Svensson approach. The method is also easy to modify by adding additional constraints for specific pricing applications.

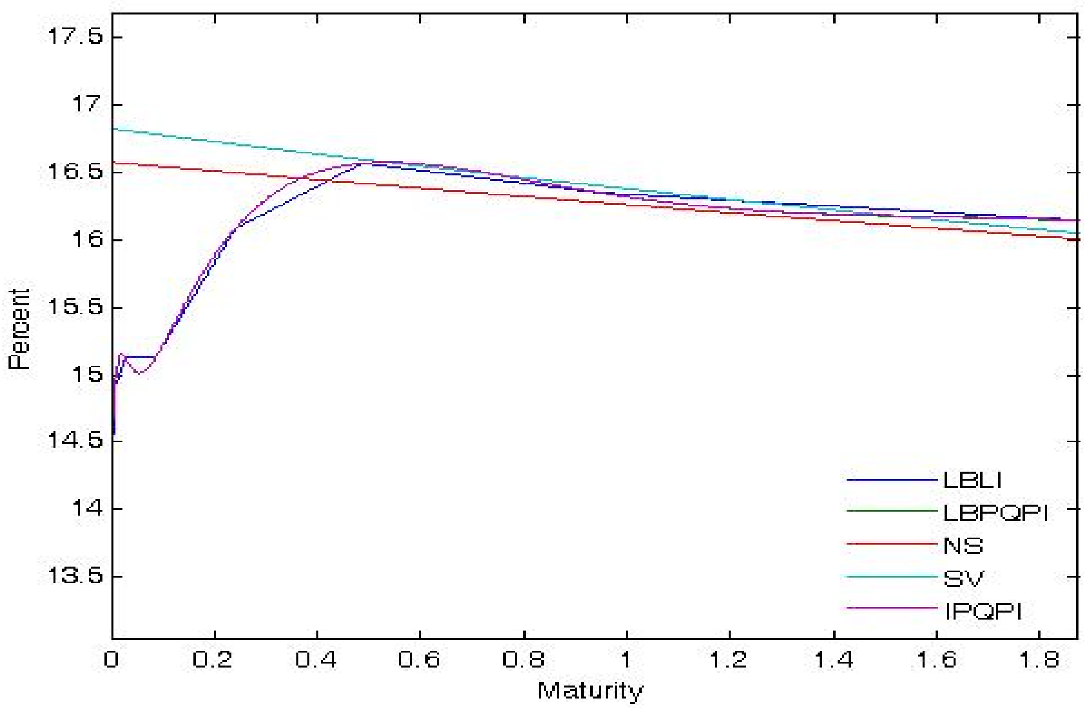

Figure 3.8: Spot rates for 8/31/1981 - early maturities for all methods.

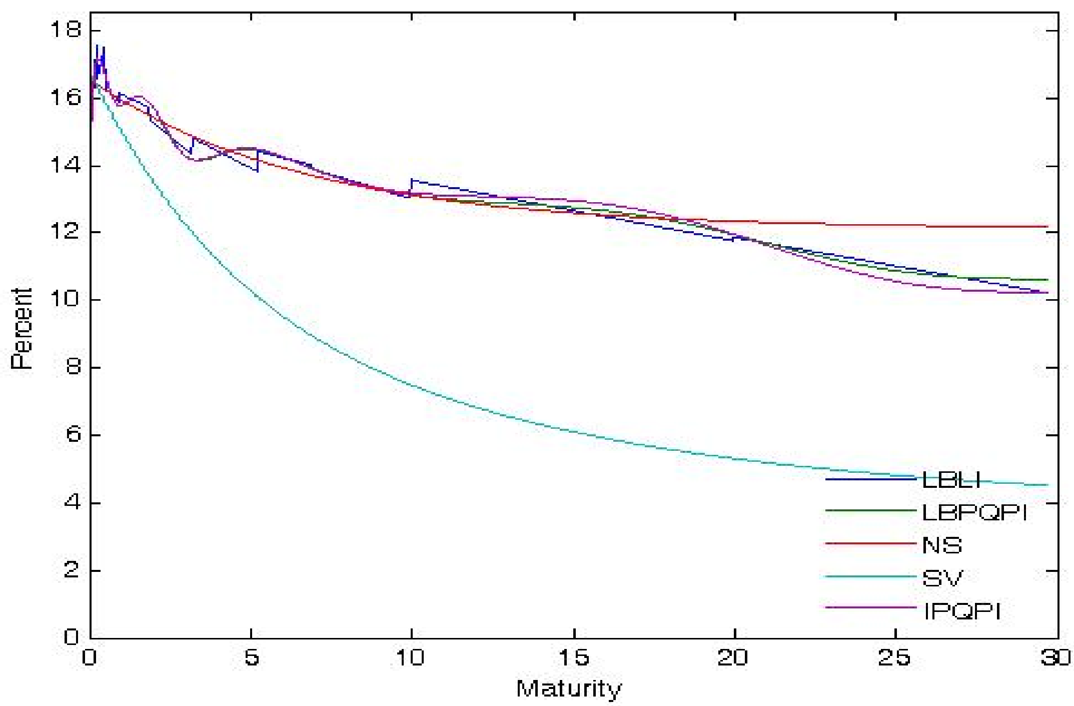

Figure 3.9: Forward curves for 8/31/1981 - all methods.

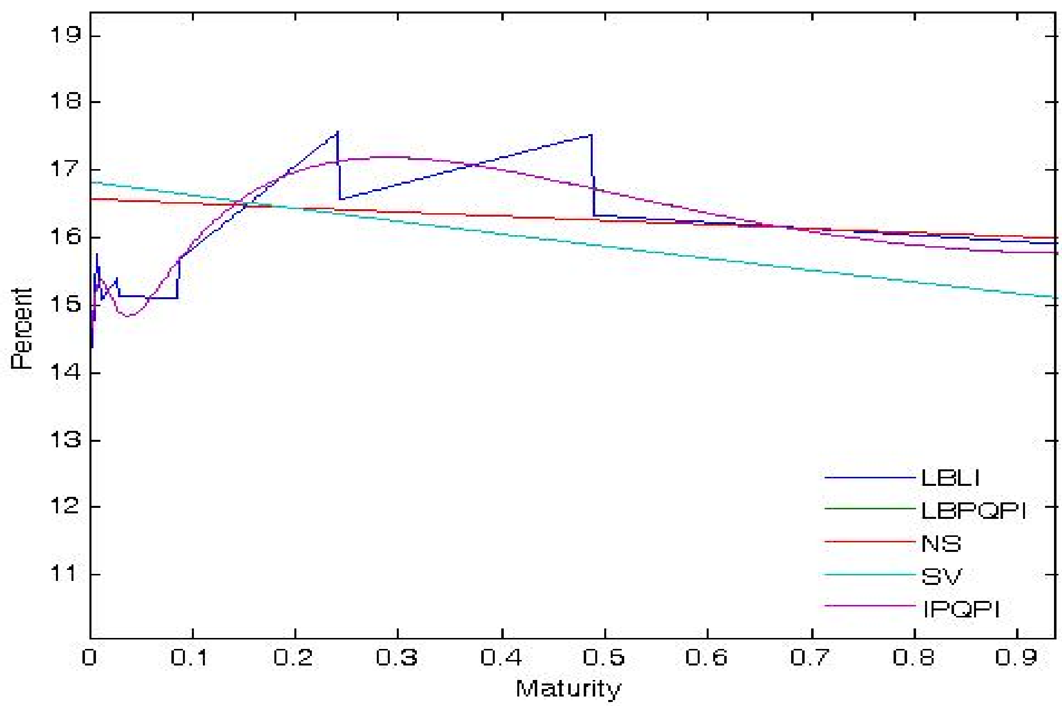

Figure 3.10: Forward curves for 8/31/1981 - early maturities for all methods.

CHAPTER 4

A GENETIC PROGRAMMING APPROACH TO FITTING THE YIELD CURVE

4.1 Introduction

Since the topic of this research is concerned with constructing a spot and forward curves, as well as the discount function for pricing purposes, it may be the case that the search for the fitting solution has been restricted thus far to a rather limited number of functions. To date, research in the field has been concentrated on several methods which we can conveniently accommodate using our standard mathematical and statistical tools. For pricing purposes we are not really concerned about *how* we got to the curve, but we are very concerned with regards to the *characteristics* of it. The need to search for a pricing function in a wider space is evident, and in light of this, the use of *genetic programming* (GP) methods to compute the forward curve is being proposed. GP has several advantages over the traditional parametric approaches I surveyed in Chapter 2. First, it is a non parametric method that is not dependant on any specific functional assumptions, that may contain specification errors. Second, GP only responds to the data, and the data shapes the result in a way a parametric approach cannot accommodate. Third, the method is relatively easy to implement.

The GP system will search in the space of continuous, smooth (twice differentiable) functions for the best functional form that produces a yield curve that fits the data. In theory, a search over a wider space of functions cannot yield a solution that will perform worse than the IPQPI method, since polynomials are included in the space searched. An additional motivation to expand the search space for an adequate function is the local stability of the forward curve function, which has serious real world implications as to portfolio hedging costs. As pointed by Hagan and West (2006), among others, a polynomial based method is sensitive to local changes in observed yields. This sensitivity can cause the entire curve to shift when small changes occur to one of the securities in the portfolio. For example, an abrupt change in the yield on the 10 year note, which is used by mortgage lenders, among others, to hedge their portfolios of commercial real estate loans, and therefore is traded heavily and is sensitive to macroeconomic related news, can shift an entire curve that is based on polynomials. Due to the nature of curve construction by polynomials, this may impose a high cost to reposition portfolios that are hedged based on this forward curve. A more stable curve, one which moves only locally to accommodate a rise or a fall in a the yield of a particular security, but does not shift much elsewhere, is a more desirable curve by practitioners.

This chapter is organized as follows: First, I give a brief overview of GP, and I review an example of curve fitting experiment (using GP) from the literature. I then discuss my methodology and the GP system I use, as well as the results. The results I obtained from the GP approach are short of my expectations. The problem proved to be difficult for the system to solve, and at the end of the chapter I discuss possible reasons for these results.

4.2 Genetic Programming

Computational science has evolved rapidly over the last 20 years. One of the fastest developing area of research is concerned with "machine intelligence" and genetic programming is one of several branches in this area of research. Poli et al. (2008) define GP as an evolutionary computation (sometimes referred to as *revolutionary algorithms*) technique that solves problems without requiring the user to know or specify the form or structure of the solution in advance. At the most abstract level GP is a systematic, domain-independent method for getting computers to solve problems automatically starting from a high-level statement of what needs to be done.

In a GP process, a population of computer programs is evolving to produce the desired outcome. That is, generation by generation, GP stochastically transforms populations of programs into new, hopefully better, populations of programs. GP is a random process, and it can never guarantee results. On the other hand, the randomness of the process may lead it to better solutions than deterministic methods. In essence, GP is a search technique that explores the space of computer programs.

Every GP system contains the following building blocks:

1. GP finds out how well a program ("population member") works by running it, and then comparing its behavior to some ideal. This comparison is quantified to give a numeric value called fitness (the "fitness score"). The ideal is defined by the user, and that is where the GP process is aiming to evolve. The fitness score is calculated by the fitness function. The fitness function is defined by the user to qualify members of the population and rank them by the degree of their similarity to the ideal.

2. The programs that receive a better (high or low, depending on the design of the experiment) fitness score are chosen to breed and produce new programs for the next generation.

3. The new generation of programs are run and compared again to the ideal, etc'.

4. The GP process stops when a certain criteria is met, such as the similarity level to ideal has been achieved,or when the user instructs the system to stop.

GP produces a syntax for the best function (this is the function that has the best fitness score in the last generation to run) it finds, and presents it in the form of a "tree". The tree contains the function that is the best-fitting member of the population in the last generation. Figure 4.1 shows a simple example of a tree which contains 9 nodes (a node is an argument or an operator in the function), and 2 "levels" (a level is a layer of arguments).

The GP process begins with an initial generation of functions, which is being generated randomly by the system. There are several techniques to accomplish this task, all of which are concerned with the size and depth of the initial *tree* (discussed below). Some methods allow the user to plant her own function in the initial population (a practice called "seeding"). The seeded individual is thought to be better than all the randomly generated initial members of the population, hence its descendants are poised to take over the evolving population within a few generations - which may lead to a lack of diversity in the GP process. I further discuss this problem below. The most common methods used to initialize the population are:

1. *The Full Method*: In this method, the initial individuals are generated so that they are perfectly balanced in the sense that all branches of each of the trees are of the same depth, which is defined by the user. However, each tree (a population member) may be of different size (number of brunches) or depth.

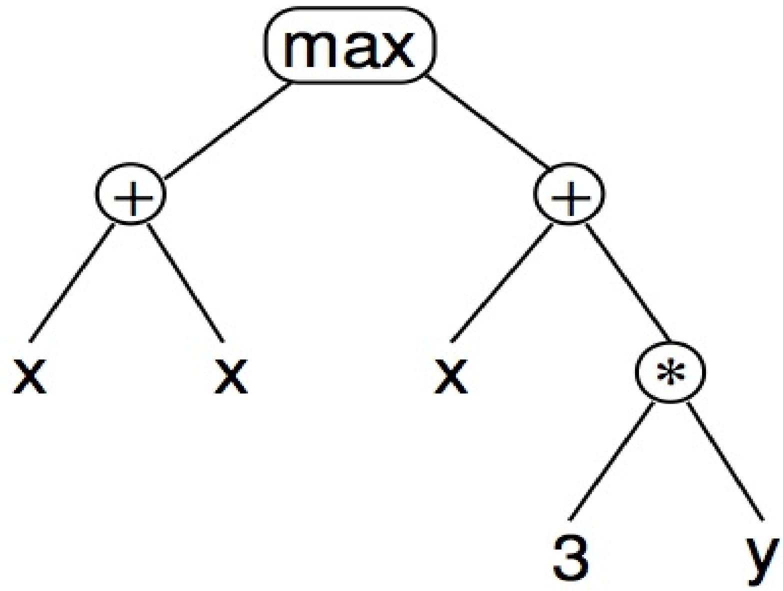

Figure 4.1: GP function "tree" output. Source: Poli et al. (2008)

2. *The Grow Method*: This method is designed to produced un-balanced trees with branches that are different lengths.

3. *Ramped Half and Half Method* This a hybrid method of the Full and Grow method. Here, some part of the initial generation is generated using the Full method, and the other part is generated using the Grow method, using a range of depth limits that are being applied randomly to population members.

The next step in the GP process is to evaluate the fitness of each member using the fitness function. The members are evaluated using a fitness function and are assigned a fitness score. Those members that are determined to be better than others, have a higher probability of producing offsprings, which will hopefully achieve a yet better fitness score. There are several methods a GP system can select members for breeding. The most common method is the *Tournament Method*. In a tournament method, a portion of the population is selected to participate in a contest (the user can determine that as well, or leave it for the system to choose the quantity of participants). The participants are compared to each other,

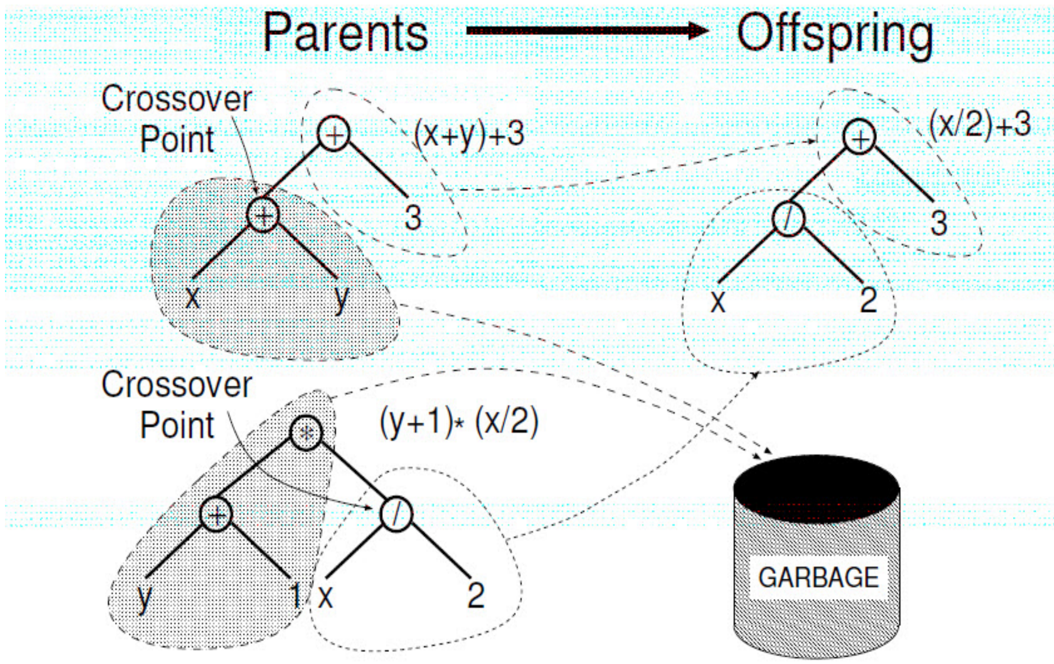

Figure 4.2: An Illustration of Crossover. Source: Poli et al. (2008)

and the one with the best fitness score is chosen to parent a new function. Parenting can happen using several processes. The primary processes that are used to create new programs and are based upon genetic operations are:

- Crossover: The creation of a child program by combining randomly chosen parts from two selected parent programs. In this case the parents are the two "winners" of the tournament - the two best members. See Figure 4.2 for an illustration of this genetic operator

- Mutation: The creation of a new child program by randomly altering a randomly chosen part of a selected parent program. Here, the is only one best function that wins the tournament, and is chosen to mutate. See Figure 4.3 for an illustration of this genetic operator

The user generally can determine the rate of mutation and the rate of cross overs in each generation (ie, what percentage of the population will be subject to each genetic operator), but she cannot decide which specific individual will be the one used for it.

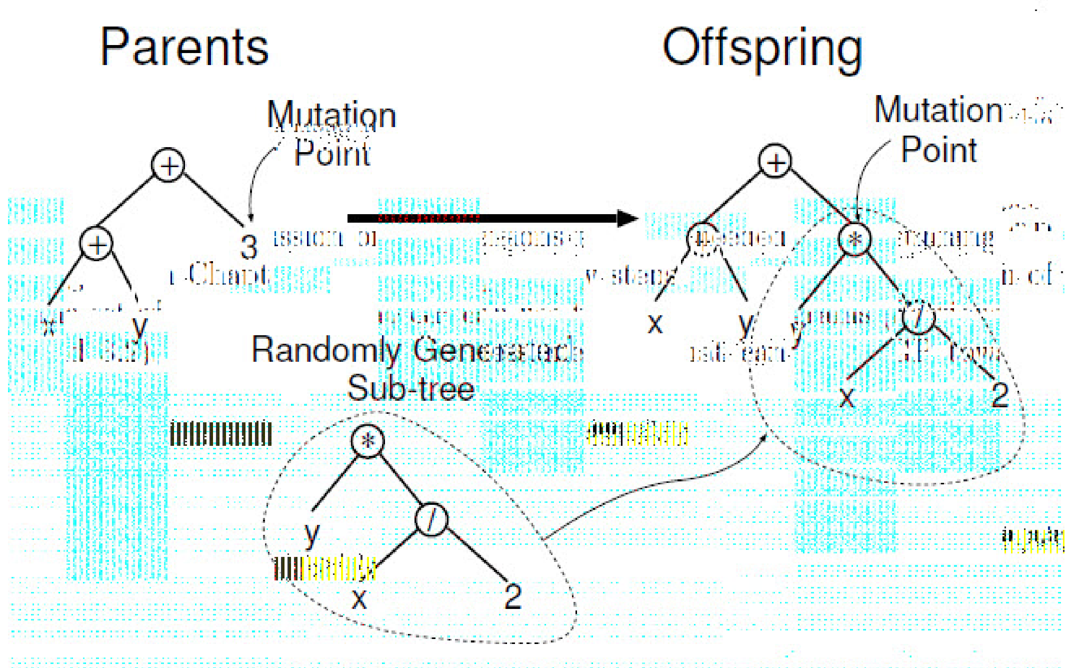

Figure 4.3: An Illustration of Mutation. Source: Poli et al. (2008)

When the genetic process is finished in a particular generation, a fitness score is calculated again, and the best(in terms of fitness) n individuals (n being the population size allowed in each generation) will move on to the next generation to repeat the process again.

GP systems are vulnerable to several issues that may cause the evolution process of the programs to either stop, or slow down significantly due to overloading of the computer resources. The first issue a good GP system design must face is the bloating issue. Tree syntaxes can get very complicated in a GP model. This is due to the fact that the search space the GP is processing is virtually unlimited. The functions that are being evaluated can have hundreds of nodes, and the syntax can have a “depth” of hundreds of “levels”. This can slow the computer process significantly, and hinder the evolution of the best-fitting programs from one generation to the next. Designers of GP systems have a built in a solution to solve this problem in the form of a mechanism whereby the user has the ability to limit the depth of the syntax, as well as control to the the size of the population that is being evaluated at each generation. There are also mechanisms to “trim” the tree by looking for operators, and function parts that are repeating themselves without adding anything to the fitness of the function. The second issue every GP system must contend with, and which is more subtle, is the genetic diversity of the population in each generation. This issue becomes more severe at advanced generations. If the function we have at generation 1500 is still too far off from the ideal ("far off" is defined by the user, and most likely will mean low degree of fitness), it may be the case that there is not enough diversity in the population to modify the function enough to get the best fitting individual to a fitness level that is satisfactory to the experiment criteria (the evolution process "got into a corner"). GP systems must be able to recognize that an evolution process is not advancing at a pace that is satisfactory, and then re-shuffle the existing population by trying new and radical changes to the functions being examined so it has a chance to find better fitting individuals eventually.

4.2.1 Symbolic Regression Example

Kamal and Eassa (2002) show how to fit data that was generated from a function, using GP. Ten observations were generated and plotted on the X/Y plane (see figure 4.4 and they use a fitness function that minimizes the sum of the square errors as follows:

$$\text{Fitness} = \frac{1}{1 + \sum_1^n [Y_i - f(X_i)]^2} \tag{4.1}$$

where Y_i is the observed value and $f(X_i)$ is the value given by the proposed function that is being evaluated. The smaller the sum of the squared errors - the larger the fitness score which implies a better fit. Figure 4.5 shows the learning that is taken place by the GP system in Kamal and Easa's research. Notice how the GP system is producing members that get better at fitting the curve. They report that GP has been able to fit the observations without an error (see Figure **??**),and the best fitting individual yielded the syntax presented in Figure 4.7, where E is the exponential function. The computed function evaluates to

$$f(x) = e^{-x}[e^x + e^{x-2x}] = [1 + e^{-2x}]^{-1} \tag{4.2}$$

and produces a perfect fitness of 1.

Their demonstration is very useful to my case, since I am concerned with a similar task - fitting a curve through a set of points. My case is more complicated since the observations are not directly observable. Recall that I am fitting the unobserved forward curve which is implied by a few observed securities along the yield curve. Thus, some preliminary work is required to "prepare" the points that will be incorporated into the GP process as the

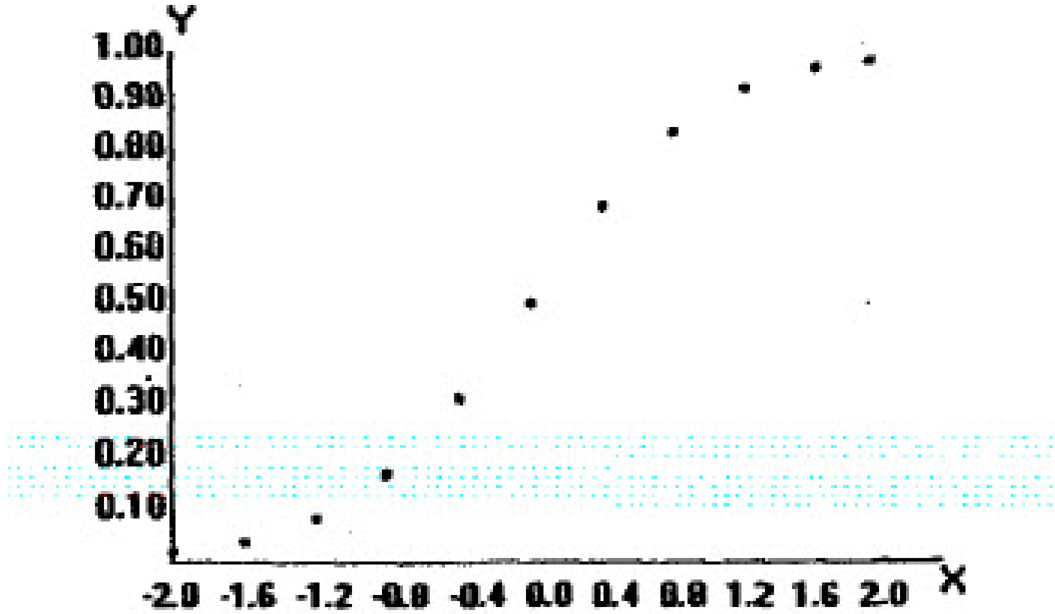

Figure 4.4: Kamal and Easa's data plot.

model-to-fit (the "ideal"). The "optimal" forward curve will not be unique and may take several different shapes while still satisfying the constraints.

4.3 Using GP to Find the Forward Curve

The GP goal in this research is to search the space of smooth functions over a large class of differentiable functions in order to find forward curves that accurately price the on-the-run bonds. Recall that the idea behind the usage of a GP system to accomplish the goal of fitting the forward curve is to remove any pre-conceived notion of what the function might look like, and to see if the current state of the research is limited by methods are easily accommodated by standard computational tools. This focus on functions that we can conveniently process may limit our ability to find a smoother function. The curve produced by GP will then be compared to the function produced by the IPQPI method described in Chapter 3 to evaluate its performance. There are several software platforms one may use to implement a GP system. I use the GPLAB toolbox (version 3) in MATLAB developed by Sara Silva (see Silva (2007)) to implement the GP experiment.

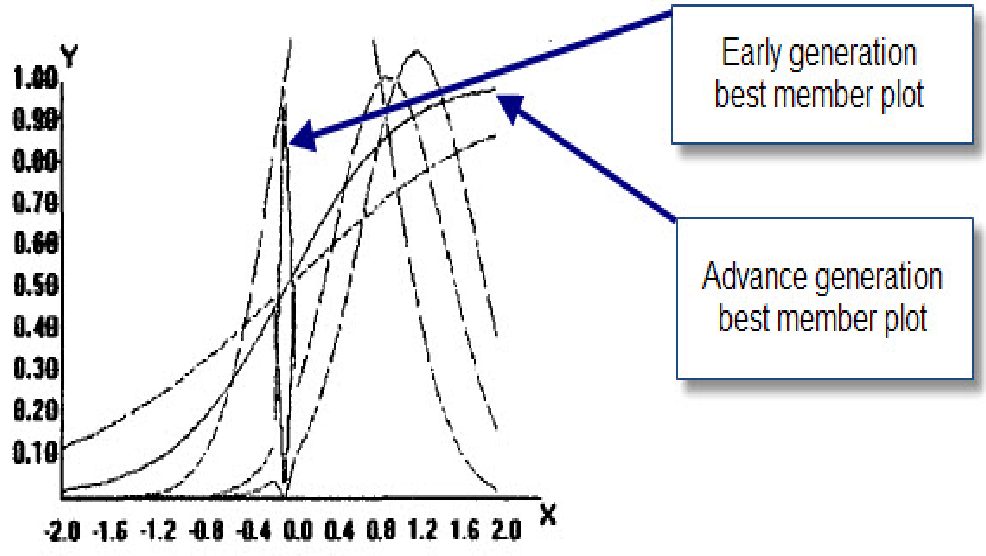

Figure 4.5: GP learning how to fit the curve in Kamal and Easa's experiment.

The main building block of any GP implementation process is defining the *ideal* to be matched. Practically, this is done by carefully designing a fitness function which will be used to assign a fitness score to each individual function. The individual functions in the population shall be ranked and either evolve or be discarded based on their fitness score. The design of the fitness function needs to penalize and reward functions so the fittest function will have the properties of a "good" forward curve (the *ideal* to be matched). These properties are: pricing ability, differentiability and smoothness, as well as initial and terminal conditions that are appropriate for the curve.

In any given generation the GP creates a population of N programs, $\{f_i(t)\}_{i=1}^{N}, t = 1, \ldots, T$, for the forward curve. To evaluate the fitness of each program I first compute the spot rate and discount function curves implied by the forward curve program. To construct the spot curve $y_i(t)$ I first set $y_i(1) = f_i(1)$ since the spot and forward curves must have the same origin. Then the spot curve can be computed recursively from the forward rate as

$$y_i(t) = \frac{1}{t}\left(f_i(t) + (t-1)y_i(t-1)\right), \; t = 2, \ldots, T. \tag{4.3}$$

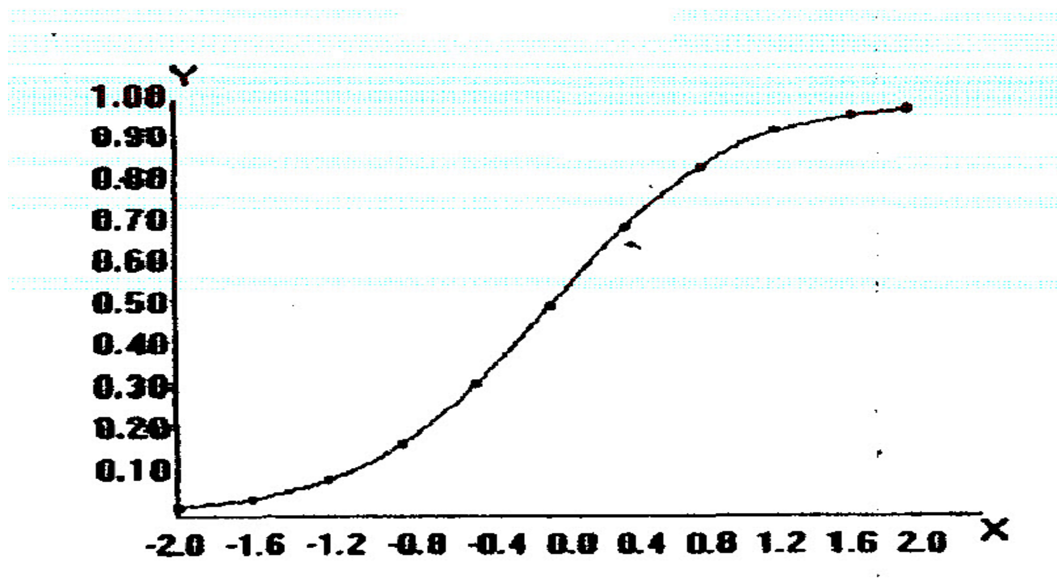

Figure 4.6: Kamal and Easa's best curve.

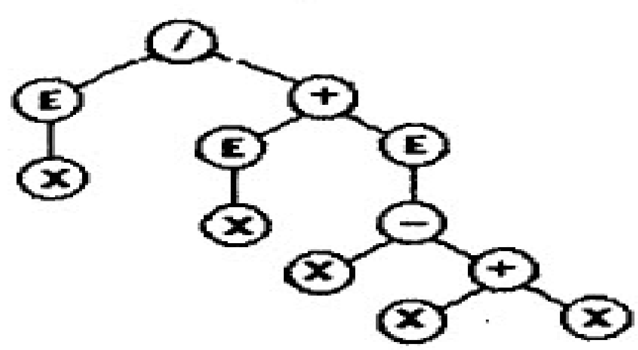

Figure 4.7: Kamal and Easa best curve's tree.

Given the spot rate, the discount function can be computed as

$$d_i(t) = \exp\left(-y_i(t)t\right), \ t = 1, \ldots, T \tag{4.4}$$

where t is measured in years to maturity.

The pricing ability of each program is evaluated as the sum of the absolute pricing errors in cents of all m bonds

$$F_1 = \sum_{j=1}^{m} 100 \left| \widehat{P}_{i,j} - P_j \right| \tag{4.5}$$

where $\widehat{P}_{i,j} = CF_j' d_i$ and CF_j is the $T \times 1$ cash flow vector of the j^{th} bond and d_i is the $T \times 1$ discount function for program i.

For short-maturity zero coupon bonds considerable changes in the spot rate imply only small changes in the asset's price. Experience has shown that, because of this, the GP has quite a lot of trouble fitting the short end of the yield curve. To help the GP, I have included in the fitness function a penalty for missing the spot rates for the zero coupon Treasury Bills

$$F_2 = \sum_{j=1}^{m_{zc}} \frac{1}{t_j} 100 \left| y(t_j) - y_i(t_j) \right| \tag{4.6}$$

where t_j is the maturity in years of the j^{th} zero coupon bill and m_{zc} is the number of zero coupon bills. Since y is measure in decimals, the scale is adjusted to percent by multiplying by 100 and the absolute error is weighted by the inverse of the maturity of the bills so that the shortest maturities have the highest weight.

Next I impose a penalty for programs that do not produce smooth forward curves

$$F_3 = \sum_{t=2}^{T-1} \left(f_i(t+1) - 2f_i(t) + f_i(t-1) \right)^2 \tag{4.7}$$

which is the same as the objective function $\int_0^T \left(f_i''(t)\right)^2 dt$ I used in the IPQPI method in the previous chapter.

The initial condition for the forward curve is enforced using the penalty term

$$F_4 = 100 \left| f_i(1) - f_0 \right| \tag{4.8}$$

where f_0 is the initial value for the forward curve from the input data for the yield curve.

The terminal condition is imposed as

$$F_5 = \sum_{s=1}^{30} 100 \left| f_i(T) - f_i(T-s) \right| \tag{4.9}$$

which penalizes the program if it produces a forward curve that is not flat over the final thirty days of the curve.

To impose non-negativity on the computed forward curves I use

$$F_6 = \#\left(\{f_i(t) < 0\}\,,\ \text{over } t = 1, \ldots, T.\right. \tag{4.10}$$

which is a simple count of the number of days for that the forward curve f_i is negative.

In early generations the GP tends to create many population members with very poor levels of fitness and it can take a considerable number of generations before those members are killed off by the cross over and mutation operations. To help speed up the process two additional terms are added in the fitness function. These terms are only used during early generations and are then dropped after the population settles into the neighbor of a reasonable forward function.

The first condition penalizes the program if it produces a forward curve whose average level over the first 1000 days of maturities (about 2.7 years) is too far from the average of the forward curve generated by the IPQPI method. Let $\bar{f}_i$ be the average value of f_i and $\bar{f}_p$ be the average value of the forward curve from IPQPI over maturities from 1 to 1000 days. Then the penalty is

$$F_7 = 100\left|\bar{f}_i - \bar{f}_p\right|. \tag{4.11}$$

Similarly, I penalize the program if it produces forward curves whose gross slope is too different from the gross slope of the IPQPI forward curve. I compute the gross slope as the average value of the forward curve over the first 1000 days of maturities minus the average value of the forward curve over the last 1000 days or the last 2.7 years of maturities along the forward curve. The penalty is

$$F_8 = 100\left|\text{slope } f_i - \text{slope } f_p\right|. \tag{4.12}$$

The final fitness function is then

$$\text{fitness}_i = \sum_{i=1}^{8} w_i F_i \tag{4.13}$$

where F_7 and F_8 are set to zero after a few generations once at least one of the programs produces a forward curve with a level and slope that is within a percent of the forward curve generated by the IPQPI method.

Since fitness is one number that is assigned to each function in the population, and there are several arguments that make up the score, a correct weighing scheme for each part of the fitness function is key to GP's success in sorting which functions are better than others. The weights w_i were chosen by experimentation and I used $w_i = 1$ for $i = 1, 2, 5 - 8$, $w_3 = 10^6$ and $w_4 = 10^3$. This puts a high weight on smoothness and the initial condition (which is hard to hit with the GP) and uniform weights on all other penalties. It is important to balance the proper weights on each of the arguments and make them as equal as possible so as to prevent GP from trying to minimize the error on one argument at the expense of another. For example, if F_3 is weighed too lightly compared to F_1 in the fitness function, then GP may produce a curve that prices correctly, but oscillates widely between coupons and maturities, producing a curve that is not reasonable. I found that the GP system I used was very sensitive to changing weights and yielded very different curves every time the fitness function was weighted differently. GP's ability to learn and retain the "best" members completely depends on the specification of the fitness function. The system is focusing on the the arguments that contribute the most to the fitness score and almost ignore the other arguments.

To illustrate the weighing problem, consider two members of a generation with fitness scores as follows: Member X has a fitness score of 32 , made up of 2 arguments, $F_1 = 1$ for pricing ability, and $F_3 = 31$ for smoothness, as follows: Fitness of X=$F_1 + F_2 = 32$. Member Y has a fitness score of 30, where $F_1 = 14$ and $F_3 = 16$. Here, member X is less likely to survive,and member Y is more likely to breed, despite the fact that X prices much better than Y (15 times better), and it is only about 2 times worse in its smoothness. Also, for Y itself, the breeding process will likely focus on reducing the F_3 argument since it is larger than F_1.

4.3.1 Results and Discussion

In this section I run the GP system on the 9/9/2009 and 8/31/1981 data sets which were discussed in Chapter 3 above. Each GP-produced curve is analyzed and then compared to its counterpart from the IPQPI method.

The 9/9/2009 run was processed using a population size of 100 individuals, for 305 generations. The tree depth was limited to 28 levels to limit bloating and aid in the reduction in processing time. The fittest individual displayed in Figure 4.8 has 268 nodes in its tree

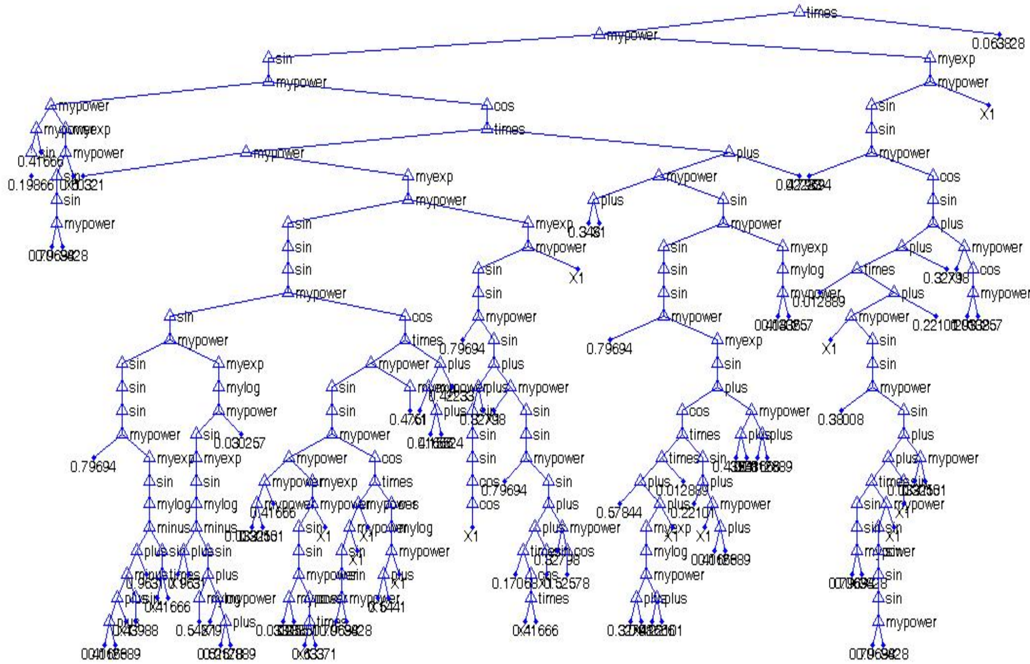

Figure 4.8: 9/9/2009 Yield Curve function produced by GP - syntax representation .

and is quite complex. The smoothness level is 2782. Figure 4.9 shows that individual's curves and we can see why the smoothness figure is relatively low. It is clear that the curve bends at the 5 year maturity somewhat abruptly and there are also some kinks at the curve following the 10 year maturity mark.

The pricing errors exhibited by GP's fittest individual for the 9/9/2009 reported in Table 4.1 are mostly tolerable (less than one penny) except for the 5 and 7 year maturities, where errors are in excess of 25 cents. Consequently, that area of the curve around the 5 and 7 year maturities is also the one that has the kink, and evidently the GP was having a difficulty accommodating it. Table 4.2 reveals the dynamics of the fitting process. Although the fitness score itself is meaningless in this case since it has no units, examining its components and their contribution to the total fitness score shows where the difficulty in the fitting process

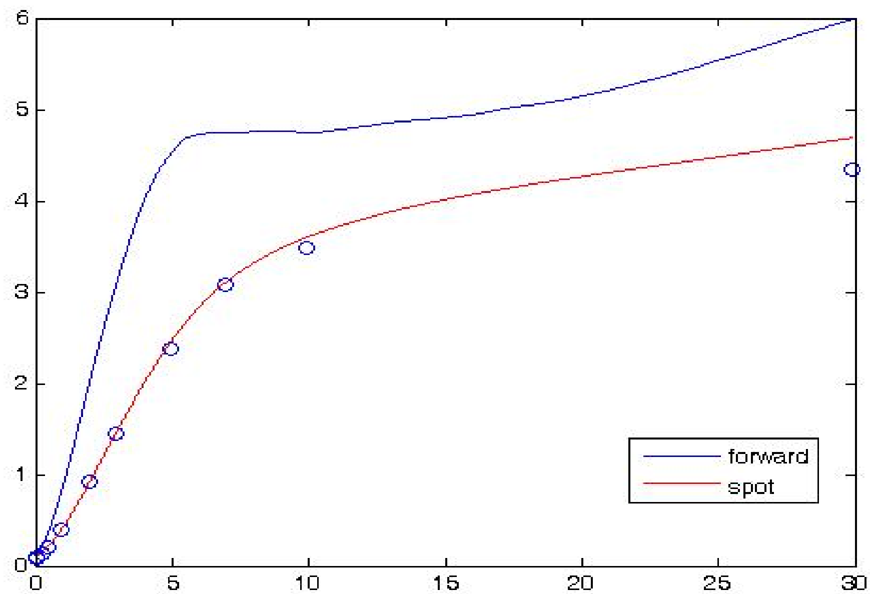

Figure 4.9: 9/9/2009 Yield and Forward Curves produced by GP .

Table 4.1: Actual and GP-estimated bond prices for the yield curve on date 9/9/2009.

Maturity	Coupon	Actual Price	Yield	GP Est. Price	Penny Error
9/9/2009	0	100.0000	0.0808	100.0000	0.0
9/17/2009	0	99.998160	0.08415	99.997538	0.0622
10/10/2009	0	99.994580	0.09	99.993219	0.1361
12/30/2009	0	99.969780	0.12982	99.965132	0.4648
3/4/2010	0	99.899780	0.20806	99.892857	0.6923
8/26/2010	0	99.626580	0.38978	99.610792	1.5788
8/31/2011	1.0	100.18111	0.92	100.183501	-0.2391
8/15/2012	1.75	100.99389	1.4443	100.991604	0.22846
8/31/2014	2.375	100.05905	2.3751	99.801678	25.7372
8/31/2016	3.0	99.605836	3.0753	99.856307	-25.0471
8/15/2019	3.625	101.43376	3.4828	101.435254	-0.1494
8/15/2039	4.5	102.96196	4.3408	102.960466	0.1494

Table 4.2: Contribution to fitness for the 9/9/2009 data.

Fitness Argument	Score	Weight	Contribution to Fitness
F_1 Price in pennies	54.4850	1	54.4850
F_2 Bill Yields (BP)	1.4982	1	1.4982
F_3 Smoothness	0.0	10^6	0.1291
F_4 Initial Condition	0.0149	10^3	14.9
F_5 Terminal Condition	0.0071	1	0.0071
F_6 Non negativity	0.0	0	0
F_7 Level different from IPQPI	0.0	0	0
F_8 Slope different from IPQPI	0.0	0	0

is coming from. GP was focusing on trying to reduce the pricing errors on the coupon bonds more than any other factor in the fitness function. Note the load on that component - which is 54.485 out of the total fitness score of 71.0019 (about 77%). The initial point constraint, $F4$, has a load of 14.9 and we can note that this translate to a 0 pennies error (in Table 4.1).

The relatively low smoothness of the curve is surprising as well. I have conjectured that since the search space for GP includes polynomials, as well as simple functions such as NS and SV, that the system cannot do worse in terms of smoothness, but it appears this is not the case in this particular example. Recalling that the forward curve is constructed from the area under the yield curve, combined with the constraints on the system placed via the fitness function, it is quit clear that the GP system can have a different solution to the curve each time it runs. In fact, at every run, the system begins the search from scratch, and it is not guaranteed that the solution one yields from that particular run would be optimal compared to the deterministic methods. The solution will be the best out of the population of that specific evolutionary system, but it may not be the best it *can* be.

To compare and contrast the IPQPI method and the GP method, I plot the two curves in Figure 4.13 and exhibit the pricing errors from both methods in Table 4.3. The curves are similar to each other in early maturities, but diverge around the 3 year mark. As noted earlier, GP has some difficulty smoothing the curve around the 5 and 10 year, and the smoothness figure is much lower than IPQPI's figure. Also GP is not able to flatten the curve as it is approaching the 30 year mark (note the GP curve's slope is positive) . The fitness function loading on the terminal condition are not coming into play since the pricing errors are still too large, and that is what is driving the evolution of the population. The GP

Table 4.3: Comparison of Pricing Errors for 9/9/2009: GP and IPQPI.

Maturity	Actual Price	GP	IPQPI
9 / 9/2009	100.000000	0.0000	0.0000
9 /17/2009	99.998160	0.0622	0.0000
10/ 1/2009	99.994580	0.1361	0.0000
12/ 3/2009	99.969780	0.4648	0.0000
3/ 4/2010	99.899780	0.6923	0.0000
8/26/2010	99.626580	1.5788	0.0000
8/31/2011	100.181110	-0.2391	-0.0237
8/15/2012	100.993890	0.22846	-0.0226
8/31/2014	100.059050	25.7372	-0.3666
8/31/2016	99.605836	-25.0471	0.0242
8/15/2019	101.433760	-0.1494	0.5797
8/15/2039	102.961960	0.1494	0.0000
Ave Abs Error		4.540405	0.0847
Max Abs Error		25.7372	0.5797
Smoothness		2782.9754	5187.5767

system is facing an optimization problem that is far more complex and complicated than the one solved by the IPQPI system. The search space is wider, the constraints/fitness function are more complicated, and the evolutionary process itself is not guaranteed to follow the same path each time the system is run.

Table 4.4 shows the pricing results of the experiment using the 8/31/1981 data, and Figure 4.11 shows the syntax representation for the best fitting individual I was able to obtain. Recall that the 1981 data represents an inverted yield curve. Examining Table 4.4 reveals that GP has failed to produce an acceptable curve that fits this data set. The pricing errors are large, and the smoothness figure is less than one tenth of that of the 9/9/2009 data. Figure 4.12 shows the curve itself. Examining the curve one learns what is the reason for the low level of smoothness it exhibits. The curve makes a sharp turn around the 1 year mark, and there are some unexplained kinks in it at the 5 to 8 year maturities, as well as between the 10 and the 14 years. Note however that the terminal condition is satisfied to a higher degree than it is with the 2009 data. TheF_5 coefficient is about 23 times smaller in the 1981 data (see Table 4.5 than it is in the 2009 data. This is also evident in the Figure 4.12 as the curve is flat as it approaches the 30 year mark.

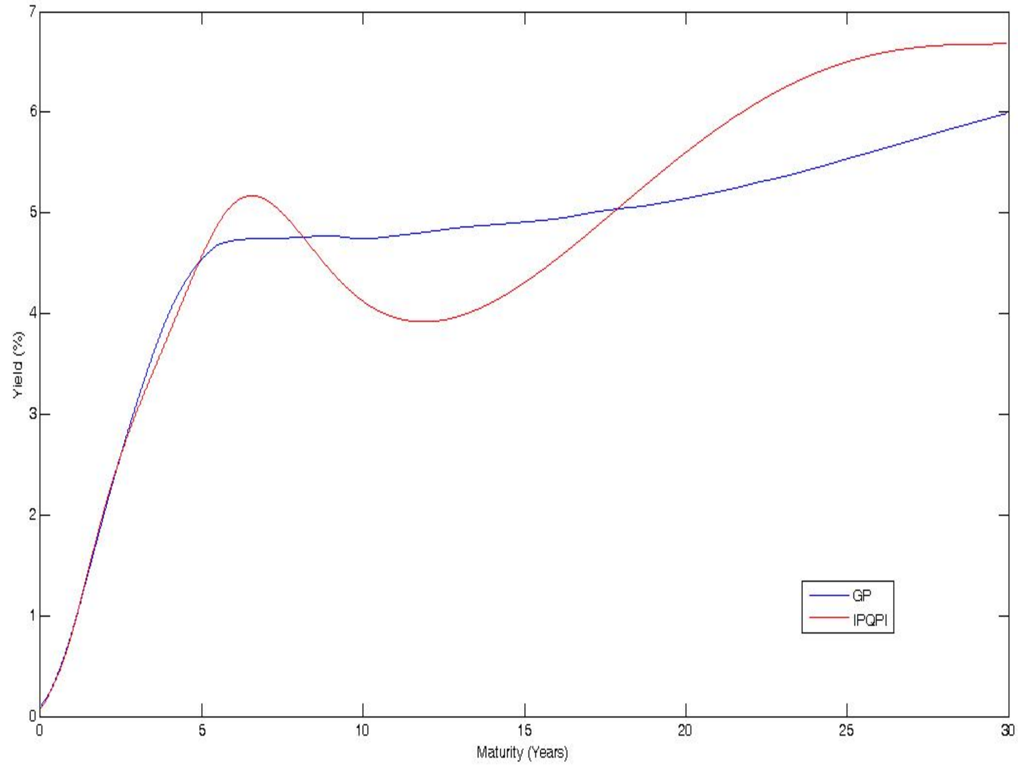

Figure 4.10: 9/9/2009 Forward Curves by GP and IPQPI.

Figure 4.13 shows the IPQPI and the GP curves overlaying, highlighting the differences between the curves. In Figure 4.14 I zoom on the early maturities. Notice that the IPQPI curve is accommodating the fluctuations in the yield in the short term effectively, while GP, which is focusing on minimizing pricing errors, is ignoring these fluctuations since they do not have a large enough impact on bill prices.

4.4 Summary

The GP experiment results were mixed at best. The method was not able to produce a smooth curve that also prices the securities with an acceptable level of pricing errors. This outcome could be the result of one of the following reasons:

Limited Computing Power and inadequate GP System During the experiment the GP

Table 4.4: Actual and GP-estimated bond prices for the yield curve on date 8/31/1981.

Maturity	Coupon	Actual Price	Yield	GP Est. Price	Penny Error
8/31/1981	0	100.0000	14.37	100.0000	0.0
9/3/1981	0	99.8775	14.53	99.841931	3.5569
9/10/1981	0	99.5867	14.91	99.560909	2.5791
10/1/1981	0	98.7247	14.91	98.690053	3.4647
11/27/1981	0	96.1989	16.28	96.207027	-0.8127
2/25/1982	0	92.2422	17.16	92.226174	1.6026
8/12/1982	0	85.6506	16.90	85.379924	27.0676
7/31/1983	15.875	99.806096	16.77	99.511201	29.4895
11/15/1984	16.0	103.75815	16.34	103.195439	56.2711
11/15/1986	13.875	97.009511	15.88	96.358619	65.0892
7/15/1988	14.0	94.663043	15.70	94.116912	54.6131
8/15/1991	14.875	97.896739	15.40	97.599552	29.7187
8/15/2001	13.375	90.519022	15.10	91.216724	-69.7702
5/15/2011	13.875	98.009511	14.74	98.589431	-57.9920

Table 4.5: Contribution to fitness for the 8/31/1981 data.

Fitness Argument	Score	Weight	Contribution to Fitness
F_1 Price in pennies	402.0274	1	402.0274
F_2 Bills Yields (BP)	330.6868	1	330.6868
F_3 Smoothness	0.0	10^6	0.0
F_4 Initial Condition	0.0003	10^3	0.3
F_5 Terminal Condition	0.0003	1	0.0003
F_6 Non negativity	0.0	0	0
F_7 Level different from IPQPI	0.0	0	0
F_8 Slope different from IPQPI	0.0	0	0

system generation processing speed slowed significantly after only a few hundreds of generations. This happened although I limited the depth level to 28. It appears the machine which was running the experiment was having a hard time completing the calculation in a reasonable amount of time. For example, when running the the 1981 data experiment, it took the system about 10 hours to run about 500 generations (between 1000 to 1500), where the speed at which the first 200 generations were run was about 10 seconds per generation. It may be the case that a better machine would be able to process the experiment better

Table 4.6: Comparison of Pricing Errors for 8/31/1981: GP and IPQPI.

Maturity	Actual Price	GP	IPQPI
8/31/1981	100.0000	0.0000	0.0000
9/3/1981	99.8775	3.5569	0.0000
9/10/1981	99.5867	2.5791	0.0000
10/1/1981	98.7247	3.4647	0.0000
11/27/1981	96.1989	-0.8127	0.0000
2/25/1982	92.2422	1.6026	0.0000
8/12/1982	85.6506	27.0676	0.0000
7/31/1983	99.806096	29.4895	0.2027
11/15/1984	103.75815	56.2711	0.2882
11/15/1986	97.009511	65.0892	0.0680
7/15/1988	94.663043	54.6131	0.1593
8/15/1991	97.896739	29.7187	0.4174
8/15/2001	90.519022	-69.7702	9.1591
5/15/2011	98.009511	-57.992	0.0000
Ave Abs Error		28.71624	0.7353
Max Abs Error		69.7702	9.1591
Smoothness		210.0438	4.2892

and achieve better results in a more reasonable time frame. Also, there are several other GP systems available to researchers. It is possible that the internal design of the GPLAB algorithms in MATLAB, which was used here, is geared towards resolving certain kinds of problems, but not others. MATLAB is an interpreted language and not compiled (such as C++, or FORTRAN), which makes it very expensive to run in terms of computing overhead costs and time of processing.

Misspecified Fitness Function The fitness function is the user's instructions to the GP system in its search for the optimal function. I have touched on it earlier in the chapter, but will do so again here - the fitness function produces one numerical value, which is a composition of several parts, each of which needs to be weighed, and added to the mix. The weights are determined by trial and error, since there are no rules to follow. The user would like to have the chosen function to satisfy *all* of the parts that make up the fitness function. A misspecified fitness function may lead the GP system down the "wrong path" of evolution. I suspect this was not the reason for the results since my system was not able to get to a fitness score that was low enough. Had the system got to a very low fitness score, and the

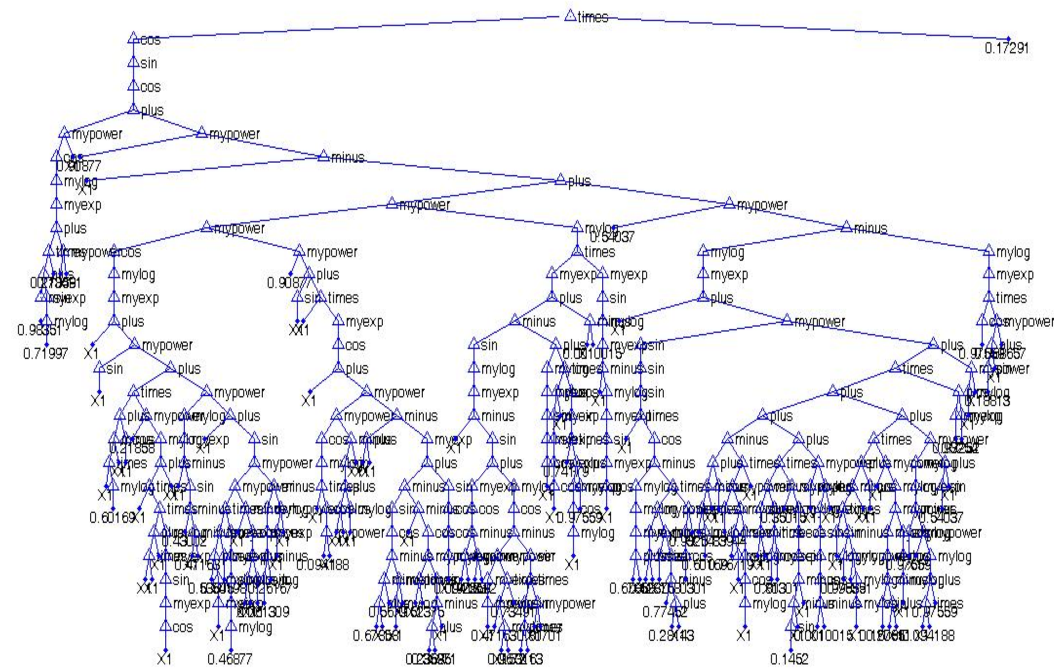

Figure 4.11: 8/31/1981 Yield Curve function produced by GP - syntax representation .

curve/function would still have the wrong properties, then a miss-specification of the fitness function would have been the prime suspect as to the reason. But my system was not able to get to that point to begin with, hence pointing the cause of the outcome to items 1 and 2 above.

Problem is just too complex It may be the case that the problem GP is facing is just too hard to solve in a limited number of generations. The unobserved nature of the spot curve, and the many forms a forward curve may take for each spot curve, combined with the large search space GP is facing is just too complicated for GP to solve within several hundreds of generations.

The IPQPI method presented in Chapter 3 however is preforming well. It is fixing the polynomials and is just searching for the optimal coefficients so the problem it faces

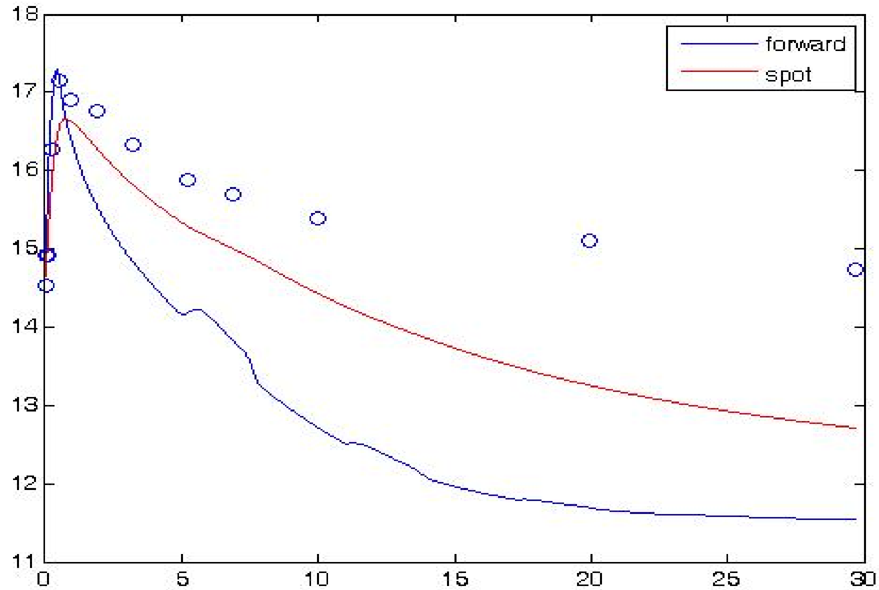

Figure 4.12: 8/31/1981 Yield and Forward Curves produced by GP .

is much narrower in scope. It produces a maximally smooth forward curve (within the family of polynomial functions) and it prices better than all the smooth methods it is being compared to. It does so while combining the stripping and smoothing routines into one comprehensive process. It is easy to program, intuitive, and can accommodate additional constraints fairly easy. The next steps in this avenue of research would be (1) designing a stabilization mechanism for the IPQPI method (to make the effect of a local bump in the data, local) and (2) Experiment with GP on different platforms.

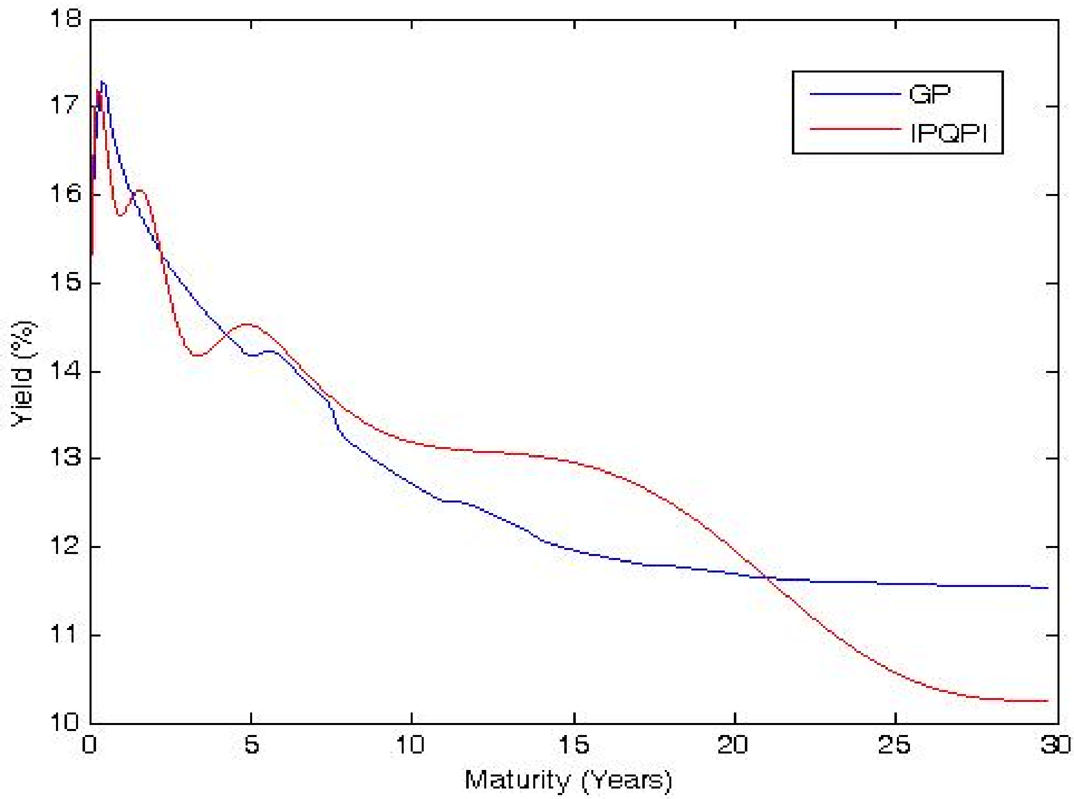

Figure 4.13: 8/31/1981 Forward Curves by GP and IPQPI.

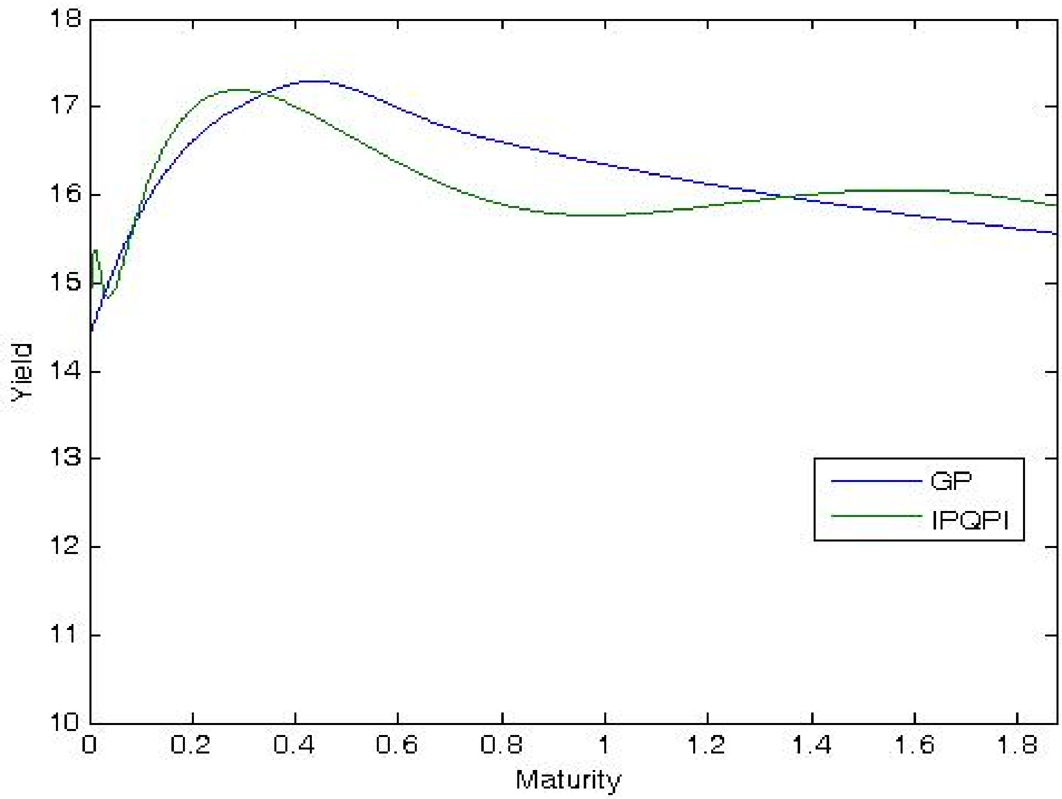

Figure 4.14: 8/31/1981 Forward Curves by GP and IPQPI - Early maturities.

REFERENCES

Adams, K., 2001. Smooth interpolation of zero curves. Algo Research Quarterly 4, 11–22. 1.3

Adams, K., van Deventer, D., 1994. Fitting yield curves and forward rate curves with maximum smoothness. Journal of Fixed Income 4(1), 52–62. 1.1

Black, F., D. E. . T. W., 1990. A one-factor model of interest rates and its applications to treasury bond options. Financial Analysts Journal 46, 33–39. 2.3

Brennan, M. J., Schwartz, E. S., 1979. A continuous-time approach to the pricing of bonds. Journal of Banking and Finance 3, 135–155. 2.2, 2.3

Campbell, J. Y., 1997. The econometrics of financial markets. Princeton University Press. 2.2

Cochrane, J. H., 2001. Asset Pricing. Princeton University Press. 1.2, 2.2

Cox, J. C., Ingersoll, Jonathan E., J., Ross, S. A., 1985. An intertemporal general equilibrium model of asset prices. Econometrica 53 (2), 363–384. 1.2, 2.2, 2.3

Culbertson, J. M., 1957. The term structure of interest rates. The Quarterly Journal of Economics 71 (4), 485–517. 2.1

Dahlquist, M., Svensson, L. E. O., 1996. Estimating the term structure of interest rates for monetary policy analysis. The Scandinavian Journal of Economics 98 (12), 163–183. 1.2

Dai, Q., Singleton, K. J., 2000. Specification analysis of affine term structure models. The Journal of Finance 55(5), 1943–1978. 2.2

Diebold, F. X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337–364. 1.2, 2.3

Duesenberry, J., 1958. Business Cycles and Econolnic Growth. McGraw-Hill. 2.1

Estrella, A., Mishkin, F. S., 1998. Predicting u.s. recessions: Financial variables as leading indicators. The Review of Economics and Statistics 80 (1), 45–61. 1.2

Fabozzi, F. J., 2000. Fixed income analysis. Frank J. Fabozzi and Associates. 1.2

Gimeno, R., Nave, J. M., Dec. 2006. Genetic algorithm estimation of interest rate term structure WP 0634 (0634). 2.3

Hagan, P. S., West, G., 2006. Interpolation methods for curve construction. Applied Mathematical Finance 13 (2), 89–129. 1.2, 2.3, 4.1

Heath, D., Jarrow, R., Morton, A., 1992. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. Econometrica 60 (1), 77–107. 2.3

Hicks, J., 1946. Value and capital; an inquiry into some fundamental principles of economic theory. Oxford University Press. 2.1

Ho, T. S. Y., Lee, S.-B., 1986. Term structure movements and pricing interest rate contingent claims. The Journal of Finance 41 (5), 1011–1029. 2.3, 2.3, 3

Hull, J., White, A., 1990. Pricing interest-rate-derivative securities. The Review of Financial Studies 3(4), 573–592. 2.2, 1

John C. Cox, Jonathan E. Ingersoll, J., Ross, S. A., 1981. A re-examination of traditional hypotheses about the term structure of interest rates. The Journal of Finance Vol. 36, No. 4, 769–799. 2.2

Jordan, J. V., Mansi, S. A., 2003. Term structure estimation from on-the-run treasuries. Journal of Banking & Finance 27. 2.2, 2.3, 2.3

Kamal, H. A., Eassa, M. H., 2002. Solving curve fitting problems using genetic programming. IEEE Melecon. 4.2.1

Keynes, J. M., 1936. The General Theory of Employment, Interest, and Money. Palgrave Macmillan. 2.1

Langetieg, T. C., 1980. A multivariate model of the term structure. The Journal of Finance 35 (1), 71–97. 2.2, 2.3

Lim, K. G., Xiao, Q., 2002. Computing maximum smoothness forward rate curves. Journanl of Statistics and Computing 12(3), 275–279. 1.2, 3.1, 3.5

Longstaff, F. A., 1990. Time varying term premia and traditional hypotheses about the term structure. The Journal of Finance 45 (4), 1307–1314. 2.2

Longstaff, F. A., Schwartz, E. S., 1992. Interest rate volatility and the term structure: A two-factor general equilibrium model. The Journal of Finance 47 (4), 1259–1282. 2.2

Lutz, F. A., 1940. The structure of interest rates. The Quarterly Journal of Economics 55 (1), 36–63. 2.1

Mansi, S. A., Phillips, J. H., 2001. Modeling the term structure from the on-the-run treasury yield curve. The Journal of Financial Research 24 (4), 545–564. 2.3

McCulloch, J. H., 1971. Measuring the term structure of interest rates. The Journal of Business 44 (1), 19–31. 2.3

McCulloch, J. H., 1975. The tax-adjusted yield curve. The Journal of Finance 30 (3), 811–830. 2.3, 1

Modigliani, F., Sutch, R., 1966. Innovations in interest rate policy. The American Economic Review 56 (1/2), 178–197. 2.1

Nelson, C. R., Siegel, A. F., 1987. Parsimonious modeling of yield curves. The Journal of Business 60 (4), 473–489. 2.3, 2.3

Poli, R., Langdon, W. B., McPhee, N. F., 2008. A Field Guide to Genetic Programming. http://www.gp-field-guide.org.uk. (document), 4.2, 4.1, 4.2, 4.3

Redlemen, R. J., 2004. Interpolating the term structure from par yield and swap curves. Journal of Fixed Income 13 (4), 89–97. 2.3

Roger J-B Wets, S. B. L. Y., 2002. Serious zero-curves. EpiSolutions Inc. 2.3

Roll, R., 1971. Investment diversification and bond maturity. The Journal of Finance 26 (1), 51–66.
URL http://www.jstor.org/stable/2325740 2.1

Sharpe, W. F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance, 19 (3), 425–442. 2.1

Shea, G. S., 1985. Interest rate term structure estimation with exponential splines: A note. The Journal of Finance 40 (1), 319–325. 2.3, 3

Shiller, J. Y. C. R. J., 1991. Yield spreads and interest rate movements: A bird's eye view. The Review of Economic Studies 58 (3), 495–514. 1.2

Silva, S., 2007. GPLAB A Genetic Programming Toolbox For Matlab. University of Coimbra, Portugal. 4.3

Svensson, L. E. O., 1995. Estimating forward interest rates with the extended nelson & siegel method. Sveriges Riksbank Quarterly Review 3, 13–26. 1.2, 2.3

Turan, B. G., Karagozoglu, A. k., 2000. Pricing eurodollar futures options using the edt term structure model: The effects of yield curve smoothing. The Journal of Futures Markets 20 (3), 293–306. 2.3

Vasicek, O. A., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 40, 319–325. 1.2, 2.2, 2.3

Vasicek, O. A., Fong, H. G., 1982. Term structure modeling using exponential splines. The Journal of Finance 37 (2), 339–348. 2.3

White, J. H. A., 1993. One-factor interest-rate models and the valuation of interest-rate derivative securities. The Journal of Financial and Quantitative Analysis 28 (2), 235–254. 2.3

BIOGRAPHICAL SKETCH

Yaniv Jerassy-Etzion

Yaniv Jerassy-Etzion was born in Israel in 1973. In the summer of 1991 he graduated from the Israeli Naval Academy and in 1993 became an officer in the Israeli Navy. In the summer of 2000 he completed his Bachelor of Science degree in Accounting and Economics at The Florida State University. Under the advisement of Dr. Paul Beaumont, he obtained both a Master of Science degree (Winter of 2001), and a Doctor of Philosophy degree in Economics (Spring of 2010) at The Florida State University.