PARAMETRIC VERSUS NON-PARAMETRIC BOND PRICING AND HEDGING MODELS

By

ARYASOMAYAJULA V. SEKHAR

Bachelor of Engineering
Delhi University
Delhi, India
1994

Post Graduate Diploma in Business Administration
S.P.Jain Institute of Management and Research
Bombay, India
1997

Submitted to the Faculty of the
Graduate College of the
Oklahoma State university
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2002

PARAMETRIC VERSUS NON-PARAMETRIC BOND

PRICING AND HEDGING MODELS

Thesis Approved:

Thesis Adviser

Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to thank my mother Usha Rani, father Narsimham Murty and fiancée Charu Varshney for the continued support and encouragement provided by them to complete this dissertation. I would also like to thank my sister Sridevi for all her moral support.

I would like to express my earnest gratitude to Dr. Tim Krehbiel for his guidance and encouragement in the completion of this dissertation. The credit for the accomplishment of this goal in time also goes to Dr. Krehbiel's investment of his precious time into this project. I would like to thank Dr. Robert Bliss for his generous help in providing the data sets. I would also like to thank Dr. Bliss for providing most of the research ideas that eventually led to this dissertation. I would like to express my thanks to Dr. Wade Brorsen for patiently listening to my problems with the implementation of most of the models and providing me with constructive solutions. I would like to thank Dr. Martin Hagan for providing me with all the help in the implementation of the Neural Networks. I would also like to thank Dr. Tom Gosnell and Dr. John Polonchek for their comments and continued guidance.

I would like to thank my colleagues Ramamurthy Dwivedula, Yong Hu and Steve Miller for their valuable suggestions and solutions to all the questions I had pertaining to software issues. Finally I would like to thank my colleagues in the Department of Finance Anjo Koeter, Jeff Whitworth, and Mounther Barakat for providing such a conducive and friendly environment in the Finance Ph.D program.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure | Page

CHAPTER 1

Introduction

The set of prices of default free bonds with maturities ranging from shortest to longest defines the term structure of interest rates. Modelling default free bond prices / term structure of interest rates has been approached in many ways. An exhaustive catalogue of the approaches has become virtually impossible. The modelling of the term structure of interest rates is important for both practitioners and academicians. A practitioner is concerned with the model's tractability, ability to price bonds, and other interest rate derivatives and last but not least, its ability to hedge the risks caused by the underlying risk factors. Whereas academic literature is concerned with understanding the factors that affect the term structure of interest rates and how to efficiently model the term structure using these factors. Financial institutions and other market participants utilize term structure models to manage their exposures to fluctuating interest rates and other risk factors, and to price fixed income securities (bonds, mortgage backed securities etc.).

The early theories of the term structure of interest rates include the pure expectations hypothesis (Lutz (1940)), the liquidity-premium theory (Hicks (1946)), and the market segmentation hypothesis (Culbertson (1957)).

The expectations theory laid more stress on the expectations of future interest rates. The state variables of the expectations theory are the spot short term rate and expectations of the spot short term rate at future dates. Rational risk neutral decision makers operating in perfect debt markets adjust borrowing and lending plans until all debt instruments outstanding have identical total returns over any given holding period, independent of their final maturity. CIR (1985) show that the implication of the general expectations hypothesis that all expected returns for all holding periods are equal, is mathematically inconsistent with Jensen's inequality[1].

As a result a restricted version of expectations theory was postulated. The Local Expectations Hypothesis states that the expected holding period returns are equal only for one specific holding period. The natural choice of holding period was the next basic (i.e. shortest) interval. The local expectations hypothesis failed to take into consideration risk aversion as it assumes local risk neutrality and hence no term premiums.

The liquidity-premium theory introduces market participant's risk preferences. The liquidity premium theory states that there is an imbalance (excess supply) of long maturity bonds. This in turn implies that the lenders must be paid positive liquidity premium to entice them to lend long term. In other words the theory implicitly assumes that all lenders have prevailingly short lending horizons and borrowers prevailingly long borrowing horizons. The liquidity theory implies that the slope of term structure is positive and is determined by the risk tolerance and excess supply of bonds. As a consequence the theory fails to explain inverted term structure.

The market segmentation theory required information about the excess demand for bonds of each maturity segment. The preferred habitat theory (a type of market segmentation theory) first set forth by Modigliani and Sutch (1966) (hereafter MS) is essentially an adaptation of the expectations theory of the term structure of interest rates to a world in which

- future rates are uncertain
- all market participants have definite preferences about the length of time for which they want to lend or borrow money (hence the name preferred maturity habitat)
- both the borrowers and lenders generally exhibit risk aversion and hence all other

[1]Jensen's inequality implies that a function of the expected value of a random variable is less than or equal to the expected value of the function of the random variable.

things being equal, would prefer to match maturities in their portfolios to their chosen horizon so as to be certain about the return or cost.

The MS model asserts that the expected change in future rates may be expressed as the difference between the current rate and a weighted average of past rates. Hence one of the testable implications of the MS model was that a distributed lag on past short term rates is an indicator of expected future short term rates. Hamburger and Latta (1969) (hereafter HLT) produce evidence inconsistent with this implication of the MS model. HLT argue that it is possible that the periods used to test the MS hypothesis- the post-Accord and in particular the Operation Twist period may contain some special factors not allowed for in the model[2]. HLT further argue unless these factors can be taken into account or supporting evidence can be presented for other periods, the case for including the lagged rates in the term structure equation as suggested by the preferred habitat theory of MS would appear to be weak.

Each of the theories, the expectations hypothesis, the liquidity premium theory, and the market segmentation hypothesis were not originally designed as dynamic pricing or hedging models. This led to the general equilibrium approach which included both the expectations of future rates and the risk preferences of the market participants. The general equilibrium approach to modelling the term structure of interest rates in a continuous time framework was pioneered by Cox, Ingersoll and Ross (1985) (hereafter CIR). The general equilibrium approach begins from assumptions concerning the dynamic properties of production processes and specification of decision maker's utility functions. Expected utility of wealth maximization by rational agents produces a relationship between the dynamics of the production process and the dynamics of the instantaneous short-term interest rate. Given the dynamics of the instantaneous short-term interest rate, bond prices must satisfy

[2]Launched at the beginning of 1961 by the incoming Kennedy administration, an attempt to twist the maturity structure of interest rates by raising yields on securities with short term to maturity while simultaneously lowering, or at least holding the line on, long term rates.

a partial differential equation to preclude arbitrage opportunities.

The market for interest rate derivatives developed significantly over the last two decades. In 1988 the notional principal outstanding for interest rate options (calls and puts) traded over the counter was \$ 279.2 billion whereas the trading volume increased to \$ 3277.8 billion in 1996. Similarly the trading volume of the interest rate options (caps, collars, floors, and swaptions) traded over the counter in 1988 was \$ 327.3 billion and increased to \$ 4190.1 billion in 1996[3]. With the substantial growth rate in the market for interest rate derivatives, people started addressing the issue of pricing of interest rate derivatives. The problem was addressed in two stages, in the first stage the bond prices of different maturities are determined and in the second stage the prices of interest rate derivatives are determined. Since the second stage is contingent on the first one, if the prices of the bonds are not determined accurately the pricing errors would compound during the second stage of pricing of interest rate derivatives. The general equilibrium models did not price the current set of bonds exactly which led to greater mispricing when applied to pricing of interest rate sensitive contingent claims.

Ho and Lee (1986) (hereafter HL) were among the first to develop a model that utilized the present term structure as an input to the model and hence was consistent with the term structure used to parameterize the model. The HL model relates current bond prices to an arbitrage free risk neutral stochastic process to determine the instantaneous short rate. The risk neutral stochastic process is useful for valuing other interest rate derivatives. Generalizing on the work by Ho and Lee (1986), Heath, Jarrow, and Morton (1992) (hereafter HJM) take a related approach. This approach models the evolution of the entire term structure. The authors show that given an initial term structure, pricing can proceed once the structure of all forward rate volatilities is provided.

In both the equilibrium models and the HJM models the derivation of the pricing formula is contingent on the parametric specification of either the underlying factors such as

[3]Source: Bank for International Settlements

the short term rate or the forward rates. A misspecification of the stochastic process for the above stated factors will lead to systematic pricing and hedging errors for derivative securities linked to these factors. This implies that the success or failure of these approaches in modelling the term structure is closely tied to their ability to capture the dynamics of the underlying risk factors.

In such parametric models it is quite difficult to justify the choice of one parametric specification over the other. This leads to a very serious problem of misspecification. Even if a model fits the term structure of interest rates well in sample, it may produce significant out of sample errors. Aït-Sahalia (1996) rejects "...every parametric model of the spot rate (previously) proposed in the literature" by comparing the marginal density implied by each model with that estimated from the data. Backus, Foresi, and Zin (1995) and Canabarro (1995) further show that the misspecification of the underlying interest rate model can lead to serious pricing and hedging errors.

In this research, a relatively new approach (non-parametric pricing technique using neural networks) is taken in which the data is allowed to determine both the dynamics of the underlying factors and the pricing relationship with minimal restrictions on the underlying factors and the pricing model. This approach again has the advantage that it prices the bonds at the current time exactly because the prices of the bonds today are used as inputs to the network. At the same time this approach does not impose the parametric restrictions as imposed by the HJM models.

Though an exhaustive categorization of the different approaches to the modelling of the terms structure of interest rates is virtually impossible, the term structure models can be broadly categorized as equilibrium, HJM, and network models. Despite the importance of term structure modelling, there is a lack of consensus on what approach would be ideal from a pricing point of view. In other words what approach should one pick for an application such as the pricing of bonds? The study addresses this issue and is concerned with the application of the three approaches to pricing and hedging zero-coupon, default free bonds.

1. Equilibrium Models

2. Heath Jarrow and Morton Models and

3. Network Models.

The CIR model was the first equilibrium model to be developed. The CIR model is a one factor model and all one factor models impose the restriction that the bond price changes of different maturity bonds are perfectly correlated. Observation of historical price changes are inconsistent with this implication. The one factor models also have a drawback that they are able to explain only a few shapes of the yield curve (downward sloping, upward sloping, and one hump).

In light of the one factor model's weaknesses, the study also considers a two factor general equilibrium model, the Longstaff and Schwartz (1992) (hereafter LS) model. The two factor LS model considers the short term rate and its volatility as the two underlying risk factors. The importance of second factor in the LS model, the short rate's volatility, is identified in a study by Chan, Karolyi, Longstaff, and Sanders (1992) (hereafter CKLS). CKLS show that the volatility of the short rate is not constant and is dependent on the level of short term rate. The resulting bond pricing relationships associated with CIR and LS models are of closed form.

Given that equilibrium models are unable to price the bonds of different maturities at the current date exactly, the HJM models are considered. The HJM model in the most general form takes the current term structure as an input. The resulting HJM model prices are fully consistent with the current term structure. Due to the intractability of the general HJM models Ritchken and Sankarasubramanian (1995) (hereafter RS) imposed restrictions on the forward rate volatility that make the model more tractable. With the restrictions imposed by RS two different kinds of models are possible. The generalized Vasicek (hereafter HJM-GV) model assumes a constant short rate volatility and the other model assumes variable short rate volatility (hereafter HJM-RS). The findings of CKLS motivate inclusion of both types of HJM in this study. The study includes both the HJM-GV and HJM-RS models. The HJM-RS model allows short rate volatility to be a function of the level of short rates. In other words any model that has a variable short rate volatility falls under the HJM-RS model category.

Both the equilibrium models and HJM models impose some kind of parametric restrictions on the underlying risk factors such as the short term rate, forward rate etc. Considering the fact that very little is known about the identity and dynamic structure of the actual underlying factors, the study considers the non-parametric approach using network models.

In the category of the network models the most commonly used and standard network model is the multi layer perceptron (hereafter MLP). Cybenko (1988) and Hornik (1989) demonstrate that one-hidden-layer MLPs can represent to arbitrary precision most classes of linear and non-linear continuous functions with bounded inputs and outputs. For an application to the term structure modelling, bond prices are non-linear functions of the risk factors and the inputs (the risk factors) and outputs (bond prices) are bounded. Results of the Cybenko and Hornik study suggest that an MLP should be able to model the term structure to arbitrary precision. Girosi and Poggio (1990) (hereafter GP) show that radial basis functions (hereafter RBF) have the "best" approximation property i.e. there is always a choice for the parameters that is better than any other possible choice. This property is a strength of RBF network because if an RBF with a particular training algorithm does not perform well in an application, there is always a way to improve the performance of the RBF. This gives an opportunity to keep improving the performance of the RBF network till one achieves the desired level of accuracy. This property in contrast is not shared by the MLP network. As a result the study also includes the RBF network in the study to compare the performance of the MLP versus RBF network for a bond pricing and hedging application.

In the network models the underlying factors are defined as the inputs and bond prices are the outputs. After sufficient training of the network, the network can be used to price the bonds for an out-of-sample period analogous to using a pricing formula in a parametric procedure.

A term structure model is assessed not only by its ability to price bonds but also by its ability to hedge the underlying risk factors. In other words can a risk manager use the model to minimize the risk associated with the movements (parallel and non-parallel) in the yield curve? Traditionally, hedging strategies are based on simple duration or convexity. Hedging strategies of this sort are valid if and only if the movements in the term structure are parallel and the volatilities of different maturity interest rates are of similar magnitude.

Another traditional approach to the hedging of a fixed income portfolio utilizes principal components analysis (hereafter PCA). PCA is one form of broader class of models that fall under the dimension reduction approach. The PCA approach essentially assumes that changes in all interest rates in the term structure are generated from a multivariate normal distribution.

PCA analysis extracts the common factors, principal components, that explain the greatest portion of the total volatility of the term structure's rates from the variance-covariance matrix of rate changes. The resulting model describes the change in the rate of a specific maturity as a linear combination of the principal components and the principal components' coefficients for that maturity.

Though there is extensive literature on each of the parametric pricing approaches there is little literature devoted to the relative pricing and hedge performance.

The objectives of this study are:

- Determine the relative performance of the parametric and non-parametric bond pricing models in their ability to price zero coupon bonds.

- Determine if the parametric models and the non-parametric models improve the hedging effectiveness over the traditional duration and principal components based hedge positions.

The study tests pricing and hedge effectiveness using a rolling window approach. Using monthly data on zero coupon bond prices from 1983 – 1995 each model is applied to price zero coupon bonds in the first month of 1996. Then the window is rolled forward by a month i.e. the first month of 1996 is added in the estimation window, the first month of 1983 is dropped and model prices are estimated for the second month of 1996. The rolling window approach produces sets of bond prices corresponding to the term structure for that month. The approach is repeated till last month of 1999 i.e. forty-eight windows.

Two portfolios are used to test the hedging effectiveness, the portfolio that is to be hedged and the portfolio that is used to hedge. The weights of the portfolio to be hedged are known but the weights of the hedge portfolio for the first month of 1996 are determined by estimating model specific hedge portfolio weights using the data from February 1983 through the January 1996 estimation period. Then by looking at the actual prices of the securities in the second month of 1996, both the hedge portfolio and the portfolio to be hedged are repriced. Ideally the return on the hedge portfolio should be the same as the return on the portfolio to be hedged. The difference in the returns of the two portfolios is defined as the hedging error. The process is repeated using the rolling window approach till last month of 1999. Hence in the hedging applications the number of windows over which the hedge error is measured is forty-seven as opposed to forty-eight in the pricing applications. This approach is applied for each model throughout the sample period. The summary statistics such as mean and variability of the hedging errors over the forty-seven months are used to test relative hedging effectiveness.

The study finds that in terms of pricing of zero coupon bonds the non-parametric pricing approach using multi layer perceptron produces smaller pricing errors than other models considered. This has important implications with regards to the pricing applications of interest rate sensitive securities especially bonds. The superior performance of the MLP network implies that other models are misspecified in terms of parametric restrictions, stochastic assumptions, volatility structure, etc. Errors due to model misspecification are compounded when a misspecified model is used to price interest rate options and other derivative contracts.

The study produces evidence that the HJM models (HJM-GV and HJM-RS) produce smaller pricing errors than the equilibrium models (CIR and LS). The HJM models use the information about the whole yield curve (interest rates of different maturities) during the estimation period whereas the equilibrium models use information about only one point of the yield curve (the short rate) during the estimation period. The movements in the yield curve are not perfectly correlated and hence the HJM models that use the information on the whole yield curve are able to capture the imperfect correlation amongst the movements of rates of different maturities better than the equilibrium models.

The study also produces evidence that amongst the HJM models the HJM-RS model produces smaller pricing errors than the HJM-GV model. The evidence presented in this study supports the contention that a model that allows for varying spot rate volatility is more suitable for the pricing of bonds than a model that specifies the spot rate volatility as constant.

In the hedging applications the study finds that the MLP network and the HJM-RS model with varying spot rate volatility perform better than the CIR model, hedge positions based on duration, and hedge positions based on principal components analysis. The study also highlights the fact that the CIR model is quite effective in hedging a portfolio that has a shorter duration but is ineffective in hedging a portfolio that has a longer duration. This result is consistent with the observed larger pricing errors for longer maturity zero coupon bonds produced by the CIR model.

The evaluation of relative hedging performance has important implications from a practitioner's point of view. The HJM-RS (with varying spot rate volatility) model and the MLP network are more difficult to implement than the simple Principal Components based hedge positions. The study finds significant evidence that the extra cost in implementation of the HJM-RS and network models is worth the effort (the benefits in terms of better pricing and hedging performance outweigh the costs of their implementation). This is demonstrated by the lesser mean and variance of the hedge errors produced by HJM-RS model and MLP network when compared to the other models. The results illustrate that the hedge performance of principal component hedge positions is adversely affected by outliers in the estimation sample. The hedge effectiveness of parametric and network models do not deteriorate with the presence of outliers in the estimation sample to the same extent as the hedge effectiveness of the principal component method.

Though the duration and principal component hedges are easy to implement, the study identifies their weaknesses by examining a scenario where the yield curve was subject to a significant shock that produced simultaneous parallel shift and change in the curvature of the term structure. During the fall of 1998 Russia defaulted on its sovereign debt and this caused a decrease in the yields of U.S. Treasury bonds of longer maturity. Hedge performance of each model is examined surrounding the Russian debt crisis during the fall of 1998.

Overall the HJM-RS model and the MLP network produce smaller pricing and hedging errors than other models. The inputs provided to the MLP network are identical to the inputs required by HJM-RS model. This implies that the network has been able to learn the relationship among the risk factors of the HJM model and the term structure of interest rates better than the relationship specified by the HJM model. This ability of the network models to uncover uncertain relationships as demonstrated by this study is one of the major advantages of network models. Till date in the financial markets there are many assets whose relationship with the underlying risk factors is yet not known exactly. Neural networks are useful in such circumstances.

The MLP's smaller pricing errors relative to those produced by HJM-RS implies that the HJM-RS model can be improved. Approaches to improve the HJM-RS model identified by this study include a relaxation of the assumption that the errors in the measurement of the yields are independent and identically, and inclusion of another stochastic factor besides the single stochastic short rate.

CHAPTER 2

Literature Review

One of the early theories of the term structure of interest rates is the expectations hypothesis. The expectations hypothesis would include any model that implies that the expectations play a pivotal role in explaining the relation between yields on bonds of different maturities in an uncertain economy. The expectation hypothesis states that under rational behavior, perfect markets, and negligible transaction costs, all debt instruments outstanding must have identical total returns over any given holding period, irrespective of their final maturity. The implications of the expectations hypothesis as identified by Modigliani and Sutch (1967) are as follows

- At any point in time there is a unique relationship between the yield of a security with m periods to maturity and the expected one-period yield in each of the following m periods.
- In the case of long term bonds with a market value close to par the relation between the yield and the expected future short rates can be closely approximated by a simple average of these m future short rates.
- If the expected future short rates are known with certainty, then in frictionless markets there would be a unique current and future structure of interest rates consistent with rational market behavior.
- All market participants including borrowers and lenders irrespective of their intended borrowing/investment horizon would be indifferent to the maturity structure of their assets or liabilities.

Cox Ingersoll and Ross (1985)(hereafter CIR) show that the implication of the expectation hypothesis that all expected returns for all holding periods are equal is mathematically inconsistent with Jensen's inequality. This led to the local expectations hypothesis that implied that the expected holding period returns are equal only for one specific period and the natural choice for the holding period was the next short period. Even the local expectations hypothesis suffered from the flaw that it failed to take into account the risk aversion of market participants.

Another theory proposed by Hicks (1946) restricted the expectations hypothesis and was called the liquidity-premium theory. This theory places more importance on the risk preferences of the market participants. This states that to overcome liquidity preference the term premia must be positive and increase with maturity in a monotonically non-decreasing manner. This implication of the liquidity premium theory is inconsistent with an inverted term structure.

In contrast the unrestricted theory proposed by Modigliani and Sutch (1966) (hereafter MS) was the preferred-habitat theory that interprets term premia as habitat-displacement allowances. This theory needs unobservable information about the distribution of borrower and lender habitats before it can restrict the sign or magnitude of term premia in different maturity sectors. This theory falls under a broader category of market segmentation hypothesis initiated by Culbertson (1957). The market segmentation theory maintains that the term structure of interest rates reflects the relative demands of different investors with a variety of relative inflexible portfolio preferences. The MS model implies that the expected change in future rates may be expressed as the difference between the current rate and a weighted average of past rates. Hence one of the testable implications of the MS model was that a distributed lag on past short term rates is an indicator of expected future short term rates. Hamburger and Latta (1969) (hereafter HLT) produce empirical evidence that is inconsistent with this implication of the MS model. HLT use monthly and quarterly observations of the yields on three month Treasury bills and long term U.S. Government bonds for the period 1951-1965 to compare the MS model with that of Wood's model (1964). The Wood model assumes that at any point in time the best estimate of next period's short term rate is the current rate, past rates are irrelevant. HLT argue that it is possible that the periods used to test the MS hypothesis- the post-Accord and in particular the Operation Twist period may contain some special factors not allowed for in the model. HLT further argue that unless these factors can be taken into account or supporting evidence can be presented for other periods, the case for including the lagged rates in the term structure equation as suggested by the preferred habitat theory would appear to be weak.

The expectations hypothesis, liquidity premium, and market segmentation hypotheses were each found lacking for the purpose of pricing and hedging bond portfolios. The inability of each of these theories to explain the term structure of interest rates led to the general equilibrium approach pioneered by Cox, Ingersoll, and Ross (1985) (hereafter CIR). This approach encompasses all the previous theories of the term structure and argues that expectations, risk aversion, investment alternatives, and preferences about the timing of consumption all play a role in determining the term structure of interest rates. At the same time CIR model is free from the flaws that the previous theories of term structure had i.e. it does not impose risk neutrality on the market participants and considers risk aversion of the market participants and investment alternatives in deriving the bond pricing relationship.

Affine models are a subset of the equilibrium term structure models. The CIR model is an example of an affine model. A term structure model is said to be affine if the logarithm of the bond yields are linear or affine in terms of the model's state variables. Brown and Schaefer (1991) and Duffie and Kan (1993) clarify the assumptions necessary for an affine yield model. Duffie and Kan show that the risk adjusted drift of the short term rate, the expected change in the short rate less the covariance of the short rate with the stochastic discount factor, and the variance of the short term rate must both be affine to produce an affine yield model. The models such as Vasicek (1977), CIR, Pearson, and Sun (1994), LS (1992), and Hull, and White (1993, 1996) satisfy these requirements, but some other continuous time models such as that of Brennan and Schwartz (1979) do not. Affine models were first investigated by Brown and Schaefer (1994). Duffie and Kan (1994) developed a general theory and showed how the state variables in an affine model can be re-interpreted as pure discount bond yields.

The market for interest rate derivatives developed significantly over the last two decades. In 1988 the notional principal outstanding for interest rate options (calls and puts) traded over the counter was $ 279.2 billion whereas the trading volume increased to $ 3277.8 billion in 1996. Similarly the trading volume of the interest rate options (caps, collars, floors, and swaptions) traded over the counter in 1988 was $ 327.3 billion and increased to $ 4190.1 billion in 1996[3]. With the substantial growth rate in the market for interest rate derivatives, people started addressing the issue of pricing of interest rate derivatives. The problem was addressed in two stages, in the first stage the bond prices of different maturities are determined and in the second stage the prices of interest rate derivatives are determined. Since the second stage is contingent on the first one, if the prices of the bonds are not determined accurately the pricing errors would compound during the second stage of pricing of interest rate derivatives. The general equilibrium models did not price the current set of bonds exactly which led to greater mispricing when applied to pricing of interest rate sensitive contingent claims.

Given the simplicity of a one factor model and the flaw of the equilibrium models that they do not price the current set of bonds exactly, a new approach using the binomial tree was developed which took the current term structure as given. Black, Derman, and Toy (1990) (hereafter BDT) developed a single factor model using the information about the current term structure and estimated volatilities to construct a tree of possible future short rates. This tree in turn can be used to value other interest rate derivatives. BDT start by assuming that changes in all bond yields are perfectly correlated, expected returns on all securities over one-period are equal, short rates at any time are lognormally distributed, and

[3]Source: Bank for International Settlements

there are no taxes or trading costs. This is in contrast with the affine yield approach that assumes that $(1 + y(t, 1))$ is conditionally lognormal and not the one-period yield, $y(t, 1)$ (yield on a one-period bond measured at time t). The single factor underlying the BDT model is the short rate, the annualized one-period interest rate. The model takes as inputs an array of yields on zero-coupon Treasury bonds for different maturities and an array of the estimated yield volatilities for the same bonds. BDT call the first array the current yield curve and the second array the current volatility curve. BDT develop a tree of future short rates consistent with the current yield curve, the current volatility curve and the absence of arbitrage opportunities. In the BDT model choosing the yield curve and volatility curve implies the cap curve. A cap curve gives for each maturity, the price of an at-the-money differential cap[4]. So the BDT model had the drawback that the cap curve implied by the BDT model was not consistent with the observed cap curve. BDT are able to match the outputs, yield curve, and volatility curve but not the cap curve.

The inconsistency of the cap curve implied by the BDT model with the observed cap curve led to the Black and Karasinki (1991) (hereafter BK) model. BK also assume that the one-period yield $y(t, 1)$ is conditionally lognormal and develop a similar approach to pricing of interest rate sensitive securities. BK model by construction matches all the outputs, the yield curve, the volatility curve, and the cap curve with the observable market prices.

Both the BDT and BK models can be effectively used to price interest rate derivatives but not zero coupon bonds as the current set of zero coupon prices is an input to both the models. Both the BDT and BK models are single factor models and suffer from the same drawback as all other single factor models i.e. they assume that the changes in the prices of bonds of different maturities are perfectly correlated.

Ho and Lee (1986) (hereafter HL) were among the first to suggest the application of the binomial tree approach to pricing of interest rate sensitive securities. BDT and BK

[4]A differential cap pays at a rate equal to the difference (if positive) between the short rate and the strike price. For a particular maturity, an at-the-money cap has a strike equal to the forward rate for that maturity.

were inspired by the same motivation as HL in their objective of pricing of interest rate sensitive securities. HL assume a discrete trading economy i.e. the market clears at discrete points in time separated by regular intervals. The authors also assume that the market is frictionless, the bond market is complete and the short term rate follows an arithmetic brownian motion with a constant volatility and a deterministic time-varying drift. In this economy HL assume that the zero coupon bond price curve fluctuates randomly over time according to a binomial process. HL model was again a single factor model and so implied that the returns on bonds of different maturities are perfectly correlated. HL model was originally presented as a whole yield curve model (a model that uses the information on the whole yield curve at current time).

The Heath Jarrow and Morton (1992) (hereafter HJM) model is a direct descendent of the HL model. HJM generalized on the work of Ho and Lee and developed a model in which the entire set of forward rates evolves simultaneously with respect to a set of volatility curves. The HJM framework encompasses a wide variety of term structure models. Effectively the framework can include any term structure model that has a continuous forward rate curve. HJM started off by taking the current term structure as given and then addressed the issue of pricing of interest rate contingent claims[5].

Another widely accepted and more recent approach is presented in the paper by Brace, Gatarek, and Musiela (1997) (hereafter BGM). BGM analyze a class of term structure models with volatility of lognormal type that falls under the general HJM framework. Under the HJM framework modeling lognormally distributed interest rates avoids the problem of negative interest rates but these rates explode (interest rates tend to infinity) with positive probability implying zero bond prices and hence arbitrage opportunities. The corresponding forward rates in the BGM framework do not explode, and are positive and mean revert-

[5]The study uses the restricted version of HJM with restrictions imposed by Ritchken and Sankarasubramanian (1995) and the empirical procedure of how the model can be applied to the pricing of bonds is explained in section 4.3 in detail.

ing. BGM argue that under an arbitrage-free setting forward rates over consecutive time intervals are related to one another and cannot all be lognormal under one arbitrage-free measure. But each forward rate is lognormal under the forward arbitrage-free measure rather than under one (spot) arbitrage-free measure.

The BGM model provides justification for use of the Black futures formula to price caps and floors. Since in most markets caps and floors form the largest component of an average swap derivatives book, BGM approach has been applied quite extensively. A cap/floor is a strip of caplets/floorlets each of which is a call/put option on a forward rate. The BGM approach also takes the current term structure as given and hence cannot be applied to the pricing of zero coupon bonds at a certain date.

Both the equilibrium models and HJM models fall under the category of parametric pricing models. In such models, often there is no economic rationale for the choice of one parametric specification over the other. This can lead to a very serious problem of misspecification. Even if a model fits the term structure of interest rates well in sample, it may produce significant out of sample errors. Aït-Sahalia (1996) rejects "...every parametric model of the spot rate (previously) proposed in the literature" by comparing the marginal density implied by each model with that estimated from the data. Backus, Foresi, and Zin (1995) and Canabarro (1995) further show that the misspecification of the underlying interest rate model can lead to serious pricing and hedging errors.

Given the fact that very little is known about the validity of the parametric restrictions imposed by the parametric models on the underlying risk factors, a non-parametric approach to the modelling of the term structure of interest rates is the next logical step.

The non-parametric approach using neural networks has been applied to many securities. Kryzanowski, Galler, and Wright (1993) (hereafter KGW) and Zirilli (1997) have shown that network models can recognize stocks providing superior future returns and the ones providing inferior future returns. KGW use a Boltzmann Machine for sorting stocks. A Boltzmann Machine is an artificial neural network that uses simulated annealing (a technique of stochastic optimization) to set the states of the neurons during both the weight-learning and function-computing stages of its operations. The network employed by KGW learns the relationship between a company's stock return one year in the future and the most recent four years of the firm's financial data, its industry and different macroeconomic factors. Refenes, Zapranis, and Francis (1995) (hereafter RZF) and Bansal and Viswanathan (1993) explore a dynamic version of the arbitrage pricing model (APT). RZF use a multi layer perceptron and the model training algorithm used is the regular backpropagation. RZF show that even simple neural learning procedures such as the backpropagation algorithm far outperform current 'best practice' in a typical application for stock ranking within the framework of the arbitrage pricing model. The network's smooth interpolation properties allow neural models to fit well in sample and produce small out of sample errors.

Neural networks have also been recently applied to different types of problems in money management. Both Daiwa Securities Co. and NEC corporation apply neural networks software to recognize price chart patterns for Tokyo Stock Exchange listed stocks. According to Loofbourrow and Loofbourrow (1993) the best known trading system user is Brad Lewis of Fidelity Investments. Lewis has developed neural networks to manage $2 billion of Fidelity's funds in seven portfolios.

Not surprisingly neural networks have also been applied to predict exchange rates. Mehta (1995) discusses the applicability of neural networks to foreign exchange forecasts and concludes that neural networks are currently the best problem solving tool available for non-linear time series. Refenes and Zaidi (1995) state that, when applied to historical data, neural networks outperform moving averages and mean value based forecasts for U.S. Dollar and German Deutsche Mark exchange rate predictions. Neely, Weller, and Dittmar (1997) implement genetic programming, a non-linear technique in the same family as neural network approaches to determine technical trading rules for foreign exchange trading.

Given the versatility of neural networks they have also been applied to prediction of bond ratings and credit scoring. Dutta and Shekhar (1988), Moody and Utans (1995), and

Singleton and Surkan (1991, 1995) apply neural networks to predict corporate bond ratings. Jensen (1992) showed how neural networks can be applied to credit scoring.

Swanson and White (1995) use different techniques including neural networks to examine the predictive power of forward interest rates. They use criteria such as forecast mean square error, forecast direction accuracy, and forecast-based trading system profitability. They find that the forward premium has predictive power to forecast the sign of future changes in the interest rates. They compare linear models such as Mishkin's (1988) to the neural networks. Though the study gives information about the predictive power of forward rates, it does not provide any information about an ideal approach to the pricing and hedging of zero coupon bonds.

Though there have been studies that apply neural networks to many financial applications, neural networks have not yet been applied to the pricing and hedging of fixed income securities. The relative pricing and hedging performance of neural network, equilibrium and HJM models is an empirical question.

This study compares the price and hedge performance of the three categories of term structure models, the equilibrium models, the HJM models, and the network models. The equilibrium models are considered since these are one of the first approaches to the modelling of the term structure of interest rates and are widely used by researchers and practitioners. One of the primary drawbacks of the equilibrium models was found to be the inability of equilibrium models to accurately price zero coupon bonds. This drawback of the equilibrium models led to another approach to the modelling of interest rates in which the current term structure was an input to the model ensuring that the bonds at the current time are exactly priced. Hence the relative performance of HJM models with respect to the equilibrium models is of interest to both practitioners and academics. Both the equilibrium models and the HJM models impose parametric restrictions on the underlying risk factors and little is known about the validity of these restrictions given the observed market data. Since very little is known about the misspecification of the underlying risk factors in any parametric models the class of network models is the third category of models considered in the study. The study uses the two standard neural networks (MLP and RBF) to price and hedge zero coupon bonds. The remainder of this chapter presents a review of the literature relevant to each type of model.

2.1 The Equilibrium Models

The equilibrium approach was triggered by CIR (1985). The model has been extensively applied in practise due to its tractability (easily implementable as it gives closed form solutions for bond prices). Testing the CIR model has generated a substantial literature.

For instance, Brown and Dybvig (1986) (hereafter BD) empirically test the CIR model using Treasury Bills, Bonds and Notes. BD use the data on prices of U.S. Treasury Bills, Bonds and Notes traded from 1952 to 1983. Using only data on the bond prices the authors estimate both the instantaneous default free interest rate and the variance of changes in that rate. They compare such estimates implied by the prices of a cross section of bonds trading at a point of time with estimates obtained from studying the time series of short term interest rates. In other words they compare the implied instantaneous interest rate to the short term Treasury Bill rate and look at the mean and variance of the difference. BD conclude that the CIR model systematically overestimates short term interest rates and the model fits Treasury Bills better than other Treasury Issues. They also conclude that the model significantly overprices premium issues and underprices discount issues.

Jordan and Kuipers (1997) (hereafter JK) use the cross sectional approach. In particular, the observed price of a discount bond at time t is assumed to equal the model price plus a random error (price fitting). They assume that the errors are distributed normally with zero mean and a constant variance. Given these assumptions they employ non-linear least squares to estimate the parameters of the Vasicek (1977), Merton (1973) and CIR (1985) models. They conclude that the CIR and Vasicek models describe the yield curves observed in the Treasury market with minimal error on average.

All single state variable models imply that price changes of bonds of different maturities are perfectly correlated. This is inconsistent with observed term structure dynamics. This resulted in the investigation of more general bond pricing models, multi-factor models which have more flexibility.

The multi-factor models address the issue that the changes in prices of bonds of different maturities are not perfectly correlated. Brennan and Schwartz (1979) specify the second factor as the long term rate and Longstaff and Schwartz (1992) specify the volatility of the short term rate as the second factor. Longstaff and Schwartz (1993) implement their model and apply it to the pricing of STRIPS (separate trading of registered interest and principal securities). Longstaff and Schwartz estimate the parameters by using the one-month U.S. Treasury bill rates from January 1964 through December 1989 and apply the model for the date of November 9, 1992. The authors find pricing errors of the order of 0.122 % for the one year bond and 1.322 % for the 29 year bond. Though the one year bond pricing error is reasonable the pricing error on the 29 year bond is quite high from a practitioner's point of view. Brennan and Schwartz (1979) apply their model to a sample of Canadian government bonds for the period $1964 - 1977$ and evaluate the predictive ability of the model. They find that the root mean square prediction error for bond prices was of the order of 1.5 % which is again quite high from a practitioner's point of veiw.

2.2 The Heath Jarrow and Morton Models

Heath Jarrow and Morton (1992) show that once the structure of forward rate volatilities is specified and an initial term structure is given, the prices of the zero coupon bonds in the future can be derived. In the HJM framework the process by which the term structure evolves is path dependent i.e. given an initial term structure one needs to know not only the risk factors or their proxies but also the entire path by which they evolved since their initialization. This makes the model computationally intractable since the description of the movements of term structure requires an almost infinite number of state variables.

In practice, the volatility function of the forward rates is usually considered to be Markov. A stochastic process is said to be Markovian with respect to a set of variables if the evolution of the process from time t is a function of only the observations of all the variables at time t and not the path that the variables took from an initial date to time t. Ritchken and Sankarasubramanian (1995) (hereafter RS) consider a class of interest rate processes in which spot rate volatilities can fluctuate according to the levels of the spot rate. RS further show that if the volatility structure of forward rates is given by an exponentially declining function of the time to maturity, the evolution of the spot rate is Markovian. With this, given an initial term structure, RS show that there exists a two state variable representation of the evolution of future bond prices.

2.3 The Network Models

The networks used to model the term structure in the study are the radial basis functions (hereafter RBF) and multi layer perceptrons (hereafter MLP) . The artificial neural networks are considered usually as a non-parametric technique. The networks do not make any distributional assumptions about the data though a functional form of the network does exist.

Hutchinson, Lo and Poggio (1994) (hereafter HLP) demonstrate that learning networks can be used successfully for both learning and estimating a pricing formula for options. The authors apply different networks (MLP, RBF, and projection pursuit regressions (hereafter PPR)) to daily call option prices on $S\&P$ 500 futures from 1987 $-$ 1991 and compare each of the network's pricing and delta hedging performance to the naive Black-Scholes (hereafter BS) model. Their performance measure is the coefficient of determination, R^2, of the model values compared to the true option prices for the out of sample data. The other measure considered is the "tracking error" of a replicating portfolio designed to delta-hedge an option position, using the model in question to calculate the hedge ratio or delta.

If the option pricing model used to calculate the delta is correctly identified and if one can costlessly and continuously maintain the hedge portfolio, then at expiration the combined value of the stock and bond positions should exactly offset the value of the option. The difference between the terminal value of the call and the terminal combined value of the stock and bond positions serves as a measure of the model's hedge performance. Since it is impossible to continuously hedge in practice, there will always be an error due to discreteness which the authors call as "tracking error".

The authors test the performance of the networks in two phases. In the first phase the authors generate option prices using the BS model by performing Monte Carlo simulation and train all the three networks MLP, RBF, and PPR using the data generated by the BS model. Using average out-of-sample R^2 values the authors compare the performance of the three networks relative to the BS model. The average values of R^2 reported by the authors in the first phase were 99.95%, 99.08%, 99.48%, and 100% for RBF, PPR, MLP, and BS respectively. The value of 100% for the BS model is not surprising since the data is generated using the BS model. The results indicate that all three networks on average have an R^2 value above 99% implying that all the three networks do a good job of learning the BS model.

In the second phase the authors apply the three networks MLP, RBF, and PPR to daily call option prices on $S\&P$ 500 futures from $1987-1991$ and compare each of the network's pricing and delta hedging performance to the naive BS model. The average out-of-sample R^2 values reported by the authors for the MLP, RBF, and PPR were 95.53 %, 93.26 %, and 96.56 % respectively. For the BS model the authors report an average R^2 of 84.56 %. The authors compare the absolute tracking error of each of the network MLP, RBF, and PPR with the absolute tracking error of the BS model. The statistic used by the authors is a paired t-test. The null hypothesis is that the average difference of the two tracking errors is zero, and the (one-sided) alternative hypothesis is that the difference is positive, i.e., the learning-network tracking error is smaller. The t-statistics reported for the MLP vs. BS, RBF vs. BS, and PPR vs. BS were 3.7818, 2.1098, and 2.0564 respectively.

Overall the authors find that the MLP is able to better learn the BS formula as compared to both the RBF and the PPR. MLP also does better when applied to the daily call option prices on $S\&P$ 500 futures from $1987 - 1991$. The study of HLP strongly indicates that the networks MLP, RBF, and PPR can learn a non-linear relationship like the BS model very well (from the phase one study of HLP) and at the same time uncover unknown non-linear relationships between the outputs and inputs (from the phase two of HLP study). This makes these networks strong candidates for an application to the pricing and hedging of bonds in which the relationship between the bonds and the underlying risk factors is highly non-linear and unknown. Overall the authors find that the MLP is able to better learn the BS formula as compared to both RBF and PPR, and MLP also does better when applied to the daily call option prices on $S\&P$ 500 futures from $1987 - 1991$.

2.4 The Hedging Applications

Empirical investigation of Bond hedging applications have examined simple immunization strategies based on duration and convexity. The duration of a bond is dependent on the first derivative of the price of a bond with respect to the interest rate whereas convexity is dependent on the second derivative of the bond with respect to the interest rate.

Gultekin and Rogalski (1984) (hereafter GR) using bond data conclude that "the data are not consistent with the hypothesis that price and volatility of Treasury securities is adequately measured by simple duration". GR use six non-stochastic measures of duration $D1$ through $D6$. $D1$ is Macaulay's (1938) duration, $D2$ is the one suggested by Bierwag (1977), $D3 - D5$ are the ones proposed by Cooper (1977), and $D6$ is Khang's (1979) duration. Each of the duration measures considered by GR presumes a specific characteristic movement of the term structure as shown in Table 2.1[6].

GR consider the duration measures D1 through D7 and a four factor model using a

[6]The details of each of the durations considered by GR are explained in detail in Appendix G

Table 2.1: Measures of Duration considered by Gultekin and Rogalski(1984)

The different measures of duration as considered by GR have the following implications for the changes in the yields.

Measure	Significance
D1	Permits the changes in only the level of yields
D2-D6	Allow for additive and multiplicative yield-curve movements i.e. changes in both level and the slope of yield curve. D2-D6 differ only in the implicit degree of slope and curvature of yield curves. D2 assumes the steepest yield curves while D6 assumes the flattest.
D7	A stochastic duration proposed by Cox, Ingersoll, and Ross (1979) that starts with the proposition that term structure cannot be expected to change according to any known and fixed pattern.

factor analysis. GR use a general equation

$$r(m) = -(\Delta y)Dk(m) \tag{2.1}$$

in which for a given change in the term structure, $-(\Delta y)$, the price change in a security with maturity m, $r(m)$, is directly related to its duration $Dk(m)$, with $k = 1, 2, \ldots, 7$ denoting the duration measure under consideration. GR use equation (2.1) to test the following three hypotheses

- The relation between security price changes and duration is linear
- Duration is a complete measure of risk i.e. duration incorporates the effect of maturity and coupon differences on price volatility
- Capital markets for bonds are efficient

The model of period-by-period changes used by GR to test the three hypotheses was

$$r_\tau(m) = \tilde{\gamma}1_\tau + \tilde{\gamma}2_\tau Dk_{\tau-1}(m) + \tilde{\gamma}3_\tau Dk^2_{\tau-1}(m) + \tilde{\gamma}4_\tau C_\tau(m) + \tilde{e}_\tau(m) \qquad (2.2)$$

The subscript τ refers to the period τ, so that $r_\tau(m)$ is the one-period continuously compounded price change on security m. $C_\tau(m)$ is the coupon on security m, and $Dk_{\tau-1}(m)$ denotes the duration measure under consideration with $k = 1, 2, \ldots, 7$. The coefficients $\tilde{\gamma}1_\tau$, $\tilde{\gamma}2_\tau$, $\tilde{\gamma}3_\tau$, and $\tilde{\gamma}4_\tau$ vary stochastically over time. If the hypothesis that Duration is a complete measure of risk is true that would imply that the expected value of the interest rate change $\tilde{\gamma}2_\tau$ is statistically significant. The authors include the term $Dk^2_{\tau-1}(m)$ to test linearity. Linearity would presume that $E(\tilde{\gamma}3_\tau) = 0$. GR include the term involving $C_\tau(m)$ in equation (2.2) to measure whether duration normalizes coupon differences. Completeness assumes that $E(\tilde{\gamma}4_\tau) = 0$ and the intercept term in equation (2.2) is included by the authors to measure the level of interest rates. The third hypothesis that the capital markets for bonds are efficient implies that $\tilde{\gamma}1_\tau$, $\tilde{\gamma}2_\tau$, $\tilde{\gamma}3_\tau$, $\tilde{\gamma}4_\tau$, and $\tilde{e}_\tau(m)$ should be uncorrelated through time. The disturbances are assumed to have zero mean and to be independent of all other variables in equation (2.2). The variables $\tilde{\gamma}1_\tau$, $\tilde{\gamma}2_\tau$, $\tilde{\gamma}3_\tau$, $\tilde{\gamma}4_\tau$, and $\tilde{e}_\tau(m)$ are assumed to follow approximately a multivariate normal distribution. GR perform tests of the implications for duration measures $D1$ to $D7$ for holding periods of 1, 3, 6, and 12 months by applying the model of equation (2.2) to U.S. Treasury bonds for the period January 1947 through December 1976. Applying the model in equation (2.2) to the data the authors rejected the hypothesis that the relation between returns and duration is linear. GR fail to reject the hypothesis that additional measures of risk systematically affect average price changes, signifying that duration is an incomplete measure of risk.

GR also construct Treasury portfolios holding maturity approximately constant each month to test a multiple-factor model (the factors are derived by performing factor analysis on bond portfolio returns). The authors systematically construct constant maturity portfolios to consist of all securities within a maturity range. Return and duration measures over any maturity range are computed as a weighted average of all the individual returns and durations that fall within the range. GR select specific maturity ranges to avoid any empty portfolios, constructing 12 portfolios during the 18-year period 1959 through 1976. Thirty-day increments are taken from 60 days up to a half year, 90-day increments up to 1 year, 360-day increments thereafter to about 6 years, and all securities with maturities beyond $2,160$ days. The authors factor analyze the 12 Treasury portfolio return series to obtain factor loadings of up to four factors. The model used by the authors to test the multi-factor model is

$$r_\tau(p) = \tilde{\gamma}1_\tau + \tilde{\gamma}2_\tau \lambda 1(p) + \tilde{\gamma}3_\tau \lambda 2(p) + \tilde{\gamma}4_\tau \lambda 3(p) + \tilde{\gamma}5_\tau \lambda 4(p) + \tilde{e}_\tau(p) \qquad (2.3)$$

where $r_\tau(p)$ is the one-period continuously compounded return on portfolio $p = 1, 2, \ldots, 12$; and $\lambda k(p)$ denotes the factor loadings with $k = 1, 2, 3, 4$. A one-factor model assumes that $\tilde{\gamma}3_\tau$, $\tilde{\gamma}4_\tau$, and $\tilde{\gamma}5_\tau$ in equation (2.3) are zero; a two-factor model sets $\tilde{\gamma}4_\tau$, and $\tilde{\gamma}5_\tau$ to zero and $\tilde{\gamma}5_\tau$ is zero for a three-factor model. The authors compute means and standard deviations of R^2 values adjusted for degrees of freedom obtained by period-by-period regressions using the multiple factor model in equation (2.3) for the time period 1959 through 1976. GR also compute comparable means and standard deviations of R^2 adjusted for degrees of freedom for duration $D1$ using equation (2.2) by assuming $\tilde{\gamma}3_\tau$ and $\tilde{\gamma}4_\tau$ to be zero. GR find that over the period 1959 through 1976 the average $\bar{R}^2$ for duration D1 was 0.577 and one, two, three, and four factors had an average $\bar{R}^2$ of 0.550, 0.588, 0.645, and 0.759 respectively. The authors also find that the standard deviations of the $\bar{R}^2$ values are systematically smaller as the number of factors is increased. Based on this evidence, GR conclude that a four-factor model (factors derived from factor analysis) explains about 18% more return variability than duration $D1$.

The GR study was one of the first studies that compared the traditional duration based hedges with the hedges based on a four factor model (the factors are extracted using factor analysis) and concluded that a multi factor model is required for the hedging of a fixed income portfolio. Litterman and Scheinkman (1991) (hereafter LS) find that there are three principal common influences on the variations in bond returns represented by the zero-coupon yield curve. LS examine the bond returns of 13 Treasury issues by using the weekly prices from February 22, 1984, through August 17, 1988. The bonds were maturing at dates that varied from May 15, 1989 to May 15, 2016. The authors observe that the percent of variance explained by three principal components for the weekly returns of the 13 bonds is approximately 99%. At the same time the authors also compare each bond's return with the returns on a duration-hedged portfolio composed of all the 13 bonds. The weights used by LS in the duration hedged portfolio were inversely proportional to the durations of the bonds, so that each made an approximately equal contribution to the total return of the portfolio. The authors compare the principal components approach to the duration approach based on the percent of variance of excess returns on the 13 bonds explained and unexplained by both approaches. LS find that the three-factor hedge reduces the residual variance (unexplained variance) by an average of 28%.

Bliss (1987) finds that hedging based on factor decomposition (using principal components) is superior to hedging based on traditional methods. Bliss uses three portfolios of bonds constructed on February 15 1996 to test the hedging effectiveness of the factor model versus Macaulay duration. The three portfolios are as follows:

- Portfolio 1: A single twenty-year 8 percent coupon bond (paying coupons semiannually)

- Portfolio 2: Equal numbers of one-year and twenty-year 8 percent coupon bonds

- Portfolio 3: Long positions in one-unit each of one-year and twenty-year zero coupon bonds, together with a short position in one-unit of a ten-year zero-coupon bond

Bliss computes the factor loadings by using the data on changes in Fama and Bliss (1997) yields for ten different maturities $(.25, .5, 1, 2, 3, 5, 7, 10, 15,$ and 20 years to maturity) for the time period November 1982 through December 1995. Bliss forms two hedge portfolios for each of the three portfolios. One is a Macaulay duration matched hedge portfolio (consisting of two zero coupon bonds of adjacent (six months apart) maturity) and the other is a factor durations matched hedge portfolio (consisting of four zero coupon bonds of 1, 5, 10, and 20 years to maturity, in amounts chosen to match the price and all three factor durations of the portfolio being hedged). Bliss reprices each of the three portfolios and the associated two hedge portfolios on March 15, 1996. An ideal hedge portfolio would have the same return over the period from February 15 to March 15 as the portfolio it is hedging. Bliss finds that for portfolio 1 the hedging errors using Macaulay's duration and factor based hedging are 0.29%, and −0.10% respectively. For portfolio 2 the hedging errors using Macaulay's duration and factor based hedging are 1.03%, and −0.06% respectively. For portfolio 3 the hedging errors using Macaulay's duration and factor based hedging are 1.07%, and 0.0% respectively.

The findings of GR, LS, and Bliss indicate that immunization strategies using simple duration are not capable of hedging a fixed income portfolio and immunization strategies based on procedures such as principal components analysis do a better job in hedging such a portfolio. Given the evidence from the literature the next issue that arises is how would immunization strategies based on specific term structure models perform with respect to the strategies based on simple duration and principal components analysis.

There is extensive literature on the testing of the equilibrium and HJM models and their application to the term structure of interest rates. Aït-Sahalia (1996) rejects "...every parametric model of the spot rate (previously) proposed in the literature" by comparing the marginal density implied by each model with that estimated from the data. Backus, Foresi, and Zin (1995) and Canabarro (1995) further show that the misspecification of the underlying interest rate model can lead to serious pricing and hedging errors. In other words there is little evidence whether the parametric restrictions imposed by each of the two categories of models equilibrium models and HJM models are true or not. One way to test the parametric restrictions is to compare the two approaches to a non-parametric estimation procedure and see which of the three approaches is more suitable to term structure modelling.

The study includes the network models as a non-parametric estimation procedure to compare the relative performance of each of the two approaches, the equilibrium models, and the HJM models with respect to the network models. Besides the pricing of bonds a term structure model is often used by a practitioner in hedging a fixed income portfolio. Traditionally duration and principal components analysis have been used to hedge a fixed income portfolio. The study also addresses the issue of whether the traditional approaches to the hedging application are superior to designing model (based) hedges using the three categories of models, the equilibrium models, the HJM models, and the network models.

CHAPTER 3

Theoretical Frameworks

The three different categories of the term structure models considered in the study have significantly different theoretical framework. The equilibrium models start from the assumptions of the underlying economy and from the assumptions about the stochastic evolution of the one or more exogenous factors or state variables in the economy and assumptions about the preferences of a representative investor. The general equilibrium conditions are used to endogenously derive the interest rates and the prices of all contingent claims. The equilibrium models because of their assumption that preferences of a representative investor play a role in explaining the term structure, require the estimation of the market price of risk. The HJM models impose structure directly on the evolution of the forward rate curve and do not require an estimation of the market price of interest rate risk. The structure imposed on the evolution of the forward rate curve imposes a stochastic spot rate process with multiple stochastic factors influencing the term structure. Then these models are derived from the necessary and sufficient conditions for the absence of arbitrage. The network models do not make any assumptions about the processes of the risk factors at all. The only assumption made by the network models is that the outputs or the bond prices (in this case) are a Borel integrable function of the inputs or risk factors. A Borel integrable function is a function with finite number of discontinuities.

The rest of the chapter presents the theoretical framework for each of the three different approaches to the modelling of term structure.

3.1 The Equilibrium Models

To illustrate the candidate bond pricing models in the equilibrium framework, let the discount bond price $F(x_1, x_2, \tau)$be a function of the two stochastic variables (risk factors) x_1, x_2, and time to expiration τ. The Cox, Ingersoll, and Ross (1985) (hereafter CIR) is a special case of this general two factor model and the Longstaff and Schwartz (1992) (hereafter LS) model specifies the two factors as the short term rate and its volatility. Let the general forms for the two stochastic processes of the general two factor model be given by

$$dx_i = \beta_i(x_1, x_2, \tau)dt + \sum_{j=1}^{2} \eta_{ij}(x_1, x_2, \tau)dz_j, i = 1, 2 \tag{3.1}$$

Where dz_j are two independent Wiener processes. Applying Ito's lemma produces the dynamics for the rate of change of the discount bond price.

$$\frac{dF}{F} = \mu(x_1, x_2, \tau)dt + S_1(x_1, x_2, \tau)dz_1 + S_2(x_1, x_2, \tau)dz_2 \tag{3.2}$$

where the mean of the process is

$$\begin{aligned} \mu(x_1, x_2, \tau) &= \frac{1}{F}\left[F_1\beta_{11} + F_2\beta_{22} + \frac{1}{2}F_{11}(\eta_{11}^2 + \eta_{12}^2)\right] \\ &+\frac{1}{F}\left[\frac{1}{2}F_{22}(\eta_{21}^2 + \eta_{22}^2) + F_{12}cov(\Delta x_1, \Delta x_2) - F_3\right] \end{aligned} \tag{3.3}$$

$$F_i = \frac{\partial F}{\partial x_i}, i = 1, 2$$

$$F_3 = \frac{\partial F}{\partial \tau}$$

$$S_i(x_1, x_2, \tau) = \frac{F_i \eta_{ii}}{F}, i = 1, 2$$

The stochastic process determining the return on a portfolio with proportions p_1, p_2 and p_3 invested in discount bonds having time to maturities τ_1, τ_2 and τ_3 is given by

$$\frac{dP}{P} = \left[\sum_{i=1}^{3} p_i\mu_i(\tau_i)\right] dt + \left[\sum_{i=1}^{3} p_i S_1(\tau_i)\right] dz_1 + \left[\sum_{i=1}^{3} p_i S_2(\tau_i)\right] dz_2 \tag{3.4}$$

For the return on the portfolio to be non-stochastic, coefficients of the stochastic terms dz_1 and dz_2 must equal zero.

$$\left[\sum_{i=1}^{3} p_i S_j(\tau_i)\right] = 0, j = 1, 2 \tag{3.5}$$

To prevent arbitrage the return on the portfolio must equal the risk free rate.

$$\sum p_i \left[\mu(\tau_i) - r\right] = 0 \tag{3.6}$$

Equations (3.2) and (3.3) have a solution if and only if

$$[\mu(\tau_i) - r] = \lambda_1(x_1, x_2, \tau)S_1(\tau_i) + \lambda_2(x_1, x_2, \tau)S_2(\tau_i), \quad i = 1, 2 \tag{3.7}$$

holds for all maturities, $t(.)$. Substituting for μ, S_1, and S_2 produces a partial differential equation that must be satisfied by all derivatives.

$$\begin{aligned} F_1\beta_{11} + F_2\beta_{22} + \frac{1}{2}F_{11}(\eta_{11}^2 + \eta_{12}^2) + \frac{1}{2}F_{22}(\eta_{21}^2 \\ +\eta_{22}^2) + F_{12}cov(\Delta x_1, \Delta x_2) - F_3 - \lambda_1 F_1 \eta_{11} - \lambda_2 F_2 \eta_{22} &= Fr \end{aligned} \tag{3.8}$$

Solving (3.8) subject to the boundary condition, $F(x_1, x_2, \tau = 0) = 1$ produces the discount bond pricing relationship.

The underlying economic structure of the single-factor Cox, Ingersoll, and Ross (1985) model produces mean reverting dynamics for the instantaneous short-term interest rate. The stochastic process (3.1) becomes

$$\beta_{11} = \kappa(\theta - r)$$

$$\eta_{11} = \sigma\sqrt{r}$$

$$dr = \kappa(\theta - r)dt + \sigma\sqrt{r}dz_1 \tag{3.9}$$

The CIR model further restricts the partial differential equation (3.8) and the functional form of the market price of risk for the single factor

$$x_1 = r, F_2 = 0, F_{22} = 0, F_{12} = 0, \lambda_1 = \lambda\frac{\sqrt{r}}{\sigma}$$

$$\lambda_2 = 0, \beta_{11} = \kappa(\theta - r), \eta_{11} = \sigma, x_2 = \eta_{12} = \eta_{21} = \eta_{22} = \beta_{22} = 0$$

with (σ, λ) constants. Solving (3.8) subject to the boundary condition produces the well known CIR discount bond pricing relationship given as

$$F(t,T) = A(t,T)\exp(-B(t,T)r) \tag{3.10}$$

Where

$$B(t,T) = \frac{2(\exp(\gamma(T-t)-1)))}{(\gamma+\kappa)(\exp(\gamma(T-t)-1))+2\gamma}$$

and

$$A(t,T) = \left[\frac{2\gamma(\exp((\gamma+\kappa)(T-t)/2))}{(\gamma+\kappa)(\exp(\gamma(T-t)-1))+2\gamma}\right]^{2\kappa\theta/\sigma^2}$$

Longstaff and Schwartz (1992) derive linear relationships between the expected return and volatility of the underlying physical production technology and the instantaneous short-term interest rate and the rate's volatility. The mean reverting specification of the returns to the physical production process produces mean reverting processes for the instantaneous short-term rate and a mean reverting process for the variance of the short term interest rate. In the LS model expected returns are driven by two economic factors, X and Y. The first factor X represents the component of expected returns that is unrelated to production uncertainty, while Y represents the component common to both, the expected returns from production and the volatility of production returns. An advantage of this specification is that expected returns and production volatility are not required to be perfectly correlated. The dynamics of X and Y are governed by mean reverting stochastic processes

$$dX = (a - bX)dt + c\sqrt{X}dz_2$$

$$dY = (d - eY)dt + f\sqrt{Y}dz_3$$

$$a, b, c, d, e, f > 0,$$

z_2 and z_3 are scalar Wiener processes. The stochastic processes (3.1) become

$$\beta_{11} = (\theta_{11} - \kappa_{11}r - \kappa_{12}V), \eta_{11} = \alpha\sqrt{X}, \eta_{12} = \beta\sqrt{Y}$$

$$\beta_{22} = (\theta_{22} - \kappa_{21}r - \kappa_{22}V), \eta_{21} = \alpha^2\sqrt{X}, \eta_{22} = \beta^2\sqrt{Y}$$

Where

$$\theta_{11} = \alpha\gamma + \beta\eta, \kappa_{11} = \frac{\beta\delta - \alpha\xi}{\beta - \alpha}, \kappa_{12} = \frac{\xi - \delta}{\beta - \alpha}$$

$$\theta_{22} = \alpha^2\gamma + \beta^2\eta, \kappa_{21} = \frac{\alpha\beta(\delta - \xi)}{\beta - \alpha}, \kappa_{22} = \frac{\beta\xi - \alpha\delta}{\beta - \alpha}$$

The resulting mean reverting processes implied for the instantaneous short-term interest rate and the volatility of the instantaneous short-term rate are

$$dr = (\theta_{11} - \kappa_{11}r - \kappa_{12}V)dt + \alpha\sqrt{x}dz_2 + \beta\sqrt{y}dz_3 \tag{3.11}$$

$$dV = (\theta_{22} - \kappa_{21}r - \kappa_{22}V)dt + \alpha^2\sqrt{x}dz_2 + \beta^2\sqrt{y}dz_3 \tag{3.12}$$

The LS model imposes restrictions on the partial differential equation (3.8) and the functional form of the market prices of risk

Rescaling

$$x = X/c^2, y = Y/f^2, x_1 = X, x_2 = Y, \lambda_1 = 0, \lambda_2 = \lambda\sqrt{y}, cov(\Delta X, \Delta Y) = 0$$

$$\beta_{11} = (\frac{a}{c^2} - bx), \beta_{22} = (\frac{d}{f^2} - ey), \eta_{11} = \sqrt{x}, \eta_{22} = \sqrt{y}$$

The partial differential equation (3.8) becomes;

$$F_1(\frac{a}{c^2} - bx) + F_2(\frac{d}{f^2} - ey) + \frac{1}{2}F_{11}x + \frac{1}{2}F_{22}y - F_3 - F_2y = Fr \tag{3.13}$$

Since z_2 is uncorrelated with z_1 and z_3 , changes in X cannot be hedged. The partial differential equation (3.13) is solved subject to the maturity condition that the bond value equals one when $\tau = 0$, and then the change of variables is made to r and V. This gives the following equilibrium value for the value of a riskless unit discount bond

$$F(r, V, \tau) = A^{2\gamma}(\tau)B^{2\eta}(\tau)\exp(\kappa\tau + C(\tau)r + D(\tau)V) \tag{3.14}$$

where

$$A(\tau) = \frac{2\phi}{(\delta + \phi)(\exp(\phi\tau) - 1) + 2\phi},$$

$$
\begin{aligned}
B(\tau) &= \frac{2\psi}{(\delta+\psi)(\exp(\psi\tau)-1)+2\psi}, \\
C(\tau) &= \frac{\alpha\phi(\exp(\psi\tau)-1)B(\tau)-\beta\psi(\exp(\phi\tau)-1)A(\tau)}{\phi\psi(\beta-\alpha)}, \\
D(\tau) &= \frac{\psi(\exp(\phi\tau)-1)A(\tau)-\phi(\exp(\psi\tau)-1)B(\tau)}{\phi\psi(\beta-\alpha)},
\end{aligned}
$$

and

$$
\begin{aligned}
\nu &= \xi+\lambda, \\
\phi &= \sqrt{2\alpha+\delta^2}, \\
\psi &= \sqrt{2\beta+\nu^2}, \\
\kappa &= \gamma(\delta+\phi)+\eta(\nu+\psi).
\end{aligned}
$$

3.2 The Heath Jarrow and Morton Models

The study follows the methodology implemented by Bliss and Ritchken (1996) (hereafter BR) to empirically investigate whether the movements in interest rates can be adequately described by the restricted version of the HJM (Arbitrage Based) model. The theory behind these models comes from the work of HJM (1992) who generalized on the work by Ho and Lee (1986). Ritchken and Sankarasubramanian (1995)(RS hereafter) identify the necessary and sufficient conditions on volatility structures that permit the term structure to be characterized by two state variables regardless of the structure for spot interest rate volatility.

HJM start by considering a continuous trading economy with a trading interval $[0, \tau]$ for a fixed $\tau > 0$. HJM characterize the uncertainty in the economy by the probability space (Ω, F, Q) where Ω is the state space with ω as a sample point, F is the σ-algebra representing measurable events, and Q is a probability measure. The information in the HJM framework evolves over the trading interval according to the augmented, right continuous, complete filtration $F_t : t \in [0, \tau]$ generated by $\Re_n$ valued Brownian motion z_t. Under the probability measure Q HJM specify the instantaneous forward rate process as

$$df_t(T) = \mu_f(t, T, w)dt + \sigma_f(t, T, w)dz_t, t < T, \tag{3.15}$$

Because of their dependence on ω, μ and σ are path-dependent. From here on the reference to ω will be dropped for simplicity. Hence unless specified otherwise all the functions in this framework are dependent on ω causing all these functions to be path dependent. The process simultaneously evolves each forward rate $f_t(T)$. Given a non-random initial forward curve $f_0(T)$, any subsequent time-t forward rate curve will be a function of the forward rate process, $df(t, T)$ and its evolution from time 0 to time t:

$$f_t(T) = f_0(T) + \int_0^t \mu_f(s, T)ds + \int_0^t \sigma_f(s, T)dz_s$$

From the forward rate process equation (3.15), the processes followed by the spot rate and the pure discount bonds can be written as

$$r_t(T) = \frac{1}{T - t}\int_t^T f_t(s)ds$$

$$B_t(T) = \exp - \int_t^T f_t(s)ds$$

Using these functional relationships the integrated form of the short rate process is

$$r_t = f_0(t) + \int_0^t \mu(s, t)ds + \int_0^t \sigma(s, t)dz_s$$

and the price of the bond evolves as

$$\frac{dB_t(T)}{B_t(T)} = \mu_p(t, T)dt + \sigma_p(t, T)dz_t$$

where the $\mu_p(t, T)$and $\sigma_p(t, T)$ are functions of the specification of the forward rate dynamics.

HJM (1992) use no arbitrage arguments to derive restrictions on the permissible forms that the drift term of the forward rate process can take:

$$df_t(T) = \sigma_f(t, T)\left[\lambda(t) - \sigma_p(t, T)\right]dt + \sigma_f(t, T)dz_t$$

where

$$\lambda(t) = \frac{\mu_p(t,T) - r(t)}{\sigma_p(t,T)}$$

is the market price of interest rate risk and

$$\sigma_p(t,T) = -\int_t^T \sigma_f(t,s)ds$$

Though the forward rate volatility process may be quite general, the forward rate drift is restricted to a particular function of the volatility structure. The forward rate process, in turn, implies the following relation between current and future bond prices:

$$\begin{aligned} B_t(T) &= \left\{\frac{B(0,T)}{B(0,t)}\right\} \\ &\exp\left\{\int_t^T \left[\int_0^t \sigma_f(u,s)\sigma_p(u,s)du\right] ds - \int_t^T \left[\int_0^t \sigma_f(u,s)[\lambda(u)du + dw(u)]\right] ds\right\} \end{aligned}$$

The HJM (1992)analysis presented so far is quite general and is also not very useful from a practitioner's view point. Firstly due to the dependence of the bond prices on the market price of risk which is a function of the risk preferences and is difficult to measure. Secondly the evolution of the term structure as stated above will not be Markovian with respect to a finite collection of variables. This implies that at any time the knowledge of only the state variables at that time is not sufficient to characterize the term structure at that time. The information about the path taken by the state variables from the initialization date to that point in time is required. This is due to the dependence of all the functions above on the sample point ω as stated above. These issues cause impediments in developing efficient numerical procedures for pricing interest rate claims. Though numerical procedures do exist for such path dependent functions, the implementation of the discrete approximations become almost impractical for long dated interest rate contingent claims. Amin and Morton (1994) test various path dependent HJM models using short maturity options on euro-dollar futures and an algorithm with ten or fewer time steps. They find that the path dependent tree develops a greater density of terminal nodes than a path-independent tree of the same number of time steps, and suggest that this makes the path-dependent trees converge more quickly. Heath, Jarrow, and Morton and Spindel (1992) report similar efficiencies for pricing five-year swaptions.

Later on Caverhill (1994), Hull and White (1993) and Ritchken and Sankarasubramanian (1995) show that if the volatility structure of forward rates has the form:

$$\sigma_f(t,T) = \sigma k(t,T), \tag{3.16}$$

where k(t,T) is a deterministic function satisfying the following semigroup property:

$$k(t,T) = k(t,u)k(u,T) \qquad t \leq u \leq T$$

$$k(u,u) = 1,$$

then, given the initial term structure, any single point on the term structure at date t is sufficient to explain the full yield curve at that date. The usual choice for the single point is the short term rate and its dynamics are path independent and the structure for the $k(t,T)$ is the exponentially dampened structure

$$k(t,T) = \exp -(\kappa(T-t)). \tag{3.17}$$

The first type of volatility structures considered in the study are generalized Vasicek(hereafter GV) structures that follow equation (3.16). The name follows from the fact that the spot rate volatility is independent of time and is constant. Under the GV model, forward rate volatilities are exponentially declining in their maturities, and the future value of the state variable is normally distributed. Empirical tests of the volatility structures performed by Flesaker (1992) and Amin and Morton (1994) find no support for the Ho and Lee (1986) model and the GV model was found to be useful in that it was capable of generating abnormal returns in particular trading strategies.

The second type of forward rate volatility structures considered in the study are the ones that allow for variable spot rate volatility. Ritchken and Sankarasubramanian (1995) consider a class of interest rate processes in which spot rate volatilities can fluctuate according to the levels of the spot rate. Let $\sigma_r(t,\cdot)$ represent the volatility of the instantaneous spot rate, $r(t)$, at date t. Such structure could depend on any term structure information available at time t. Ritchken and Sankarasubramanian then show that if the volatility structure of forward rates is given by

$$\sigma_f(t,T) = \sigma_r(t,\cdot)k(t,T) \tag{3.18}$$

where $k(t,T)$ again is a deterministic function satisfying the semigroup property, then, given an initial term structure, there exists a two state variable representation of the evolution of future interest rates.

For volatility structures that do not use the time varying parameters, the only feasible (for the evolution of the spot rate to be markovian) $k(t,\cdot)$ function is again the deterministic, exponentially dampened function given by equation (3.17), and thus

$$\sigma_f(t,T) = \sigma_r(t,\cdot)\exp -\kappa(T-t)$$

In other words, the forward rate volatility, normalized by the spot interest rate volatility, must be a deterministic, exponentially dampened function of maturity

$$\frac{\sigma_f(t,T)}{\sigma_r(t,\cdot)} = \exp -(\kappa(T-t)) \tag{3.19}$$

For the volatilities belonging to the two state variable class, Ritchken and Sankarasubramanian show that there is a simple analytical linkage between interest rate term structures at dates s and t.

Based on the notation used by BR (1996) the bond prices are indexed at date t by their maturities rather than their maturity dates. Let

$$F_t(t+m) = \exp\left\{-\int_t^{t+m} f_t(u)du\right\}$$

represent the time-t price of an m-maturity pure discount bond that pays a dollar at date t+m. Given an initial term structure, $F_0(\cdot)$ at time 0, the price of a bond, at any future date, t, must be defined in terms of its forward price at date 0, the short interest rate at date t, and a variable that captures the history of the path of interest rates from s to t as follows:

$$F_t(t+m) = \left[\frac{F(0,t+m)}{F(0,t)}\right] \exp\left[-\frac{1}{2}\beta^2(m)\phi(t) + \beta(m)\psi(t)\right] \tag{3.20}$$

where

$$\beta(m) \equiv \frac{1}{\kappa}(1-e^{-\kappa m});$$

$$\phi(t) \equiv \int_0^t \sigma_f^2(,t)ds;$$

$$\psi(t) \equiv f(0,t) - r(t).$$

The non-existence of the above conditions would imply the violation of the no arbitrage assumption. In this representation given the initial term structure, the entire term structure at a subsequent time t can be reconstructed once $\phi(t)$ and $\psi(t)$ are known. Neither of these factors depend on maturity m. Thus this model is a two state variable model, even though there is only one underlying stochastic factor. Equation (3.20) identifies the two state variables as the ex post forward premium on the spot interest rate, $\phi(t)$, and the "integrated variance" factor, $\psi(t)$. Ritchken and Sankarasubramanian characterize the dynamics of the two state variables,$\phi(t)$ and $\psi(t)$, in terms of their current values and the forward rate curve at an earlier date 0. Specifically, interest rate claims can be priced as if the local expectations applied, if the dynamics of the state variables are taken as

$$d\psi(t) = [\kappa\psi(t) + \phi(t)]dt + \sigma_r(t,\cdot)dw(t)$$

$$d\phi(t) = [\sigma_r^2(t,\cdot) - 2\kappa\phi(t)]dt$$

The dynamics of the instantaneous spot interest rate under the Ritchken and Sankarasubramanian assumption is given by

$$dr(t) = \left[\kappa\psi(t) + \phi(t) + \frac{d}{dt}f(0,t)\right]dt + \sigma_r(t,\cdot)dw(t).$$

The evolution of the spot rate process turns out to be Markovian as it depends only on the two state variables ϕ and ψ at time t. This is in contrast with the general models under the HJM framework in which the spot interest rate process cannot described by a Markov process with a finite number of state variables.

Hence the two models considered in this study are the ones that are presented in equation (3.20). For the GV type of volatility structures the state variable $\phi(t)$ is deterministic. From equation (3.16) and the value of $\phi(t)$ from equation (3.20) GV structures imply

$$\begin{aligned}\phi(t) &= \int_0^t \sigma^2 \exp -(2\kappa(t-s))ds \\ \Rightarrow \phi(t) &= \frac{\sigma^2}{2\kappa}(1-\exp(-2\kappa t)). \qquad (3.21) \\ & \qquad (3.22)\end{aligned}$$

The GV volatility structures (hereafter HJM-GV) have only one state variable $\psi(t)$ as the state variable $\phi(t)$ is deterministic. For the HJM-GV model the prices of bonds are given by

$$F_t(t+m) = \left[\frac{F(0,t+m)}{F(0,t)}\right] \exp\left[-\frac{1}{2}\beta^2(m)\phi(t) + \beta(m)\psi(t)\right] \qquad (3.23)$$

where

$$\beta(m) \equiv \frac{1}{\kappa}(1-e^{-\kappa m});$$

$$\phi(t) \equiv \frac{\sigma^2}{2\kappa}(1-\exp(-2\kappa t));$$

$$\psi(t) \equiv f(0,t) - r(t).$$

Whereas the forward rate volatility structures that allow for variable spot rate volatility (hereafter HJM-RS) have two state variables $\psi(t)$ and $\phi(t)$. The prices of bonds in the case of HJM-RS model are given by the equation (3.20).

3.3 The Network Models

The third approach to the term structure modelling considered in the study is the non-parametric pricing approach using learning networks. Though there are different types of networks that exist the ones used in the study are Radial Basis Functions and Multi Layer Perceptrons. Studies in the past have also shown some type of equivalence (convergence to similar outputs) between different learning networks. For instance, Maruyama, Girosi, and Poggio (1991) show an equivalence between Multi Layer Perceptrons (hereafter MLP)and Radial Basis Functions (hereafter RBF). The authors show that if the inputs are normalized then the two networks MLP and RBF produce the same results. Girosi, Jones, and Poggio (1993) prove that a wide class of approximation schemes can be derived from regularization theory, including RBF networks and some forms of Projection Pursuit Regressions and MLP networks. At the same time researchers like Ng and Lippman (1991) argue that the practical differences in using each method, e.g., in running time or memory used, may be more important than model accuracy.

The term Artificial Neural Networks(ANN hereafter) comes from the network's analogy to the biological structure of Neural Networks. The human brain consists of a huge number of highly connected elements called neurons. The neurons have three major components that are relevant from the ANN point of view. The components are the dendrites, the cell body and the axon. The dendrites are tree-like receptive networks of nerve fibers that carry electrical signals into the cell body. The cell body essentially sums and scales these incoming signals. The axon is a single long fiber that carries the signal from the cell body to other neurons. The point of contact between an axon of one cell and a dendrite of another cell is called a synapse. A neural network's performance is primarily dependent on the functional form specified by the arrangement of neurons and the strengths of the individual synapses, determined by a complex chemical process.

A single input neuron is a good point to start with, to comprehend the concept of the neural networks. The scalar weight w is multiplied by the scalar p to form wp, one of the terms that is sent to the summer. Another input, 1, is multiplied by a bias b and then passed to the summer. The summer output n, also referred to as the net input, is passed through an activation function (essentially a transfer function) f, which produces the scalar neuron output a. The weight w corresponds to the strength of a synapse, the cell body is represented by the summation and the activation function. The neuron output a represents the signal on the axon. The neuron output is calculated as

$$a = f(wp + b)$$

A single layer perceptron is a perceptron that has more than one neuron to which the inputs are fed.

Rosenblatt's(1958) learning rule and the Least Mean Squares(LMS) algorithm of Widrow and Hoff(1960) were designed to train single-layer perceptron like networks. It was realized that the single layer networks suffer from the disadvantage that they are only able to solve linearly separable classification problems. As a consequence both Rosenblatt and Widrow proposed multilayer networks that could overcome such problems but were not able to extend their algorithms to train such networks. It was in the mid 1980s that the backpropagation algorithm was rediscovered and widely publicized independently by Rumelhart, Hinton and Williams (1986), Parker (1985), and Cun (1985). The algorithm was popularized by its inclusion in the book *Parallel Distribution Processing* that described the work of the Parallel Distributed Processing Group led by psychologists Rumelhart and McClelland. This triggered a torrent of research in neural networks. This backpropagation algorithm is used in the study for both the pricing as well as the hedging applications using the MLP network. The Multi Layer Perceptron, trained by the backpropagation algorithm, is currently the most widely used neural network and is one of the two networks used in the study.

The Multi Layer perceptron

A Multi Layer perceptron is more than one single layer perceptron cascaded together. The output of the first layer is the input to the second layer and the output of the second layer is the input to the third layer and so on. The MLP in this study has two layers one hidden layer and one outer layer. Let the weight matrix for the first layer be $\mathbf{W}^1$ and the weight matrix for the second layer be denoted by $\mathbf{W}^2$ and the input matrix be denoted by

$\mathbf{p}$. Let f^1, and f^2 denote the activation functions in the layers 1, and 2 respectively. Let $\mathbf{b}^1$, and $\mathbf{b}^2$ denote the bias matrices in the layers 1, and 2 respectively, and a^2 denote the final output. Then the final output of the network is given by

$$a^2 = f^2\left(\mathbf{W}^2 f^1\left(\mathbf{W}^1\mathbf{p} + \mathbf{b}^1\right) + \mathbf{b}^2\right) \tag{3.24}$$

The activation function also known as the transfer function used in the layer 1, f^1 is the tansigmoid function. For a given input x the tansigmoid function transfers it to an output y given by

$$y = \frac{1 - \exp(-x)}{1 + \exp(x)}. \tag{3.25}$$

whereas the transfer function used in layer 2 is the pure linear function that transfers an input x into an output y given by

$$y = x. \tag{3.26}$$

Alternatively the pure linear function preserves the input and does not change it's value. The study uses the equation (3.24) as a bond pricing relationship once all the parameters of the network have been estimated using Bayesian Regularization algorithm explained in detail in Appendix A.

The Radial Basis Functions

Besides the MLP the second type of network the study uses is the RBF. An RBF network is a two layer or a three layer (depending on whether the inputs are fed to the radial basis functions or they are processed before feeding) feed forward network whose output nodes form a linear combination of the basis functions in the hidden layers that are usually the Gaussian kernels and multiquadratics as defined in Appendix B. The radial basis function networks were initially used to solve the problem of interpolation i.e. fitting a curve exactly through a set of points. Gradually the RBF formulation has been extended by researchers to perform the broader task of approximation (Broomhead and Lowe (1988), Moody and Darken (1989), and Poggio and Girosi (1990)). Poggio and Girosi show how RBFs can be derived from the classical regularization problem in which some unknown function is to be approximated given a sparse data set and some smoothness constraints. Lay and Hwang (1993) state that an RBF network can be regarded as an improved alternative of a probabilistic neural network (PNN). In a probabilistic neural network, a symmetric RBF kernel is placed on each training data set so that the unknown density can be well interpolated and approximated.

An RBF network can be regarded as linear in the parameters if all the RBF centers and the non-linearities are fixed in the hidden layer. So the hidden layer performs a fixed non-linear transformation with no adjustable parameters and essentially it maps the input space onto an output space of the hidden layer. The output layer then performs a linear combination of this output space and the only training or adjustable parameters are the weights of the linear combiner. A schematic of the RBF network with n inputs and a scalar output is shown in Figure (3.1). Such a network essentially implements a mapping $f_r : R_n \to R$ according to

$$f_r(x) = \lambda_0 + \sum_{i=1}^{n_r} \lambda_i \phi(\|x - c_i\|) \tag{3.27}$$

where $x \in R^n$ is the input vector, $\phi(\cdot)$ is the non-linear function of the hidden layer from R^+ to R,$\|\cdot\|$ is the Euclidean norm, λ_i $0 \leq i \leq n_r$ are the weights or parameters, $c_i \in R^n$, $1 \leq i \leq n_r$ are known as RBF centers, and n_r are the number of centers. In the RBF networks the non-linear function and the centers are fixed and the only adjustable parameters are the weights of the second layer. The $\phi(\cdot)$ used for the study is the Gaussian function that is

$$\phi(\nu) = \exp\left(\frac{-\nu^2}{\beta^2}\right) \tag{3.28}$$

Once the number of centers are chosen and the coefficients of the equation (3.27) determined using the orthogonal least squares algorithm explained in detail in Appendix B, the functional form of equation (3.27) is used to get the bond prices of the required maturities.

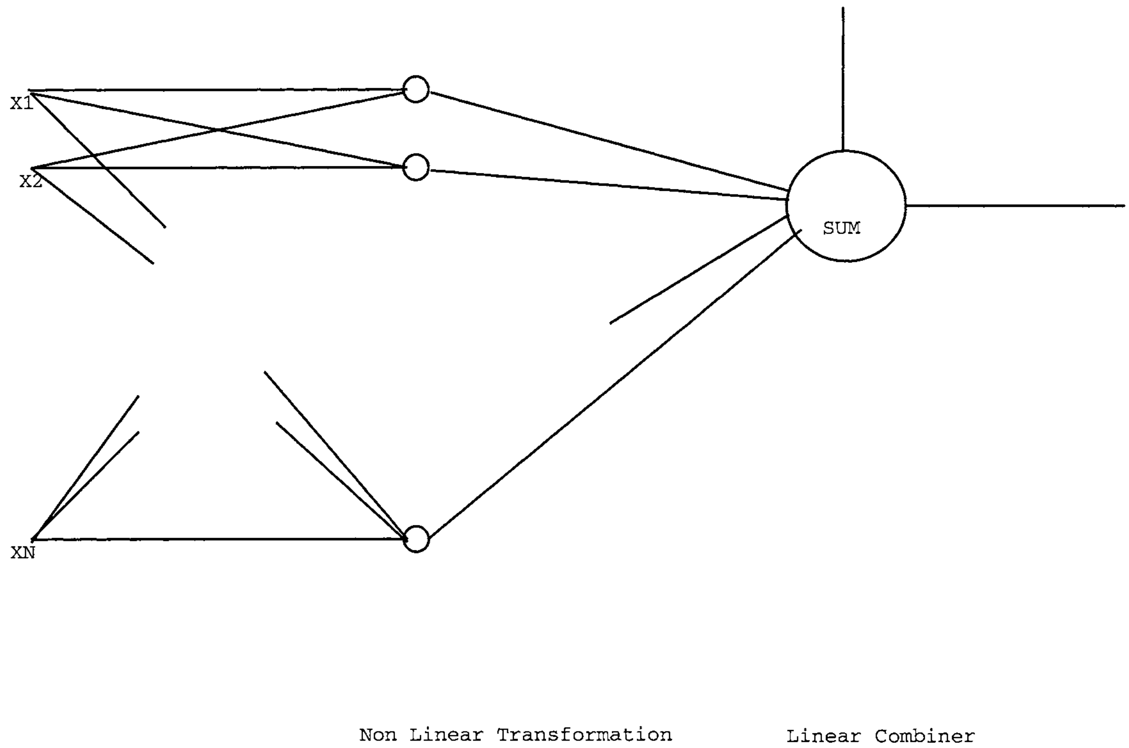

Figure 3.1: A Radial Basis Function

At this point some comparisons between the two types of networks used, MLP and RBF is worthwhile.The RBF networks possess a property that the choice of the parameters cannot be claimed as the optimal choice. In this particular case the study can use different estimation procedures to estimate the weights of the second layer and probably could improve the ability of the network's approximation ability. This result is shown by Girosi and Poggio (1990) that RBFs have the "best" approximation property i.e. there is always a choice for the parameters that is better than any other possible choice. This property in contrast is not shared by the MLP network. The performance of an RBF network depends on

- the selection of the centers
- estimation of the weights of the second layer
- choice of the parameter b (usually learned while training the network). This b is

related to how much of an area overlap one wants between two adjacent Gaussian functions.

- choice of the tolerance value in the orthogonal least squares algorithm (also learnt during the training of the network)

The neurons that are the centers of the Gaussian functions have localized receptive fields in the sense that they respond only to inputs close to the centers. This is also in sharp contrast to the MLP networks which use the tansigmoid function that has a global response. So the networks with localized receptive fields are quite advantageous when the new data is coming from only a small sub sample of the whole population and the centers or neurons in that sub sample region will only respond to such inputs. It is quite difficult to predict if the term structure modelling has this particular property or not because that depends on the type of shock the yield curve is subject to in the out of sample period as well as the two points on the yield curve that are being used as inputs. On compact data the training times for RBFs are shorter than the training times on other standard MLP networks. A compact data set is characterized by lesser number of inputs. In this study the number of inputs is always two and hence this study deals with only compact data sets. At the same time RBF has a disadvantage that if the data set are not compact and the number of dimensions is huge i.e. the number of inputs keeps increasing, then the number of centers required to span the whole input space is sufficiently large and this increases the complexity of the network exponentially.

The RBF network trained in this study can be alternatively trained by changing the way the weights of the second layer are determined. They are determined using the linear least squares regression. The data can be checked if the data of the outputs of the first layer conforms to the assumptions of the linear least squares model or not. That is whether the errors are normally distributed, the variance of the error terms is constant, the outputs and the error terms are not correlated, if there is any auto correlation etc. So one way to improve the performance of an RBF network is to test these assumptions and correct for them if any of them is violated and then estimate the weights of the second layer. Other areas in which the RBF networks performance can be improved is the choice of the convergence criteria for minimizing the generalizing error that an RBF network would do. The error caused by applying a neural network to the out of sample period is known as the generalizing error.

3.4 The Hedging Applications

There are different ways one can hedge a portfolio of fixed income securities. Traditionally hedging strategies were based on Macaulay's duration, convexity, and principal components analysis (hereafter PCA). Duration is based on the first order derivative of the bond price with respect to the interest rate and convexity is based on the second order derivative. PCA is based on the decomposition of the variance covariance matrix of the changes in yields of differing maturities into three principal components that are orthogonal i.e. uncorrelated by construction. Besides these traditional measure one could also design a specific model based hedging strategy because a hedging application would also depend on the number of underlying risk factors.

Hedging based on Macaulay's duration is a single factor hedging and is very simple to implement. Litterman and Scheinkman (1991) argue that there are three principal components that explain the variations in the yield curve. After Litterman and Scheinkman (1991) there is some literature that highlights some of the drawbacks of estimating a PCA model. Nunes and Webber (1997) find that leaving the jumps in the data set before estimating a PCA model does affect the results and James and Webber (2000) suggest it is necessary to remove the jumps before estimating the PCA model. Rebonato (1996) finds that "...any yield curve model that uses principal components as driving factors is constrained to displaying a sigmoid-like correlation structure. This feature is not a result of the particular assumptions of the specific models, but a general consequence of the low dimensionality of these approaches". Rebonato argues that PCA methods that use 2 or 3 components are incapable of producing an exponential like correlation structure amongst the instantaneous forward rates for maturities 1 through 10 years.

Finally hedging strategy can be based on hedging all the underlying risk factors specific to a model. So a hedging strategy can be designed in the following three ways

- Devise a hedging strategy that is based on Macaulay's duration
- Devise a hedging strategy based on PCA methods in which the factors are uncorrelated to one another by construction
- Devise a hedging strategy that hedges all the underlying risk factors specific to a model (dependent on what model is used to explain the term structure).

Till date ambiguity persists relating to which of these approaches should one pick for the hedging of a fixed income portfolio? Using the three approaches to hedge, the study addresses the issue of what approach is ideal for the hedging of a fixed income portfolio and if the third approach is better than the first two then what model should one pick to design the hedging strategy?

CHAPTER 4

Empirical Methodology

4.1 Data Requirements

The data requirements for the study are the prices of zero coupon bonds. The data used for the study are the unsmoothed Fama-Bliss (1987) yields (analogous to bootstrapping to create a spot rate curve) for the time period $1983 - 1999$[7]. The unsmoothed Fama-Bliss method extracts forward rates by an iterative procedure. Fama-Bliss (hereafter FB) extend the discount rate function each step by computing the forward rate necessary to price successively longer maturity bonds given the discount rate function fitted to the previously included issues. FB use a series of filters to throw out the suspicious quotes. The resulting discount rate function exactly prices the included bonds. FB use the mean price and the weighting is irrelevant because fitted-price errors of bonds remaining after the filtering step are all zero. The resulting discount rate function is piecewise linear (jagged) with the number of parameters equal to the number of included issues.

The data used in the study includes a time series of zero coupon bond yields (extracted with the unsmoothed FB method) over the period of $1983 - 1999$. The unsmoothed Fama and Bliss (1987) yields are converted into prices using the discount function $\exp(-(m)y)$ (m is the maturity of the bond and y is the yield). The data includes the yields on the target maturities $.25, .5, 1, 2, 3, 5, 7, 10, 15,$ and 20 years on the first trading day of every month during the time period January 1983 through December 1999. The summary statistics of the estimated term structure used are listed in Table 4.1.

[7] I would like to thank Robert R. Bliss for kindly providing the programs on the estimation of the term structures.

Table 4.1: Summary Statistics of the Estimated Term Structure(1983 – 1999)

The mean and variance of the zero coupon yields for different maturities as estimated by the unsmoothed FB method in percentage and percentage square respectively.

Time to Maturity (years)	Mean	Variance
.25	6.76	7.36
.5	6.99	7.39
1	7.22	6.87
2	7.49	6.16
3	7.67	5.62
5	7.89	5.11
7	8.05	4.77
10	8.13	4.54
15	8.41	4.47
20	8.47	4.56

The study also requires a proxy for the instantaneous short term rate for the Cox Ingersoll and Ross (1985) (hereafter CIR) model. To be consistent with the use of the data for the other models the unsmoothed FB yield for the bond that has .25 year to maturity is used as a proxy for the instantaneous short term rate.

4.2 The Equilibrium Models

4.2.1 The CIR model

CIR (1985) pioneered the Equilibrium approach to the modelling of the term structure of interest rates. They start from the description of the underlying economy and from the assumptions about the stochastic evolution of the one or more exogenous factors or state variables in the economy. The assumptions made by CIR are as follows

- There is a finite number of constant stochastic returns to scale production technologies that produce a single good that can be allocated to either consumption or investment

- There is a fixed number of identical individuals who maximize the time additive expected utility of consumption function by selecting optimal consumption and investment function

- All investment is done by firms and the individuals invest all of their unconsumed wealth in the shares of those firms. The values of the firms follow multivariate diffusion process.

- The joint process for the firm values and the state variables completely describe the state of the system.

- There are perfect competitive markets for continuous trading in the firm's shares and a variety of contingent claims, as well as for instantaneous risk free borrowing and lending.

- Equilibrium in this economy gives the market clearing interest rate, prices for the contingent claims, and the total production and consumption plans.

- The authors also assume that temporal change in production opportunities is described by a single state variable, the means and variances of the rates of return of the production processes are proportional to this variable and the assumption about the stochastic differential equation describing the development of the variable.

Given these assumptions CIR derive the dynamic process for the instantaneous short term rate relationship, equation(3.9) and the resulting bond pricing relationship, equation (3.10). Hence in this model the anticipations, risk aversion, investment alternatives, and preferences about the timing of consumption all play a role in determining the bond prices.

The methodology used to estimate the CIR model is the one used by Chan, Karolyi, Longstaff, and Sanders (1992) (hereafter CKLS). To estimate the CIR model the unsmoothed FB yield extracted for a bond that matures in .25 years is used as the proxy for the instantaneous short-term rate. The model is estimated using Hansen's (1982) Generalized Method of Moments estimation to estimate parameters of the instantaneous short-term interest rate process, equation (3.9). The continuous time model is estimated using the discrete time analog

$$\begin{aligned} r_t - r_{t-1} &= \alpha + \beta r_{t-1} + \epsilon_t \\ \epsilon_t^2 &= \sigma^2 r_{t-1} + \eta_t \end{aligned} \tag{4.1}$$

From the set of equations (4.1) it follows that the relation of its parameters to that of equation (3.9) is

$$\begin{aligned} \kappa &= -\beta \\ \theta &= \frac{\alpha}{\kappa} \end{aligned} \tag{4.2}$$

(4.3)

For the GMM estimation the following orthogonality conditions are used

$$g_t(\theta) \equiv [\epsilon_t,\ \epsilon_t r_{t-1},\ \eta_t,\ \eta_t r_{t-1}]'$$
$$E[g_t(\theta)] = 0,$$

where $\theta \equiv (\alpha, \beta, \gamma)$ and ϵ_t and η_t are defined in the set of equations (4.1). An estimate of θ is obtained by choosing $\hat{\theta}$ to minimize the quadratic

$$J_T(\theta) = g_T(\theta)' W_T(\theta) g_T(\theta),$$

where

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta)$$

is the sample average of the realizations of g_t, and $W_T(\theta)$ is a positive definite weighting matrix. The sample average g_T will converge to zero for large T, under the null hypothesis when evaluated at the true values of θ. Hansen (1982) shows that W_T can be consistently estimated by

$$W_T(\hat{\theta}) = \left[\frac{1}{T}\sum_{t=1}^{T} g_t(\hat{\theta}) g_t(\hat{\theta})'\right]^{-1}$$

as long as $\hat{\theta}$ is a consistent estimate of θ and the $g_T(\theta)$ is serially uncorrelated. It turns out that the estimation of $W_T(\theta)$ requires an estimate of θ and hence an interactive procedure can be used as suggested by Hansen (1982). The $W_T(\hat{\theta})$ can be initialized to an identity matrix and the $\hat{\theta}$ can be estimated by minimizing the function $g_T(\theta)' I g_T(\theta)$. This estimate of $\hat{\theta}$ is used to compute the new $W_T(\hat{\theta})$ and this in turn is used to estimate the new $\hat{\theta}$. The iterative procedure is repeated till $\hat{\theta}s$ converge. The software used to write the programs was MATLAB.

CKLS estimate the parameters of the short rate process as specified in the set of equations (4.1) using the generalized method of moments. These parameters are estimated by using the annualized one-month U.S. Treasury bill yield from June 1964 to December 1989 (306 observations). The estimates of the parameters are shown in Table 4.2.

Table 4.2: CIR Parameter Estimates (CKLS)

The parameters of the short rate process as specified in the set of equations (4.1) and estimated by GMM in the CKLS study are stated. These parameters are estimated by using the annualized one-month U.S. Treasury bill yield from June 1964 to December 1989 (306 observations).

Parameter	Estimate	Standard Error
α	0.0189	0.0201
β	-0.2339	0.3544
σ^2	0.0073	0.0010

Table 4.3: CIR Parameter Estimates

The parameters of the short rate process as specified in set of equations (4.1) and estimated by GMM are stated. These parameters are for the last window that uses the data on the short term rate from December 1986 till November 1999 (156 months).

Parameter	Estimate	Standard Error
α	0.0476	0.0684
β	-0.0087	0.0129
σ^2	0.0098	0.0018
λ	-0.1432	NA

Since the study uses a rolling window approach, the parameters are re-estimated at the beginning of every month starting January 1996 to December 1999. The parameters for the last window that uses the data on the short term rate from December 1986 till November 1999 (156 months) and their standard errors are stated in Table 4.3. An additional measure the market price of interest rate risk is needed to price a unit discount bond. The ask price (implied by the ask discount rate) of Treasury Bills maturing in one year from the first trading day of each month beginning January 1996 till December 1999 is used. The market price of interest rate risk at each date is implied by minimizing the squared difference between the CIR predicted price for the Bill using parameters estimated for that date and the actual price. The estimate of the market price of risk for the last window is also stated in Table 4.3. The values of the estimated parameters with the implied values of the market prices of risk for all the windows are plugged into the pricing equation (3.10) to get the predicted price of unit discount bonds.

From the parameters α, β, and σ^2 shown in Tables 4.3 and 4.2 one can derive the parameters of the short rate process, equation (3.9) using the set of equations (4.2). The parameters shown in Tables 4.3 and D.1 are similar to the ones in Table 4.2 except the parameter β. The parameter β appears to be of a different order due to a difference in scaling (the study uses the interest rates of $x\%$ as x instead of $.01(x)$).

4.2.2 A numerical example of the CIR model

A numerical example is included to illustrate the empirical methodology of the CIR. Considering the first window i.e. using the data on FB yield for the .25 year bond from January 1983 till December 1995 the parameters α, β, and σ of equation (4.1) are estimated using Generalized Method of Moments procedure. The estimates are shown in Table 4.4. The parameters of the short rate process κ, and θ are calculated using the set of equations (4.2).

The values of the parameters are plugged into the pricing equation (3.10) to get the

Table 4.4: CIR Parameter Estimates (example)

The parameters of the short rate process as specified in the set of equations (4.1) and estimated by GMM are stated. The parameters of the short rate process as specified in equation (3.9) are calculated using the set of equations (4.2). These parameters are for the last window that uses the data on the short term rate from January 1983 to December 1995 (156 months).

Parameter	Estimate	Standard Error
α	0.0202	0.0717
β	-0.0052	0.0124
σ^2	0.0149	0.0023
λ	-0.03841	NA
κ	0.0052	NA
θ	0.0389	NA

price of the bond of a particular maturity. The prices as produced by the pricing relation, equation (3.10) for the first trading day of January 1996 for all the different maturities are shown in Table 4.5.

4.2.3 The LS model

Longstaff and Schwartz (1992) (hereafter LS) develop a two factor general equilibrium model of the term structure of interest rates using the CIR (1985) framework. The authors assume that all physical investment is performed by a single stochastic constant returns to scale technology which produces a good that is either consumed or reinvested in production. LS assume that the expected returns are driven by two economic factors, one represents the component of returns that is unrelated to production uncertainty, while the other represents the component common to both the expected returns from production and the volatility of production returns. The development of the processes for both the state

Table 4.5: CIR Predicted Bond Prices

The values of the parameters from Table 4.4 are plugged into the pricing equation (3.10) to get the price of the bond of a particular maturity. The prices as produced by the pricing relation, equation (3.10) for the first trading day of January 1996 for all the different maturities are shown as CIR-Price. The unsmoothed Fama and Bliss (1987) yields for the first trading day of January 1996 are converted into prices using the discount function $\exp(-(m)y)$ (m is the maturity of the bond and y is the yield) are shown as FB-Price.

Maturity	FB-Price	CIR-Price
.5	0.9758	0.9751
1	0.9523	0.9505
2	0.9072	0.9024
3	0.8602	0.8564
5	0.7714	0.7729
7	0.6823	0.7028
10	0.5709	0.6229
15	0.4049	0.5433
20	0.2853	0.5039

variables is assumed to be a stochastic differential equation of the form of equation (3.9). The authors use the two factor general equilibrium model of the term structure of interest rates to derive closed from expressions for discount bond prices and discount bond option prices.

The methodology used to implement the LS model is based on the study by Longstaff and Schwartz (1993). To estimate the LS model, the stochastic processes of short term rate and its volatility given by equations (3.11) and (3.12) are proxied by a discrete GARCH model, Bollerslev (1986) as

$$\begin{aligned} r_{t+1} - r_t &= \alpha_0 + \alpha_1 r_t + \alpha_2 V_t + \epsilon_{t+1} && (4.4) \\ \epsilon_{t+1} &\approx N(0, V_t) \\ V_t &= \beta_0 + \beta_1 r_t + \beta_2 V_{t-1} + \beta_3 (\epsilon)^2 && (4.5) \end{aligned}$$

This specification allows unexpected changes in r to be conditionally heteroscedastic through their dependence on the value of V. In turn, the volatility V follows an autoregressive process since its current value depends on its lagged value. This specification closely resembles the continuous time dynamics of r and V. Because of this the parameters of the GARCH model need not map directly into the parameters of the continuous time process. In compliance with the rolling window approach the parameters of the set of equations (4.5) are estimated using the time series of short term interest rates over the forty-eight rolling windows. The proxy for the short term interest rate used is the FB yield on the .25 year bond. The estimates of the parameters for the equation for the last window i.e. from December 1986 till November 1999 (156 months) and their standard errors are stated in Table 4.6.

The residuals from the above estimated equation coupled with the coefficients of the variance equation of the set of equations (4.5) for the GARCH model are used to get a time series of volatility estimates from the time series of interest rates. With the time series estimates of interest rates and the volatilities the following steps are taken to estimate the

Table 4.6: LS Parameter Estimates

The parameters of the variance equation of the discrete GARCH model as in the set of equations (4.5) are stated. These parameters are for the last window that uses the data on the short term rate from December 1986 till November 1999 (156 months).

Parameter	Estimate	Standard Error
β_0	0.00000019133	0.0000010304
β_1	0.0000204	0.0000169
β_2	0.3282	0.134
β_3	0.5536	0.1615

parameters of the LS model.

- compute the mean and variance of the time series of interest rates,
- compute the mean and variance of the time series of volatilities,
- compute the maximum and the minimum value of the ratio V/r for contemporaneous values of these factors.

The six statistics that determine the LS bond pricing model are given as

$$\begin{aligned}
\alpha &= \min\left(\frac{V_t}{r_t}\right) \\
\beta &= \max\left(\frac{V_t}{r_t}\right) \\
\delta &= \frac{\alpha(\alpha+\beta)(\beta E[r]-E[V])}{2(\beta^2(Var[r]-Var[V]))} \\
\gamma &= \frac{\delta(\beta E[r]-E[V])}{\alpha(\beta-\alpha)} \\
\xi &= \frac{\beta(\alpha+\beta)(E[V]-\alpha E[r])}{2(Var[V]-\alpha^2 Var[r])} \\
\eta &= \frac{\xi(E[V]-\alpha E[r])}{\beta(\beta-\alpha)}
\end{aligned} \tag{4.6}$$

The values of the calculated statistics by Longstaff and Schwartz (1993) are shown in Table 4.7.

Table 4.7: LS Statistics (Longstaff and Schwartz (1993))

The values of α, β, δ, γ, ξ, and η in the set of equations (4.6) as calculated in the study by Longstaff and Schwartz (1993) are stated. The authors use the data on the one month U.S. Treasury bills from January 1964 to December 1989.

statistic	Calculated Value
α	0.001149
β	0.1325
δ	0.05658
γ	3.0493
ξ	3.998
η	0.1582

The values of the statistics in the set of equations (4.6) as calculated in the study for the last window that uses the data on the short term rate from December 1986 till November 1999 (156 months) are stated. The proxy used for the short rate is the Fama and Bliss (1987) yield on the .25 year bond.

The differences in the parameters shown in Tables 4.8 and D.2 could be due to the different data sets used in both cases.

The values of the calculated statistics with the implied values of the market prices of interest rate risk (using the CIR model) for all the windows are plugged into the pricing equation (3.14) to get the predicted price of a unit discount bond.

4.2.4 A numerical example of the LS model

A numerical example is included to illustrate the empirical methodology of the LS model. Considering the first window i.e. using the data on FB yield for the .25 year

Table 4.8: LS Statistics

The values of α, β, δ, γ, ξ, and η in the set of equations (4.6) for the last window that uses the data on the short term rate from December 1986 till November 1999 (156 months) are stated.

statistic	Calculated Value
α	0.00003699
β	0.00182
δ	0.004669
γ	6.5099
ξ	0.0951
η	0.1377

Table 4.9: LS Parameter Estimates (example)

The parameters of the variance equation of the discrete GARCH model as in the set of equations (4.5) are stated. These parameters are for the first window that uses the data on the short term rate from January 1983 to December 1995(156 months).

Parameter	Estimate	Standard Error
β_0	0.000000010537	0.000001599
β_1	0.0000764	0.0000471
β_2	0.2639	0.3399
β_3	0.2726	0.1321

Table 4.10: LS Statistics (example)

The values of α, β, δ, γ, ξ, and η in the set of equations (4.6) for the last window that uses the data on the short term rate from January 1983 till December 1995(156 months) are stated.

statistic	Calculated Value
α	0.0001054
β	0.001026
δ	0.008326
γ	4.7169
ξ	0.05168
η	0.1993

bond from January 1983 to December 1995 the parameters of variance equation of the set of equations (4.5) are estimated using GARCH estimation procedure. The estimates are shown in Table 4.9. The values of the statistics for the first window are calculated using the set of equations (4.6). The statistics are stated in Table 4.10.

The values of the calculated statistics from Table 4.10 with the implied values of the market prices of risk (using the CIR model) ($\lambda = -0.03841$) for this window are plugged into the pricing equation (3.14) to get the predicted price of a unit discount bond of a particular maturity. The prices of the bonds of different maturities for this window are shown in Table 4.11.

4.3 The Heath Jarrow and Morton Models

Heath Jarrow and Morton (1992) (hereafter HJM) designed a methodology to price interest rate sensitive contingent claims given the prices of all zero coupon bonds. Their methodology imposes structure directly on the evolution of the forward rate curve and does not require an estimation of the market price of interest rate risk. The authors also

Table 4.11: LS Predicted Bond Prices

The values of the parameters from Table 4.10 are plugged into the pricing equation (3.14) to get the price of the bond of a particular maturity. The prices as produced by the pricing relation, equation (3.14) for the first trading day of January 1996 for all the different maturities are shown as LS-Price. The unsmoothed Fama and Bliss (1987) yields for the first trading day of January 1996 are converted into prices using the discount function $\exp(-(m)y)$ (m is the maturity of the bond and y is the yield) are shown as FB-Price.

Maturity	FB-Price	LS-Price
.5	0.9758	0.9753
1	0.9523	0.9511
2	0.9072	0.9044
3	0.8602	0.8597
5	0.7714	0.7764
7	0.6823	0.7005
10	0.5709	0.5996
15	0.4049	0.4612
20	0.2853	0.3537

impose a stochastic spot rate process with multiple stochastic factors influencing the term structure. The model is derived from the necessary and sufficient conditions for the absence of arbitrage. The HJM model specifies a general continuous time stochastic process for the evolution of the forward rate curve across time given the initial forward rate curve. The authors use the insights of Harrison and Kreps (1979) to characterize the conditions on the forward rate process to ensure that the process is consistent with an arbitrage free economy which in turn implies that there is a unique equivalent martingale probability measure. Under these conditions the authors state that the markets are complete and use the procedures in Harrison and Pliska (1981) to price contingent claims.

The HJM framework examined in this study is the restricted version of HJM suggested by Ritchken and Sankarasubramanian (1995). The restrictions imposed by Ritchken and Sankarasubramanian are as specified in equation (3.19) that the forward rate volatility is the spot rate volatility times an exponentially decaying function. Under these restrictions the model that allows for the variability of the spot rate volatility is called HJM-RS model from hereon and the one that has a constant spot rate volatility is called the HJM-GV (generalized Vasicek) model.

4.3.1 The HJM-RS model

The empirical methodology used to estimate the HJM models is based on the study by Bliss and Ritchken (1996) (hereafter BR). Let $y_s(t, t+m)$ represent the continuously compounded annualized forward yield over the time $[t, t+m]$ measured at date $s \leq t$. So the yield on a m-maturity pure discount bond paying a dollar at time $t+m$ and costing $B_t(t+m)$ at date t will be

$$y_s(t, t+m) = -\frac{1}{m} \ln B_t(t+m)$$

and the forward rate observed at time $s < t$ for the period t to t+m is

$$y_s(t, t+m) = +\frac{t+m-s}{m} y_s(s, t+m) - \frac{t-s}{m} y_s(s, t)$$

The equation (3.20) rewritten in yield form is

$$y_t(t,t+m)m = y_s(t,t+m)m + \left[-\frac{1}{2}\beta^2(m)\phi(t) + \beta(m)\psi(t)\right] \tag{4.7}$$

In other words the yield at time t equals the sum of its original forward yield and a deviation that is fully determined at date t by the two state variables, $\psi(t)$ and $\phi(t)$. Let

$$\Delta y_t(t,t+m) = y_t(t,t+m) - y_s(t,t+m) \tag{4.8}$$

denote the deviation between the actual yield at date t and the original forward yield at previous observation date s. The deviation is referred to as "forward rate change". By substituting (4.8) into (4.7)

$$m\Delta y_t(t,t+m) = \left\{-\frac{1}{2}\beta^2(m)\phi(t) + \beta(m)\psi(t)\right\} \tag{4.9}$$

Bliss and Ritchken suggest a transformation of the variables $\phi(t)$ and $\psi(t)$, since the two variables are not readily observed. By selecting any two distinct maturities τ_1 and τ_2 and observing the $\Delta y_t(t,t+\tau_1)$ and $\Delta y_t(t,t+\tau_2)$ the equations

$$\tau_1\Delta y_t(t,t+\tau_1) = \left\{-\frac{1}{2}\beta^2(\tau_1)\phi(t) + \beta(\tau_1)\psi(t)\right\}$$

$$\tau_2\Delta y_t(t,t+\tau_2) = \left\{-\frac{1}{2}\beta^2(\tau_1)\phi(t) + \beta(\tau_2)\psi(t)\right\}$$

can be inverted to solve for $\phi(t)$ and $\psi(t)$. Substituting the two back in equation (4.9)

$$\Delta y_t(t,t+m) = \Delta y_t(t,\tau_1)H_1(m;\tau_1,\tau_2) + \Delta y_t(t,\tau_2)H_2(m;\tau_1,\tau_2) \tag{4.10}$$

Where

$$H_1(m;\tau_1,\tau_2) \equiv \frac{\tau_1\beta(m)[\beta(\tau_2)-\beta(m)]}{m\beta(\tau_1)[\beta(\tau_2)-\beta(\tau_1)]}$$

$$H_2(m;\tau_1,\tau_2) \equiv \frac{\tau_2\beta(m)[\beta(m)-\beta(\tau_1)]}{m\beta(\tau_2)[\beta(\tau_2)-\beta(\tau_1)]}$$

Hence the equation (4.9) that involves unobservable state variables $\phi(t)$ and $\psi(t)$, is transformed into equation (4.10) that involves the observable state variables $\Delta y_t(t,t+\tau_1)$ and $\Delta y_t(t,t+\tau_2)$.

In other words, if the no arbitrage condition holds, and the restriction of the volatility structure of (3.19) holds, then given the forward rate change in any two maturities τ_1 and τ_2, the forward rate change of any arbitrary maturity can be explained. Consequently under such circumstances the equation (4.10) must hold.

It is first recognized that annualized yields are measured with error (caused due to the estimation of the term structure from actual bond prices) and assume that

$$y_t(t, t+m) = y_t^m(t, t+m) + \epsilon(t, t+m) \tag{4.11}$$

Where $y_t^m(t, t+m)$ is the measured yield on a m maturity pure discount bond at date t, and $\epsilon(t, t+m)$ is the error in the measurement. The errors are assumed to be independent and identically distributed normal random variables with mean 0 and variance η^2.

The assumed error structure for the measurement error dictates a very specific error structure for the forward rate change. Substituting equation (4.11) in (4.8) and simplifying

$$\Delta y_t^m(t, t+m) = y_t^m(t, t+m) - y_s(t, t+m) + \epsilon(t, t+m)$$

Also since

$$\begin{aligned} y_s(t, t+m) &= +\frac{t+m-s}{m} y_s(s, t+m) - \frac{t-s}{m} y_s(s, t) \\ \Delta y_t(t, t+m) &= y_t^m(t, t+m) - \frac{t+m-s}{m} y_s(s, t+m) + \frac{t-s}{m} y_s(s, t) + \epsilon(t, t+m) \end{aligned}$$

Substituting for $y_s(s, t+m)$ and $y_s(s, t)$

$$\begin{aligned} \Delta y_t^m(t, t+m) &= y_t^m(t, t+m) - \frac{t+m-s}{m} y_s^m(s, t+m) + \frac{t-s}{m} y_s^m(s, t) \\ &- \frac{t+m-s}{m} \epsilon(s, t+m) + \frac{t-s}{m} \epsilon(s, t) + \epsilon(t, t+m) \end{aligned}$$

Simplifying

$$\Delta y_t^m(t, t+m) = y_t^m(t, t+m) - y_s^m(t, t+m) - \frac{t+m}{m} \epsilon(s, t+m) + \frac{t}{m} \epsilon(s, t) + \epsilon(t, t+m) \tag{4.12}$$

Substituting

$$\epsilon^*(t,t+m) \equiv \epsilon(t,t+m) - \frac{t+m}{m}\epsilon(s,t+m) + \frac{t}{m}\epsilon(s,t)$$

in equation (4.12)

$$\Delta y_t^m(t,t+m) = y_t^m(t,t+m) - y_s^m(t,t+m) + \epsilon^*(t,t+m) \tag{4.13}$$

The $\epsilon^*(t,\cdot)$s in turn are normally distributed with mean 0. For the covariances if $m = n$

$$cov[\epsilon^*(t,t+m), \epsilon^*(t,t+n)] = Var[\epsilon] + [(\frac{t+m}{m})^2(Var[\epsilon])] + [(\frac{t}{m})^2(Var[\epsilon])]$$

Simplifying

$$cov[\epsilon^*(t,t+m), \epsilon^*(t,t+m)] = 2\eta^2(1 + \frac{t}{m} + (\frac{t}{m})^2)$$

The covariances if $m \neq n$ are

$$cov[\epsilon^*(t,t+m), \epsilon^*(t,t+n)] = \eta^2(\frac{t}{m})(\frac{t}{n})$$

Since the $\epsilon(t,\cdot)$s are independently and identically distributed and the ϵ^* is just a linear combination of normally distributed variables.

Substituting equation (4.13) into equation (4.10) yields

$$\Delta y_t^m(t,t+m) = \Delta y_t^m(t,t+\tau_1)H_1(m) + \Delta y_t^m(t,t+\tau_2)H_2(m) + \epsilon^{**}(t,t+m) \tag{4.14}$$

Where

$$\epsilon^{**}(t,t+m) \equiv \epsilon^*(t,t+\tau_1)H_1(m) + \epsilon^*(t,t+\tau_2)H_2(m) - \epsilon^*(t,t+m)$$

A similar analysis as above shows that the $\epsilon^{**}(t,\cdot)$s are again normally distributed with mean 0 and covariances given by.

$$cov[\epsilon^{**}(t,t+m), \epsilon^{**}(t,t+n)] = \begin{cases} \alpha'(m)\Sigma(m,n)\gamma(n) & \text{if } m \neq n \\ \alpha'(m)\Sigma(m,m)\gamma(m) & \text{if } m = n \end{cases}$$

Where

$$\alpha'(m) = (H_1(m), H_2(m), -1, 0)$$

$$\gamma'(m) = (H_1(n), H_2(n), 0, -1)$$

and

$$\sum(m,n) = \eta^2 \begin{pmatrix} 2(1+\frac{t}{\tau_1}+(\frac{t}{\tau_1})^2) & \frac{t^2}{\tau_1\tau_2} & \frac{t^2}{\tau_1 m} & \frac{t^2}{\tau_1 n} \\ \frac{t^2}{\tau_1\tau_2} & 2(1+\frac{t}{\tau_2}+(\frac{t}{\tau_2})^2) & \frac{t^2}{\tau_2 m} & \frac{t^2}{\tau_2 n} \\ \frac{t^2}{\tau_1 m} & \frac{t^2}{\tau_2 m} & 2(1+\frac{t}{m}+(\frac{t}{m})^2) & \frac{t^2}{mn} \\ \frac{t^2}{\tau_1 n} & \frac{t^2}{\tau_2 n} & \frac{t^2}{mn} & 2(1+\frac{t}{n}+(\frac{t}{n})^2) \end{pmatrix}$$

Hence the HJM-RS model has two unknown parameters, κ and η^2 to be estimated from the data.The two parameters are estimated by the maximum likelihood estimation procedure using the variance covariance structure developed.

Bliss and Ritchken (1996) (hereafter BR) estimate the parameters of the HJM-RS model for the years 1982 to 1994 using each year's data. BR use data on the change in forward rates for the months of January through December for a particular year to estimate the parameters of that year. The estimates of the parameters and their standard errors for each of the 13 years in their study are shown in Table 4.12.

The estimates of the parameters κ and η in the study for the last window and their standard errors are stated in Table 4.13. The last window uses the data on the change in forward rates for all the maturities from December 1998 to November 1999 (12 months).

If one compares the results of Table 4.13 and the results of Table D.4 with the results of Table 4.12 the ranges of the values of the parameters are similar. The estimates of the parameters κ and η vary from 0.1127 to 0.1952 and 0.033 to 0.0582 respectively in this study for the time period January 1996 to December 1999. Whereas the range of the estimates of the parameters κ and η vary from 0.09 to 0.28 and 0.04 to 0.20 respectively in the study by Bliss and Ritchken for the time period 1982 to 1994.

4.3.2 A numerical example of the HJM-RS model

A numerical example is included to illustrate the empirical methodology of the HJM-RS model. Considering the first window i.e. using the data on bond prices of all the

Table 4.12: HJM-RS Parameter Estimates (Bliss and Ritchken)

The parameters κ and η of the equation (4.10) and their standard errors as estimated by Bliss and Ritchken (1996) for the years 1982 through 1994 are stated. These parameters are estimated by using the data on the change in forward rates for all the maturities from January to December of that particular year.

Year	κ	Standard Error	η	Standard Error
1982	0.276	0.079	0.207	0.016
1983	0.081	0.013	0.072	0.006
1984	0.178	0.022	0.084	0.007
1985	0.104	0.014	0.067	0.005
1986	0.094	0.011	0.069	0.005
1987	0.170	0.026	0.071	0.005
1988	0.091	0.017	0.043	0.003
1989	0.349	0.039	0.053	0.004
1990	0.098	0.013	0.049	0.004
1991	0.164	0.034	0.048	0.004
1992	0.287	0.036	0.058	0.004
1993	0.207	0.022	0.046	0.004
1994	0.210	0.038	0.067	0.005

Table 4.13: HJM-RS Parameter Estimates

The parameters of the equation (4.10) are stated. These parameters are for the last window that uses the data on the change in forward rates for all the maturities from December 1998 to November 1999 (12 months).

Parameter	Estimate	Standard Error
κ	0.1952	0.023
η	0.0466	0.0032

Table 4.14: HJM-RS Parameter Estimates (example)

The parameters of the equation (4.10) are stated. These parameters are for the window that uses the data on the change in forward rates for all the maturities from January 1995 to December 1995 (12 months).

Parameter	Estimate	Standard Error
κ	0.1738	0.0211
η	0.0582	0.0039

maturities considered $(.25, 1, 2, 3, 5, 7, 10, 15,$ and $20)$ from January 1995 till December 1995 the parameters κ and η are estimated using maximum likelihood estimation procedure with the variance covariance structure developed in equation (4.14). The estimates are shown in Table 4.14

The objective at hand is to price a bond of 2 year maturity given the prices of .5 and 5 year maturity. This implies $m = 2$, $\tau_1 = .5$ and $\tau_2 = 5$. The yields on the bonds in this example are all in percentages. The betas are calculated using the relation

$$\beta(t) = \frac{1}{\kappa}(1 - e^{-\kappa t})$$

This results in $\beta(2) = 1.6894$, $\beta(.5) = 0.4789$ and $\beta(5) = 3.3408$.

In this particular case the $\Delta y_t(t, t+m)$ denotes the difference of the forward rate for the m year yield measured on December 1995 and the actual spot rate for the m year yield measured in January 1996. The notation is the same for time τ_1 and τ_2.

Using the equation (4.10) $H_1(m; \tau_1, \tau_2)$ and $H_2(m; \tau_1, \tau_2)$ are calculated and substituted to get the $\Delta y_t(t, t+m)$. The values of $H_1(m; \tau_1, \tau_2) = 0.5089$, $H_2(m; \tau_1, \tau_2) = 0.5347$ and $\Delta y_t(t, t+m) = -0.2475$. The values of $\Delta y_t(t, t+\tau_1)$ and $\Delta y_t(t, t+\tau_2)$ as observed from December 1995 to January 1996 are -0.3218 and -0.1565 respectively. This implies the $\Delta y_t(t, t+m)$ as predicted by the model for the 2 year bond is -0.2475. The forward rate for the 2 year bond in December 1995 is 5.1579. The model predicts that the yield on

Table 4.15: HJM-RS Predicted Bond Prices

The values of the parameters from Table 4.13 are plugged into the equation (4.10) to get the change in forward yield for a particular maturity. The change in forward yield is then converted into the price of a bond. The prices for the first trading day of January 1996 for all the different maturities are shown as HJM-RS-Price. The unsmoothed Fama and Bliss (1987) yields for the first trading day of January 1996 are converted into prices using the discount function $\exp(-(m)y)$ (m is the maturity of the bond and y is the yield) are shown as FB-Price.

Maturity	FB-Price	HJM-RS-Price
.5	0.9758	NA
1	0.9523	0.9525
2	0.9072	0.9065
3	0.8602	0.8604
5	0.7714	NA
7	0.6823	0.6937
10	0.5709	0.5837
15	0.4049	0.4355
20	0.2853	0.3062

the 2 year bond in January 1996 as $5.1579 - 0.2475 = 4.9104$. This is compared to the smoothed yield on a 2 year bond on January 1996 i.e. 4.87. The two yields the smoothed and the model predicted are converted into prices using the discount function $\exp(-(m)y)$ (m is the maturity of the bond and y is the yield) and the difference is called the pricing error for the HJM-RS model. The prices of the bonds as predicted by the HJM-RS model for the first trading day of January 1996 are shown in Table 4.15.

4.3.3 The HJM-GV model

The second model in the Heath Jarrow and Morton framework considered in the study is the HJM-GV model. As opposed to the HJM-RS model, in this model the volatility of the spot rate is a constant and does not depend on the level of spot rate. In such a case, $\phi(t)$ becomes

$$\phi(t) = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t}) \tag{4.15}$$

that is deterministic. Hence this model has only the state variable $\psi(t)$ requiring a single benchmark maturity τ, to be used. Again assume $y_s(t, t+m)$ represent the continuously compounded annualized forward yield over the time $[t, t+m]$ measured at date $s \leq t$. So the yield on a m-maturity pure discount bond paying a dollar at time $t+m$ and costing $B_t(t+m)$ at date 't' will be

$$y_s(t, t+m) = -\frac{1}{m} \ln B_t(t+m)$$

and the forward rate observed at time $s < t$ for the period t to t+m is

$$y_s(t, t+m) = +\frac{t+m-s}{m} y_s(s, t+m) - \frac{t-s}{m} y_s(s,t)$$

The equation (3.20) rewritten in yield form is

$$y_t(t, t+m)m = y_s(t, t+m)m + \left[-\frac{1}{2}\beta^2(m)\phi(t) + \beta(m)\psi(t)\right] \tag{4.16}$$

Where $\phi(t)$ is given by equation (4.15). In other words the yield at time 't' equals the sum of its original forward yield and a deviation that is fully determined at date 't' by the sole state variable, $\psi(t)$. Let

$$\Delta y_t(t,t+m) = y_t(t,t+m) - y_s(t,t+m) \tag{4.17}$$

denote the deviation between the actual yield at date t and the original forward yield at previous observation date s. The deviation is referred to as "forward rate change". By substituting (4.17) into (4.16)

$$m\Delta y_t(t,t+m) = \left\{-\frac{1}{2}\beta^2(m)\phi(t) + \beta(m)\psi(t)\right\} \tag{4.18}$$

As done earlier with the HJM-RS model, a transformation of the variable $\psi(t)$ is done, since the variable is not readily observed. By selecting any single maturity τ and observing the $\Delta y_t(t,t+\tau)$ the equation

$$\tau\Delta y_t(t,t+\tau) = \left\{-\frac{1}{2}\beta^2(\tau)\phi(t) + \beta(\tau)\psi(t)\right\}$$

can be inverted to solve for $\psi(t)$. Substituting it back in equation (4.18)

$$\Delta y_t(t,t+m) = \Delta y_t(t,\tau)h(m) + \sigma^2 k(m) \tag{4.19}$$

Where

$$h(m) \equiv \frac{\tau\beta(m)}{m\beta(\tau)}$$

$$k(m) \equiv \frac{\beta(m)[\beta(\tau) - \beta(m)](1 - e^{-2\kappa t})}{4\kappa m}$$

Hence the equation (4.18) that involves unobservable state variable $\psi(t)$, is transformed into equation (4.19) that involves the observable state variable $\Delta y_t(t,t+\tau)$.

Again similar to the analysis of the HJM-RS model it is first recognized that annualized yields are measured with error and assume that

$$y_t(t,t+m) = y_t^m(t,t+m) + \epsilon(t,t+m) \tag{4.20}$$

Where $y_t^m(t,t+m)$ is the measured yield on a m maturity pure discount bond at date t, and $\epsilon(t,t+m)$ is the error in the measurement. The errors are assumed to be independent and identically distributed normal random variables with mean 0 and variance η^2.

The assumed error structure for the measurement error dictates a very specific error structure for the forward rate change. Substituting equation (4.11) in (4.8) and simplifying

$$\Delta y_t^m(t, t+m) = y_t^m(t, t+m) - y_s(t, t+m) + \epsilon(t, t+m)$$

Also since

$$\begin{aligned} y_s(t, t+m) &= +\frac{t+m-s}{m} y_s(s, t+m) - \frac{t-s}{m} y_s(s, t) \\ \Delta y_t(t, t+m) &= y_t^m(t, t+m) - \frac{t+m-s}{m} y_s(s, t+m) + \frac{t-s}{m} y_s(s, t) + \epsilon(t, t+m) \end{aligned}$$

Substituting for $y_s(s, t+m)$ and $y_s(s, t)$

$$\begin{aligned} \Delta y_t^m(t, t+m) &= y_t^m(t, t+m) - \frac{t+m-s}{m} y_s^m(s, t+m) + \frac{t-s}{m} y_s^m(s, t) \\ &- \frac{t+m-s}{m} \epsilon(s, t+m) + \frac{t-s}{m} \epsilon(s, t) + \epsilon(t, t+m) \end{aligned}$$

Simplifying

$$\Delta y_t^m(t, t+m) = y_t^m(t, t+m) - y_s^m(t, t+m) - \frac{t+m}{m} \epsilon(s, t+m) + \frac{t}{m} \epsilon(s, t) + \epsilon(t, t+m) \tag{4.21}$$

Substituting

$$\epsilon^*(t, t+m) \equiv \epsilon(t, t+m) - \frac{t+m}{m} \epsilon(s, t+m) + \frac{t}{m} \epsilon(s, t)$$

in equation (4.21)

$$y_t^m(t, t+m) = y_t^m(t, t+m) - y_s^m(t, t+m) + \epsilon^*(t, t+m)$$

The $\epsilon^*(t, \cdot)$s in turn are normally distributed with mean 0. For the covariances if $m = n$

$$cov[\epsilon^*(t, t+m), \epsilon^*(t, t+n)] = Var[\epsilon] + [(\frac{t+m}{m})^2 (Var[\epsilon])] + [(\frac{t}{m})^2 (Var[\epsilon])]$$

Simplifying

$$cov[\epsilon^*(t, t+m), \epsilon^*(t, t+m)] = 2\eta^2 (1 + \frac{t}{m} + (\frac{t}{m})^2)$$

The covariances if $m \neq n$ are

$$cov[\epsilon^*(t,t+m),\epsilon^*(t,t+n)] = \eta^2(\frac{t}{m})(\frac{t}{n})$$

Since the $\epsilon(t,\cdot)$s are independently and identically distributed and the ϵ^* is just a linear combination of normally distributed variables.

Substituting equation (4.21) into equation (4.19) yields

$$\Delta y_t^m(t,t+m) = \Delta y_t^m(t,t+\tau)h(m) + \sigma^2 k(m) + \epsilon^{**}(t,t+m) \tag{4.22}$$

Where

$$\epsilon^{**}(t,t+m) \equiv \epsilon^*(t,t+\tau)h(m) - \epsilon^*(t,t+m)$$

Again $\epsilon^{**}(t,\cdot)$s are normally distributed with mean 0 and covariances given by.

$$cov[\epsilon^{**}(t,t+m),\epsilon^{**}(t,t+n)] = \begin{cases} \alpha^{*'}(m)\Sigma^*(m,n)\gamma^*(n) & \text{if } m \neq n \\ \alpha^{*'}(m)\Sigma^*(m,m)\gamma^*(m) & \text{if } m = n \end{cases}$$

Where

$$\alpha^{*'}(m) = (h(m),-1,0)$$

$$\gamma^{*'}(m) = (h(n),0,-1)$$

and

$$\sum(m,n) = \eta^2 \begin{pmatrix} 2(1+\frac{t}{\tau}+(\frac{t}{\tau})^2) & \frac{t^2}{\tau m} & \frac{t^2}{\tau n} \\ \frac{t^2}{\tau m} & 2(1+\frac{t}{m}+(\frac{t}{m})^2) & \frac{t^2}{mn} \\ \frac{t^2}{\tau n} & \frac{t^2}{mn} & 2(1+\frac{t}{n}+(\frac{t}{n})^2) \end{pmatrix}$$

Hence this model requires the estimation of the parameters κ, η^2 and σ^2. The three parameters are estimated by the maximum likelihood estimation procedure using the variance covariance structure developed above. The estimates of the parameters and their standard errors are stated in Table 4.16.

Table 4.16: HJM-GV Parameter Estimates

The parameters of the equation (4.18) are stated. These parameters are for the last window that uses the data on the change in forward rates for all the maturities from December 1998 till November 1999 (12 months).

Parameter	Estimate	Standard Error
κ	0.0681	0.0234
η	0.0685	0.0051
σ	0.6736	0.1374

Table 4.17: HJM-GV Parameter Estimates (example)

The parameters of the equation (4.18) are stated. These parameters are for the first window that uses the data on the change in forward rates for all the maturities from January 1995 till December 1995 (12 months).

Parameter	Estimate	Standard Error
κ	0.0079	0.0184
η	0.0768	0.0059
σ	0.3005	0.0868

4.3.4 A numerical example of the HJM-GV model

A numerical example is included to illustrate the empirical methodology of the HJM-GV model. Considering the first window i.e. using the data on bond prices of all the maturities considered ($.25, 1, 2, 3, 5, 7, 10, 15,$ and 20) from January 1995 till December 1995 the parameters κ, η, and σ are estimated using maximum likelihood estimation procedure with the variance covariance structure developed in equation (4.22). The estimates are shown in Table 4.17

The objective at hand is to price a bond of 2 year maturity given the price of .5 year maturity. This implies $m = 2$, $\tau = .5$. The yields on the bonds in this example are all in

Table 4.18: HJM-GV Predicted Bond Prices

The values of the parameters from Table 4.16 are plugged into the equation (4.18) to get the change in forward yield for a particular maturity. The change in forward yield is then converted into the price of a bond. The prices for the first trading day of January 1996 for all the different maturities are shown as HJM-GV-Price. The unsmoothed Fama and Bliss (1987) yields for the first trading day of January 1996 are converted into prices using the discount function $\exp(-(m)y)$ (m is the maturity of the bond and y is the yield) are shown as FB-Price.

Maturity	FB-Price	HJM-GV-Price
.5	0.9758	NA
1	0.9523	0.9527
2	0.9072	0.9077
3	0.8602	0.8629
5	0.7714	NA
7	0.6823	0.7019
10	0.5709	0.5949
15	0.4049	0.4486
20	0.2853	0.3182

percentages. The betas are calculated using the relation

$$\beta(t) = \frac{1}{\kappa}(1 - e^{-\kappa t})$$

This results in $\beta(2) = 0.2498$, $\beta(.5) = 0.4990$.

In this particular case the $\Delta y_t(t, t+m)$ denotes the difference of the forward rate for the m year yield measured on December 1995 and the actual spot rate for the m year yield measured in January 1996. The notation is the same for time τ_1.

Using the equation (4.18) $h(m;\tau)$ is calculated and substituted to get the $\Delta y_t(t, t+m)$. The value of $h(m;\tau) = 1.000987$ and $\Delta y_t(t, t+m) = -0.3144$. The value of $\Delta y_t(t, t+\tau)$ as observed from December 1995 to January 1996 is -0.3218. This implies the $\Delta y_t(t, t+m)$ as predicted by the model for the 2 year bond is -0.3144. The forward rate for the 2 year bond in December 1995 is 5.1579. The model predicts that the yield on the 2 year bond in January 1996 as $5.1579 - 0.3144 = 4.8435$. This is compared to the smoothed yield on a 2 year bond on January 1996 i.e. 4.87. The two yields the smoothed and the model predicted are converted into prices and the difference is called the pricing error for the HJM-GV model. The prices of the bonds as predicted by the HJM-GV model for the first trading day of January for different maturities are shown in Table 4.18.

4.4 The Network Models

Non-parametric techniques have become popular recently in an effort to reduce the number of arbitrary parametric restrictions imposed on the underlying process. A non-parametric approach that used an artificial neural network makes an assumption that the output is a Borel integrable function of the underlying state variables (inputs). In other words the function is allowed to have a finite number of discontinuities and bond prices as a function of the two inputs are assumed to satisfy this condition. This is a reasonable assumption because all the models considered in the study, the affine models, and most of the other term structure models satisfy this relation between the bond prices and the risk

Table 4.19: MLP Predicted Bond Prices

The values of the parameters from the set of equations (4.23) are plugged into the equation (3.24) to get the output of the network or the bond price for a particular maturity. The prices for the first trading day of January 1996 for all the different maturities are shown as MLP-Price. The unsmoothed Fama and Bliss (1987) yields for the first trading day of January 1996 are converted into prices using the discount function $\exp(-(m)y)$ (m is the maturity of the bond and y is the yield) are shown as FB-Price.

Maturity	FB-Price	MLP-Price
.5	0.9758	NA
1	0.9523	0.9519
2	0.9072	0.9055
3	0.8602	0.8583
5	0.7714	NA
7	0.6823	0.6906
10	0.5709	0.5676
15	0.4049	0.4051
20	0.2853	0.3034

factors.

4.4.1 The Multi Layer Perceptrons

In the case of a Multi Layer Perceptron (hereafter MLP) the first task at hand is to choose the number of layers and the number of neurons in each layer. In this study the MLP considered has two layers and number of neurons in the first layer are 22 and 1 in the second layer. The transfer function used in the first layer is the tansigmoid function as given by equation (3.25) and the transfer function used in the second layer is the pure linear function as given by equation (3.26).

The algorithm used is based on the usage of the flexible models i.e. MLPs by restricting the complexity of the models using Bayesian methods. Bayesian methods have become quite popular in the neural network literature since David Mackay (1992). This is because the neural networks since their inception have been notorious for overfitting the data i.e. they used to fit the in sample data well but once applied to the out of sample data the network used to produce large errors. The Bayesian methods tend to reduce this overfitting problem. The Bayesian Regularization algorithm used to train the network is explained in detail in Appendix A. In the relation between the inputs and outputs as given by equation (3.24) the following are the dimensions of the different matrices

$$\begin{aligned}
\mathbf{p}^1 &\longrightarrow 2 \times 1 \\
\mathbf{W}^1 &\longrightarrow 22 \times 2 \\
\mathbf{b}^1 &\longrightarrow 22 \times 1 \\
\mathbf{W}^2 &\longrightarrow 1 \times 22 \\
\mathbf{b}^2 &\longrightarrow 1 \times 1 \\
a^2 &\longrightarrow 1 \times 1
\end{aligned} \tag{4.23}$$

With the dimensions of each matrix in equation (3.24) as given in the set of equations (4.23), the number of parameters required to be estimated from the data are 89 (the elements of $\mathbf{W}^1$, $\mathbf{b}^1$, $\mathbf{W}^2$, and $\mathbf{b}^2$).

To apply the MLP to the pricing of the zero coupon bonds consistent with the HJM framework two maturities are chosen at first. The zero coupon bond prices of the two maturities on a monthly basis are fed to the network for the time period $1983 - 1995$. During that time period the third maturity bond price that is to be forecast is also fed to the network as a target. During the training period $1982 - 1995$ based on a function of the difference between the output of the network and the target value (explained in detail in the Appendix A) the network estimates all the parameters in the set of equations (4.23). In essence the

Table 4.20: RBF Predicted Bond Prices

The values of the parameters estimated using the data from January 1983 through December 1995 are plugged into the equation (3.27) to get the price of a bond. The prices for the first trading day of January 1996 for all the different maturities are shown as RBF-Price. The unsmoothed Fama and Bliss (1987) yields for the first trading day of January 1996 are converted into prices using the discount function $\exp(-(m)y)$ (m is the maturity of the bond and y is the yield) are shown as FB-Price.

Maturity	FB-Price	RBF-Price
.5	0.9758	NA
1	0.9523	0.9529
2	0.9072	0.9025
3	0.8602	0.8504
5	0.7714	NA
7	0.6823	0.6659
10	0.5709	0.5619
15	0.4049	0.4086
20	0.2853	0.2763

pricing formula in equation (3.24) is used to price the zero coupon bond in the out of sample period i.e. first month of 1996 after estimating all the parameters using the data from 1983 – 1995. The bond prices so obtained for the first trading day of January 1996 are shown in Table 4.19. Then the rolling window approach is taken and the network is trained using the data from second month of 1983 to first month of 1996 to price bonds in the second month of 1996. This is carried on till the out of sample pricing is done for the last month of the year 1999.

4.4.2 The Radial Basis Functions

The first task at hand in the case of Radial Basis Functions (hereafter RBF) is to choose the non-linear function for the first layer. There has been extensive research in the literature that supports the result that the choice of the non-linear function does not affect the performance of the RBF network. However the performance of this network is highly sensitive to the choice of the centers. All the set of data points is a potential set of centers but as the number of data points increases it is fairly obvious that choosing all the data points as centers is not a reasonable choice. Whenever the training data is huge, to overcome such a problem some kind of clustering algorithm must be applied to the training data to reduce the number of deployed kernels (hidden neurons) and, at the same time improve the approximation ability of the network. In many applications including the studies by Micchelli (1986) and Powell (1987) these centers are chosen to be a subset of the data points. In some of the published literature the centers are arbitrarily chosen from the set of all the data points. Such a method is not a reasonable choice as it causes poor performance, a large size of the network and numerical ill conditioning. Numerical ill conditioning is caused by the near linear dependency due to some centers being too close. In essence the choice of centers is subject to the curse of dimensionality and the performance of the RBF network relies heavily on the choice of the centers.

This study uses the orthogonal least squares algorithm as a solution to the center selection problem (the algorithm was first suggested by Chen, Cowan, and Grant(1991)). The algorithm is explained in detailed in the Appendix B. For the RBF the number of parameters to be estimated depends on the number of centers that are chosen from the set of all possible centers (the set of all values of the inputs in the training period). If the number of centers is c then the RBF has to estimate $(c + 2)$ number of parameters.

To apply the RBF network to the pricing of zero coupon bonds a similar procedure as stated for the MLP is undertaken.The zero coupon bond prices of the two maturities (consistent with the HJM framework there is no market price of risk) on a monthly basis are fed to the network for the time period January 1983 through December 1995 as inputs. During that time period the third maturity bond price that is to be forecast is also fed to the network as a target output. During the training period 1983 – 1995 based on the difference between the output of the network and the target value (explained in detail in Appendix B) the network estimates all the parameters. For the first window i.e. to price bonds on the first trading day of January 1996 the number of centers using the orthogonal least squares algorithm is found to be forty-eight. The number of parameters that need to be estimated are 50 (the 49 λs and the β in the gaussian function as given in equation (3.28)). Once the parameters are estimated using the data from January 1983 through December 1995 the pricing formula as given in equation (3.27) is used to price the zero coupon bond in the out of sample period i.e. January 1996. The different prices as predicted by the RBF network are shown in Table 4.20. This is repeated for all the maturity bond prices to be forecast during the out of sample period. Then the rolling window approach is taken and the network is trained using the data from February 1983 through January 1996 to price bonds on the first trading day of February 1996. This is carried on until the out of sample period is the first trading day of December 1999.

CHAPTER 5

The Hedging Applications

The values of assets of financial institutions are sensitive to fluctuations in interest rates. Many of the major financial institutions are involved in the trading of interest rate derivatives and market making. This highlights the importance of the ability of a financial institution to manage its exposure to fluctuating interest rates. The ability of a model to hedge a portfolio of zero coupon bonds would strongly indicate what type of a model can be used by a financial institution to hedge its exposure to fluctuating interest rates. A model that accurately hedges interest rate exposure for a portfolio of zero coupon bonds can be used to manage interest rate risk for a subset of financial institutions' activities. The study addresses this issue by applying all the different models considered in the study to the hedging of two portfolios to test which one of them is able to better hedge a portfolio of zero coupon bonds of different maturities.

The methodology used for the testing of the hedging effectiveness is based on the formation of theoretical hedge portfolios comprising zero coupon bonds of different maturities and then comparing the change in hedge portfolio value with the change in value of the hedged portfolio. The methodology consists of two portfolios

- A portfolio A to be hedged and
- A portfolio B that is used to hedge the previous portfolio.

The portfolio A to be hedged is a portfolio with fixed proportions invested in zero coupon bonds of different maturities. The proportions to be invested in the zero coupon bonds of portfolio B are determined using the sensitivities of the value of a zero coupon bond with respect to the risk factors specific to a model. Beginning the first trading day of January 1996 for a particular model the proportions to be invested in each of the bonds of portfolio B are determined using the parameters estimated for the model from the data (January 1983 through December 1995) and the sensitivities as predicted by the model. On the first trading day of February 1996 the two portfolios A, and B are repriced. Then the returns on the two portfolios A, and B are calculated over the one month period January 1996 to February 1996. The absolute value of the difference in the return of the two portfolios A, and B over the one month period is called the hedging error. If the position is completely hedged the difference ought to be zero. This process is repeated by rolling the window forward by each month i.e. using the data from February 1983 through January 1996 to estimate the parameters and forming the hedge portfolio B in February 1996. For the second window the hedge error is calculated by measuring the return on the two portfolios over the period February 1996 to March 1996. This process of rolling the window forward is repeated till the last window of hedge error measurement is from November 1999 to December 1999.

The sample set of all the possible choices of the two portfolios to be hedged is quite large given the different maturity bonds that exist. For instance if one assumes that there are only 10 different maturities that are traded, then the total number of possible choices is $(2^{10} - 1) = 1023$.

A hedging strategy for a fixed income portfolio is primarily dependent on two factors

- What type of unexpected changes in the yield curve can occur ?
- How would these unexpected changes in the yield curve affect the values of the assets in the portfolio ?

To answer such questions traditionally the fluctuations in the term structure have been categorized as

- Parallel shifts
- Change in the slope or steepness of the yield curve and

- Change in the curvature of the yield curve.

The two portfolios chosen for the testing methodology in the study possess assets that are particularly sensitive to different types of shifts in the yield curve. The study chooses two portfolios P-I and P-II. P-I is a portfolio formed of long positions in one-unit each of bonds of maturity 1, and 20 years and a short position in a bond of maturity 10 years. Portfolio P-I is formed of long and short positions in bonds of mixed maturities. This portfolio's value is expected to be more sensitive to changes in the curvature of the yield curve and the slope of the yield curve than the changes in the level of interest rates. The second portfolio chosen is P-II, a portfolio formed by taking long positions of one-unit each in bonds of maturities 10, 15, and 20 years. This portfolio's value would be very sensitive to the levels of the interest rates (especially at the long end of the yield curve) as well as being sensitive to the slope and curvature of the yield curve.

To get an idea about how the values of the two portfolios react to the different types of shifts of the yield curve the study identifies different instances in the out of sample period January 1996 through December 1999 when there were shifts in the yield curve that can be categorized as parallel shifts, change in steepness and change in curvature. The change in the values of the portfolios P-I and P-II over the three instances would indicate the sensitivities of the portfolio to different types of shifts in the yield curve.

In the sample period from January 1996 to February 1996 the term structure exhibits a parallel shift. The Figure 5.1 illustrates this parallel shift.

From October 1997 to November 1997 the term structure exhibits a change in steepness. The Figure 5.2 illustrates this change in steepness.

From August 1998 to September 1998 the term structure exhibits a twist in the yield curve. Figure 5.3 illustrates this change. During this period there were parallel shifts at the short end of the term structure (less than 2 years) and changes in both the steepness and curvature of the term structure for maturities greater than 2 years.

The values of the portfolios P-I and P-II to be hedged are

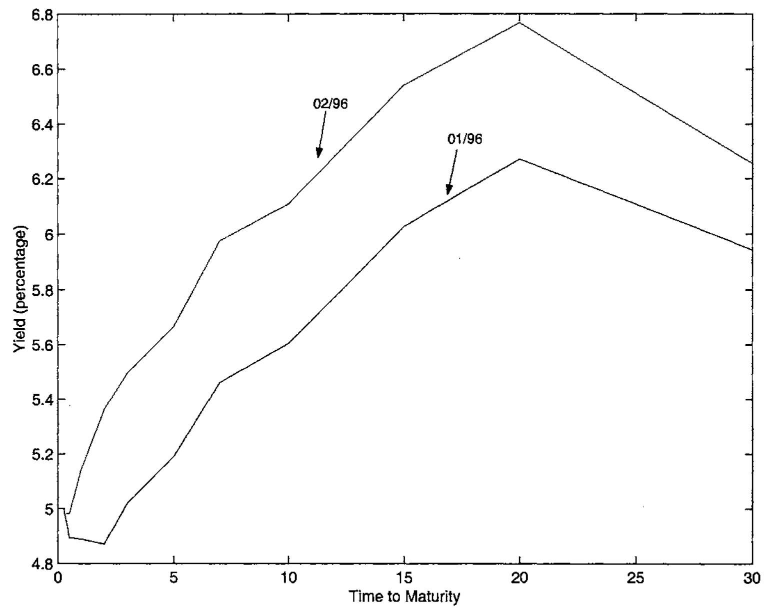

Figure 5.1: Parallel Shift

$$V_{P-I} = P^1 + P^{20} - P^{10}, \tag{5.1}$$

and

$$V_{P-II} = P^{10} + P^{15} + P^{20}. \tag{5.2}$$

Where P^i is the price of a bond with i years left to maturity. V_{P-I} and V_{P-II} denote the values of the portfolios P-I and P-II. The change in the value of the portfolios during the three sample periods of parallel, steepness and curvature shifts are shown in Table 5.1. As anticipated the portfolio P-II is more sensitive to changes in the levels of the term structure that is evident from the change in value of portfolio P-II during the first time period ($01/96 - 02/96$ when there was a parallel shift in the term structure) -6.7481 that is approximately 33 times the change in the value of portfolio P-I during the same time. Also the portfolio P-I chosen due to its sensitivity to the change in the curvature of the yield curve is evident from the change in the value of portfolio P-I during the third time period ($08/98 - 09/98$). Also the results from Table 5.1 strongly indicate that the two portfolios

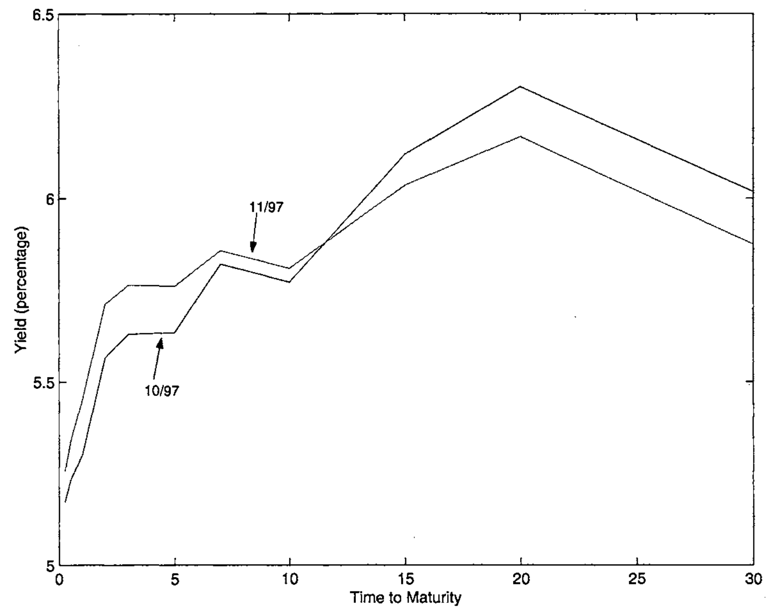

Figure 5.2: Steepness Shift

Table 5.1: Changes in the Portfolio's Value

The monthly change in the value of the portfolio to be hedged in percentage are calculated.

Time	ΔV (P-I)%	ΔV (P-II)%
01/96-02/96	0.2112	-6.7481
10/97-11/97	1.2718	0.8638
08/98-09/98	4.3269	4.4264

P-I and P-II are equally sensitive to changes in the slope and curvature of the term structure but P-II is significantly more sensitive to changes in the levels of interest rates than portfolio P-I.

The following assumptions have been made to test the hedging effectiveness

- The value of the hedge portfolio and the portfolio to be hedged is matched at the beginning of every month
- The portfolio is rebalanced every month

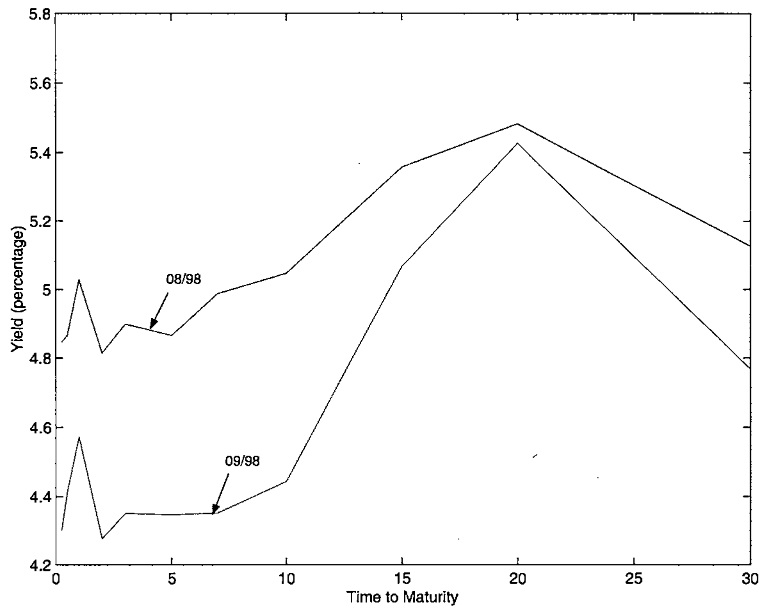

Figure 5.3: Curvature Shift

- The transaction costs are assumed to be negligible
- There are no restrictions on short selling and assets are infinitely divisible

The rest of the analysis with respect to hedging effectiveness is explained by considering only portfolio P-I. The analysis is exactly similar when considering the portfolio P-II except the composition of the hedged portfolio is different.

5.1 The Equilibrium Models

5.1.1 The CIR Model

The Cox, Ingersoll, and Ross (1985) (hereafter CIR) model considered here is a one factor model with the short term interest rate as the single risk factor. The objective of the hedging strategy is to create a hedge portfolio whose return due to a change in the risk factor is identical to the return on the hedged portfolio. The hedge portfolio must contain two assets to hedge the risk. To form a hedge portfolio is again as difficult as choosing a portfolio to be hedged. For instance from a choice of 10 maturities a portfolio of two securities can be picked in 45 ways. The hedge portfolio picked consists of proportions invested in two bonds of 2 and 3 years to maturity. The portfolios of other combinations such as 2 and 7 years to maturity, 3 and 7 years to maturity were tried and the results were not significantly different. But the author admits that it is impossible to claim that the results are invariant to all possible choices of the hedge portfolio.

Let x_i be the proportion invested in a zero coupon bond with i years to maturity and P^i be the price of the bond with maturity i, where $i = 2, 3$. Taking the partial derivative of the equation (5.1) on both sides

$$V_r = P_r^1 + P_r^{20} - P_r^{10} \tag{5.3}$$

The constraint on the portfolio weights of the hedge portfolio is that the value of the portfolio to be hedged and the hedge portfolio should have the same value at time t, that implies

$$x_2 P^2 + x_3 P^3 = P^1 + P^{20} - P^{10} \tag{5.4}$$

where P_r^i defines the partial derivative of the Bond price of maturity i with respect to the short term interest rate. Since the purpose of the hedge portfolio is to hedge the risk of the portfolio from time t to $t + \Delta t$, the return on the two portfolios in this time period should be the same. This would imply

$$x_2 P_r^2 + x_3 P_r^3 = P_r^1 + P_r^{20} - P_r^{10} \tag{5.5}$$

The functional form of the price of a bond in the CIR model is given by equation (3.10). The proportions to be invested in the two and three year maturities of the hedge portfolio can be solved for using the equations (5.4) and (5.5). This yields x_2 and x_3 as

$$x_2 = \left\{ \frac{V_r}{P_r^2} - \frac{P_r^3(V - V_r)}{(P_r^2 P^3 - P_r^2 P^2)} \right\} \tag{5.6}$$

and

$$x_3 = \left\{ \frac{\left[V - \frac{V_r P^2}{P_r^2}\right]}{\left[P^3 - \frac{P_r^3 P^1}{P_r^2}\right]} \right\} \tag{5.7}$$

The partial derivative of the price of the bond with respect to the short term rate can be obtained from the functional form of equation (3.10) i.e.

$$\left(\frac{\partial P^i}{\partial r} \right) = P^i(-B(\tau_i))$$

Based on the parameters estimated from the estimation window of $1983 - 1995$ the x's as described above are estimated and the return on the two portfolios is measured over the period January 1996 to February 1996. This process is repeated on the first business day of each month through December 1999. This approach produces forty-seven monthly observations of hedge performance.

To produce the first observation of hedge performance the data on the short term interest rates from February 1983 to January 1996 is used to estimate the parameters of the interest rate process. Using these parameters the sensitivities of the bond prices are calculated and are used to calculate the proportions invested in the bonds of the hedge portfolio. In February 1996 the hedge portfolio is marked to market using the bond prices of the two and three year maturities and the portfolio to be hedged is marked to market using the bond prices of $1, 10$ and 20 year maturities. From January to February the returns on the two portfolios are calculated in percentage and the absolute difference between the two returns is called hedging error and used as a measure of hedging performance.

5.1.2 The LS Model

In the case of the Longstaff and Schwartz (1992) (hereafter LS) model there are two risk factors short term interest rate r_t and its volatility σ. The hedge portfolio must contain three bonds to hedge these risks. The hedge portfolio used comprises of different proportions invested in three zero coupon bonds of $2, 3$, and 7 year maturities. The hedging error again is defined as the difference in the return on the portfolio to be hedged and the return on the hedge portfolio.

Let x_i be the proportion invested in a zero coupon bond with i years to maturity and P^i be the price of the bond with maturity i, where $i = 2, 3$, and 7. Taking the partial derivative of the equation (5.1) on both sides first with respect to r and σ yields

$$V_r = P_r^1 + P_r^{20} - P_r^{10} \tag{5.8}$$

$$V_\sigma = P_\sigma^1 + P_\sigma^{20} - P_\sigma^{10} \tag{5.9}$$

The constraint on the hedge portfolio proportions is that the return on the portfolio to be hedged and return on the hedge portfolio should be the same, which implies

$$\begin{aligned} P^1 + P^{20} - P^{10} &= x_2 P^2 + x_3 P^3 + x_7 P^7 \\ P_r^1 + P_r^{20} - P_r^{10} &= x_2 P_r^2 + x_3 P_r^3 + x_7 P_r^7 \\ P_\sigma^1 + P_\sigma^{20} - P_\sigma^{10} &= x_2 P_\sigma^2 + x_3 P_\sigma^3 + x_7 P_\sigma^7 \end{aligned} \tag{5.10}$$

Solving the set of equations (5.10) for the values of x_2, x_3 and x_7 gives the proportions to be invested in each of the assets. The values are

$$x_1 = \frac{-P^3 P_\sigma^7 V_r + P^3 V_\sigma P_r^7 - P^7 P_r^3 V_\sigma + P^7 P_\sigma^3 V_r - V P_\sigma^3 P_r^7 + V P_r^3 P_\sigma^7}{-P^2 P_\sigma^3 P_r^7 + P^2 P_\sigma^7 P_r^3 + P^3 P_\sigma^2 P_r^7 - P^7 P_\sigma^2 P_r^3 + P^7 P_\sigma^3 P_r^2 - P^3 P_\sigma^7 P_r^2}$$

$$
\begin{aligned}
x_2 &= \frac{-P^2 P^7_\sigma V_r + P^2 V_\sigma P^7_r + P^7 P^2_\sigma V_r + P^2_r P^7_\sigma V - V_\sigma P^2_r P^7 - V P^7_r P^2_\sigma}{-P^2 P^3_\sigma P^7_r + P^2 P^3_r P^7_\sigma + P^3 P^2_\sigma P^7_r - P^2_\sigma P^3_r P^7 + P^3_\sigma P^2_r P^7 - P^2_r P^3 P^7_\sigma} \quad (5.11) \\
x_3 &= \frac{-P^3_r P^2_\sigma V + P^2 P^3_r V_\sigma - P^2 P^3_\sigma V_r + P^2_r P^3_\sigma V + P^2_\sigma P^3 V_r - P^2_r P^7 V_\sigma}{-P^2 P^3_\sigma P^7_r + P^2 P^7_\sigma P^3_r + P^3 P^2_\sigma P^7_r - P^7 P^2_\sigma P^3_r + P^7 P^3_\sigma P^2_r - P^3 P^7_\sigma P^2_r}
\end{aligned}
$$

The partial of the price of the bond with respect to the short term rate and its volatility can be obtained from the functional form of equation (3.14) as

$$\left(\frac{\partial P^i}{\partial r}\right) = P^i(C(\tau_i))$$

and

$$\left(\frac{\partial P^i}{\partial \sigma}\right) = P^i(D(\tau_i))$$

.

Using the parameters estimated from the estimation window of February 1983 - January 1996 the hedge portfolio proportions as described above are calculated and the hedge portfolio is formed.

5.2 The Heath Jarrow and Morton Models

5.2.1 The HJM-RS Model

For the Heath Jarrow and Morton (hereafter HJM) model with varying spot rate volatility (hereafter HJM-RS) the yield of a bond with a certain maturity is a function of the two state variables, integrated variance and the ex-post forward premium. These state variables are unobservable. Though the state variables are unobservable the sensitivity of a bond's price to each state variable is observable at a point in time t. Using the equation (3.20) the first partial of a bond price with respect to the first state variable (integrated variance $\phi(t)$) is

$$\frac{\partial B_t(t+m)}{\partial \phi(t)} = \left[\frac{B(0,t+m)}{B(0,t)}\right] \exp\left[-\frac{1}{2}\beta^2(m)\phi(t) + \beta(m)\psi(t)\right] (-\frac{1}{2}\beta^2(m))$$

where

$$\beta(m) \equiv \frac{1}{\kappa}(1 - e^{-\kappa m});$$

$$\phi(t) \equiv \int_0^t \sigma_f^2(,t)ds;$$

$$\psi(t) \equiv f(0,t) - r(t).$$

This can be rewritten as

$$\frac{\partial B_t(t+m)}{\partial \phi(t)} = B_t(t+m)(-\frac{1}{2}\beta^2(m)) \tag{5.12}$$

Similarly the partial of the bond price with respect to the second state variable $\psi(t)$ is

$$\frac{\partial B_t(t+m)}{\partial \psi(t)} = B_t(t+m)(\beta(m)) \tag{5.13}$$

The hedge portfolio must contain three bonds to hedge these risks. The hedge portfolio contains three zero coupon bonds with maturities 2, 3, and 7 years.

Let x_i be the proportion invested in a zero coupon bond with τ_i years to maturity and P_i be the price of the bond with maturity τ_i, where $i = 1, 2,$ and 3. Taking the partial derivative of the equation (5.1) on both sides first with respect to $\phi(t)$ and $\psi(t)$ yields

$$V_{\phi(t)} = P^1_{\phi(t)} + P^{20}_{\phi(t)} - P^{10}_{\phi(t)} \tag{5.14}$$

$$V_{\psi(t)} = P^1_{\psi(t)} + P^{20}_{\psi(t)} - P^{10}_{\psi(t)} \tag{5.15}$$

The dependence of the two variables on t will be omitted for simplicity of presentation. The hedge portfolio proportions are chosen so that the return on the portfolio to be hedged and return on the hedge portfolio are equal i.e.

$$
\begin{aligned}
P^1 + P^{20} - P^{10} &= x_2 P^2 + x_3 P^3 + x_7 P^7 \\
P^1_\phi + P^{20}_\phi - P^{10}_\phi &= x_2 P^2_\phi + x_3 P^3_\phi + x_7 P^7_\phi \\
P^1_\psi + P^{20}_\psi - P^{10}_\psi &= x_2 P^2_\psi + x_3 P^3_\psi + x_7 P^7_\psi
\end{aligned}
\tag{5.16}
$$

Solving the set of equations (5.16) for the values of x_2, x_3 and x_7 gives the proportions to be invested in each of the assets. The values are

$$
\begin{aligned}
x_2 &= \frac{-P^3 P^7_\psi V_\phi + P^3 V_\psi P^7_\phi - P^7 P^3_\phi V_\psi + P^7 P^3_\psi V_\phi - V P^3_\psi P^7_\phi + V P^3_\phi P^7_\psi}{-P^2 P^3_\psi P^7_\phi + P^2 P^7_\psi P^3_\phi + P^3 P^2_\psi P^7_\phi - P^7 P^2_\psi P^3_\phi + P^7 P^3_\psi P^2_\phi - P^3 P^7_\psi P^2_\phi} \\
x_3 &= \frac{-P^2 P^7_\psi V_\phi + P^2 V_\psi P^7_\phi + P^7 P^2_\psi V_\phi + P^2_\phi P^7_\psi V - V_\psi P^2_\phi P^7 - V P^7_\phi P^2_\psi}{-P^2 P^3_\psi P^7_\phi + P^2 P^3_\phi P^7_\psi + P^3 P^2_\psi P^7_\phi - P^2_\psi P^3_\phi P^7 + P^3_\psi P^2_\phi P^7 - P^2_\phi P^3 P^7_\psi} \\
x_7 &= \frac{-P^3_\phi P^2_\psi V + P^2 P^3_\phi V_\psi - P^2 P^3_\psi V_\phi + P^2_\phi P^3_\psi V + P^2_\psi P^3 V_\phi - P^2_\phi P^7 V_\psi}{-P^2 P^3_\psi P^7_\phi + P^2 P^7_\psi P^3_\phi + P^3 P^2_\psi P^7_\phi - P^7 P^2_\psi P^3_\phi + P^7 P^3_\psi P^2_\phi - P^3 P^7_\psi P^2_\phi}
\end{aligned}
\tag{5.17}
$$

The partial of the price of the bond with respect to the state variables are substituted from equations (5.12) and (5.13). Based on the parameters estimated from the estimation window ending in January 1996 the hedge portfolio proportions as described above are calculated and the hedge portfolio is formed.

5.2.2 The HJM-GV Model

For the Heath Jarrow and Morton (1992) (hereafter HJM) model with the generalized Vasicek volatility i.e. constant spot rate volatility (hereafter HJM-GV) the yield of a bond with a certain maturity is a function of the one state variable, the ex-post forward premium. This risk factor is unobservable. Though the state variable is unobservable the sensitivity of a bond price to the state variable is observable at a certain point in time t. The first partial of a bond price with respect to the state variable $\psi(t)$ is

$$\frac{\partial B_t(t+m)}{\partial \psi(t)} = \left[\frac{B(0,t+m)}{B(0,t)}\right] \exp\left[-\frac{1}{2}\beta^2(m)\phi(t) + \beta(m)\psi(t)\right] (\beta(m)) \tag{5.18}$$

where

$$\beta(m) \equiv \frac{1}{\kappa}(1 - e^{-\kappa m});$$
$$\phi(t) = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t});$$
$$\psi(t) \equiv f(0,t) - r(t).$$

This can be rewritten as

$$\frac{\partial B_t(t+m)}{\partial \psi(t)} = B_t(t+m)(\beta(m)) \tag{5.19}$$

The hedge portfolio must contain two assets to hedge the risk. The hedge portfolio contains zero coupon bonds of 2 and 3 year maturities.

Let x_i be the proportion invested in a zero coupon bond with i years to maturity and P^i be the price of the bond with maturity i, where $i = 2$, and 3. Taking the partial derivative of the equation (5.1) on both sides first with respect to $\psi(t)$ yields

$$V_{\psi(t)} = P^1_{\psi(t)} + P^{20}_{\psi(t)} - P^{10}_{\psi(t)} \tag{5.20}$$

The hedge portfolio proportions are chosen so that the return on the portfolio to be hedged and return on the hedge portfolio are equal i.e.

$$\begin{aligned} P^1 + P^{20} - P^{10} &= x_2 P^2 + x_3 P^3 \\ P^1_\psi + P^{20}_\psi - P^{10}_\psi &= x_2 P^2_\psi + x_3 P^3_\psi \end{aligned} \tag{5.21}$$

Solving the set of equations (5.22) for the values of x_2 and x_3 gives the proportions to be invested in each of the assets. The values are

$$
\begin{aligned}
x_2 &= \frac{VP^3_\psi - P^3V_\psi}{P^2P^3_\psi - P^3P^2_\psi} \\
x_3 &= \frac{V_\psi P^2 - VP^2_\psi}{P^2P^3_{\psi(t)} - P^3P^2_\psi}
\end{aligned} \tag{5.22}
$$

The partial of the price of the bond with respect to the state variable is substituted from equation (5.18). Based on the parameters estimated from the estimation window ending in January 1996 the hedge portfolio proportions as described above are calculated and the hedge portfolio is formed.

5.3 The Network Models

5.3.1 The MLP Network

For the MLP and RBF networks the two risk factors are the two inputs that are given to the network, $P^{.5}$ and P^5 where P^i denotes the price of a zero coupon bond with maturity i years. The objective of the hedging strategy is to hedge the risk due to those risk factors. The hedge portfolio must contain three bonds to hedge these risks. The hedge portfolio contains three zero coupon bonds with maturities 2, 3, and 7 years.

Let x_i be the proportion invested in a zero coupon bond with i years to maturity and P^i be the price of the bond with maturity i, where $i = 2, 3$, and 7. Taking the partial derivative of the equation (5.1) on both sides first with respect to $P^{.5}$ and then with respect to P^5 yields

$$V_{P.5} = P^1_{P.5} + P^{20}_{P.5} - P^{10}_{P.5} \tag{5.23}$$

$$V_\sigma = P^1_{P5} + P^{20}_{P5} - P^{10}_{P5} \tag{5.24}$$

The constraint on the hedge portfolio proportions is that the return on the portfolio to be hedged and return on the hedge portfolio should be the same, which implies

$$
\begin{aligned}
P^1 + P^{20} - P^{10} &= x_2 P^2 + x_3 P^3 + x_7 P^7 \\
P^1_{P.5} + P^{20}_{P.5} - P^{10}_{P.5} &= x_2 P^2_{P.5} + x_3 P^3_{P.5} + x_7 P^7_{P.5} \\
P^1_{P5} + P^{20}_{P5} - P^{10}_{P5} &= x_2 P^2_{P5} + x_3 P^3_{P5} + x_7 P^7_{P5}
\end{aligned}
\tag{5.25}
$$

Solving the set of equations (5.25) for the values of x_1, x_2 and x_3 gives the proportions to be invested in each of the assets. The values are

$$
\begin{aligned}
x_2 &= \frac{-P^3 P^7_{P5} V_{P.5} + P^3 V_{P5} P^7_{P.5} - P^7 P^3_{P.5} V_{P5} + P^7 P^3_{P5} V_{P.5} - V P^3_{P5} P^7_{P.5} + V P^3_{P.5} P^7_{P5}}{-P^2 P^3_{P5} P^7_{P.5} + P^2 P^7_{P5} P^3_{P.5} + P^3 P^2_{P5} P^7_{P.5} - P^7 P^2_{P5} P^3_{P.5} + P^7 P^3_{P5} P^2_{P.5} - P^3 P^7_{P5} P^2_{P.5}} \\
x_3 &= \frac{-P^2 P^7_{P5} V_{P.5} + P^2 V_{P5} P^7_{P.5} + P^7 P^2_{P5} V_{P.5} + P^2_{P.5} P^7_{P5} V - V_{P5} P^2_{P.5} P^7 - V P^7_{P.5} P^2_{P5}}{-P^2 P^3_{P5} P^7_{P.5} + P^2 P^3_{P.5} P^7_{P5} + P^3 P^2_{P5} P^7_{P.5} - P^2_{P5} P^3_{P.5} P^7 + P^3_{P5} P^2_{P.5} P^7 - P^2_{P.5} P^3 P^7_{P5}} \\
x_7 &= \frac{-P^3_{P.5} P^2_{P5} V + P^2 P^3_{P.5} V_{P5} - P^2 P^3_{P5} V_{P.5} + P^2_{P.5} P^3_{P5} V + P^2_{P5} P^3 V_{P.5} - P^2_{P.5} P^7 V_{P5}}{-P^2 P^3_{P5} P^7_{P.5} + P^2 P^7_{P5} P^3_{P.5} + P^3 P^2_{P5} P^7_{P.5} - P^7 P^2_{P5} P^3_{P.5} + P^7 P^3_{P5} P^2_{P.5} - P^3 P^7_{P5} P^2_{P.5}}
\end{aligned}
$$

From the training algorithm in the Appendix A the backpropagation at the last iteration gives the derivatives of the output w.r.t. the weights and biases using chain rule in equations (A.4) and (A.5). Using the same algorithm (based on backpropagation) the derivatives with respect to the inputs are found by substituting the partial of the net input with respect to the weights and biases with the partial of the net input with respect to the inputs. These derivatives are used as the sensitivities of the prices of the bonds (used as outputs) w.r.t. the prices of the two target maturity bonds.

Based on the network trained from the estimation window of February 1983 - January 1996 the x's as described above are estimated and the hedge portfolio is formed.

5.3.2 The RBF Network

For the radial basis function (hereafter RBF) networks just like in MLPs, the two risk factors are the two inputs that are given to the network. So the complete procedure for the hedging is exactly the same as that for the MLP and it differs only in the derivatives that are substituted from the training algorithm.

The partial derivative of the network output with respect to the network input that is required for the calculation of the hedge portfolio proportions are calculated using equation (5.26) using the same notation as used by equation (3.27) for the pricing of bonds.

$$\frac{\partial f_r(x)}{\partial x} = \sum_{i=1}^{n_r} \lambda_i \cdot \frac{\partial \phi(||x - c_i||)}{\partial x} \tag{5.26}$$

The partial of the gaussian function ϕ with respect to the input is calculated using the relation

$$\frac{\partial \phi(||x - c_i||)}{\partial x} = \exp\left(\frac{-(||x - c_i||)^2}{\beta^2}\right) \cdot \left(\frac{(x - c_i)}{(||x - c_i||)}\right) \tag{5.27}$$

5.4 Traditional Hedging Measures

5.4.1 Duration Based Hedging

Traditionally Macaulay's duration has been used to hedge a portfolio of zero coupon bonds. Duration based hedging is very simple to implement as compared to the model based hedging suggested so far. It is an empirical question if the extra complexity involved in the implementation of the model based hedges improves hedge performance relative to duration based hedging.

The hedge errors for the duration based hedging are constructed in a similar fashion. The hedge portfolio must contain two zero coupon bonds to hedge the risk. The hedge portfolio contains two zero coupon bonds with maturities 2, and 3 years. If the value and the durations of the two portfolios (hedge portfolio and the portfolio to be hedged) are to be matched, then this would imply

$$\begin{aligned} P^1 + P^{20} - P^{10} &= x_2 P^2 + x_3 P^3 \\ 1P^1 + 20P^{20} - 10P^{10} &= 2x_2 P^2 + 3x_3 P^3 \end{aligned} \tag{5.28}$$

Since the duration of a zero coupon bond is its maturity and a duration of a portfolio of bonds is the weighted average of the durations of the individual bonds. Solving the set of equations 5.28 for the values of x_2 and x_3 yield their values as

$$x_2 = \frac{3(P^1 + P^{20} - P^{10}) - (1P^1 + 20P^{20} - 10P^{10})}{P^2}$$

$$x_3 = \frac{(1P^1 + 20P^{20} - 10P^{10}) - 2(P^1 + P^{20} - P^{10})}{P^3}$$

Consistent with all the other testing procedures the hedge portfolio proportions are calculated on the first trading day of January 1996 and the hedging error is calculated in February 1996.

5.4.2 The Principal Component Analysis

Principal components analysis (hereafter PCA) has also been used to hedge fixed income portfolios. Though PCA is not as simple as Macaulay's duration it is simpler than all the model based hedges to implement. The study compares the hedge errors from PCA with the other model based hedges. If PCA performs better, then given it's simplicity, a hedging strategy based on PCA would be preferable.

The estimation procedure for the PCA is explained in Appendix C in detail. In the case of the factors extracted by principal component analysis it turns out that the hedging strategy can be designed in a simple way. By construction, the factors are all independent of one another and hence they are not correlated to one another. The change in a zero coupon interest rate y_t, for maturity t, is related to the factor shocks, F_i where $i = 1, 2, \ldots m$ (m is the number of factors)

$$dy_t = \sum_{i=1}^{m} L_{i,t} F_i$$

The change in the value of a portfolio of interest rate-sensitive cash flows, P is a function of changes in each interest rate indicated by

$$\begin{aligned} dP &= \sum_{t=1}^{M} \frac{\partial P}{\partial y_t} dy_t = \sum_{t=1}^{M} (-t) CF_t \exp(-ty_t) dy_t \\ &= \sum_{t=1}^{M} \left[tCF_t \exp(-ty_t \sum_{i=1}^{m} L_{i,t} F_i \right]. \end{aligned}$$

Dividing throughout by P and rearranging gives

$$\begin{aligned} \frac{dP}{P} &= -\sum_{i=1}^{m} \left[\sum_{t=1}^{M} tw_t L_{i,t} \right] F_i \\ &= -\sum_{i=1}^{m} S_i F_i \end{aligned}$$

The S_i measures the sensitivity of the portfolio to the i^{th} factor. The factor sensitivities combine linearly. So to devise a hedging strategy on a general level

$$S_i^p = \sum_{j=1}^{N} x_j S_i^j,$$

where the x_j are the proportions invested in each security. The hedging strategy in this case comprises the choice of the hedge portfolio proportions such that the factor sensitivities of the portfolio to be hedged and the factor sensitivities of the hedge portfolio are identical.

The factors are estimated using the procedure explained in Appendix C and are shown in Table 5.2 for the last window of December 1999. The data on the changes in the yields for the maturities $.25, .5, 1, 2, 3, 5, 7, 10, 15,$ and 20 years is used from December 1986 to November 1999.

The percentage of the variance explained by each of the three factors in the last window i.e. for the period December 1987 to November 1999 are shown in Table 5.3.

Table 5.2: The Three Principal Components

The three principal components for the maturities $.25, .5, 1, 2, 3, 5, 7, 10, 15$, and 20 years using the data from December 1986 to November 1999 on the change in yields are stated.

Maturity	Component 1	Component 2	Component 3
.25	0.297	0.6851	0.1613
.5	0.2913	0.3851	0.2499
1	0.3121	0.057	0.36
2	0.31	-0.1965	0.3262
3	0.3008	-0.3076	0.2668
5	0.3055	-0.3132	0.1094
7	0.2929	-0.2865	0.0045
10	0.3055	-0.1438	-0.23
15	0.289	-0.0703	-0.3385
20	0.2794	-0.0084	-0.3828

Table 5.3: Explained Variance

The explained variance as a proportion of the total due to each principal component based on the eigen values of the variance covariance matrix of the change in yields for all the maturities are stated.

Factors	Explained Variance	Cumulative Explained Variance	Unexplained Variance
Factor 1	.8359	.8359	.1641
Factor 2	.0987	.9346	.0654
Factor 3	.0306	.9652	.0348

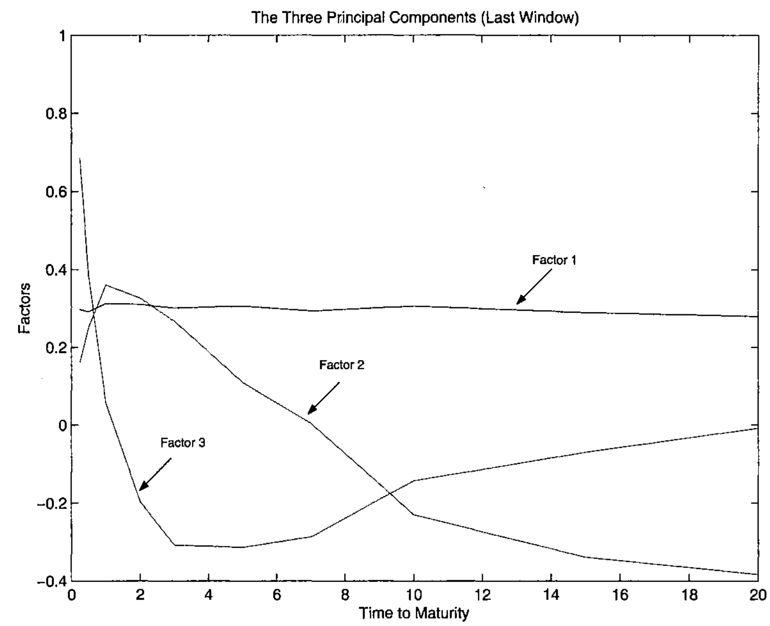

Figure 5.4: Principal Components

The results in Table 5.3 suggest that the first factor explains 83.59 percent of the total variance of the changes in the yields for all the bonds considered in the last window. The second factor explains 9.87 percent of the total variance and the two factors together explain 93.46 percent of the total variance. The third factor individually explains only 3.06 percent of the total variance.

The plot of the three factors versus the time to maturity of the zero coupon bonds as shown in Figure 5.4 clearly indicates that the first factor affects all maturities equally. This factor has been associated with changes in term structure levels by Litterman and Schienkman (1991). The second factor is related to the changes in the steepness of the term structure and the third factor is related to the changes in the curvature of the term structure. From the Figure it is evident that to some extent the second factor and the third factor do affect the steepness and curvature of the yield curve respectively. For the purpose of this study if the three factors of the PCA are actually able to capture the three different types of unexpected shifts in the yield curve, levels, steepness, and curvature then a hedging strategy based on the principal components analysis should perform the best whenever all three types of shifts in the yield curve occur.

Based on the principal components estimated from the estimation window of January 1983 through December 1995 the hedge portfolio proportions as described above are calculated to form the hedge portfolio in January 1996. In February 1996 the hedging errors are measured and this window is rolled forward till December 1999.

CHAPTER 6

Results

6.1 The Pricing Applications

For the Cox Ingersoll and Ross (1985) (hereafter CIR) model the parameters of the short rate process κ, θ, and σ in equation (3.9) are calculated from the parameters of the set of equations (4.1) using the set of equations (4.2). The parameters of the set of equations (4.1) are estimated using Generalized Method of Moments. The proxy used for the short term rate is the unsmoothed Fama and Bliss (1987) (hereafter FB) yield extracted for a bond that matures in .25 years. An additional measure the market price of interest rate risk is needed to price a unit discount bond using the CIR model. The ask price (implied by the ask discount rate) of Treasury Bills maturing in one year from the first trading day of each month beginning January 1996 till December 1999 is used. The market price at each date is implied by minimizing the squared difference between the CIR predicted price for the Bill using parameters estimated for that date and the actual price. On the first trading day of January 1996 the parameters of the model are estimated using the data on the short term rate from January 1983 through December 1995. Given the short term rate and the market price of risk on the first trading day of January 1996 zero coupon bonds of different maturities are priced using the parameters estimated and equation (3.10).

For the Longstaff and Schwartz (1992) (hereafter LS) model the stochastic processes of short term rate and its volatility given by equations (3.11) and (3.12) are proxied by a discrete GARCH model as in the set of equations (4.5). On the first trading day of January 1996 the parameters of the model are estimated using the data on the short term rate from January 1983 through December 1995. The proxy used for the short term rate is the unsmoothed Fama and Bliss (1987) (hereafter FB) yield extracted for a bond that matures in .25 years. The market price of risk used for the LS model on a particular date is the same used for the CIR model for that particular date. Given the short term rate, the GARCH estimate of volatility as predicted by the estimated GARCH model, and the market price of risk on the first trading day of January 1996 zero coupon bonds of different maturities are priced using the parameters estimated and equation (3.14).

In the case of Heath Jarrow and Morton (1992) (hereafter HJM) models the data on the changes in the forward rates for the 10 maturities (.25, .5, 1, 2, 3, 5, 7, 10, 15, and 20 years) for the time period January 1995 to December 1995 are used to estimate the parameters of the equation (4.10), for the HJM model with Ritchken and Sankarasubramanian (1995) (hereafter HJM-RS) restrictions and equation (4.19) for the HJM model with a constant spot rate volatility i.e. generalized Vasicek (hereafter HJM-GV) model. Change in forward rates as defined here for a particular maturity m is the forward rate measured for m maturity bond in December minus the spot rate for the m maturity bond in January. Using the data on the changes in the forward rate of the two benchmark maturities (.5 and 5 yr.) for the month of January 1996 and the estimated parameters, the change in forward rates for all the other maturities are calculated. Change in forward rates for all the other maturities are in turn converted into bond prices for different maturities.

Using 13 years of data on the whole term structure for HJM would be tantamount to using $1248(12 \times 13 \times 8)$ observations as opposed to 156 in the equilibrium models. This imposes the restriction that the cross sectional variance covariance matrix of the change in forward rates of different maturities is constant through time from January 1983 through 1995. These restrictions on the variance covariance structure of the changes in yields of different maturity bonds over a period of 13 years are not consistent with the model leading to non-convergence of the likelihood functions of the HJM models to a unique maximum. As a result the HJM models use only one year's data on the whole term structure as compared to equilibrium models that use 13 years of data but only on one point on the term structure (short term rate).

For the Network models the data on the changes in the forward rates for the 2 benchmark maturities (0.5 and 5 years)over the period of January 1983 through December 1995 are used to train the networks to be implemented for the first trading day of January 1996. On the first trading day of January 1996 using the trained network and the changes in forward rates for the two benchmark maturities from December 1995 to January 1996, the changes in forward rates for all the other maturities are predicted. These are converted into bond prices for maturities $.25, .5, 1, 2, 3, 5, 7, 10, 15$, and 20 years.

The difference between the price as predicted by a model and the price implied by the FB yield for a particular maturity is defined as the pricing error. Then the window is rolled over by throwing out January of 1983 and including the January of 1996 to predict prices for February of 1996. This process is repeated to produce forty-eight pricing error observations till December 1999. The vector of forty-eight pricing errors is used to construct the Mean Absolute Deviation (hereafter MAD) and Root Mean Square Error (hereafter RMSE) for each maturity. The MAD and the RMSE for the different maturities over the forty-eight windows are presented in Table 6.1 and Table 6.2. The errors are in cents on a bond with a face value of one-dollar.

To compare the models amongst one another and determine if the differences are statistically significant, the modified Diebold Mariano (1995) (hereafter MDM) statistic is calculated.

For a pair of models in a set of a step ahead pricing error, let the errors be (e_{1t}, e_{2t}) where $t = 1, \ldots, 48$. In this case the pricing performance is judged on the squared error function $g(e) = e^2$. The null hypothesis of equality of expected pricing performance is

$$E[g(e_{1t}) - g(e_{2t})] = 0$$

The Diebold Mariano statistic (1995) for such a case is

$$S_{dm} = (\hat{V}(\bar{d}))^{-\frac{1}{2}}\bar{d},$$

Table 6.1: Mean Absolute Deviation (cents)

The mean absolute deviations are calculated for the six models during the sample period January 1996 through December 1999 for the forty-eight months. The errors are in cents on a bond with a face value of one-dollar.

Maturity	CIR	LS	HJM-RS	HJM-GV	MLP	RBF
.25	0.0183	0.0011	0.0198	0.0204	0.0177	0.0196
1	0.1149	0.3118	0.0804	0.0862	0.0718	0.1163
2	0.4474	0.9456	0.1007	0.2081	0.0872	0.2617
3	0.9434	1.5500	0.1020	0.3404	0.0919	0.4537
7	5.2787	3.6664	0.6116	0.9697	0.6207	0.7857
10	8.5037	3.9415	0.5275	0.9906	0.4677	1.0694
15	8.8293	6.1221	2.5784	2.8044	2.6478	1.9553
20	8.8113	6.1042	0.9606	1.3274	1.0116	1.2383

Table 6.2: Root Mean Squared Error(cents)

The root mean squared errors are calculated for the six models during the sample period January 1996 through December 1999 for the forty-eight months. The errors are in cents on a bond with a face value of one-dollar.

Maturity	CIR	LS	HJM-RS	HJM-GV	MLP	RBF
.25	0.0205	0.0012	0.0247	0.0246	0.0215	0.0250
1	0.1424	0.3538	0.1018	0.1116	0.0874	0.1557
2	0.5969	1.0692	0.1279	0.2622	0.1222	0.3966
3	1.3533	1.7762	0.1361	0.4225	0.1334	0.8615
7	7.1748	4.1186	0.6626	1.2140	0.6784	1.8906
10	10.9092	4.4966	0.6713	1.2547	0.6099	1.8906
15	11.0637	6.6277	2.8708	3.1703	2.9557	3.6718
20	11.2179	6.5897	1.2689	1.7072	1.3001	1.7561

where

$$d_t = g(e_{1t}) - g(e_{2t}),$$

$$\bar{d} = \frac{1}{48}\sum_{t=1}^{48} d_t$$

$$V(\bar{d}) \approx \frac{}{(48)^2}\sum_{t=1}^{48}(d_t - \bar{d})^2$$

Under the null hypothesis (that the expected value of the difference between the squared errors of the two models is zero) the statistic S_{dm} has an asymptotic standard normal distribution. Harvey, Leybourne and Newbold (1997) suggest an approximately unbiased estimator of the variance of $\bar{d}$ which is

$$S_{mdm} = \left[\frac{n + 1 - 2h + n^{-1}h(h-1)}{n}\right]^{1/2} S_{dm} \tag{6.1}$$

Where h is the number of steps ahead forecast and n is the number of observations. For this study h is 1 and n is forty-eight. In this application the MDM statistic can be compared with critical values from the Student's t distribution with forty-seven degrees of freedom rather than the standard normal distribution.

The comparison of term structure models is done using the MDM statistic in two phases. In the first phase the MDM test is performed within each class of models Equilibrium (CIR vs. LS), HJM (HJM-RS vs. HJM-GV), and Neural Networks (MLP vs. RBF). In phase two the superior models from each class are compared against one another.

The MDM statistic for each of the three categories Equilibrium models, HJM models and Network models are presented in Table 6.3, Table 6.4 and Table 6.5 respectively. The differences of the squares of the error are calculated in the way suggested by the nomenclature of the comparison. For instance the statistic for the CIR-LS model is calculated by subtracting the square of the error for the LS model from that for the CIR model. This implies that if the statistic is positive the errors for CIR model were greater than the errors for LS model and vice versa. The null hypothesis for the tests is that the difference between the squared errors of the two models is zero and the alternate hypothesis is that the difference is different from zero. These values can be compared to critical values from Student's t distribution with forty-seven degrees of freedom, to gauge the statistical significance.

Table 6.3: MDM Statistic(Equilibrium Models)

The MDM statistic is calculated for all the maturities using equation (6.1) for the equilibrium models CIR and LS and the sample period January 1996 through December 1999 for the forty-eight months. (* 1% level, ** 5% level, *** 10 % level)

Maturity	CIR-LS
.25	6.972*
1	-7.333*
2	-4.340*
3	-2.001**
7	2.493*
10	3.541*
15	2.975*
20	3.322*

The results from Table 6.3 show that for maturities 2 and 3 years CIR model performs significantly better than the LS model at a level of 1% and 5% respectively[8]. Whereas the LS model outperforms the CIR model for maturities $7, 10, 15$, and 20 years at a level of 1%. This is not surprising given the structure of the two models. CIR model is a single factor (3 month yield) model whereas the two factor LS model specifies both the 3 month yield and its volatility as risk factors. This leads to the poor performance of the CIR model in pricing bonds of longer maturities. A surprising finding in Table 6.3 is that the CIR also misprices the .25 year bond significantly when compared to the LS model. Both the CIR

[8]The 1 year bond result is not considered because the CIR model uses the 1 year bond price to extract the market price of interest rate risk

and LS model have the short term rate (the yield on a .25 year bond) as an input but the LS model also has the volatility of the short term rate as the second input. For the .25 year bond the results indicate that a model that considers the short rate volatility does better than the one that does not. Neither equilibrium model is clearly superior for all maturities, so both the models are taken to phase two for further comparisons.

Table 6.4: MDM Statistic(HJM Models)

The MDM statistic is calculated for all the maturities using equation (6.1) for the HJM models HJM-RS and HJM-GV and the sample period January 1996 through December 1999 for the forty-eight months.(* 1% level, ** 5% level, *** 10 % level)

Maturity	RS-GV
.25	0.124
1	-2.151**
2	-4.775*
3	-4.788*
7	-3.730*
10	-3.965*
15	-1.556
20	-3.360*

The results in Table 6.4 show that for all maturity bonds except the .25, 1, and 15 year bonds the HJM-RS model significantly outperforms the HJM-GV model at a level of 1%. HJM-RS model outperforms HJM-GV model in the case of 1 year bond at a level of 5%. In this category clearly the HJM-RS model outperforms the HJM-GV model. These findings are consistent with the findings of Bliss and Ritchken (1996). HJM-GV model implies constant spot rate volatility whereas the HJM-RS allows for the variability of the spot rate volatility with time. The statistically smaller pricing errors of HJM-RS model are consistent with evidence that the spot rate volatility is not constant.

Table 6.5: MDM Statistic(Network Models)

The MDM statistic is calculated for all the maturities using equation (6.1) for the network models MLP and RBF and the sample period January 1996 through December 1999 for the forty-eight months. (* 1% level, ** 5% level, *** 10 % level)

Maturity	RBF-MLP
.25	0.988
1	1.866***
2	1.666***
3	1.476
7	1.120
10	1.736***
15	0.708
20	1.344

The results in the Table 6.5 indicate that on average, the pricing errors of an RBF network are greater than those for an MLP network. But the difference is significant at only the 10% level for 1, 2, and 10 year maturities. In this category the MLP network emerges as the winner in terms of it ability to price bonds. Different types of Neural Networks are suitable to model different types of data. This implies that an MLP network (using Bayesian Regularization as the training algorithm) is better suited to capture the dynamics of the term structure of interest rates than an RBF network (using the orthogonal least squares training algorithm).

During phase two the term structure models CIR, LS, HJM-RS, and MLP are compared against one another. The results for the comparison of CIR, HJM-RS and MLP models are shown in Table 6.6.

The first column of Table 6.6 indicates that the MLP network performs significantly better than the CIR model for the 3, 10, 15, and 20 year bonds at a level of 1%. The market

Table 6.6: MDM Statistic (CIR, HJM-RS, and MLP)

The MDM statistic is calculated for all the maturities using equation (6.1) for the three matched pairs of CIR-MLP, HJMRS-MLP, and HJMRS-CIR and the sample period January 1996 till December 1999 for the forty-eight months. (* 1% level, ** 5% level, *** 10 % level)

Maturity	CIR-MLP	HJMRS-MLP	HJMRS-CIR
.25	-0.423	1.882***	1.411
1	-3.946*	1.915***	4.101*
2	0.948	0.393	-0.820
3	3.128*	0.132	-3.129*
7	2.459**	-0.549	-2.536**
10	3.522*	1.733***	-3.504*
15	3.020*	-2.261**	-3.062*
20	4.209*	-0.765	-4.215*

price of risk used for the CIR model was implied from the 1 year treasury bonds and hence it is not surprising that the CIR model prices the 1 year bond the best. The CIR model being a single factor model imposes restrictions on the yield curve that all movements in the different interest rates are perfectly correlated. This indicates that the MLP network succeeds in capturing the imperfect correlation between the movements of interest rates of different maturities.

The second column of Table 6.6 implies that the MLP performs significantly better than the HJM-RS model for the .25, 1, and 10 year bonds at a level of 10%. The HJM-RS model outperforms the MLP network significantly only for the 15 year bond. At this stage it is appropriate to mention that the Neural Networks were trained using the data such that it is consistent with the HJM category models. In other words the inputs to the network were not learnt from any kind of training of the network. They were decided even before the design of the network, to make the informational set used by the network consistent with that of the HJM models. This implies that this might not be the best possible performance of MLP network trained using Bayesian Regularization. Despite this, the network performs better than the HJM-RS model. The better performance of the MLP network than the HJM-RS model strongly indicates that a non-parametric approach using the MLP (trained by Bayesian Regularization) can be effectively applied to the modelling of the term structure of interest rates.

The third column of Table 6.6 indicates that the HJM-RS model outperforms the CIR model significantly for the 3, 10, 15, and 20 year bonds at 1% significance level. Analogous to the CIR and MLP comparison the CIR outperforms the HJM-RS model for the 1 year bond at the 1% significance level.

When compared to the LS model the MLP and HJM-RS models both have significantly smaller pricing errors than the LS model at a level of 1% for all except the 3 month securities. This is not surprising because the 3 month yield is an input to the LS model and hence the LS model prices it the best. These statistics are reported in Table 6.7.

Table 6.7: MDM Statistic (LS, HJM-RS, and MLP)

The MDM statistic is calculated for all the maturities using equation (6.1) for the matched pairs LS-MLP and HJMRS-LS and the sample period January 1996 till December 1999 for the forty-eight months. (* 1% level, ** 5% level, *** 10 % level)

Maturity	LS-MLP	LS-HJMRS
.25	-5.7917*	-5.2189*
1	7.4801*	7.1671*
2	8.3082*	8.2925*
3	8.3078*	8.2948*
7	8.6883*	8.6607*
10	7.8486*	7.7767*
15	9.0003*	9.0545*
20	9.9742*	9.9482*

6.2 The Hedging Applications

Relative hedging performance is measured using two portfolios P-I and P-II to be hedged. P-I is a portfolio with long positions of one-unit each in a 1 year and 20 year bonds and a short position of one-unit in a 10 year bond. P-II is a portfolio with long positions in 10, 15, and 20 year bonds.

For the CIR model the hedge portfolio contains bonds with maturities 2 and 3 years. For the LS model the hedge portfolio contains bonds with maturities 2, 3, and 7 years. In the case of HJM-RS model the hedge portfolio contains bonds with maturities 2, 3, and 7 years. For the HJM-GV model the hedge portfolio contains bonds with maturities 2 and 3 years. For the Network models the hedge portfolio contains bonds of maturities 2, 3 and 7 years.

The same estimation windows as used in the measurement of price errors are considered to estimate the hedge portfolio proportions. The only difference is that to form a hedge

Table 6.8: Hedging Errors

In February 1996 the hedging error for a particular model is measured as the difference in return of the hedge portfolio and the portfolio to be hedged over the period January 1996 to February 1996. The two portfolios to be hedged are P-I and P-II. P-I is a portfolio of one-unit long positions in each of 1 year and 20 year bonds and one-unit short position in 10 year bond. P-II is a portfolio of one-unit long positions in each of the 10, 15, and 20 year bonds. The mean and variance of the absolute value of this difference over the period of February 1996 to December 1999 i.e. forty-seven months are calculated. The errors are in percentages i.e. an error of 0.7286% implies a difference in the returns on the two portfolios of 0.007286.

Statistic	CIR	LS	HJMGV	HJMRS	MLP	RBF	PCA2	PCA3	DUR
Mean(P-I)	0.6189	1.726	0.7626	0.6496	0.5986	2.4288	0.6580	1.6044	0.7951
Mean(P-II)	1.5084	1.8858	1.6287	0.7841	0.8250	2.5655	1.0892	1.6524	1.1166
Var.(P-I)	0.653	5.2828	0.905	0.591	0.237	5.505	0.659	7.729	1.002
Var.(P-II)	1.361	5.6309	1.395	0.685	0.389	9.213	1.519	3.835	0.734

portfolio in January 1996 the data from February 1983 to January 1996 is used. Then the first hedging error is measured in February 1996. This leads to forty-seven windows for hedging as opposed to forty-eight in the case of pricing. The difference of the return on the hedge portfolio and the return on the portfolio to be hedged is defined as the hedging error. In the case of hedging not only the mean of the hedging error but also the variance would be of interest to a practitioner. Duration and Principal Components Analysis (2 and 3 factor) models are used to compare with Equilibrium, HJM, and Network models as an indicator of how well the models capture different facets of the term structure such as the changes in the level, steepness, and curvature of the term structure. The mean and variances of the hedging errors over the forty-seven windows sample for all models are stated in Table 6.8.

Two criteria are used to compare the performance of any two models. Firstly if the mean of the hedging error for the model is significantly less than the other model then the model with the lower mean is said to have performed better. Secondly the one with the lesser variance is said to be better.

From the results in Table 6.8 it is evident that the models with the least mean and variance for the two portfolios to be hedged are HJM-RS and MLP network. Amongst the CIR and LS models CIR has lesser mean and variance for both the portfolios. Amongst the HJM-GV and HJM-RS models the HJM-RS model has a lesser mean and variance. Amongst the MLP and RBF models clearly MLP has lesser mean and variance of for both the portfolios. The results clearly indicate that the models HJM-RS and MLP perform the best in terms of the means and variances of the hedging errors. Though these models have the least hedging errors it is not clear from the results of Table 6.8 if the differences between any two models are statistically significant or not.

The term structure shift in 1998 offers the opportunity to evaluate model performance and the sensitivity of the estimation procedures to market shocks. The shift in the term structure in latter half of 1998 consisted of changes in all three aspects of the yield curve, level, steepness, and curvature. Given such a shock it behooves to compare the performance of the models during the pre-shock period (1996 – 97) versus their performance during the post-shock period (1998 – 99) to gauge how well each of these models reacts to such unanticipated shocks to the term structure. The hedging performance of the models in the first two years (1996 – 97) and in the last two years (1998 – 99) are shown in Table 6.9.

From the Table(6.9) it is evident that the PCA3 model works well from February 1996 through December 1997 in hedging portfolio P-I. From the fifth column of Table 6.9 it is evident that the average hedging errors of the PCA3 model increase by 651.75% which is the maximum rate of increase when compared to all the other models. In the second period since the data from each recent month is also added to the data that is used to compute the principal components (rolling window approach) the incorporation of the data for the latter half of 1998 deteriorates the performance of the PCA3 model. From Table 6.10 the findings are repeated when applied to the hedging of portfolio P-II. The PCA3 model has

Table 6.9: Split Hedging Errors (Portfolio P-I)

The mean and variance of the hedging errors over the periods I and II are calculated. The portfolio to be hedged is P-I. P-I is a portfolio of one-unit long positions in each of 1 year and 20 year bonds and one-unit short position in 10 year bond. The period I is from February 1996 to December 1997 and period II is from January 1998 to December 1999. The errors are in percentages i.e. an error of 0.5185% implies a difference in the returns on the two portfolios of 0.005185. The fifth column of change in means is the change in the average hedging error from period I to period II in percentage.

Statistic	Mean (I)	Variance(I)	Mean(II)	Variance(II)	Δ means
CIR	0.4096	0.0923	0.8194	1.1328	100.04
LS	1.6473	7.7075	1.8016	3.1811	9.37
HJM-GV	0.5667	0.2433	0.9503	1.5014	67.68
HJM-RS	0.4565	0.1122	0.8346	1.0023	82.82
MLP	0.5528	0.1725	0.6425	0.3056	16.22
RBF	2.6854	8.8791	1.6621	1.9814	-38.11
PCA2	0.4421	0.0779	0.8649	1.1531	95.66
PCA3	0.3707	0.0855	2.7867	12.3955	651.75
DUR	0.4744	0.2119	1.1025	1.5994	132.38

Table 6.10: Split Hedging Errors (Portfolio P-II)

The mean and variance of the hedging errors over the periods I and II are calculated. The portfolio to be hedged is P-II. P-II is a portfolio of one-unit long positions in each of the 10, 15, and 20 year bonds. The period I is from February 1996 to December 1997 and period II is from January 1998 to December 1999. The errors are in percentages i.e. an error of 0.5185% implies a difference in the returns on the two portfolios of 0.005185. The fifth column of change in means is the change in the average hedging error from period I to period II in percentage.

Statistic	Mean (I)	Variance(I)	Mean(II)	Variance(II)	Δ means
CIR	1.1602	0.7310	1.8421	1.7849	58.772
LS	0.6735	0.5124	1.6162	2.4544	139.98
HJM-GV	1.8625	0.8700	1.8625	1.8403	34.51
HJM-RS	0.5880	0.2023	0.9720	1.1002	65.29
MLP	0.8646	0.3767	0.9385	0.8436	8.55
RBF	3.5622	17.4372	3.1317	28.3251	-12.09
PCA2	0.5417	0.3337	1.6139	2.1336	197.90
PCA3	0.6495	0.3618	2.6136	5.3545	302.37
DUR	0.8633	0.3326	1.3593	1.0237	57.44

the highest rate of increase in percentage of errors (302.37%) from pre-shock period to post-shock period. The significant impact of outliers on PCA3 model hedging performance was illustrated by Nunes and Webber (1997) who argued that leaving the jumps in the data set when calculating covariance of yield changes does deteriorate PCA3 hedge performance. James and Webber (2000) suggest it is necessary to remove the jumps before estimating a PCA model.

Though the optimality of the PCA model is not dependent on normality, if normality assumption is true then the principal component loadings and the scores become maximum likelihood estimators with the additional desirable property of asymptotic efficiency. Alternatively one can say that unbiased, asymptotically efficient estimates are obtained using principal components analysis when the data is generated by multivariate normal distribution. PCA model is sensitive to departure from the assumptions of normality of the underlying state variables (changes in yields of different maturities in this application), homogeneity of sample and missing data. PCA is also sensitive to the changes in the variances of the components (due to the embedded homoskedasticity assumption). Since the PCA analysis utilizes Euclidean norms and inner products it is expected to be more sensitive to outliers (non-normal) and the larger the outlier, the affect is disproportionately greater.

Surprisingly in the case of RBF model the hedging errors go down. This is because this type of a network is by design a locally receptive network. This implies that the data point (that has the shock) responds only when the out of sample data point is close to it otherwise it does not. So the behavior of the yield curve at a particular instance of time does not determine the future performance unless there is a similar shock in the future. Though from the results it might appear that this is a desirable property but it is not the case. An RBF type of network would imply that if there is a shock of similar magnitude in the future to the benchmark maturities (inputs that are 0.5 and 5 years) the rest of the curve would behave in a similar way (if that point in the estimation data set explained a certain amount of variance of the errors of the network that is greater than the tolerance value set during the training period). In practice the term structure of interest rates does not possess such properties of regularity and hence it is not a desirable property.

Analogous to the testing done in the case of pricing of the bonds the hedging effectiveness is also compared in different phases.

- In phase I the MDM test statistic is used to determine the model generating superior hedge performance within the three classes of term structure models. Phase I MDM test statistics are reported in Table 6.11.

- In phase II the MDM test statistic is used to determine the model generating superior hedge performance amongst all model classes. Phase II test statistics are reported in Table 6.12.

- In phase III the MDM test statistic is used to compare term structure model hedge performance with duration based hedge performance. Phase III MDM test statistics are reported in Table 6.13.

- In phase IV the MDM test statistic is used to compare term structure model hedge performance with two factor Principal Component Analysis hedge performance. Phase IV test statistics are reported in Table 6.14.

- In phase V the MDM test statistic is used to compare term structure model hedge performance with three factor Principal Component Analysis hedge performance. Phase V test statistics are reported in Table 6.15.

The first column in Table 6.11 indicates that CIR model generates significantly smaller hedge errors than the LS model at the 1% significance level for the portfolio P-I but the difference is not significant for the portfolio P-II. Also from Table 6.8 the variance of the CIR model is substantially less than that of LS model for both the portfolios. This is an interesting finding because the CIR model is easier than the LS model to implement and it

Table 6.11: Hedge Errors (The three classes of models)

The MDM statistic is calculated for the different models using equation (6.1) for the matched pairs CIR-LS, HJMGV-HJMRS, and MLP-RBF in the sample period February 1996 to December 1999 for the forty-seven months. The two portfolios to be hedged are P-I and P-II. P-I is a portfolio of one-unit long positions in each of 1 year and 20 year bonds and one-unit short position in 10 year bond. P-II is a portfolio of one-unit long positions in each of the 10, 15, and 20 year bonds.(* 1% level, ** 5% level, *** 10 % level)

Statistic	CIR-LS	HJMGV-HJMRS	MLP-RBF
MDM (P-I)	-1.944***	1.540	-2.774*
MDM (P-II)	-1.465	4.047*	-2.336**

outperforms the LS model when used to hedge a portfolio of zero coupon bonds with short duration. One must be cautious while interpreting these results because of the methodology used to implement the LS model in this study. In this study the market price of risk for the LS model was implied using a bond of one year maturity and then that value of the market price of risk was used to price all the other bonds. This is in contrast to what the authors Longstaff and Schwartz (1993) suggest. LS suggest implying the term structure parameter which includes the market price of risk for each maturity and then use that value to price contingent claims.

In a term structure hedging application one would expect a model that has an appropriate specification for the forward rate volatility to perform better than a model that does not. The HJM-RS model outperforms the HJM-GV model at a level slightly above 10 percent for the portfolio P-I and a level of 1% for the portfolio P-II. The second column in Table 6.11 indicates that the specification suggested by HJM-RS model (spot rate volatility is varying with time) produces smaller hedging errors than HJM-GV (constant spot rate volatility). In the case of Network models the hedge error results are consistent with the results from evaluation of pricing errors. The MLP network is able to hedge the fixed income portfolios P-I and P-II better than an RBF network.

Table 6.12: Hedge Errors (CIR, HJM-RS, and MLP)

The MDM statistic is calculated for the different models using equation (6.1) for the matched pairs CIR-HJMRS, HJMRS-MLP, and MLP-CIR in the sample period February 1996 through December 1999 for the forty-seven months. The two portfolios to be hedged are P-I and P-II. P-I is a portfolio of one-unit long positions in each of 1 year and 20 year bonds and one-unit short position in 10 year bond. P-II is a portfolio of one-unit long positions in each of the 10, 15, and 20 year bonds. (* 1% level, ** 5% level, *** 10 % level)

Statistic	CIR-HJMRS	HJMRS-MLP	CIR-MLP
MDM (P-I)	0.261	0.799	0.742
MDM (P-II)	4.178*	0.469	3.056*

The results in the first and third columns of Table 6.12 indicate that on average the HJM-RS and MLP models generate smaller average hedge errors (not significant) than the CIR model for the portfolio P-I and the difference is significant at a level of 1% for the portfolio P-II. The significant deterioration of the performance of the CIR model when used to hedge portfolio of zero coupon bonds of long duration (P-II) is clearly illustrated in Tables 6.11 and 6.12. P-II contains long positions in three long term bonds (which the CIR model misprices the most) and hence has a longer duration. For instance in the case of first window the duration of portfolio P-I is 1.42 whereas the duration of portfolio P-II is 13.87. The CIR model uses only information from the time series of the short term rate (the 3 month FB yield in this study) during the estimation period. Given the fact that CIR model does not take into account the dynamics of long term rates it is expected to misprice the long term bonds the most. This is illustrated in the results of the pricing errors of the CIR model as shown in Tables 6.1 and 6.2 that increase as the maturity of the bond increases. In contrast the HJM-RS model incorporates the variance covariance matrix of the interest rates of different maturities during the estimation period and the MLP network is fed with the long term rate as a target output during the estimation period. The fact that HJM-RS and

Table 6.13: Hedge Errors (CIR, HJM-RS, MLP, Duration, PCA2, and PCA3)

The MDM statistic is calculated for the different models using equation (6.1) for the matched pairs of duration based hedging with PCA2, PCA3, CIR, HJMRS, and MLP each, in the sample period February 1996 to December 1999 for the forty-seven months. The two portfolios to be hedged are P-I and P-II. P-I is a portfolio of one-unit long positions in each of 1 year and 20 year bonds and one-unit short position in 10 year bond. P-II is a portfolio of one-unit long positions in each of the 7, 10, 15, and 20 year bonds. (* 1% level, ** 5% level, *** 10 % level)

Statistic	DUR-PCA2	DUR-PCA3	DUR-CIR	DUR-HJMRS	DUR-MLP
MDM (P-I)	1.6321	-1.4707	1.861***	1.587	1.198
MDM (P-II)	-0.7651	-2.0289**	-2.3221**	1.3129	2.3622**

MLP models take into account how the long term rates behaved in the estimation period leads to lesser pricing errors as compared to the CIR model when applied to the pricing of long term bonds. This was illustrated by the measurement of pricing errors in Tables 6.1 and 6.2. The MDM test statistics in the second column of Table 6.12 indicate that the MLP network has lesser hedging errors than the HJM-RS model on average for both portfolios but the differences are not significant.

Table 6.13 introduces three new models, Duration, two factor PCA (hereafter PCA2), and three factor PCA (hereafter PCA3) models. The first column indicates that the differences between the PCA2 and duration based models are not significant for both portfolios P-I and P-II. The difference between the Duration and PCA3 models indicates that the Duration model has lesser hedging errors than the PCA3 model on average for the portfolio P-I and for portfolio P-II the difference is significant at a level of 5%. The findings of Table 6.13 are counterintuitive given the evidence from the findings of Litterman and Scheinkman (1991) (hereafter LS) and Bliss(1997). From the results of Table 6.9 and 6.10 the mean hedge errors for both portfolios in the case of PCA2 and PCA3 models are lesser than the mean hedge errors in the case of Duration model for the time period before the shock of 1998 i.e. for the years 1996 and 1997. These findings are in accordance with the findings of LS and Bliss. The hedge errors for both PCA2 and PCA3 models increase in the after shock period of 1998 and 1999 and are higher than the hedge errors for Duration models. The data used by LS are the weekly Treasury prices from February 22, 1984, through August 17, 1988 and the data used by Bliss are FB yields from November 1982 through December 1995. In both the cases the data did not have shocks of the nature of August 1998 which can be categorized as outliers. The CIR model (one factor model)has significantly lesser hedging errors than the duration based hedges for portfolio P-I at a level of 10%. It is interesting to see that for portfolio P-II the results are reversed exactly i.e. the hedging errors for CIR model are significantly higher than for the duration based model at a level of 5%. This is in conformance with the findings of the CIR model's performance relative to other models. The CIR model does not do a good job at all in hedging portfolio P-II that comprises bonds of maturities 10, 15, and 20 years. The HJM-RS model has smaller mean hedging errors than the duration based hedges and the differences are not significant for the portfolios P-I and P-II. Relative to the MLP network also duration has hedging errors greater than the MLP network and the differences are not significant for portfolio P-I but are significant for portfolio P-II at a level of 5%. Both the HJM-RS model and the MLP network model outperform the traditional Duration based models.

In phase IV the two factor PCA2 model is introduced and compared with all the other models in terms of hedging of the two portfolios P-I and P-II. From the Table 6.14 it is evident that the HJM-RS model and MLP network have lesser hedging errors than the PCA2 model but the differences are not significant for portfolio P-I but for portfolio P-II HJM-RS model has significantly less hedging errors at a level of 5 % and MLP has significantly less hedging errors at a level of 10 %. Also the difference between the CIR and the PCA2 model is not significant for portfolios P-I and P-II.

From the first column of Table 6.15 hedging errors of CIR model are less than the PCA3 model but the differences are not significant for both the portfolios P-I and P-II. The PCA3

Table 6.14: Hedge Errors (CIR, HJM-RS, MLP, and PCA2)

The MDM statistic is calculated for the different models using equation (6.1) for the matched pairs PCA2-PCA3, PCA2-CIR, PCA2-HJMRS, and PCA2-MLP in the sample period February 1996 through December 1999 for the forty-seven months. The two portfolios to be hedged are P-I and P-II. P-I is a portfolio of one-unit long positions in each of 1 year and 20 year bonds and one-unit short position in 10 year bond. P-II is a portfolio of one-unit long positions in each of the 10, 15, and 20 year bonds. (* 1% level, ** 5% level, *** 10 % level)

Statistic	PCA2-PCA3	PCA2-CIR	PCA2-HJMRS	PCA2-MLP
MDM (P-I)	-1.5723	-0.9425	0.8960	0.8292
MDM (P-II)	-2.4180**	-1.4296	2.7216**	1.7376***

Table 6.15: Hedge Errors (CIR, HJM-RS, MLP, and PCA3)

The MDM statistic is calculated for the different models using equation (6.1) for the matched pairs PCA3-CIR, PCA3-HJMRS, and PCA3-MLP in the sample period February 1996 till December 1999 for the forty-seven months. The two portfolios to be hedged are P-I and P-II. P-I is a portfolio of one-unit long positions in each of 1 year and 20 year bonds and one-unit short position in 10 year bond. P-II is a portfolio of one-unit long positions in each of the 7, 10, 15, and 20 year bonds. (* 1% level, ** 5% level, *** 10 % level)

Statistic	PCA3-CIR	PCA3-HJMRS	PCA3-MLP
MDM (P-I)	1.5840	1.5884	1.6668
MDM (P-II)	1.5169	2.7191*	2.3644**

models when compared to the equilibrium model CIR, the only information that is used by the CIR model from the term structure is short term rate. Whereas the PCA method uses the information on the whole term structure. So shocks to term structure in estimation sample will necessarily have greater affect on PCA models.

The MLP network also performs better than the PCA3 model and the difference in the hedge errors for the two models from Table 6.15 is insignificant for portfolio P-I and is significant at a level of 5% for portfolio P-II . In comparison with the PCA method the network models use only two benchmark maturity yields as inputs. This serves to reduce impact of term structure shocks on network performance. In comparison to the PCA3 model the MLP model's hedging errors increase only by 16.22% and 8.55% for the portfolios P-I and P-II respectively as illustrated in Tables 6.9 and 6.10. The objective function used to train the MLP network serves to dampen the impact of outliers. The objective function, equation(A.18) is the sum of the squared errors and squared weights weighed inversely with their corresponding variances. In contrast the PCA method utilizes only the variance covariance matrix (squared errors). Due to this the hedge errors of the MLP network are not as sensitive to term structure shocks as are the PCA models.

In the case of the HJM-RS model and PCA3 model the difference is statistically insignificant for portfolio P-I but is significant for portfolio P-II at a level of 1 %. It is interesting though that HJM models use similar information and also make similar assumptions as the PCA models and still are significantly less sensitive to shocks than the PCA method. The reasons for this are that the PCA method combines the factors linearly and is therefore more sensitive to shocks than the HJM models that combine the state variables in a non-linear fashion. Rebonato (1996) finds that "...any yield curve model that uses principal components as driving factors is constrained to displaying a sigmoid-like correlation structure. This feature is not a result of the particular assumptions of the specific models, but a general consequence of the low dimensionality of these approaches". Rebonato argues that it is very difficult to get an exponentially declining correlation structure of the instantaneous forward rates for maturities ranging from 1 through 10 years using PCA model with 2 or 3 factors. The HJM-RS model imposes an exponentially decaying forward rate volatility structure. This restriction imposes an exponential like correlation structure on the instantaneous forward rates in these models. The ability of HJM-RS model to explain the exponential correlation structure amongst the instantaneous forward rates explains the better performance of the HJM-RS models when compared to the PCA models in terms of hedging applications.

6.3 A Sub Sample

A term structure model is often gauged by its flexibility to capture different types of unexpected shifts in the yield curve. A term structure model A can be said to capture different facets of the term structure better than another model B if A produces smaller hedge errors when applied to a portfolio of zero coupon bonds in a situation when the term structure suffers a shock. The term structure is said to suffer a shock when all three aspects of the yield curve, level, steepness, and curvature undergo a change. To substantiate the findings that the HJM-RS and the MLP models are able to capture more facets of the term structure of interest rates (such as twists etc.) than the other models, a sub-sample of this time period is investigated. During the latter half of 1998 a host of events triggered shifts in the yield curve that can be classified as twists.

In the last week of August 1998 Russia announced the terms of a $40 billion forced debt restructuring. Russia was on the brink of a financial collapse at this time. Many international investors, disappointed by the treatment of foreign bondholders by Russia dumped foreign stocks and expressed their confidence in U.S. dollar and Treasury bonds. By October 31^{st} the 30-year Treasury bond fell to 5.33 percent, the lowest level since October 1968. The yield curve became almost flat at that time.

The hedging performance of different models is analyzed during the time period July 1998 through December 1998 to judge which of these models is able to best capture the

Table 6.16: Russian Crisis (Portfolio P-I)

The hedging errors (absolute value of the difference between the return on the hedge portfolio and the portfolio to be hedged during the last month) as a percentage of the average pricing error (over the whole sample period of January 1996 through December 1999) for the six models CIR, HJM-RS, MLP, Duration, PCA2, and PCA3 during the time period July 1998 through December 1998 are presented. The portfolio to be hedged is P-I containing one-unit long positions in each of 1 year and 20 year bonds and one-unit short position in 10 year bond.

Date	CIR	LS	HJM-RS	MLP	Duration	PCA2	PCA3
980731	0.30	1.09	0.46	1.15	0.49	0.58	2.71
980831	0.91	0.87	1.04	0.98	0.84	0.57	1.44
980930	8.32	5.15	7.41	0.43	7.84	7.85	1.11
981030	1.39	1.17	1.40	2.92	0.81	1.04	1.59
981130	2.67	1.99	2.62	0.32	2.21	2.96	0.74
981231	2.29	2.24	2.53	1.40	1.37	2.54	2.20

dynamics of the term structure during this period. Tables 6.16 and 6.17 contain the hedging errors (the difference in the returns of the hedge portfolio and the portfolio to be hedged) as a percentage of the average pricing error (over the whole sample period of January 1996 through December 1999) for models identified as producing superior hedge performance in the previous section.

The results from the Table 6.16 and Table 6.17 clearly indicate that the pricing errors of all the models were high on September 30^{th}. But the MLP network and HJM-RS models have the least hedging errors as compared to all the others. These findings indicate that an MLP network and HJM-RS model can be used to hedge a bond portfolio in an environment where not only the level of the term structure but also its slope and curvature are changing. Amongst the two models HJM-RS and MLP, the MLP network produces lesser hedging errors during the period.

Table 6.17: Russian Crisis (Portfolio P-II)

The hedging errors (absolute value of the difference between the return on the hedge portfolio and the portfolio to be hedged during the last month) as a percentage of the average pricing error (over the whole sample period of January 1996 through December 1999) for the six models CIR, HJM-RS, MLP, Duration, PCA2, and PCA3 during the time period July 1998 through December 1998 are presented. The portfolio to be hedged is P-II containing one-unit long positions in each of the 10, 15, and 20 year bonds.

Date	CIR	LS	HJM-RS	MLP	Duration	PCA2	PCA3
980731	1.55	1.07	0.19	0.17	1.74	1.34	3.61
980831	1.06	1.89	1.65	2.62	0.25	2.16	0.42
980930	3.74	4.36	5.33	0.49	1.21	5.48	5.52
981030	1.21	0.52	2.23	2.19	2.59	0.69	3.50
981130	1.69	3.01	4.65	2.55	2.24	4.52	4.68
981231	0.26	1.50	1.05	0.11	0.06	1.16	0.44

CHAPTER 7

Summary and Conclusions

The accuracy of parametric bond pricing models is highly dependent on the parametric specification of the underlying risk factors. In the non-parametric pricing approach the data are allowed to determine the pricing relationship with few restrictions on the parametric specification of the underlying risk factors.

This study's investigation of zero coupon bond pricing and hedging over the period 1996 through 1999 window illustrates that the Heath Jarrow and Morton model with the restrictions of Ritchken and Sankarasubramanian (1995) (hereafter HJM-RS) and the multi layer perceptron (hereafter MLP) produce statistically smaller price and hedge errors. The equilibrium models start with the assumptions about the economy and then develop processes for the risk factors. Particularly the Cox, Ingersoll, and Ross (1985) (hereafter CIR) model uses the parameters of the short term rate process and the Longstaff and Schwartz (1992) (hereafter LS) model uses the parameters of the short term rate process and the short term rate volatility's process to determine bond prices of all maturities. HJM-RS model's feature that it uses the information of the whole yield curve along with it's flexibility (in terms of its specification of the spot rate volatility and the two state variables) lead to its smaller price and hedge errors than the CIR and LS model. The study also finds that the CIR model's hedge errors increase as the duration of the portfolio is increased. The non-parametric approach to the modelling of term structure of interest rates using MLP with Bayesian Regularization (hereafter BR) captures the dynamics of the term structure of interest rates well despite the fact that its information set is restricted to the one used by the HJM models.

Another interesting finding relates to the performance of the MLP network vs. the radial basis function (hereafter RBF) network. Amongst the Network models the MLP outperforms the RBF network. The performance of the RBF network, its characteristics and the training algorithm used here (orthogonal least squares) suggest that the RBF network trained with this algorithm is not the right choice for an application to the pricing or hedging of bonds, whereas MLP with BR algorithm is more appropriate for such an application.

The most common problem with the non-parametric estimation procedures is that they tend to overfit the data. In other words the estimation works very well for in sample data but does not work well for the out of sample data. The BR algorithm used to train the MLP network uses an objective function that penalizes both squared errors and squared weights of the network (complexity of the network). The results of the study suggest that the BR algorithm minimizes the overfitting problem. The outperformance of the MLP network relative to the RBF network can be attributed to either the BR algorithm or the success of the MLP network as a whole to the RBF network as a whole for a term structure application. Further investigation is required to pinpoint which one of the two reasons is actually contributing to the better performance of the MLP network. The MLP network uses BR algorithm whereas RBF uses orthogonal least squares. In this particular application the number of parameters for the MLP network were 89 whereas the RBF network had 50 parameters. As a result the training times for the MLP network were much greater than the RBF networks. On a Pentium-III 933MHz processor with 252 MB of RAM the MLP network took 50 seconds to train each network (for a particular window) whereas the RBF took only 15 seconds to train each network.

The evidence from the previous chapter implies that the model that produces the least pricing errors from the parametric pricing category is HJM-RS and its counterpart in the non-parametric pricing category is the MLP network. Amongst these two models the HJM-RS produces greater pricing errors. This is despite the fact that the inputs to the Network were chosen such that they were consistent with the information set of the HJM models.

This implies that the network has been able to learn the relationship between the risk factors as specified by the HJM framework and the term structure of interest rates. This ability of the network models as demonstrated by this and other studies (Hutchinson, Lo, and Poggio (1995)) is one of the major advantages of network models. Till date in the financial markets there are many assets whose relationship with the underlying risk factors is yet not known exactly. Other potential neural network applications include the pricing of interest rate and other options, prediction of exchange rates, credit scoring, and prediction of credit card default rates. The outperformance of the MLP over the HJM-RS model also implies that the HJM-RS model can be improved. Approaches to improve the HJM-RS model identified by this study include a relaxation of the assumption that the errors in the measurement of the yields are distributed i.i.d., and inclusion of another stochastic factor besides the single stochastic short rate.

The relative performance of the models used in hedging applications is consistent with bond pricing applications though the differences are not significant. The models HJM-RS, and MLP perform better than the traditional Duration based models and models based on principal components analysis. This indicates that in addition to parallel shifts of the term structure, these models can be used to effectively hedge against changes in the term structure's steepness and curvature.

The principal components analysis model with three factors (hereafter PCA3) model was introduced because the three components correspond closely to levels, steepness and curvature of the yield curve. The HJM-RS model and the MLP network both produce significantly smaller hedge errors than the PCA3 model.

The PCA3 model is found to be very sensitive to jumps in the yield curve relative to all the other models considered. This is clearly evident from the results of Tables 6.9 and 6.10 where the PCA3 model has the highest rate of increase (651.75% and 302.37%) in the hedge errors from the pre-shock period to the post-shock period for both portfolios. A PCA3 model is specifically sensitive to departure from the assumptions of normality of the underlying state variables, homogeneity of sample and missing data. PCA3 model is also sensitive to the changes in the variances of the components (due to the embedded homoskedasticity assumption). Since the PCA analysis utilizes Euclidean norms and inner products it is expected to be more sensitive to outliers (non-normal) and the larger the outlier, the affect is disproportionately greater.

To summarize the conclusions of the study

- The HJM-RS model and the MLP Network model outperform the CIR model, LS model, and HJM-GV model in the pricing and hedging of zero coupon bonds.

- An MLP network with the training algorithm of Bayesian Regularization captures the dynamics of the term structure of interest rates better than an RBF network with Orthogonal Least Squares algorithm.

- The models MLP network, and HJM-RS on an average produce smaller hedge errors with lesser variability when compared to the Duration based hedges.

- The MLP network and HJM-RS model produce statistically smaller hedging errors than application of three factor principal components. The hedge performance of PCA3 model is more sensitive to shocks to the yield curve than the hedge performance of MLP and HJM-RS models

- The MLP network and the HJM-RS model produce the smallest hedge errors than all the other models during the Russian crisis of August 1998.

- Overall a non-parametric approach using a Multi Layer Perceptron (with the Bayesian Regularization) can be added to the vast set of approaches to the modelling of term structure of interest rates and can be effectively applied to the pricing and hedging of zero coupon bonds.

Though every attempt has been made to implement models in a consistent fashion, input data series and estimation periods are not identical for all models considered. The equilibrium models and network models use data from 1983 – 1995 for the first window whereas the HJM models use the data from 1995. Furthermore HJM models utilize estimation data for all maturities.

The conclusions of this study can be a result of the usage of the different data sets. But due to the unique estimation procedure of each of the three approaches neither can the estimation period for the HJM models be increased nor the estimation period for the equilibrium models and the network models be decreased. The study reports tests of joint hypotheses.

Based on this study the possible areas of further research are as follows

- Extensive testing of the hedging performance of one model versus the other using portfolios of coupon bonds and other interest rate derivatives. This would provide information on what model is able to hedge a fixed income portfolio using interest rate derivatives.

- Change the algorithm used in the training of the RBF network to analyze if such a network is able to capture the dynamics of the term structure

- Change the way the HJM-RS model is implemented by considering a two factor two state variable model and compare its performance with respect to other models.

BIBLIOGRAPHY

[1] Abken, Peter and Hugh Cohen 1994 "Generalized Method of Moments Estimation of Heath Jarrow-Morton Models of Interest-Rate Contingent Claims," *Working Paper 94-8, Federal Reserve Bank of Atlanta, August.*

[2] Amin, Kaushik and Andrew Morton 1994 "Implied Volatility Functions in Arbitrage -Free Term Structure Models," *Journal of Financial Economics 35, 193-234.*

[3] Au, K.T. and D.C.Thurston 1994 "Exchange Options on Bonds," *Working Paper, University of New South Wales 63, 1-8.*

[4] Au, K.T. and D.C.Thurston 1994 "Markovian Term Structure Movements," *Working Paper, University of New South Wales 61, 1-13.*

[5] Backus, David K., Silverio Foresi, and Stanley E.Zin 1995 "Arbitrage opportunities in arbitrage free models of bond pricing," *Working paper, New York University.*

[6] Survey released by *Bank for International Settlements.*

[7] Bansal, Ravi and Viswanathan 1993 "No Arbitrage and Arbitrage Pricing: A New Approach," *Journal of Finance 48 No. 4 September, 1231-1262.*

[8] Baxter M., A. Rennie *1996 "Financial Calculus," Cambridge University Press.*

[9] Bierwag, Gerald O., George G. Kaufman and Alden Toevs *1983 "Innovations in Bond Portfolio Management," Greenwich, Conn.: JAI Press.*

[10] Bjork T. 1996 "Interest Rate Theory-CIME Lectures," *Working Paper, Stockholm School of Economics, 1-90.*

[11] Bjork T. *1998 "Arbitrage Theory in Continuous Time," Oxford University Press.*

[12] Rebonato R. *1998 "Interest Rate Option Models,"* 2^{nd} *Edition, Wiley.*

[13] Black, F. 1995 "Interest Rates as Options," *Journal of Finance 50, 1371-1376.*

[14] Black, F., E. Derman, and W. Toy 1990 "A One-Factor Modle of Interest Rates and its Application to Treasury Bond Options," *Financial Analyst's Journal Jan-Feb, 33-39.*

[15] Black, F. and P. Karasinski 1991 "Bond and Option Pricing When Short Rates are Lognormal," *Financial Analyst's Journal July-Aug, 52-59.*

[16] Bliss, Robert R. 1994 "Testing Term Structure Estimation Methods," *Working Paper, Federal Reserve Bank of Atlanta, April.*

[17] Bliss, Robert R. and Peter Ritchken 1996 "Empirical Tests of Two State Variable Heath-Jarrow-Morton Models," *Journal of Money Credit and Banking 28(3), 452-476.*

[18] Bliss, Robert R. and Ehud I. Ronn 1995 "The Implied Volatility of U.S. Interest Rates: Evidence from Callable U.S. Treasuries," *Federal Reserve Bank of Atlanta Working paper 95-12, November.*

[19] Bliss, Robert R. and David C.Smith 1998 "The elasticity of interest rate volatility: Chan, Karolyi, Longstaff and Sanders Revisited," *Federal Reserve Bank of Atlanta Working paper 97-13a, March.*

[20] Bliss, Robert R. 1997 "Movements in the Term Structure of Interest Rates," *Federal Reserve Bank of Atlanta Economic Review Fourth Quarter.*

[21] Bollerslev, Tim 1986 "Generalized Autoregressive Conditional Heteroskedasticiy," *Journal of Econometrics 31, 307-327.*

[22] Brace A., D. Gatarek, and M. Musiela 1997 "The Market Model of Interest Rate Dynamics," *Mathematical Finance 7, 127-155.*

[23] Brennan, Michael and Eduardo Schwartz 1979 "A Continuous Time Approach to the Pricing of Bonds," *Journal of Banking and Finance 3, 133-155.*

[24] Brown, S.J. and P.H.Dybvig 1986 "The Empirical Implications of the CIR Theory of the Term Structure of Interest Rates," *Journal of Finance 41, 617-630.*

[25] Brown, R.H. and S.M. Schaefer 1994 "Interest Rate Volatility and the Shape of the Term Structure," *Philosophical Transactions of the Royal Society of London A 347, 563-576.*

[26] Canabarro, Eduardo 1995 "Where do one-factor interest rate models fail?," *Journal of Fixed Income 5, 31-52.*

[27] Carcano Nicola and Silverio Foresi 1997 "Hedging against Interest Rate Risk: Reconsidering Volatility Adjusted Immunization," *Journal of Banking and Finance 21, 127-141.*

[28] Caverhill, Andrew 1994 "When Is the Short Rate Markovian?" *Mathematical Finance 44, 305-312.*

[29] Chan, Ka-Kung, C.G. Andrew Karolyi, Francis Longstaff, and Anthony Sanders 1992 "An Empirical Comparison of the Alternative Models of Term Structure of Interest Rates," *Journal of Finance 47, 1209-1228.*

[30] Chen, S., C.F.N. Cowan, and P.M. Grant 1991 "Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks," *IEEE Transactions on Neural Networks, Vol. 2, No. 2, March.*

[31] Cohen, Hugh and David Heath 1992 "A New Method of Testing Pricing Models as Applied to Forward Interest Rate Models," *Federal Reserve Bank of Atlanta Working paper 92-16, November.*

[32] Constantinides, G.M. 1992 "A Theory of the Nominal Term Structure of Interest Rates," *Review of Financial Studies 5, 531-552.*

[33] Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross 1981 "A Reexamination of Traditional Hypothesis about the Term Structure of Interest Rates," *Journal of Finance 36, 769-799.*

[34] Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross 1985a "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica 53, 363-384.*

[35] Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross 1985b "A Theory of the Term Structure of Interest Rates," *Econometrica 53, 385-407.*

[36] Culbertson, J.M.1957 "The Term Structure of Interest Rates," *Quarterly Journal of Economics 71, 485-517.*

[37] Cun, Y.Le, 1985 "Une Procedure d'apprentissage pour reseau a seuil assymetrique," *Cognitiva 85, 599-604.*

[38] Cybenko, G., 1988 "Approximations by superpositions of a sigmoidal function," *Technical report 856, University of Illinois, Department of Electrical and Computer Engineering.*

[39] Dai, Q. and K.J. Singleton 1998 "Specification Analysis of Affine Term Structure Models," *Working Paper.*

[40] Diebold, F.X. and R.S. Mariano 1995 "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics, 13, 253-263.*

[41] Donaldson, R. Glen and Mark Kamstra 1996 "A New Dividend Forecasting Procedure Rejects Bubbles in Asset Prices: The Case of 1929's Stock Crash," *Review of Financial Studies 9 No. 2, 333-383.*

[42] Duffie, Darrell *1996 "Dynamic Asset Pricing Theory," Princeton Press, Princeton, New Jersey.*

[43] Duffie, Darrell and R. Kan 1994 "Multi-factor Term Structure Models," *Philosophical Transactions of the Royal Society of London A 347, 577-586.*

[44] Duffie, Darrell and R. Kan *1996 "A Yield-Factor Model of Interest Rates," Mathematical Finance 6, 379-406.*

[45] Dutta, Soumitra and Shashi Shekhar 1988 "Bond Rating: A Non-Conservative Application of Neural Networks," *Proceedings of the IEEE International Conference on Neural Networks 2, 443-450.*

[46] Fama, Eugene and Robert R. Bliss 1987 "The Information in Long Maturity Forward Rates," *American Economic Review 77, 680-692.*

[47] Flesaker, Bjorn 1992 "Testing the Heath-Jarrow-Morton/Ho-Lee Model of Interest Rate Contingent Claim Prices," *University of Illinois at Urbana-Champaign Working paper.*

[48] Foresee, F.D. and M.T.Hagan 1997 "Gauss-Newton Approximation to Bayesian Regularization," *Proceedings of the 1997 International Conference on Neural Networks, Houston, Texas.*

[49] Gibbons, Michael R. and Krishna Ramaswamy 1993 "A Test of the Cox, Ingersoll, and Ross Model of the Term Structure," *The Review of Financial Studies 6(3), 619-658.*

[50] Girosi,F., M.Jones and T.Poggio 1993 "Priors, Stabilizers and Basis Functions: From Regularization to Radial, Tensor and Additive Splines," *Artificial Intelligence Memo 1430, Massachusetts Institute of Technology, Artificial Intelligence Laboratory.*

[51] Girosi,F., and T.Poggio 1990 "Networks and the Best Approximation Property," *Biological Cybernetics 63, 169-176.*

[52] Glasserman, P., and X. Zhao 1998 "Arbitrage-free Discretization of Lognormal Forward Libor and Swap Rate Models," *Working Paper, Columbia Business School, 1-33.*

[53] Gultekin, N. Bulent and Richard J. Rogalski 1984 "Alternative Duration Specification Measurement of Basis Risk: Empirical Tests," *Journal of Business 57, 241-264.*

[54] Hagan, M.T., H.B.Demuth and M.Beale *1996 "Neural Network Design, Boston:PWS Publishing Co.*

[55] Harrison, J.M., and D.M.Kreps 1979 "Martingales and Arbitrage in Multiperiod Security Markets," *Journal of Economic Theory 20, 381-408.*

[56] Harrison, J.M., and S.Pliska 1981 "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and their Applications 11, 215-260.*

[57] Harvey, David, Stephen Leybourne, and Paul Newbold 1997 "Testing the equality of Prediction Mean Squared Errors," *International Journal of Forecasting 13, 281-291.*

[58] Hawley, Delvin D., John D. Johnson, and Dijjotam Raina 1990 "Artificial Neural Systems: A New Tool for Financial Decision-Making," *Financial Analysts Journal 46 (November/December), 63-72.*

[59] Heath, David, Robert Jarrow and Andrew Morton 1990 "Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation," *Journal of Financial and Quantitative Analysis 254, 419-440.*

[60] Heath, David, Robert Jarrow and Andrew Morton 1992 "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claim Valuation," *Econometrica 60, 77-105.*

[61] Heath, David, Robert Jarrow and Andrew Morton 1990 "Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation," *Journal of Financial and Quantitative Analysis 254, 419-440.*

[62] Heath, David, Robert Jarrow, Andrew Morton and Mark Spindel 1992 "Easier Done Than Said," *Risk 5, 77-80.*

[63] Hertz, J., A. Krogh, and R.G. Palmer 1991 "Introduction to the Theory of Neural Computation," *Santa Fe Institute Studies in Sciences of Complexity. Addison Wesley.*

[64] Hicks, J.R. *1939 "Value and Capital,* 2^{nd} *Edition, Oxford University Press, Oxford.*

[65] Hicks, J.R. *1946 "Value and Capital,* 2^{nd} *Edition, London, Oxford University Press, Oxford.*

[66] Ho, Thomas and Sang Bin Lee 1986 "Term Structure Movements and the Pricing of Interest Rate Contingent Claims," *Journal of Finance 41(5), 1011-1030.*

[67] Hornik, K., 1989 "Multilayer feedforward networks are universal approximators," *Neural Networks 2, 359-366.*

[68] Hull, John C. and Alan White 1993 "Bond Option Pricing Based on a Model for the Evolution of the Bond Prices," *Advances in Futures and Options Research 6,1-13.*

[69] Hutchinson, J.M. 1993 "A Radial Basis Function Approach to Financial Time Series Analysis," *Technical Report 1457, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.*

[70] Hutchinson,J., Andrew Lo and T.Poggio 1994 "A Non-Parametric Approach to the Pricing and Hedging of Derivative Securities via Learning Networks," *Journal of Finance 49, 851-889.*

[71] James, Jessica and Nick Webber *2000 "Interest Rate Modelling," John Wiley and Sons, Ltd.*

[72] Jamshidian F. 1996 "Bond, Futures, and Option Valuation in the Quadratic Interest Rate Model," *Applied Mathematical Finance 3, 93-115.*

[73] Jensen, Herbert L. 1992 "Using Neural Networks for Credit Scoring," *Managerial Finance 18 (No. 6), 15-26.*

[74] Jordan, Bradford and David Kuipers 1997 "On the Performance of Affine Term Structure Models: Evidence from the U.S. Treasury STRIPS Market," *Working paper.*

[75] Kryzanowski, Lawrence, Michael Galler, and David W. Wright 1993 "Using Artificial Neural Networks to Pick Stocks," *Financial Analyst's Journal 49 No. 4 July-Aug., 21-67.*

[76] Litterman, Robert, and *José* Scheinkman 1991 "Common Factors Affecting Bond Returns," *Journal of Fixed Income 1,54-61 June.*

[77] Longstaff, Francis and Eduardo Schwartz 1992 "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance 47(4),1259-1282.*

[78] Longstaff, Francis and Eduardo Schwartz 1993 "Implementation of the Longstaff-Schwartz Interest Rate Model," *Journal of Fixed Income September, 7-14.*

[79] Loofbourrow, John and Tod Loofbourrow 1993 "Neural Networks for Trading," *AI in Finance (Special Report), 42-51.*

[80] Lutz, F.A. 1940 "The Structure of Interest Rates," *Quarterly Journal of Economics 55 November, 36-63.*

[81] Macaulay, Fredrick R., *1938 "The Movements of Interest Rates, Bond Yields and Stock Prices in the United States Since 1856 (National Bureau of Economic Research, New York.*

[82] Mackay,D.J.C. 1992 "Bayesian Interpolation," *Neural Computation 4,415-447.*

[83] Maruyama, M., F.Girosi and T.Poggio 1991 "A Connection between GRBF and MLP." *Artificial Intelligence Memo 1291, Massachusetts Institute of Technology, Artificial Intelligence Laboratory.*

[84] Mehta, Mahendra *1995 "Foreign Exchange Markets," In Apostolos-Paul Refenes, Ed., Neural Networks in the Capital Markets, Chichester, England, John Wiley & Sons, 177-198.*

[85] Micchelli, C.A. 1986 "Interpolation of Scattered data: distance matrices and conditionally positive definitive functions," *Construct. Approx., vol. 2, 11-22.*

[86] Mikkelson, P. 1999 "Pricing Interest Rate and Foreign Exchange Derivatives in a Consistent LIBOR Market." *Working Paper, Aarhus School of Business, 1-37.*

[87] Modigliani,F., and R.Sutch 1966 "Innovations in Interest Rate Policy," *American Economic Review 56, 178-197.*

[88] Moody, John and Jaochim Utans *1995 "Architecture Selection Strategies for Neural Networks," in A.P. Refenes, Ed., Neural Networks in the Capital Markets, Chichester, England, John Wiley & Sons, 277-300.*

[89] Musiela, M., and M. Rutkowski 1997 "Continuous Time Term Structure Models: Forward Measure Approach," *Finance and Stochastics 1, 261-291.*

[90] Musiela, M., and M. Rutkowski *1997 "Martingale Methods in Financial Modelling," Springer, Applications of Mathematics 36.*

[91] Neely, Christopher, Paul Weller, and Rob Dittmar 1997 "Is Technical Analysis in Foreign Exchange Market Profitable? A Genetic Programming Approach," *Journal of Financial and Quantitative Analysis 32 (No. 4 December), 405-426.*

[92] Ng, K. and R.Lippman 1991 "A Comparative Study of the Practical Characteristics of Neural Network and Conventional Pattern Classifiers, in R.Lippman, J.Moody and D.Touretsky, eds." *Advances in Neural Information Processing Systems 3 (Morgan-Kaufman, San Mateo, California).*

[93] Nunes,J. and N.J.Webber 1997 "Low Dimensional Dynamics and the Stability of HJM Term Structure Models," *Working Paper, University of Warwick,1-27.*

[94] Parker,D.B., 1995 "Learning Logic: Casting the Cortex of the Human Brain in Silicon," *Technical Report TR-47, Center for Computational Research in Economics and Management Science, MIT, Cambridge, MA.*

[95] Pearson, N. and T. S. Sun 1994 "An Empirical Examination of the Cox, Ingersoll and Ross Model of the Term Structure of Interest Rates using the Method of Maximum Likelihood," *Journal of Finance 54, 929-959.*

[96] Pedersen, M.B. 1998 "Calibrating LIBOR Market Models," *Working Paper, SimCorp A/S, Copenhagen, 1-26.*

[97] Poggio,T. and F.Girosi 1990 "Networks for Approximation and Learning," *Proceedings of IEEE 78, 1481-1497.*

[98] Powell,M.J.D. 1987 "Radial Basis Functions approximations to polynomials," *in Proc. 12th Biennial Numerical Analysis Conf. (Dundee), 223-241.*

[99] Priaulet, P. 1997 "Structure Par Terme Des Taux D'intérêt: Reconstitution, Modélisation et Couverture," *Thèse de Doctorat en Sciences Économiques, Université Paris IX – Dauphine, June.*

[100] Refenes, Apostolos-Paul, Achileas Zapranis, and Gavin Francis *1995 "Modelling Stock Returns in the Framework of APT," in A.P. Refenes, Ed., Neural Networks in the Capital Markets, Chichester, England, John Wiley & Sons, 101-125.*

[101] Refenes, Apostolos-Paul and Ado Zaidi *1995 "Managing Exchange Rate Prediction Strategies with Neural Networks," in A.P. Refenes, Ed., Neural Networks in the Capital Markets, Chichester, England, John Wiley & Sons, 213-219.*

[102] Ritchken, Peter and L. Sankarasubramanian 1995 "Volatility Structure of Forward Rates and the Dynamics of the Term Structure," *Mathematical Finance 5, 55-72.*

[103] Rosenblatt, F. 1958 "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review 65, 386-408.*

[104] Rumehlart, D.E., G.E.Hinton and R.J.Williams 1986 "Learning Representations by Backpropogating Errors," *Nature 323, 533-536.*

[105] Rumehlart, D.E. and J.L.McClelland, eds. *1986 "Parallal Distributed Processing: Explorations in the Microstructure of Cognition 1, Cambridge, MA: MIT Press.*

[106] Rutkowski, M. 1997 "A Note on the Flesaker-Hughston Model of the Term Structure of Interest Rates," *Applied Mathematical Finance 4, 151-163.*

[107] Samuelson, Paul A. 1945 "The Effect of Interest Rate Increases on the Banking System," *American Economic Review, March.*

[108] Sargent, T.J. 1972 "Rational Expectations and the Term Structure of Interest Rates," *Journal of Money, Credit and Banking, February, 4-Part I: 74-97.*

[109] Singleton, J. Clay and Alvin J. Surkan *1995 "Bond Rating with Neural Networks," in A.P. Refenes, Ed., Neural Networks in the Capital Markets, Chichester, England, John Wiley & Sons, 301-307.*

[110] Singleton, J. Clay and Alvin J. Surkan 1991 "Modelling the Judgement of Bond Agencies," *Journal of the Midwest Finance Association 20, 72-80.*

[111] Swanson, Norman R., and Halbert White 1995 "A Model Selection Approach to Assessing the Information in the Term Structure using Linear Models and Artificial Neural Networks," *Journal of Business and Economic Statistics Volume 13 Issue 3 July, 265-276.*

[112] Vasicek, O. 1977 "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics 5, 177-188.*

[113] Widrow, B. and M.E.Hoff 1960 "Adaptive Switching Circuits," *IRE WESCON Convention Record, New York: IRE Part 4, 96-104.*

[114] Wood, John H. 1964 "The Expectations Hypothesis, the Yield Curve, and Monetary Policy," *Quarterly Journal of Economics, August, 457-70.*

[115] Zirilli, Joseph S. *1997 "Financial Prediction Using Neural Networks," London International Thompson Computer Press.*

APPENDIX A

The Algorithm for the Training of MLP

The first procedure to train the MLPs was evidently found in Werbos[1974]. The algorithm was rediscovered in the 80's independently by Rumelhart, Hinton and Williams[1986], Parker[1985] and Cunn[1985]. The Backpropogation (hereafter backprop) algorithm is based on the steepest descent or steepest ascent algorithm based on the objective of whether the function is to be minimized or maximized. Generally the objective function is to minimize squared errors and hence the algorithm used is steepest descent.

As discussed in the chapter, theoretical framework, for the MLPs the outputs of one layer are the inputs to the next layer. The following note on the backprop is based on the discussion in Hagan, Demuth and Beale[1996]. Let the number of neurons in an MLP be S and the number of layers be M, and the number of inputs be R. The functional form to describe the output input relation can be written as

$$a^{m+1} = f^{m+1}(W^{m+1}a^m + b^{m+1}) \tag{A.1}$$

for

$$m = 0, 1, 2, \ldots, M$$

The inputs of the data are initially fed to the first layer. That is

$$a^0 = p$$

This provides the initial condition for the recursive set of equations (A.1). The outputs to the last layer of the network are the network outputs:

$$a = a^M$$

The algorithm is based on the objective of minimizing a mean squared error function that is usually known as a mean squared error performance index. In this algorithm the inputs are provided to the network and the outputs are compared to the targets from the data that determine proper network behavior as

$$p_1, t_1, p_2, t_2, \ldots, p_Q, t_Q,$$

where p_q, $q = 1, \ldots, Q$ is an input to the network and t_q is the target output for the input. At every step as each input is applied to the network the parameters are updated using the target outputs. The network parameters are adjusted so as to minimize the sum of squared errors given as

$$F(x) = \sum_{q=1}^{Q} e'_q e_q = \sum_{q=1}^{Q} (t_q - a_q)'(t_q - a_q)$$

where x is a vector of all the network weights and biases, t_q is a vector of target outputs of the proper network, a_q is the vector of outputs of the network, and e_q is the vector of the error terms that is the difference between the target output and the actual output of the network. The algorithm starts with a stochastic approximation to the error function by replacing it with the error on the latest target and later on updating the parameters at each iteration. The expectation of the squared error is replaced by the squared error at iteration k as

$$\hat{F}(x) = e(k)'e(k) = (t(k) - a(k))'(t(k) - a(k))$$

The steepest descent algorithm applied to approximate the mean squared error is

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \alpha \frac{\partial \hat{F}}{\partial w_{i,j}^m}, \tag{A.2}$$

$$b_i^m(k+1) = b_i^m(k) - \alpha \frac{\partial \hat{F}}{\partial b_i^m}, \tag{A.3}$$

where the first index of the weights and the biases denote the neuron they correspond to, and the second index corresponds to the respective input. The superscript corresponds to the layer they belong to. The parameter α is often termed as the learning rate.
At this stage it turns out that the partials are to be estimated at each iteration. For this purpose efficient use of the chain rule of calculus is made. If it were a single layer, the partial could be computed directly since the error is a direct function of the weights of the layer. But in an MLP, the error is not a direct function of the weights of the hidden layers. Using the chain rule the equations (A.2) and (A.3) can be written as

$$\frac{\partial \hat{F}}{\partial w_{i,j}^m} = \frac{\partial \hat{F}}{\partial n_i^m} \times \frac{\partial n_i^m}{\partial w_{i,j}^m}, \tag{A.4}$$

$$\frac{\partial \hat{F}}{\partial b_i^m} = \frac{\partial \hat{F}}{\partial n_i^m} \times \frac{\partial n_i^m}{\partial b_i^m}, \tag{A.5}$$

where n_i^m denotes the net input to layer m's neuron i. The terms on the right hand side of these equations can be computed easily since the output is a function of the net input to the layer and the net input is a direct function of the weights and biases of that layer. The net input to layer m is given by

$$n_i^m = \sum_{j=1}^{S^{M-1}} w_{i,j}^m a_j^{m-1} + b_i^m$$

This implies

$$\frac{\partial n_i^m}{\partial w_{i,j}^m} = a_j^{m-1}, \quad \frac{\partial n_i^m}{\partial b_i^m} = 1$$

Let

$$s_i^m \equiv \frac{\partial \hat{F}}{\partial n_i^m},$$

The term s is derived from the sensitivity of $\hat{F}$ to changes in the i^{th} element of the net input vector in layer m. With this the equations (A.4) and (A.5) can be rewritten as

$$\frac{\partial \hat{F}}{\partial w_{i,j}^m} = s_i^m a_j^{m-1} \tag{A.6}$$

$$\frac{\partial \hat{F}}{\partial b_i^m} = s_i^m \tag{A.7}$$

$$\tag{A.8}$$

Now the complete algorithm can be summarized as

$$\begin{aligned} w_{i,j}^m(k+1) &= w_{i,j}^m(k) - \alpha s_i^m a_j^{m-1} \\ b_i^m(k+1) &= b_i^m(k) - \alpha s_i^m \end{aligned}$$

In matrix notation

$$\begin{aligned} \mathbf{W}^m(k+1) &= \mathbf{W}^m(k) - \alpha \mathbf{s}^m (\mathbf{a}^{m-1})' \\ \mathbf{b}^m(k+1) &= \mathbf{b}^m(k) - \alpha \mathbf{s}^m \end{aligned}$$

The next task in this algorithm is to calculate the so called sensitivities $\mathbf{s}^m$ at each layer. The sensitivity at each layer is computed from the sensitivity at the next layer. To derive the sensitivities the Jacobian matrix

$$\frac{\partial \mathbf{n}^{m+1}}{\partial \mathbf{n}^m} \equiv \begin{bmatrix} \frac{\partial n_1^{m+1}}{\partial n_1^m} & \frac{\partial n_1^{m+1}}{\partial n_2^m} & \cdots & \frac{\partial n_1^{m+1}}{\partial n_{S^m}^m} \\ \frac{\partial n_2^{m+1}}{\partial n_2^m} & \frac{\partial n_2^{m+1}}{\partial n_2^m} & \cdots & \frac{\partial n_2^{m+1}}{\partial n_{S^m}^m} \\ \vdots & \vdots & & \vdots \\ \frac{\partial n_{S^{m+1}}^{m+1}}{\partial n_1^m} & \frac{\partial n_{S^{m+1}}^{m+1}}{\partial n_2^m} & \cdots & \frac{\partial n_{S^{m+1}}^{m+1}}{\partial n_{S^m}^m} \end{bmatrix} \tag{A.9}$$

is used. Each term in the Jacobian matrix denotes

$$\frac{\partial n_i^{m+1}}{\partial n_j^m} = \frac{\partial \left(\sum_{l=1}^{S^m} w_{i,l}^{m+1} a_l^m + b_i^{m+1} \right)}{\partial n_j^m} = w_{i,l}^{m+1} \frac{\partial a_j^m}{\partial n_j^m} \tag{A.10}$$

$$= w_{i,j}^{m+1} \frac{\partial f^m(n_j^m)}{\partial n_j^m} = w_{i,j}^{m+1} \dot{f}^m(n_j^m), \tag{A.11}$$

where

$$\dot{f}^m(n_j^m) = \frac{\partial f^m(n_j^m)}{\partial n_j^m}$$

The equation (A.10) is true since all the other outputs, corresponding to all the neurons except the one for which $l = j$ are independent of the net input to neuron j i.e. n_j^m. The derivatives will all be zero except the one that has $l = j$. Hence the Jacobian can also be written as

$$\frac{\partial \mathbf{n}^{m+1}}{\partial \mathbf{n}^m} = \mathbf{W}^{m+1}\dot{\mathbf{F}}^m(\mathbf{n}^m), \tag{A.12}$$

where

$$\dot{\mathbf{F}}^m(\mathbf{n}^m) = \begin{bmatrix} f^m(n_1^m) & 0 & \dots & 0 \\ 0 & f^m(n_1^m) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & f^m(n_{S^m}^m) \end{bmatrix} \tag{A.13}$$

Now all the sensitivities can be obtained by the following recurrence relation

$$\begin{aligned} \mathbf{s}^m &= \frac{\partial \hat{F}}{\partial \mathbf{n}^m} = \left(\frac{\partial \mathbf{n}^{m+1}}{\partial \mathbf{n}^m}\right)' \frac{\partial \hat{F}}{\partial \mathbf{n}^{m+1}} = \dot{\mathbf{F}}(\mathbf{n}^m)(\mathbf{W}^{m+1})' \frac{\partial \hat{F}}{\partial \mathbf{n}^{m+1}} \\ &= \dot{\mathbf{F}}^m(\mathbf{n}^m)(\mathbf{W}^{m+1})'\mathbf{s}^{m+1} \end{aligned}$$

To be able to solve for all the sensitivities using the recurrence relation above, the initial condition or the sensitivity at the final layer is to be calculated. To calculate $\mathbf{s}^M$ it is nothing but the partial of the performance function (sum of squared errors) w.r.t. the net input at the M^{th} layer.

$$s_i^M = \frac{\partial \hat{F}}{\partial n_i^M} = \frac{\partial (\mathbf{t}-\mathbf{a})'(\mathbf{t}-\mathbf{a})}{\partial n_i^M} = \frac{\partial \sum_{j=1}^{s^M}(t_j - a_j)^2}{\partial n_i^m} = -2(t_i - a_i)\frac{\partial a_i}{\partial n_i^m}$$

Also

$$\frac{\partial a_i}{\partial n_i^m} = \frac{\partial a_i^M}{\partial n_i^m} = \frac{f^M(n_i^M}{\partial n_i^m} = \dot{f}^M(n_i^M)$$

This implies

$$s_i^M = -2(t_i - a_i)\dot{f}^M(n_i^M)$$

Written in matrix notation this would mean

$$\mathbf{s}^M = -2\mathbf{F}^M(\mathbf{n}^M)(\mathbf{t}-\mathbf{a})$$

Given the above set of operations the weights and biases are updated using

$$\mathbf{W}^m(k+1) = \mathbf{W}^m(k) - \alpha\mathbf{s}^m(\mathbf{a}^{m-1})', \tag{A.14}$$

$$\mathbf{b}^m(k+1) = \mathbf{b}^m(k) - \alpha\mathbf{s}^m \tag{A.15}$$

at the k^{th} iteration.

The backpropagation as described is just a steepest descent algorithm. Such algorithm converges most of the times but usually is found to be quite slow and hence is not efficient. As a consequence some variations of the same are usually used rather than the pure steepest descent with backpropogation algorithm.

The steepest descent requires calculating the second derivatives. Other procedures such as the Gauss Newton method do not require calculating second derivatives. In the steepest descent method the weights or biases (vector x) are updated with the relation

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{A}_k^{-1}\mathbf{g}_k$$

where A_k is the Hessian$g_k$is the gradient. The gradient can be written in matrix form as

$$\nabla F(\mathbf{x}) = 2\mathbf{J}'(\mathbf{x})\mathbf{e}(\mathbf{x}), \tag{A.16}$$

where

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial e_1(\mathbf{X})}{\partial x_1} & \frac{\partial e_1(\mathbf{X})}{\partial x_2} & \cdots & \frac{\partial e_1(\mathbf{X})}{\partial x_n} \\ \frac{\partial e_2(\mathbf{X})}{\partial x_1} & \frac{\partial e_2(\mathbf{X})}{\partial x_2} & \cdots & \frac{\partial e_2(\mathbf{X})}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial e_N(\mathbf{X})}{\partial x_1} & \frac{\partial e_N(\mathbf{X})}{\partial x_2} & \cdots & \frac{\partial e_N(\mathbf{X})}{\partial x_n} \end{bmatrix} \tag{A.17}$$

is the Jacobian matrix and $\mathbf{e}(\mathbf{x})$ is a vector of the error terms. Similarly Hessian matrix can be expressed in the matrix form as

$$\nabla^2 F(\mathbf{x}) = 2\mathbf{J}'(\mathbf{x})\mathbf{J}(\mathbf{x}) + 2\mathbf{S}(x)$$

where

$$\mathbf{S}(x) = \sum_{i=1}^{N} e_i(\mathbf{x})\nabla^2 e_i(x)$$

If the term $\mathbf{S}(x)$ is assumed to be small the Hessian can be approximated as

$$\nabla^2 F(\mathbf{x}) \cong 2\mathbf{J}'(\mathbf{x})\mathbf{J}(\mathbf{x})$$

Using this as the approximation to the Hessian and the term in equation (A.16) as the gradient then the steepest descent is converted into the Gauss Newton method. In this procedure at times it might be difficult to invert the matrix $\mathbf{H} = \mathbf{J}'\mathbf{J}$. For such reasons the modification $\mathbf{G} = \mathbf{H} + \mu\mathbf{I}$ is made to the Hessian such that it becomes positive definite and the matrix is invertible. This updating of the Hessian algorithm is the Levenberg-Marquardt algorithm and it is a combination of steepest descent and Gauss Newton. As μ is increased, the algorithm approaches the steepest descent algorithm with a small learning rate and as μ approaches zero the algorithm approaches the Gauss Newton method. In the neural networks literature the generalization error is referred to the error that the network produces on the unseen data. The objective in general is to produce networks with good generalization. For the network to generalize, the number of parameters of the network should be less than the number of data points. One way to improve generalization is regularization. In regularization, the performance index is not only a function of sum of squared errors but also a function of the sum of squares of the network weights. This way regularization constrains the network complexity. An example of the performance index would be

$$F = \alpha S_e + \beta s_W \tag{A.18}$$

In this if $\alpha \gg \beta$ then the network's training emphasizes more on the sum of squared errors and the network drives the errors smaller. At the same time if $\alpha \ll \beta$ the network's function is much smoother (smaller weights) at the expense of network errors. To implement regularization the primary task is to set the parameters α and β.

David Mackay[1992] has done extensive work on the application of Bayes' rule to optimizing regularization and neural networks. Some of the major results of his work that is relevant to bayesian regularization are presented.

In a Bayesian set up the weights of the network are assumed to be random variables. Once the data are available, the density function for the weights of the network are updated according to Bayes' rule that says

$$f(\mathbf{w}|D,\alpha,\beta,M) = \frac{f(D|\mathbf{w},\beta,M)f(\mathbf{w}|\alpha,M)}{f(D|\alpha,\beta,M)} \tag{A.19}$$

where D is the data set, M is the neural network model used, and $\mathbf{w}$ is the vector of network weights. The density $f(\mathbf{w}|\alpha,M)$ represents the prior knowledge of the distribution of the weights before any data is collected. The density $f(D|\mathbf{w},\beta,M)$ represents the likelihood function of the data points given the weights of the network. And the density $f(D|\alpha,\beta,M)$ is a normalization factor that ensures that the density function integrates to 1. If it is assumed that the noise in the data set has a Gaussian distribution and also the weights of the network are distributed Gaussian, the density functions can be written as

$$\begin{aligned} f(D|\mathbf{w},\beta,M) &= \frac{1}{Z_e(\beta)}\exp(-\beta S_e) \\ f(\mathbf{w}|\alpha,M) &= \frac{1}{Z_w(\alpha)}\exp(-\alpha S_w) \end{aligned}$$

where $Z_e(\beta) = (\Pi/\beta)^{n/2}$ and $Z_w(\alpha) = (\Pi/\alpha)^{N/2}$. If these densities are substituted in equation (A.19)

$$f(\mathbf{w}|D,\alpha,\beta,M) = \frac{\frac{1}{Z_w(\alpha)}\frac{1}{Z_e(\beta)}\exp(-(\beta S_e + \alpha S_w))}{f(D|\alpha,\beta,M)} \tag{A.20}$$

$$= \frac{1}{Z_F(\alpha,\beta)}\exp(-F(\mathbf{w})) \tag{A.21}$$

In this Bayesian scenario the optimal weights should maximize the posterior probability $f(\mathbf{w}|D,\alpha,\beta,M)$. Also minimizing the regularized function $F = (\beta S_e + \alpha S_w)$ is equivalent to maximizing the posterior probability.

The next step is to apply Bayes' rule to optimizing the parameters α and β. Again using Bayes' rule

$$f(\alpha,\beta|D,M)=\frac{f(D|\alpha,\beta,M)f(\alpha,\beta|M)}{f(D|M)} \tag{A.22}$$

or in words

$$Posterior=\frac{Likelihood\ \times\ Prior}{Evidence}.$$

If a uniform prior density $f(\alpha,\beta|M)$ is assumed for the parameters α and β, then maximizing the posterior is achieved by maximizing the likelihood function $f(D|\alpha,\beta,M)$. At this point it is worth noting that the likelihood function is the normalization factor for equation (A.19). Since all the densities are Gaussian so is the posterior density of equation (A.19) shown in equation (A.21). Solving the equation (A.19) for the normalization factor

$$f(f(D|\alpha,\beta,M))=\frac{f(D|\mathbf{w},\beta,M)f(\mathbf{w}|\alpha,M)}{\mathbf{w}|D,\alpha,\beta,M} \tag{A.23}$$

$$=\frac{\left[\frac{1}{Z_e(\beta)}\exp(-\beta S_e)\right]\left[\frac{1}{Z_w(\alpha)}\exp(-\alpha S_w)\right]}{\frac{1}{Z_F(\alpha,\beta)}\exp(-F(\mathbf{w}))}$$

$$\frac{Z_F(\alpha,\beta)}{Z_e(\beta)Z_w(\alpha)}\frac{\exp(-\beta S_e-\alpha S_w)}{\exp(-F(\mathbf{w}))}=\frac{Z_F(\alpha,\beta)}{Z_e(\beta)Z_w(\alpha)} \tag{A.24}$$

In equation (A.24) the terms $Z_e(\beta)$ and $Z_w(\alpha)$ are known but $Z_F(\alpha,\beta)$ is unknown. This can be estimated using a Taylor series expansion. It can be safely assumed that the objective function will have a quadratic form near the minimum point. So $F(\mathbf{w})$ can be expanded using Taylor series expansion around the minimum point of the posterior density $\mathbf{w}^{mp}$ (mp stands for most probable), where the gradient is zero. Now solving for the normalizing constant yields

$$Z_F\approx(2\pi)^{N/2}(det((\mathbf{H}^{mp})^{-1}))^{1/2}\exp(-F(\mathbf{w}^{mp})) \tag{A.25}$$

where $\mathbf{H}$ is the Hessian matrix of the objective function i.e. $\beta\nabla^2 S_e+\alpha\nabla^2 S_w$. By substituting the value of Z_F in equation (A.23) the optimal values of α and β can be solved for using first order conditions. This yields

$$\alpha^{mp}=\frac{\gamma}{2S_e(\mathbf{w}^{mp})}\ and\ \beta^{mp}=\frac{n-\gamma}{2S_w(\mathbf{w}^{mp})} \tag{A.26}$$

where $\gamma=N-2\alpha^{mp}tr(\mathbf{H}^{mp})^{-1}$ is called the effective number of parameters, and N is the total number of parameters in the network. The parameter γ is a measure of how many parameters in the neural network are effectively used in reducing the error function.

At this stage another question that arises is about the estimation of the Hessian matrix of the function $F(w)$. Foresee and Hagan[1997] suggest a Gauss Newton approximation to the Hessian. The following are the steps required for the Bayesian optimization of the regularization parameters, with the Gauss Newton approximation to the Hessian.

1. Initialize α, β and the weights.

2. Take one step of the Levenberg-Marquardt algorithm to minimize the objective function $F(w) = \beta S_e + \alpha S_w$.

3. compute the effective number of parameters $\gamma = N - 2\alpha^{mp} tr(\mathbf{H}^{mp})^{-1}$ making use of the Gauss Newton approximation to the Hessian available in the Levenberg-Marquardt training algorithm that implies

$$\mathbf{H} = \nabla^2 F(\mathbf{w}) \approx 2\beta \mathbf{J}'\mathbf{J} + 2\alpha \mathbf{I}_N$$

where $\mathbf{J}$ is the Jacobian matrix of the training set errors.

4. compute new estimates for the objective function parameters $\alpha^{mp} = \frac{\gamma}{2S_e(\mathbf{W}^{mp})}$ *and* $\beta^{mp} = \frac{n-\gamma}{2S_w(\mathbf{W}^{mp})}$.

5. Now iterate steps 1 through 3 until the algorithm converges.

   The default values used for the training algorithm are as follows

   - Maximum number of epochs to train the network - 100
   - The adjustment parameter μ in the Levenberg-Marquardt algorithm - 0.005
   - The factor used to decrease the μ whenever required - 0.1
   - The factor used to increase the μ whenever required - 10

With this the complete algorithm that would be used for the MLP is the Gauss Newton approximation to the Bayesian regularization.

The mean absolute deviations and root mean squared errors for the MLP model during the sample period January 1996 till December 1999 for the 48 months are presented in tables A.1 and table A.2 respectively for the network MLP when the number of iterations is changed.

Since the reduction in the error after the first 100 iterations is almost insignificant the results used in the study are the ones with the number of iterations set to 100.

The mean absolute deviations and root mean squared errors for the MLP model during the sample period January 1996 till December 1999 for the 48 months are presented in tables A.3 and table A.4 respectively for the network MLP when the value of the Levenberg-Mauquardt algorithm parameter μ is changed.

Again since the results for a particular value of μ are not different for another value of μ the value of μ chosen for the study is the value of 0.005.

Another issue that arises in global optimization techniques like the one used to train the MLP is whether the network converges to the same solution when one starts the training algorithm with different random starting values. To analyze how sensitive the results of the network are to different

Table A.1: Mean Absolute Deviation(variable i)

The mean absolute deviations are calculated for the MLP model during the sample period January 1996 till December 1999 for the 48 months. This is done for different values of i where i denotes the number of iterations in the training algorithm. The errors are in cents on a bond with a face value of a dollar.

Maturity	$i = 100$	$i = 200$	$i = 300$	$i = 400$
.25	0.0177	0.0177	0.0177	0.0177
1	0.0718	0.0721	0.0721	0.0721
2	0.0872	0.0873	0.0873	0.0874
3	0.0919	0.0898	0.0897	0.0896
7	0.6207	0.6206	0.6207	0.6207
10	0.4677	0.4681	0.4685	0.4685
15	2.6478	2.6496	2.6490	2.6491
20	1.0116	1.0093	1.0092	1.0092

Table A.2: Root Mean Squared Error(variable i)

The root mean squared errors are calculated for the MLP model during the sample period January 1996 till December 1999 for the 48 months. This is done for different values of i where i denotes the number of iterations in the training algorithm. The errors are in cents on a bond with a face value of a dollar.

Maturity	$i = 100$	$i = 200$	$i = 300$	$i = 400$
.25	0.0215	0.0215	0.0215	0.0215
1	0.0874	0.0876	0.0876	0.0876
2	0.1222	0.1223	0.1223	0.1224
3	0.1334	0.1314	0.1314	0.1314
7	0.6784	0.6783	0.6783	0.6783
10	0.6099	0.6110	0.6114	0.6114
15	2.9557	2.9569	2.9563	2.9563
20	1.3001	1.2979	1.2978	1.2978

Table A.3: Mean Absolute Deviation(variable μ)

The mean absolute deviations are calculated for the MLP model during the sample period January 1996 till December 1999 for the 48 months. This is done for different values of μ where μ denotes the Levenberg-Marquardt algorithm parameter. The errors are in cents on a bond with a face value of a dollar.

Maturity	$\mu = 0.003$	$\mu = 0.004$	$\mu = 0.005$	$\mu = 0.006$	$\mu = 0.007$
.25	0.0176	0.0174	0.0177	0.0178	0.0176
1	0.0739	0.0731	0.0718	0.0727	0.0728
2	0.0822	0.0871	0.0872	0.0833	0.0869
3	0.0911	0.0913	0.0919	0.0933	0.0945
7	0.6172	0.6179	0.6207	0.6267	0.6271
10	0.4671	0.4742	0.4677	0.4693	0.4575
15	2.6269	2.6560	2.6478	2.6587	2.6577
20	0.9862	0.9998	1.0116	0.9972	0.9744

Table A.4: Root Mean Squared Error(variable μ)

The root mean squared errors are calculated for the MLP model during the sample period January 1996 till December 1999 for the 48 months. This is done for different values of μ where μ denotes the Levenberg-Marquardt algorithm parameter. The errors are in cents on a bond with a face value of a dollar.

Maturity	$\mu = 0.003$	$\mu = 0.004$	$\mu = 0.005$	$\mu = 0.006$	$\mu = 0.007$
.25	0.0215	0.0212	0.0215	0.0217	0.0215
1	0.0898	0.0886	0.0874	0.0891	0.0899
2	0.1064	0.1199	0.1222	0.1141	0.1205
3	0.1234	0.1302	0.1334	0.1341	0.1360
7	0.6770	0.6770	0.6784	0.6824	0.6821
10	0.6022	0.6077	0.6099	0.6194	0.5937
15	2.9293	2.9748	2.9557	2.9781	2.9741
20	1.2667	1.2743	1.3001	1.2988	1.2647

starting values the study starts the training of the networks during the first out of sample date with 100 different starting values selected randomly. The first out of sample date is the first trading day of January 1996 and the network uses the data on the yields of bonds with .5, and 5 years to maturity, from January 1983 through 1999. The different sum of squared errors for the maturities .25, 1, 2, 3, 7, 10, 15, and 20 when the network is randomly initialized with 100 different starting values are shown in Figures A.1, A.2, A.3, A.4, A.5, A.6, A.7, and A.1 respectively.

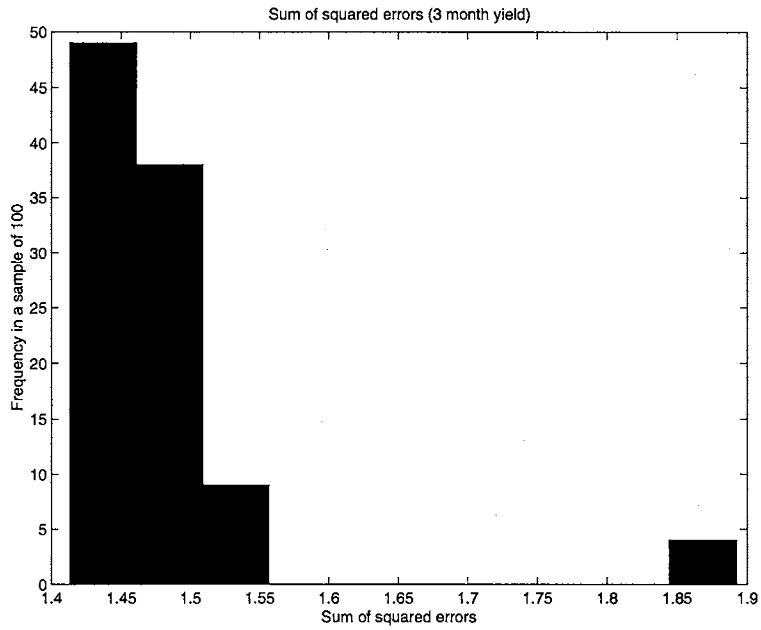

Figure A.1: Sum of Squared Errors (3 month yield)

The Figure shows the sum of the squared errors during the in sample period of January 1983 through December 1995 for the 3 month maturity yield. The network is started 100 times with different random initial values for the weights and biases and the sum of squared errors is plotted.

The different values of the changes in yields for the maturities .25, 1, 2, 3, 7, 10, 15, and 20 when the network is randomly initialized with 100 different starting values are shown in Figures A.9, A.10, A.11, A.12, A.13, A.14, A.15, and A.9 respectively.

Using the 100 different randomly selected initial values for the weights to train the same network the results from the figures above show that the final output of the network is not significantly different by using different starting values. Though every trained network's weights do not correspond so closely to one another as do the final output and the sum of squared errors. This is evident by looking at the variability of 7 out of 89 weights of the network for 3 month maturity on the first trading day of January 1996 as shown in the Table A.5.

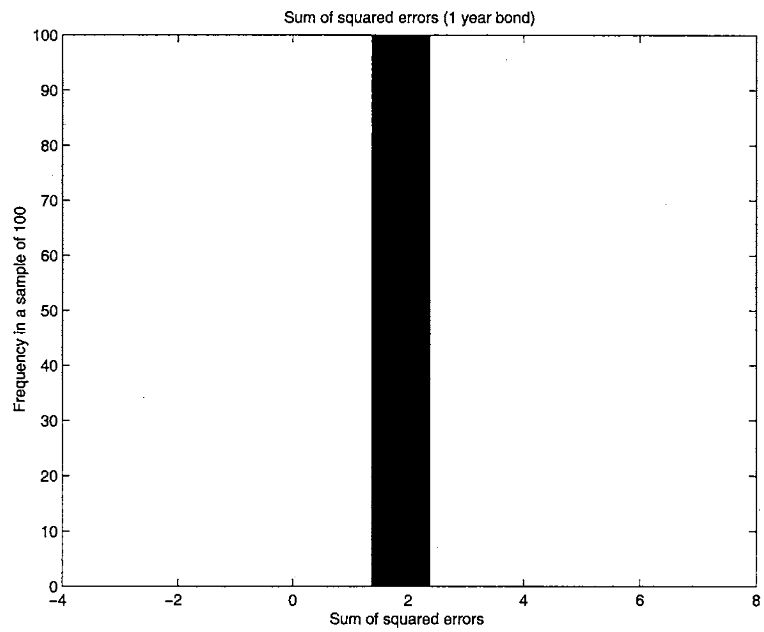

Figure A.2: Sum of Squared Errors (1 year bond)

The Figure shows the sum of the squared errors during the in sample period of January 1983 through December 1995 for the bond with 1 years to maturity. The network is started 100 times with different random initial values for the weights and biases and the sum of squared errors is plotted.

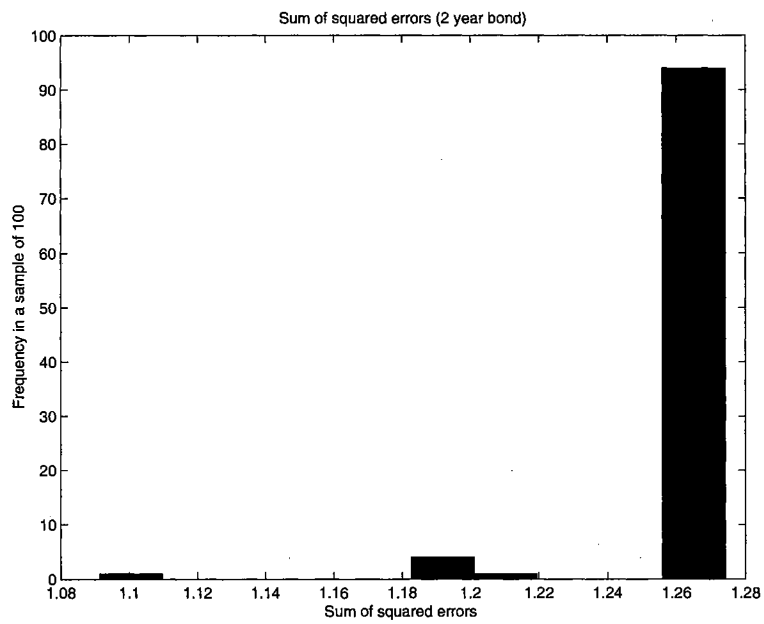

Figure A.3: Sum of Squared Errors (2 year bond)

The Figure shows the sum of the squared errors during the in sample period of January 1983 through December 1995 for the bond with 2 years to maturity. The network is started 100 times with different random initial values for the weights and biases and the sum of squared errors is plotted.

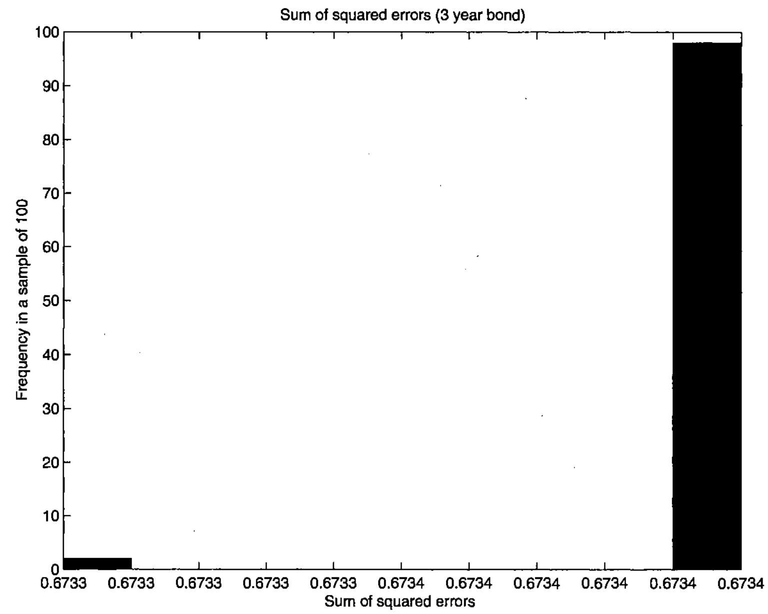

Figure A.4: Sum of Squared Errors (3 year bond)

The Figure shows the sum of the squared errors during the in sample period of January 1983 through December 1995 for the bond with 3 years to maturity. The network is started 100 times with different random initial values for the weights and biases and the sum of squared errors is plotted.

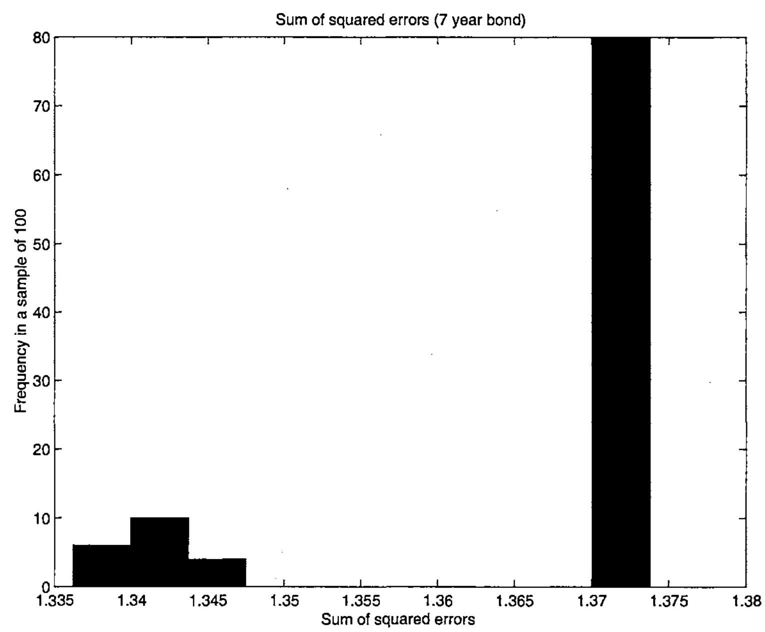

Figure A.5: Sum of Squared Errors (7 year bond)

The Figure shows the sum of the squared errors during the in sample period of January 1983 through December 1995 for the bond with 7 years to maturity. The network is started 100 times with different random initial values for the weights and biases and the sum of squared errors is plotted.

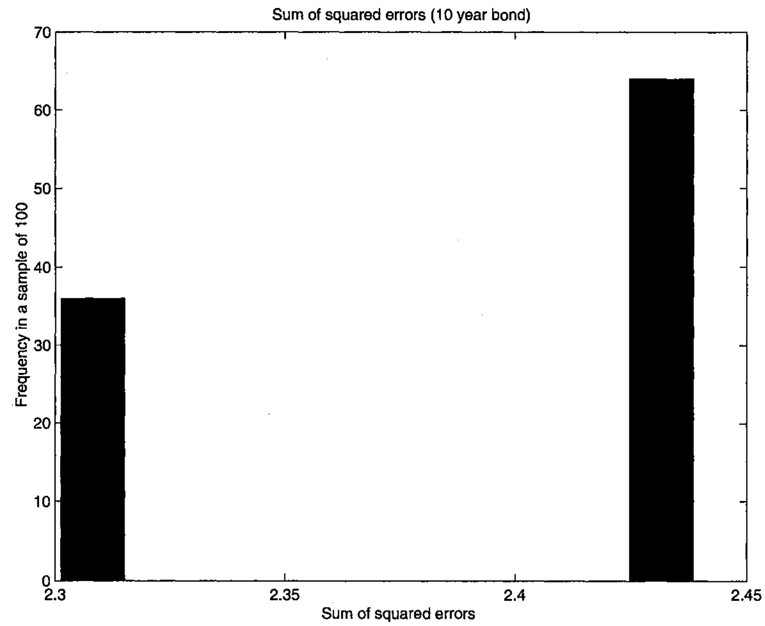

Figure A.6: Sum of Squared Errors (10 year bond)

The Figure shows the sum of the squared errors during the in sample period of January 1983 through December 1995 for the bond with 10 years to maturity. The network is started 100 times with different random initial values for the weights and biases and the sum of squared errors is plotted.

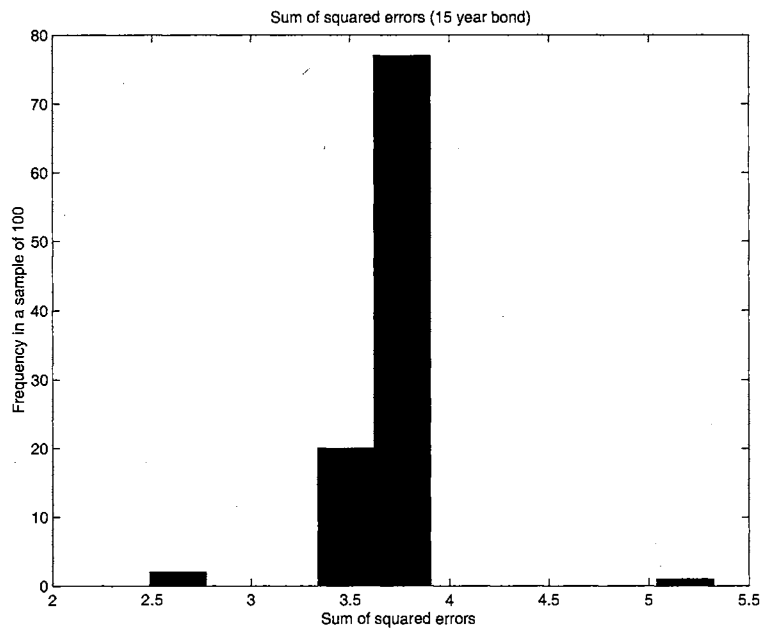

Figure A.7: Sum of Squared Errors (15 year bond)

The Figure shows the sum of the squared errors during the in sample period of January 1983 through December 1995 for the bond with 15 years to maturity. The network is started 100 times with different random initial values for the weights and biases and the sum of squared errors is plotted.

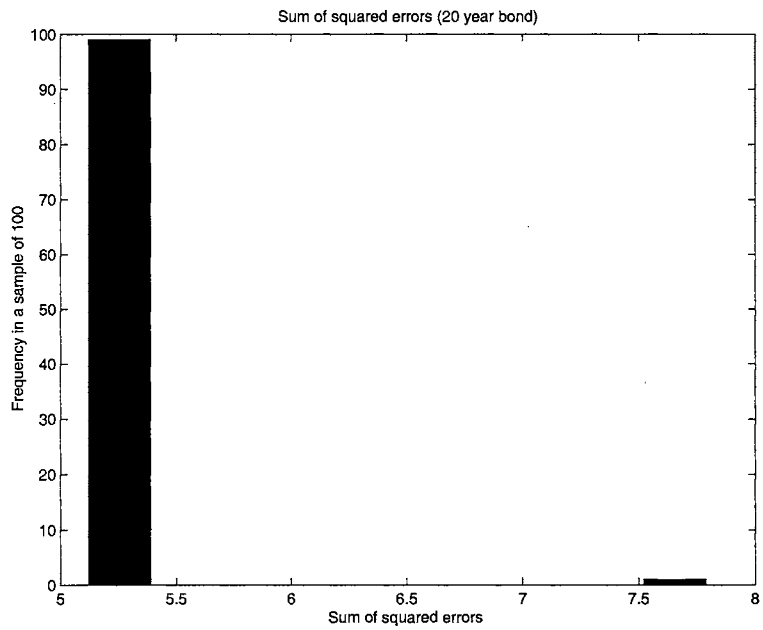

Figure A.8: Sum of Squared Errors (20 year bond)

The Figure shows the sum of the squared errors during the in sample period of January 1983 through December 1995 for the bond with 20 years to maturity. The network is started 100 times with different random initial values for the weights and biases and the sum of squared errors is plotted.

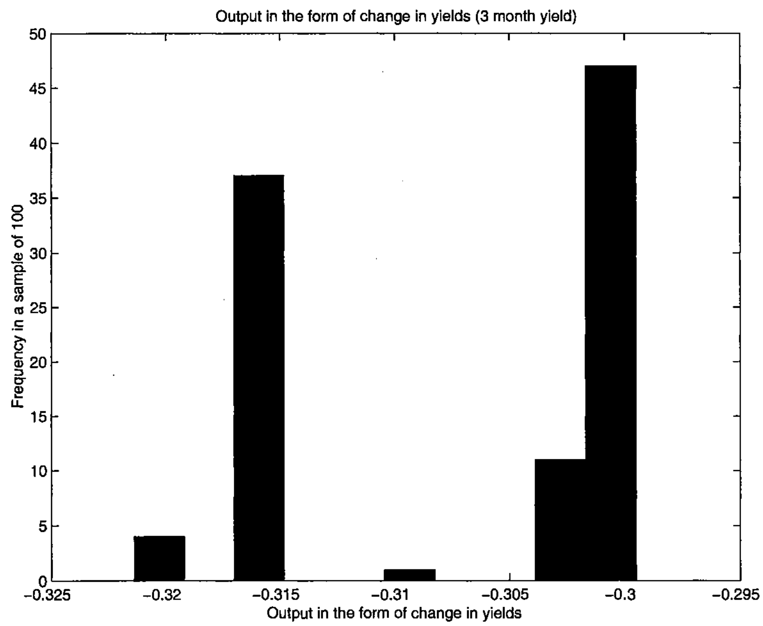

Figure A.9: Output in the form of change in yields (3 month yield)

The Figure shows the output in the form of change in yields for the 3 month maturity during the first trading day in January 1996 using the data on the yields from January 1983 through December 1995. The network is started 100 times with different random initial values for the weights and biases and the output in the form of change in yields is plotted.

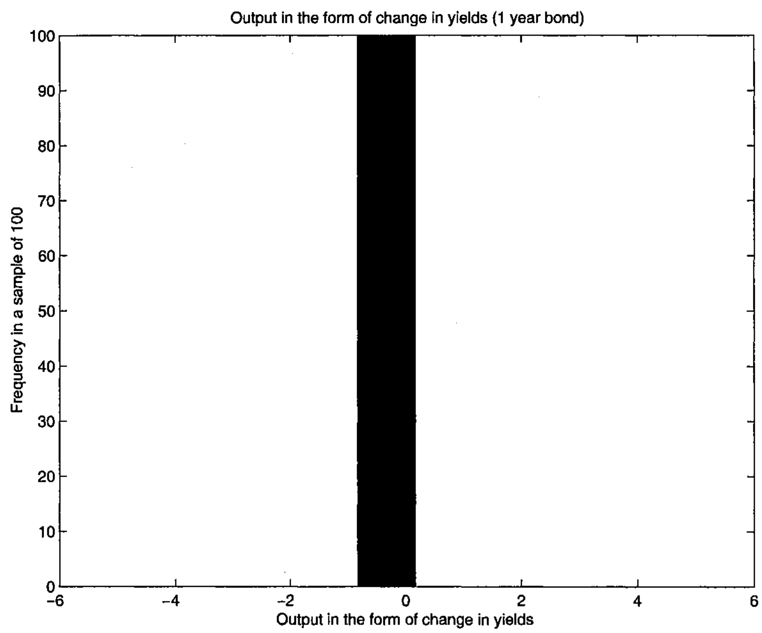

Figure A.10: Output in the form of change in yields (1 year bond)

The Figure shows the output in the form of change in yields for the 1 year maturity bond during the first trading day in January 1996 using the data on the yields from January 1983 through December 1995. The network is started 100 times with different random initial values for the weights and biases and the output in the form of change in yields is plotted.

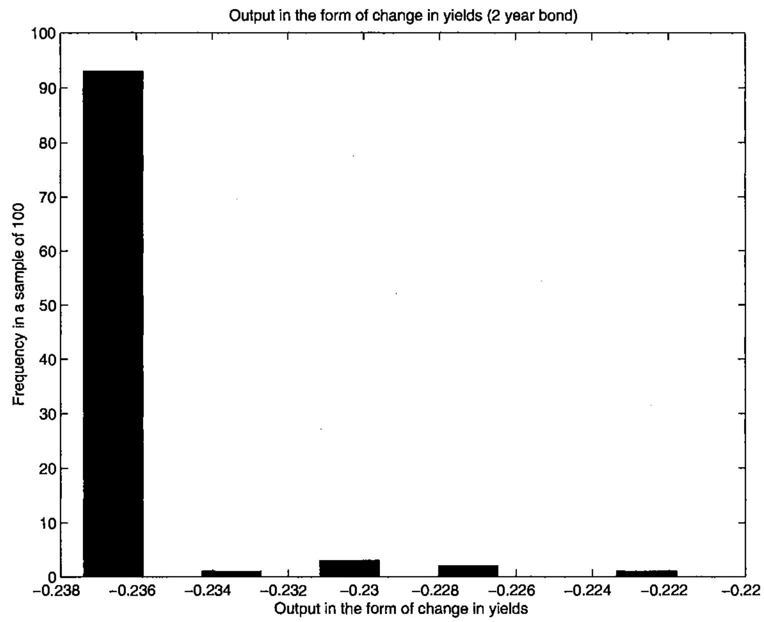

Figure A.11: Output in the form of change in yields (2 year bond)

The Figure shows the output in the form of change in yields for the 2 year maturity bond during the first trading day in January 1996 using the data on the yields from January 1983 through December 1995. The network is started 100 times with different random initial values for the weights and biases and the output in the form of change in yields is plotted.

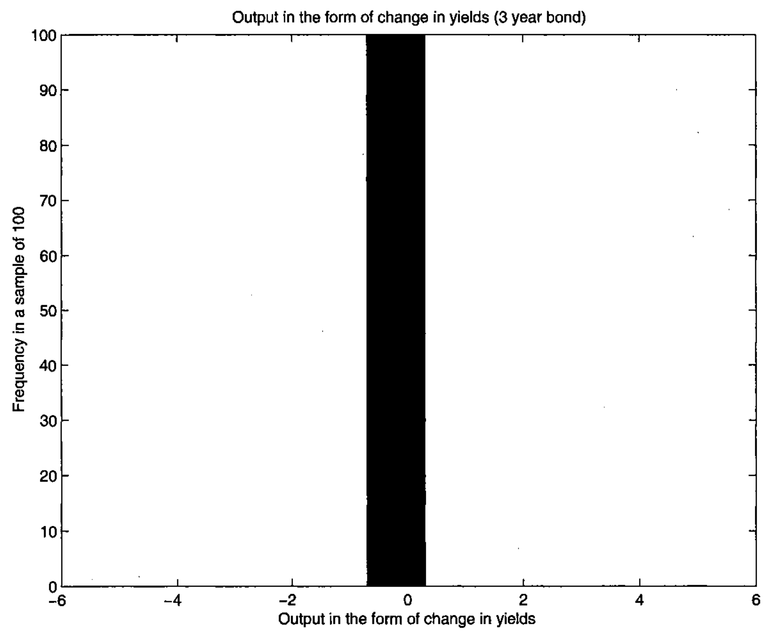

Figure A.12: Output in the form of change in yields (3 year bond)

The Figure shows the output in the form of change in yields for the 3 year maturity bond during the first trading day in January 1996 using the data on the yields from January 1983 through December 1995. The network is started 100 times with different random initial values for the weights and biases and the output in the form of change in yields is plotted.

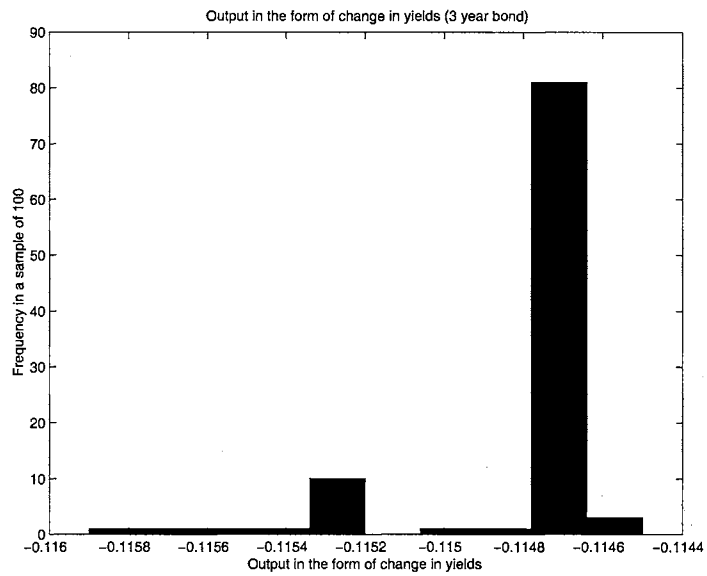

Figure A.13: Output in the form of change in yields (7 year bond)

The Figure shows the output in the form of change in yields for the 7 year maturity bond during the first trading day in January 1996 using the data on the yields from January 1983 through December 1995. The network is started 100 times with different random initial values for the weights and biases and the output in the form of change in yields is plotted.

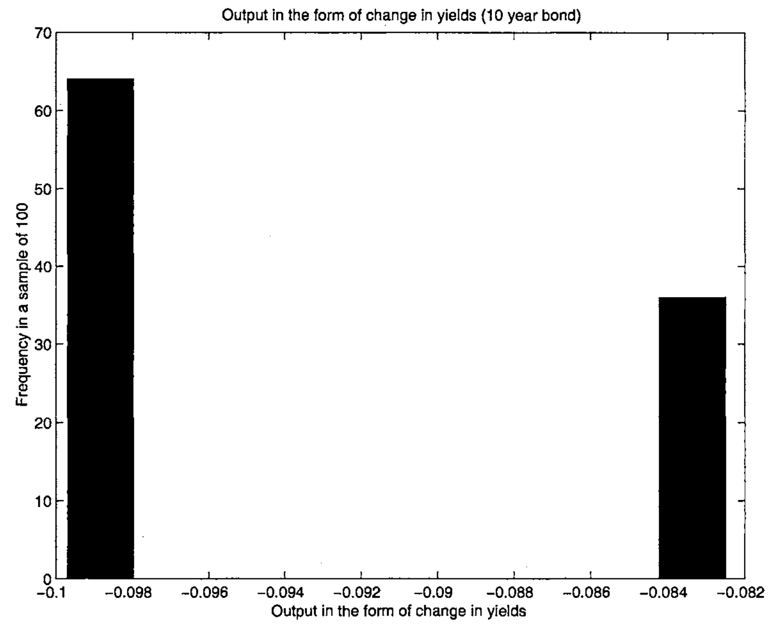

Figure A.14: Output in the form of change in yields (10 year bond)

The Figure shows the output in the form of change in yields for the 10 year maturity bond during the first trading day in January 1996 using the data on the yields from January 1983 through December 1995. The network is started 100 times with different random initial values for the weights and biases and the output in the form of change in yields is plotted.

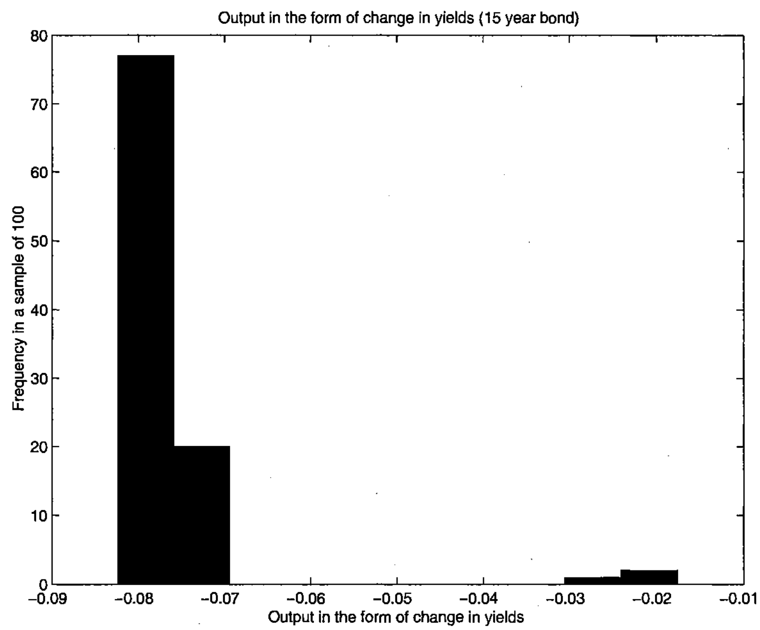

Figure A.15: Output in the form of change in yields (15 year bond)

The Figure shows the output in the form of change in yields for the 15 year maturity bond during the first trading day in January 1996 using the data on the yields from January 1983 through December 1995. The network is started 100 times with different random initial values for the weights and biases and the output in the form of change in yields is plotted.

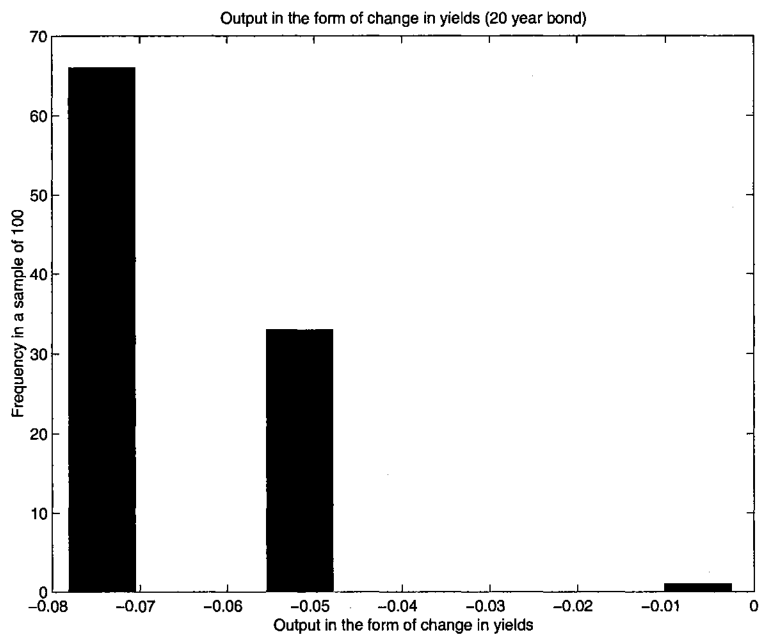

Figure A.16: Output in the form of change in yields (20 year bond)

The Figure shows the output in the form of change in yields for the 20 year maturity bond during the first trading day in January 1996 using the data on the yields from January 1983 through December 1995. The network is started 100 times with different random initial values for the weights and biases and the output in the form of change in yields is plotted.

Table A.5: Weights of the network

The weights of the network (7 out of 89) for 3 month maturity are shown for 50 random restarts.

Restart No.	w1	w2	w3	w4	w5	w6	w7
1	-0.3842	0.3903	0.3928	0.3767	0.3928	-0.3778	0.3913
2	-0.4038	0.3988	-0.3833	-0.3997	-0.8808	0.3926	-0.3928
3	-0.3876	-0.3899	0.39	-0.3899	0.3864	-0.39	-0.3896
4	0.2473	0.2461	-0.2473	0.247	-0.2471	-0.2473	-0.2473
5	0.3945	0.395	-0.395	-0.395	0.3951	-0.3946	-1.0215
6	-0.3894	-0.3885	0.3894	-0.3895	0.3869	-0.3894	-0.3865
7	0.4002	0.4012	-0.3994	0.4051	0.4051	0.4016	-0.3987
8	-0.2486	-0.2554	0.2429	0.1612	0.2637	0.1441	0.214
9	0.2512	-0.2025	-0.2499	0.2425	0.2426	-0.2081	0.2524
10	1.1061	-0.3915	-0.3939	1.0218	0.3703	-0.3932	-0.3668
11	-0.3894	0.3894	0.3894	-0.3894	-0.3813	-0.3878	-0.3889
12	-0.2277	0.2288	0.8659	0.2335	-0.2361	-0.2361	-0.236
13	-0.263	0.2769	-0.2138	-0.2788	0.2746	-0.0635	0.1303
14	-1.1056	0.3939	-0.3944	-0.3938	-0.3943	0.3944	0.3944
15	0.1061	-0.2346	0.0713	0.2433	0.2593	1.0405	-0.2582
16	0.3466	0.3458	0.346	-0.3432	0.0998	0.0045	-0.3424
17	-0.3918	-0.3918	-0.3886	-0.391	0.3866	0.3909	0.3919
18	-0.1713	0.2419	0.2419	-0.194	0.2382	0.1688	0.2401
19	0.1773	-0.1827	0.2335	-0.2205	0.2518	-0.2426	-0.2502
20	-0.3281	-1.0122	-1.4185	-0.3255	0.3035	0.3235	0.3276
21	0.3987	0.3983	-0.3988	-0.3988	0.3988	0.3987	0.3988
22	-0.3726	-0.4116	0.4117	1.1696	-0.4117	0.3812	0.4116
23	-0.2326	1.0399	0.239	0.2388	-0.239	0.2391	-0.2321
24	0.3945	-0.3929	0.388	0.3944	-0.3944	-0.3945	0.3885
25	0.2587	0.2557	0.2224	-0.2584	-0.2388	0.8676	0.177
26	-0.867	0.2416	0.2425	0.2373	0.2371	-1.3279	0.242
27	0.2253	-0.0974	0.2724	-0.2775	-0.2773	-0.1459	0.2671
28	-0.2639	-0.2616	-0.2641	-0.2628	-0.18	0.1534	0.098
29	-0.3929	0.3824	0.3613	-0.3902	0.3929	0.3929	1.0216
30	0.2362	0.2364	-1.3281	-0.2326	-0.2362	0.2195	-0.2167
31	0.4391	0.4391	-0.4391	-0.4391	-0.4391	0.4391	0.4391
32	-0.3261	-0.3262	-0.326	0.3262	-0.3162	-0.3228	0.3244
33	1.106	-0.3899	0.386	-0.3874	-0.3896	-0.3896	0.3896
34	0.392	-0.3895	-0.3925	0.3923	0.3889	-0.3927	-0.3871
35	0.3897	-0.3849	-0.3895	0.3897	0.3879	0.3897	-1.0218
36	-0.3888	0.3888	-0.3888	-0.3888	0.3888	-0.3888	0.3871
37	0.3897	-0.3897	-0.3897	-0.3896	0.3897	-0.3898	-0.3825
38	-0.2529	-0.2528	0.2527	0.2527	-0.2528	-0.2514	0.2528
39	0.2505	0.2488	0.2209	-0.2333	0.2483	-0.2178	-0.2503
40	0.3931	-0.3931	0.3931	-1.1062	0.3897	-0.3931	-0.3931
41	-0.3889	0.3849	0.3891	0.3864	0.3891	1.172	0.3884
42	-0.1765	-0.1659	0.2802	0.0452	0.2358	-0.2416	0.2794
43	0.2142	0.1055	-0.1945	0.2622	0.2292	0.2376	-0.2042
44	-0.3881	-0.3884	-0.3888	0.3887	-0.3882	-0.3888	-0.3883
45	-0.3905	0.3905	0.3883	-0.3901	-0.3905	0.3884	-0.3627
46	-0.4391	-0.439	-0.4391	0.4389	-0.4391	1.2791	0.4391
47	-0.3893	-0.3888	0.3894	-0.3892	-0.3893	-0.3893	-0.3893
48	-0.2069	-0.2572	-0.2364	0.2641	0.2615	-0.2635	-0.2636
49	-0.3985	-0.352	0.3985	-0.3986	-0.3983	-0.3986	0.8794
50	1.0402	-0.2465	-1.3279	0.2314	0.1577	0.8669	-0.2432

APPENDIX B

The Algorithm for the Training of RBF

A fixed center in the Radial Basis Function (hereafter RBF) corresponds to a given regressor in a linear regression model and the selection of centers is regarded as a problem of subset model selection.The algorithm is a method used as a forward regression procedure to select a suitable set of centers (regressors) from a large set of candidates. At each step the increment to the explained variance of the desired output is maximised.

For a RBF There are four choices for the fixed non-linear functions that are as follows:

- The thin plate spline function:

$$\phi(\nu) = \nu^2 \log(\nu)$$

- The Gaussian function :

$$\phi(\nu) = \exp\left(\frac{-\nu^2}{\beta^2}\right)$$

- The multi quadratic function:

$$\phi(\nu) = \sqrt{\nu^2 + \beta^2}$$

- The inverse multi quadratic function:

$$\phi(\nu) = \frac{1}{\sqrt{\nu^2 + \beta^2}}$$

Once the choice of the non-linear function is made, the next step in the designing of the first layer of the RBF network is the choice of the centers. At the beginning the whole set of data points are potential centers. But it is computationally quite complex to choose all the data points if the number of data points are large. So some people choose centers arbitrarily. This can lead to poor performance of the network because most of the time it does not satisfy the criterion that the centers should span the entire input space. Under such circumstances one could experience ill conditioning and techniques such as singular value decomposition have to be used. The study uses an orthogonal least squares algorithm to select the centers in order to get an adequate and parsimonious RBF. The hidden layer of the RBF performs a fixed non-linear transformation of the inputs and the outer layer of the RBF linearly combines the outputs from the hidden layer using certain weights. The only adjustable parameters are the weights of the linear combiner in the second layer which can be chosen using a special case of linear regression model. The special case of the regression model considered is

$$d(t) = \sum p_i(t)\theta_i + \epsilon(t). \tag{B.1}$$

Where the d is the desired output and is also called the dependent variable in the regression terminology, the θ_i are the set of parameters, $p_i(t)$ are known as the regressors which are the fixed non-linear transformation of the input which can be any of the non-linear functions discussed above. But for the purpose of this study the chosen function is the Gaussian function. At this point the problem of choosing the potential centers can be looked at as a potential subset selection of significant regressors from the set of all possible choices (the whole data set). The orthogonal least squares learning algorithm used in the study is as follows. For $t = 1$ to N (N to begin with is all the potential centers i.e. all data points, 156 in this study) the equation (B.1) is rearranged in matrix notation to get

$$d = P\theta + E \tag{B.2}$$

where

$$\begin{aligned} d &= [d(1)\ldots d(N)]', \\ P &= [p_1 \ldots p_M], \\ p_i &= [p_i(1)\ldots p_i(N)]', 1 \leq i \leq M \\ \theta &= [\theta_1 \ldots \theta_M]', \end{aligned}$$

and

$$E = [\epsilon(1)\ldots\epsilon(N)]'.$$

The set of vectors p_i form a set of basis vectors, and the least squares solution to the set of equation (B.1) satisfies the condition that $P\hat{\theta}$ represents the expected value of the desired output d or in other words $P\hat{\theta}$ is the projection of d onto the space spanned by these basis vectors. So the square of this projection indicates the part of the variance that is explained by the regressors. since different regressors are generally correlated so it is difficult to quantify the explained variance due to each regressor. The algorithm essentially involves transforming the set of basis vectors into a set of orthogonal basis vectors which makes it possible to quantify the individual contribution to the explaining of the variance of the dependent variable.

The regression matrix P is first decomposed into

$$P = WA$$

Where

$$A = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \dots & \alpha_{1M} \\ 0 & 1 & \alpha_{23} & \dots & \alpha_{2M} \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & & & 1 & \alpha_{(M-1)M} \\ 0 & \dots & 0 & 0 & 1 \end{bmatrix}$$

Where A is an M x M unit upper-triangular matrix and

$$W = [w_1 w_2 \dots w_M]$$

is an $N \times M$ matrix with orthogonal columns w_i that satisfy

$$W'W = H$$

where H is a diagonal matrix with elements h_i as

$$h_i = w_i' w_i = \sum_{t=1}^{N} w_i(t) w_i(t), 1 \leq i \leq M$$

The space spanned by these orthogonal vectors is essentially the same space spanned by the vectors p_i. So the linear regression equation rewritten for this set of orthogonal basis vectors is

$$d = Wg + E \tag{B.3}$$

The least squares solution to (B.3) is found by minimizing the squared errors given by

$$\begin{aligned} SE &= (d - Wg)'(d - Wg) \\ \Rightarrow SE &= (d - g'W')(d - Wg) \\ \Rightarrow SE &= d'd - 2g'W'd + g'W'Wg \end{aligned}$$

Taking the first partial of SE w.r.t. g and equating it to zero

$$\frac{\partial SE}{\partial g} = -2W'd + 2gW'W = 0$$

Solving for g

$$\hat{g} = (W'W)^{-1} W'd$$

It can be verified by looking at the second derivative that it is positive as it is just $2W'W$. So the above estimate of g minimizes the squared errors and similarly for the space spanned by the vectors p_i the estimate of $\hat{\theta}$ is given by

$$\hat{\theta} = (P'P)^{-1} P'd \tag{B.4}$$

As $P = WA$ and $W'W = H$, so substituting the same in the equation (B.4)

$$\begin{aligned} \hat{\theta} &= ((WA)'WA)^{-1}(WA)'d \\ \Rightarrow \hat{\theta} &= (A'HA)^{-1}(A'W'd) \end{aligned}$$

By premultiplying both sides of the above equation with $(A'HA)$

$$A'HA\hat{\theta} = A'W'd \tag{B.5}$$

Substituting $H = W'W$ in the estimate of $\hat{g}$

$$\hat{g} = H^{-1}W'd$$

$$\Rightarrow H\hat{g} = W'd$$

Substituting in equation (B.5)

$$A'HA\hat{\theta} = A'H\hat{g}$$

$$\Rightarrow A\hat{\theta} = \hat{g}$$

So the quantities $\hat{\theta}$ and $\hat{g}$ satisfy the triangular system as stated above. At this stage the study uses the well known classical Gram-Schmidt method that computes one column of A at a time and orthogonalizes P as follows. At every step k the k^{th} column is made orthogonal to all the previous $k-1$ orthogonalized columns and the procedure is repeated for $k = 2, \dots M$. Since the columns of W are all orthogonal to one another the sum of squared errors of d can be written as

$$d'd = \sum_{i=1}^{M} g_i^2 w_i' w_i + E'E$$

which implies that if d is the desired output or the dependent variable after its mean has been subtracted, the variance of $d(t)$ is given by

$$N^{-1}d'd = N^{-1}\sum_{i=1}^{M} g_i^2 w_i' w_i + N^{-1}E'E$$

So the term $N^{-1}\sum_{i=1}^{M} g_i^2 w_i' w_i$ is that part of the total variance of the dependent variable d that is explained by the regressors and $N^{-1}E'E$ is that part of the variance that is unexplained. Thus an error reduction ratio due to w_i can be defined as

$$[err]_i = \frac{g_i^2 w_i^T w_i}{(d'd)}, \quad 1 \leq i \leq M \tag{B.6}$$

This ratio is essentially what turns out to be a simple and effective means to seek a subset of significant regressors in a forward regression manner.

The process starts by assuming the first column of W to be the first column of P and the maximum error reduction ratio is computed by taking the following steps

$$\begin{aligned} w_1^i &= p_i \\ g_1^i &= \frac{(w_1^i)'d}{(w_1^i)'w_1^i} \\ [err]_1^i &= \frac{(g_1^i)^2(w_1^i)'w_1^i}{d'd} \end{aligned}$$

Now find

$$[err]_1^{i_1} = \max [err]_1^i, \quad 1 \leq i \leq M, \quad i \neq i_1, \ldots, i \neq i_{k-1}$$

Select

$$w_1 = w_1^{i_1} = p_{i_1}$$

After the first step from the second step onwards, for $k \geq 2$, for $1 \leq i \leq M$, and $i \neq i_1, \ldots i \neq i_{k-1}$,

$$\begin{aligned} \alpha_{jk}^i &= \frac{w_j^T p_i}{w_j' w_j}, \quad 1 \leq j < k \\ w_k^i &= p_i - \Sigma_{j=1}^{k-1} \alpha_{jk}^i w_j \\ g_k^i &= \frac{(w_k^i)'d}{(w_k^i)'w_k^i} \\ [err]_k^i &= \frac{(g_k^i)^2(w_k^i)'w_k^i}{d'd} \end{aligned}$$

And

$$[err]_k^{i_1} = \max [err]_k^i, \quad 1 \leq i \leq M, \quad i \neq i_1, \ldots, i \neq i_{k-1}$$

Now select

$$w_1 = w_k^{i_1} = p_{i_{k_1}} - \Sigma_{j=1}^{k-1} \alpha_{jk}^i w_j$$

The procedure is repeated and performed till the tolerence limit is reached, i.e.

$$1 - \sigma_{j=1}^{M_s} [err]_j < \rho$$

The tolerance parameter ρ is an important instrument in balancing the accuracy and the complexity of the final network. The actual value of the tolerance used in the study is 0.001.

APPENDIX C

Principal Components Analysis

Let p be a vector of variables, X_t. The principal component analysis is applied to the change in yields for any two consecutive months of different maturities. For the first window the monthly change in yields for the maturities ($.25, 1, 2, 3, 5, 7, 10, 15$, and 20 years) from January 1983 through 1995 are used. Hence X_t would be the change in yields observed at time t for p different maturities. Let T be the number of observations taken (for the purpose of the first window and for all other windows it is 155). It is assumed that these observations are linearly related to $m, m < p$, underlying unobserved factors by the following relation:

$$X - \mu = FL + \epsilon \tag{C.1}$$

where the dimensions of the matrices are

$$\begin{aligned}
X - \mu &= T \times p, \\
L &= m \times p, \\
F &= T \times m, \\
\epsilon &= T \times p, \\
E(X) &= \mu, \\
E(F_t) &= 0, \\
E(\epsilon_t) &= 0, \\
cov(F_t) &= E(F_t F_t') = I(m \times m), \\
cov(\epsilon_t) &= E(\epsilon_t \epsilon_t') = \psi(p \times p).
\end{aligned}$$

where ψ is a diagonal matrix that has zeros as all the off diagonal elements. The F_ts are called factors at a certain time t and F is a matrix of the factors over the whole time period T and the Ls are called the factor loadings.

The task at hand is the estimation of the factors from the equation (C.1) whose structure implies that

$$\Sigma = E[(X - \mu)'(X - \mu)] = LL' + \psi$$

At this point the first m principal components of the estimated variance-covariance matrix, Σ are used to construct L. Then ψ is estimated from $\Sigma - LL'$ by setting the off diagonal elements to zero. The solution L is not unique. If T is any orthogonal matrix, then $L* = LT$ is also a solution because

$$\Sigma = L^*L^{*'} + \psi = LTT'L' + \psi = LIL' + \psi = LL' + \psi.$$

The rotation also affects the factors but does not affect the ψ. This in turn permits the rotation of the original solution until the factor loadings have meaningful economic interpretation.

C.1 The Two-Factor Model

The rotation matrices in the case of the two-factor model are

$$T1 = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{bmatrix}$$

$$T2 = \begin{bmatrix} \sin\theta_2 & \cos\theta_2 \\ -\cos\theta_2 & \sin\theta_2 \end{bmatrix}$$

Since $T1$ and $T2$ are orthogonal their product $T = T1.T2$ is also orthogonal. So T is the orthogonal matrix used to rotate the matrix L and the parameters θ_1 and θ_2 are estimated using a non-linear optimizer to minimize the variance of the first column of $L*$. This gives the final principal components matrix $L*$.

C.2 The Three-Factor Model

The rotation matrices in the case of the three-factor model are

$$T1 = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 & 0 \\ -\sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$T2 = \begin{bmatrix} \cos\theta_2 & 0 & \sin\theta_2 \\ 0 & 1 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 \end{bmatrix}$$

$$T3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_3 & \sin\theta_3 \\ 0 & -\sin\theta_3 & \cos\theta_3 \end{bmatrix}$$

Since T1, T2 and T3 are orthogonal their product $T = T1.T2.T3$ is also orthogonal. So T is the orthogonal matrix used to rotate the matrix L and the parameters θ_1, θ_2 and θ_3 are estimated using a non-linear optimizer to minimize the variance of the first column of L*. This gives the final principal components matrix L*.

APPENDIX D

Summary Statistics of the Estimates

Table D.1: CIR Estimates(Equation (3.9))

The parameters of the short rate process as specified in set of equations (4.1) and estimated by GMM. The parameters of the short rate process as specified in equation (3.9) are calculated using the set of equations (4.2). The summary statistics of the parameters of equation(**??**) are stated. These parameters are summarized over the out of sample time period January 1996 through December 1999.

Statistic	κ	θ	σ
Maximum	0.0277	5.6098	0.1221
Minimum	0.0041	3.5122	0.0985
Median	0.0111	4.6178	0.1061

Table D.2: LS Estimates(Equation (4.5))

The summary statistics of the parameters of the variance equation of the discrete GARCH model as in the set of equations (4.5) are stated. These parameters are summarized over the out of sample time period January 1996 through December 1999.

Statistic	$\beta 0$	β_1	β_2	β_3
Maximum	0.0000002960	0.0000764000	0.4726	0.6082
Minimum	0.0000000105	0.0000153000	0.2298	0.2726
Median	0.0000000105	0.0000203000	0.4045	0.4893

Table D.3: HJM-GV Estimates(Equation (4.22))

The summary statistics of the parameters of the equation (4.18) are stated. These parameters are summarized over the out of sample time period January 1996 through December 1999.

Statistic	κ	η	σ
Maximum	0.0967	0.0867	0.8622
Minimum	-0.0389	0.0553	0.1474
Median	-0.0178	0.0659	0.2711

Table D.4: HJM-RS Estimates(Equation (4.14))

The summary statistics of the parameters of the equation (4.10) are stated. These parameters are summarized over the out of sample time period January 1996 through December 1999.

Statistic	κ	η
Maximum	0.1952	0.0582
Minimum	0.1127	0.033
Median	0.1494	0.0414

Table D.5: Hedging Errors (Summary Statistics)

The summary statistics of the absolute value of the hedge errors produced for the portfolio P-I for all the models considered are presented. Portfolio P-I consists of one-unit long positions in 1 year and 20 year bonds and one-unit short position in a 10 year bond. The hedge errors are summarized over the out of sample time period February 1996 through December 1999.

Statistic	CIR	LS	HJM-GV	HJM-RS	MLP	RBF	DUR	PCA2	PCA3
Max.	0.0359	0.1406	0.0602	0.0481	0.0219	0.0957	0.0624	0.0628	0.0402
Min.	0.00006	0.0008	0.0001	0.0004	0.0003	0.0004	0.0002	0.00003	0.0004
Med.	0.0052	0.0404	0.0047	0.0045	0.0048	0.0128	0.0058	0.0091	0.0051

Table D.6: Hedge Portfolio Proportions (Summary Statistics) - CIR

The summary statistics of the proportions to be invested in the portfolio of zero coupon bonds of 2 and 3 years to maturity according to the CIR model are presented. The portfolio is used to hedge portfolio P-I. Portfolio P-I consists of one-unit long positions in 1 year and 20 year bonds and one-unit short position in a 10 year bond. The hedge portfolio proportions are summarized over the out of sample time period February 1996 through December 1999.

Statistic	x_2	x_3
Maximum	3.20	0.60
Minimum	0.14	-2.61
Median	1.57	-0.87

Table D.7: Hedge Portfolio Proportions (Summary Statistics) - LS

The summary statistics of the proportions to be invested in the portfolio of zero coupon bonds of 2,3, and 7 years to maturity according to the LS model are presented. The portfolio is used to hedge portfolio P-I. Portfolio P-I consists of one-unit long positions in 1 year and 20 year bonds and one-unit short position in a 10 year bond. The hedge portfolio proportions are summarized over the out of sample time period February 1996 through December 1999.

Statistic	x_2	x_3	x_7
Maximum	83.26	138.34	24.78
Minimum	-102.79	-106.92	-36.37
Median	27.84	-33.81	6.72

Table D.8: Hedge Portfolio Proportions (Summary Statistics) - HJM-GV

The summary statistics of the proportions to be invested in the portfolio of zero coupon bonds of 2 and 3 years to maturity according to the HJM-GV model are presented. The portfolio is used to hedge portfolio P-I. Portfolio P-I consists of one-unit long positions in 1 year and 20 year bonds and one-unit short position in a 10 year bond. The hedge portfolio proportions are summarized over the out of sample time period February 1996 through December 1999.

Statistic	x_2	x_3
Maximum	2.84	1.65
Minimum	-0.79	-2.23
Median	0.89	-0.15

Table D.9: Hedge Portfolio Proportions (Summary Statistics) - HJM-RS

The summary statistics of the proportions to be invested in the portfolio of zero coupon bonds of 2, 3 and 7 years to maturity according to the HJM-RS model are presented. The portfolio is used to hedge portfolio P-I. Portfolio P-I consists of one-unit long positions in 1 year and 20 year bonds and one-unit short position in a 10 year bond. The hedge portfolio proportions are summarized over the out of sample time period February 1996 through December 1999.

Statistic	x_2	x_3	x_7
Maximum	4.35	-2.82	0.73
Minimum	3.33	-4.38	0.10
Median	3.50	-3.09	0.21

Table D.10: Hedge Portfolio Proportions (Summary Statistics) - MLP

The summary statistics of the proportions to be invested in the portfolio of zero coupon bonds of 2,3, and 7 years to maturity according to the MLP model are presented. The portfolio is used to hedge portfolio P-I. Portfolio P-I consists of one-unit long positions in 1 year and 20 year bonds and one-unit short position in a 10 year bond. The hedge portfolio proportions are summarized over the out of sample time period February 1996 through December 1999.

Statistic	x_2	x_3	x_7
Maximum	43.10	27.55	5.44
Minimum	-20.99	-47.66	-5.79
Median	1.34	-0.32	-0.31

Table D.11: Hedge Portfolio Proportions (Summary Statistics) - RBF

The summary statistics of the proportions to be invested in the portfolio of zero coupon bonds of 2,3, and 7 years to maturity according to the RBF model are presented. The portfolio is used to hedge portfolio P-I. Portfolio P-I consists of one-unit long positions in 1 year and 20 year bonds and one-unit short position in a 10 year bond. The hedge portfolio proportions are summarized over the out of sample time period February 1996 through December 1999.

Statistic	x_2	x_3	x_7
Maximum	17.68	16.84	6.10
Minimum	-15.05	-17.90	-9.34
Median	0.49	0.31	-0.08

Table D.12: Hedge Portfolio Proportions (Summary Statistics) - Duration

The summary statistics of the proportions to be invested in the portfolio of zero coupon bonds of 2 and 3 years to maturity according to the Duration model are presented. The portfolio is used to hedge portfolio P-I. Portfolio P-I consists of one-unit long positions in 1 year and 20 year bonds and one-unit short position in a 10 year bond. The hedge portfolio proportions are summarized over the out of sample time period February 1996 through December 1999.

Statistic	x_2	x_3
Maximum	1.37	1.13
Minimum	-0.33	-0.65
Median	0.78	-0.04

Table D.13: Hedge Portfolio Proportions (Summary Statistics) - PCA2

The summary statistics of the proportions to be invested in the portfolio of zero coupon bonds of 2,3, and 7 years to maturity according to the PCA2 model are presented. The portfolio is used to hedge portfolio P-I. Portfolio P-I consists of one-unit long positions in 1 year and 20 year bonds and one-unit short position in a 10 year bond. The hedge portfolio proportions are summarized over the out of sample time period February 1996 through December 1999.

Statistic	x_2	x_3	x_7
Maximum	9.54	-4.77	4.91
Minimum	4.37	-13.15	0.91
Median	5.61	-6.67	1.79

Table D.14: Hedge Portfolio Proportions (Summary Statistics) - PCA3

The summary statistics of the proportions to be invested in the portfolio of zero coupon bonds of 2,3,5, and 7 years to maturity according to the PCA3 model are presented. The portfolio is used to hedge portfolio P-I. Portfolio P-I consists of one-unit long positions in 1 year and 20 year bonds and one-unit short position in a 10 year bond. The hedge portfolio proportions are summarized over the out of sample time period February 1996 through December 1999.

Statistic	x_2	x_3	x_5	x_7
Maximum	4.11	-0.01	0.58	0.89
Minimum	1.85	-3.84	-2.21	-0.49
Median	2.40	-1.43	-0.62	0.21

APPENDIX E

Pricing Errors - RMSE (Sub-Periods)

Table E.1: Pricing Errors - Root Mean Squared Error(Jan. 1996 - Dec. 1996)

The Root Mean Squared Errors produced by the different models considered for bonds of different maturities during the sub sample from January 1996 through December 1996 are presented.

Maturity	CIR	LS	HJM-GV	HJM-RS	MLP	RBF
.25	0.0228	0.0012	0.0200	0.0194	0.0177	0.0239
1	0.1594	0.3912	0.0865	0.0641	0.0727	0.1011
2	0.6249	1.2525	0.3333	0.1574	0.1091	0.1914
3	1.4413	2.1738	0.5202	0.1524	0.0952	0.2285
7	7.7547	5.2212	1.6835	0.9176	0.9678	0.9463
10	10.9113	6.1808	1.5117	0.5556	0.5984	0.4449
15	11.4582	7.7719	2.7175	1.9456	2.0315	1.8245
20	13.3716	7.4848	1.5499	1.0033	1.0509	0.9825

Table E.2: Pricing Errors - Root Mean Squared Error(Jan. 1997 - Dec. 1997)

The Root Mean Squared Errors produced by the different models considered for bonds of different maturities during the sub sample from January 1997 through December 1997 are presented.

Maturity	CIR	LS	HJM-GV	HJM-RS	MLP	RBF
.25	0.0235	0.0013	0.0259	0.0241	0.0216	0.0253
1	0.1246	0.3937	0.1264	0.1082	0.0727	0.0887
2	0.4795	1.2241	0.2393	0.0916	0.0614	0.1414
3	1.3102	1.9938	0.3910	0.1116	0.0866	0.1439
7	7.7734	4.3387	1.0436	0.5900	0.5571	0.4819
10	11.6666	4.7791	1.0371	0.2521	0.2762	0.3070
15	10.8937	5.9604	2.0026	1.6985	1.6595	1.5701
20	8.9157	5.4656	1.1425	0.7475	0.7353	0.7983

Table E.3: Pricing Errors - Root Mean Squared Error(Jan. 1998 - Dec. 1998)

The Root Mean Squared Errors produced by the different models considered for bonds of different maturities during the sub sample from January 1998 through December 1998 are presented.

Maturity	CIR	LS	HJM-GV	HJM-RS	MLP	RBF
.25	0.0174	0.0012	0.0231	0.0238	0.0235	0.0305
1	0.1369	0.2197	0.0816	0.0838	0.0792	0.1090
2	0.8704	0.4889	0.1854	0.1372	0.1359	0.1381
3	1.8343	0.7489	0.2849	0.1538	0.1371	0.1489
7	8.3423	1.5130	0.7512	0.4567	0.40933	0.9631
10	12.6398	1.5302	0.9704	0.8533	0.7598	0.7171
15	13.9458	3.8008	3.2506	2.9157	2.9477	4.2911
20	13.9849	5.2395	2.4836	2.0711	2.0695	0.8814

Table E.4: Pricing Errors - Root Mean Squared Error(Jan. 1999 - Dec. 1999)

The Root Mean Squared Errors produced by the different models considered for bonds of different maturities during the sub sample from January 1999 through December 1999 are presented.

Maturity	CIR	LS	HJM-GV	HJM-RS	MLP	RBF
.25	0.0175	0.0010	0.4533	0.0301	0.0177	0.0229
1	0.1466	0.3799	0.9744	0.1363	0.0727	0.1172
2	0.2173	1.1255	1.5019	0.1159	0.1091	0.1601
3	0.4080	1.8326	1.9312	0.1217	0.0952	0.1894
7	3.9697	4.4134	3.3869	0.5979	0.9678	0.6529
10	7.8174	4.1825	3.5282	0.8379	0.5984	0.6901
15	6.7210	8.0826	6.5103	4.2182	2.0315	4.4016
20	7.0355	7.7684	3.9776	0.7649	1.0509	0.9125

APPENDIX F

Pricing Errors - MAD (Sub-Periods)

Table F.1: Pricing Errors - Mean Absolute Deviation(Jan. 1996 - Dec. 1996)

The Mean Absolute Deviation Errors produced by the different models considered for bonds of different maturities during the sub sample from January 1996 through December 1996 are presented.

Maturity	CIR	LS	HJM-GV	HJM-RS	MLP	RBF
.25	0.0199	0.0012	0.0170	0.0164	0.0146	0.0178
1	0.1402	0.3618	0.0671	0.0514	0.0618	0.0792
2	0.5388	1.1669	0.2813	0.1195	0.0879	0.1359
3	1.0138	2.0076	0.4586	0.1103	0.0708	0.1509
7	5.4790	5.0367	1.4518	0.8890	0.9310	0.9031
10	8.2275	6.0098	1.3209	0.4168	0.4549	0.3517
15	9.2890	7.6113	2.4213	1.8459	1.9351	1.7767
20	10.9267	7.0809	1.2275	0.7724	0.8136	0.6976

Table F.2: Pricing Errors - Mean Absolute Deviation(Jan. 1997 - Dec. 1997)

The Mean Absolute Deviation Errors produced by the different models considered for bonds of different maturities during the sub sample from January 1997 through December 1997 are presented.

Maturity	CIR	LS	HJM-GV	HJM-RS	MLP	RBF
.25	0.0213	0.0013	0.0210	0.0184	0.0166	0.0180
1	0.0949	0.3567	0.1014	0.0859	0.0596	0.0756
2	0.4327	1.1520	0.1974	0.0798	0.0533	0.1122
3	1.0626	1.8721	0.3203	0.0866	0.0626	0.1090
7	5.6531	4.0876	0.8855	0.5665	0.5342	0.4553
10	8.5978	4.4594	0.8811	0.1914	0.2110	0.2215
15	8.5960	5.7088	1.8043	1.6662	1.6244	1.5156
20	8.0944	5.2318	1.0155	0.6613	0.6552	0.7256

Table F.3: Pricing Errors - Mean Absolute Deviation(Jan. 1998 - Dec. 1998)

The Mean Absolute Deviation Errors produced by the different models considered for bonds of different maturities during the sub sample from January 1998 through December 1998 are presented.

Maturity	CIR	LS	HJM-GV	HJM-RS	MLP	RBF
.25	0.0152	0.0011	0.0194	0.0190	0.0187	0.0200
1	0.1077	0.1869	0.0652	0.0681	0.0618	0.0633
2	0.6293	0.3970	0.1479	0.1157	0.0969	0.1313
3	1.4098	0.5692	0.2346	0.1214	0.1140	0.1732
7	6.4703	1.3177	0.6663	0.4153	0.3936	0.4997
10	10.0776	1.3107	0.6138	0.7730	0.6490	0.8663
15	11.3304	3.1702	2.9133	2.6361	2.6892	2.6772
20	10.9587	4.4123	2.0972	1.7912	1.8143	2.0031

Table F.4: Pricing Errors - Mean Absolute Deviation(Jan. 1999 - Dec. 1999)

The Mean Absolute Deviation Errors produced by the different models considered for bonds of different maturities during the sub sample from January 1999 through December 1999 are presented.

Maturity	CIR	LS	HJM-GV	HJM-RS	MLP	RBF
.25	0.0167	0.0010	0.0240	0.0256	0.0208	0.0255
1	0.1168	0.3420	0.1110	0.1162	0.1041	0.0816
2	0.1888	1.0664	0.2058	0.0878	0.1107	0.1033
3	0.2874	1.7513	0.3479	0.0899	0.1203	0.1268
7	3.5125	4.2236	0.8855	0.5754	0.6241	0.7607
10	7.1119	3.9863	1.1464	0.7287	0.5558	0.6246
15	6.1019	7.9982	4.0787	4.1656	4.3425	4.2247
20	5.2656	7.6917	0.9694	0.6175	0.7635	0.7461

APPENDIX G

Measures of Duration considered by Gultekin and Rogalski (1984)

The different measures of duration considered by Gultekin and Rogalski (1984) (hereafter GR) are as follows

$$
\begin{aligned}
D1 &= \left[\frac{1}{P(m)}\right] \int_0^m tC(t)\exp[-R(t)t]dt \\
D2 &= \left[\frac{1}{P(m)}\right] \int_0^m tC(t)R(t)\exp[-R(t)t]dt \\
D3 &= \left[\frac{1}{P(m)}\right] \int_0^m t^2 C(t)R(t)\exp[-R(t)t]dt \\
D4 &= \left[\frac{1}{P(m)}\right] \int_0^m t\ln(t)C(t)R(t)\exp[-R(t)t]dt \\
D5 &= \left[\frac{1}{P(m)}\right] \int_0^m t^2 C(t)\exp[-R(t)t]dt \\
D6 &= \left(\exp\left[\frac{1}{P(m)}\right] \int_0^m C(t)\ln(1+\alpha t)\exp[-R(t)t]dt\right)
\end{aligned}
$$

Where $P(m)$ is the market value of a bond with maturity m at some instant of time, $C(t)$ is the cash flow received at t, $R(t)$ is the spot rate associated with each cash flow, and α in $D6$ is a measure of the variability of long-term yields relative to short-term yields.

$D7$ was proposed by Cox, Ingersoll, and Ross (1979) (hereafter CIR) by assuming that the instantaneous nominal spot rate *dr* follows a first-order auto-regressive process

$$dr = \beta(\mu - r)dt + \sigma\sqrt{r}dt.$$

Using the spot rate process CIR express the duration $D7$ as

$$D7 = G^{-1}\left[\frac{\Sigma C(t)P(t)G(t)}{C(t)P(t)}\right]$$

where

$$
\begin{aligned}
G^{-1}[x] &= \frac{2}{\gamma}\coth^{-1}\left(\frac{2}{\gamma x} + \frac{\pi - \beta}{\gamma}\right) \\
P(t) &= F(t)\exp[-rG(t)]
\end{aligned}
$$

$$
\begin{aligned}
F(t) &= \left[\frac{2\gamma \exp[(\gamma+\beta-\pi)t/2]}{(\gamma+\beta-\pi)[\exp(\gamma t)-1]+2\gamma}\right]^{(2\beta\mu)/(\sigma^2)} \\
G(t) &= \frac{2}{[\beta-\pi+\gamma\coth(\gamma t/2)]} \\
\gamma &= [(\beta-\pi)^2+2\sigma^2]^{1/2}
\end{aligned}
$$

and π is the liquidity premium.

VITA

Aryasomayajula V. Sekhar

Candidate for the Degree of

Doctor of Philosophy

Thesis: PARAMETRIC VERSUS NON-PARAMETRIC BOND PRICING AND HEDGING MODELS

Major Field: Business Administration

Biographical:

Personal Data: Born in Vizag, India, on October 11, 1972, the son of A. N. Murty and A. Usha Rani.

Education: Received the B.E. degree from Delhi University, Delhi, India in 1994, in Mechanical Engineering; Received the Post graduate diploma in Business Administration from S.P.Jain Institute of Management and Research, Bombay, India in 1997, in Finance; Completed the requirements for the Doctor of Philosophy degree with a major in Business Administration and a specialization in Finance at Oklahoma State University in December, 2001.

Experience: Graduate Teaching Associate at Oklahoma State University from Fall 1998 to present; Management Trainee at HDFC Bank, Bombay, India from July 1997 to June 1998; Territory Manager with Hindustan Motors Limited, Delhi, India from July 1994 to June 1995.