#

A COMPARISON OF FIXED INCOME VALUATION MODELS: PRICING AND ECONOMETRIC ANALYSIS OF INTEREST RATE DERIVATIVES

by

Michael Jacobs, Jr.

A dissertation submitted to the Graduate Faculty in Business in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2001

UMI Number: 3024803

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Business in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

7/24/2001 — Kishore Tandon

Date — Chair of Examining Committee

July 25, 2001

Date — Executive Officer

Doug Howard

Joel Rentzler

Gloria Thomas

Joseph Onochie

Supervisory Committee

The City University of New York

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

A COMPARISON OF FIXED INCOME VALUATION MODELS: PRICING AND ECONOMETRIC ANALYSIS OF INTEREST RATE DERIVATIVES

by

Michael Jacobs, Jr.

Advisor: Professor Kishore Tandon

This study compares several continuous-time stochastic interest rate and stochastic volatility models of interest rate derivatives, examining these models across several dimensions: different classes of models, factor structures, and pricing algorithms. We consider a broader universe of pricing models, using improved econometric and numerical methodologies. We establish several criteria for model quality that are motivated by financial theory as well as practice: realism of the assumed stochastic process for the term structure, consistency with no-arbitrage or financial market equilibrium, consistency with financial practice, parsimony, as well as computational efficiency. This helps resolve the controversies over the stochastic process for yield curve dynamics, the models that best manage and measure interest rate risk, and theories of the term structure that are supported by empirical evidence.

We perform econometric experiments at three levels: the short interest rate, bond prices, as well as interest rate derivatives. We extend CKLS (1992) to a broader class of

single factor spot rate models and international interest rates. we find that a single-factor general parametric model (1FGPM) of the term structure, with non-linearity in the drift function, better captures the time series dynamics of US 30 Day T-Bill rates. Our results vary greatly across international markets. Building upon the work of Longstaff and Schwartz (1992), we perform a statistical analysis of the U.S. default-free term structure and identify at least three factors that capture 98% of the variation (level, slope, and curvature). We compare various term structure models on US Treasury bonds, ranging from the two-factor Cox-Ingersoll-Ross (2FCIR) to a multi-layer perceptron neural network model (MLP-ANN). Finally, we compare various interest rate bond option pricing models, in their ability to price interest rate derivatives and manage and interest rate risk. We compare the spot rate, forward-rate, and non-parametric models (e.g., multivariate kernel estimation) and extend it to a broader factor structure. We find that no one model dominates the others under various criteria.

In memory of my mother Rachel, who taught me the value of knowledge and passed away before her time.

ACKNOWLEDGMENTS

There are many people, some that I have known personally and others that I have only known through books and articles, who have contributed to this dissertation. This work would not have been possible without the wisdom, patience, and guidance of my advisor, Kishore Tandon, through my time in the doctoral program and during the completion of this dissertation. I also owe a great debt to my committee members Joel Rentzler, Joseph Onochie, Larry Tatum, and Gloria Thomas - all of whom have been very helpful to me during this time in various ways. A special thanks is due to Shulamith Gross, whose course in stochastic differential equations greatly expanded my research opportunities. Other faculty at CUNY and SUNY Stony Brook who deserve special mention for expanding my horizons and teaching me about academics include Michael Zweig, Thomas Muench, Stefan Mittnick, Ronald Anderson, David Feldman, and Salih Neftci. Finally I owe a great debt to Michel Araten, my colleague and mentor at J.P. Morgan Chase & Co., who has given me the opportunity to apply my craft in the practical realm. Any errors or deficiencies in this work are solely my responsibility and do not reflect on the contribution that the above people have made in my intellectual journey.

CONTENTS

LIST OF TABLES

LIST OF DIAGRAMS

CHAPTER 1

THE PRICING OF INTEREST RATE DERIVATIVES

1. INTRODUCTION AND DISCUSSION

The purpose of this research is to link two streams of research in financial economics, within the context of theory and management of interest rate risk. The first focuses on the term structure of interest rates, the underlying theories, as well as their econometric and market efficiency implications. We term this as the *forecasting emphasis* of the term structure literature. The second stream of literature is termed the *pricing emphasis*, in that its methodology has centered upon the design and evaluation of models that are capable of accurately explaining the market prices of interest rate dependent derivatives. Such a modeling strategy is central to any system constructed to supervise interest rate risk, whether it is Value-at-Risk, sensitivity analysis, stress testing, or scenario analysis. Unfortunately, few studies in the current literature evaluate the empirical performance of the competing valuation models on similar or international data sets. Further, few compare fundamentally different modeling approaches. These streams of literature are by no means mutually exclusive, the primary link being the characterization of the stochastic processes governing the evolution of the yield curve.

We extend this taxonomy by dividing the pricing literature into the *general equilibrium theory approach* on one hand, and the *arbitrage pricing theory approach* on the other. The former starts with the stochastic evolution of the underlying economic determinants and agent's utility maximization, thereby deriving endogenously the processes that drive the existence of both the yield curve and the derivatives dependent upon them. This class is often termed *spot rate models* in the literature (e.g., see Cox et al (1985)-henceforth CIR, Hull and White (1990, 1993), Jamshidian (1989, 1991)). In the implementation of these models, the authors follow the suggestion of Cox et al (1985) to fit the endogenous term and volatility structure to that observed through time dependency in the parameters of the stochastic factor processes. In contrast, *arbitrage-free valuation* takes as inputs either the spot or forward term structure of interest rates, and derives pricing relationships that are consistent with the absence of arbitrage, under general behavioral and technological assumptions. Due to the equivalence of the spot and forward term structures in the absence of arbitrage, this class has been termed *forward-rate models* (see Ho and Lee (1986); Heath et al (1992), henceforth HJM). These models share the stochastic structure of the term structure of interest rates and their volatilities, and therefore are not only similar to each other, but in some cases mathematically equivalent. We then compare these two approaches to the non-parametric class of models, with no distributional assumptions on the underlying stochastic processes. Examples of these are artificial neural networks, kernel estimation, as well as methods that construct a risk neutral density.

Finally, a general distinction applying to both approaches is between those studies that are concerned with *theoretical discourse*, either relating to the design of pricing models or the development of theories that are potentially subject to empirical failure, or *empirical discourse*, which is concerned with evaluating such models and econometrically testing the theories. Since both pertain potentially to either pricing or forecasting, we must refer to both in trying to link these streams of literature. Our first investigation involves the consistency of models with financial market equilibrium or the weaker requirement of a no-arbitrage economy. The assumption of no arbitrage is sometimes enough to have confidence in implementing a model, since the assumptions underlying an equilibrium are not likely to hold exactly in practice. Second, in an effort to understand how agents formulate optimal forecasts and to refine theoretical precepts using financial data, one must have a null hypothesis to begin with and as a best practice make these propositions amenable to some kind of empirical verification process.

We test a range of single and multi-factor spot and forward-rate models, and compare them to a non-parametric model. This is to determine which model, as well as which factors, best explain interest rate derivatives and best manage interest rate risk. While in numerous cases these types of models may be related, if not mathematically equivalent, they tend to diverge widely in their practical implementation. A priori, the theory often sheds little light on the number of factors that best describe the economic reality of the term structure. Few studies have attempted to address this level of generality. One notable exception is Buhler et al (1999), in which competing models are compared utilizing German warrants over the 1990-1993 period. We extend this research to international short-term rates and CBOT Treasury bond future options, testing a wide range of term structure models, including non-parametric models.

Different stochastic structures are proposed in the two model classes. Among the spot rate models, we distinguish between the single and multiple factor variants. The *single factor models* are driven by a generalization of the Chan et al (1992, henceforth CKLS) stochastic differential equation, which features a mean-reverting drift and a constant elasticity of variance diffusion function. In our framework, we allow for non-linearity in the drift as well as a displaced diffusion function, following the general parameterization of Duffie (1996), which nests several more term structure models than CKLS. These include the popular Black et al (1990, henceforth BDT) and Pearson et al (1994) models. The *multiple factor models* include those with both two and three underlying factors. The most general model is a 3-factor affine term structure formulation, where the underlying factors are associated with the level, the steepness and the curvature of the yield curve. In this class, we consider a correlated Gaussian model with multivariate log-normally distributed bond prices, a multi-factor CIR model with factors following independent square-root processes, as well as a three factor jump diffusion model. The other multiple factor models are restricted to having two factors. In the first of these, we generalize the Schaefer and Schwartz (1984) model, where we identify the two factors as the short rate and the difference between the long rate and the short-rate. The second model postulates two unobservable factors that are linearly related to the short rate, resulting in a model in which the term structure depends upon the short rate and its volatility. This is an extension of the Longstaff and Schwartz (1992) model, which is a two-factor variant of the CIR model. The commonality among these models is that they price interest rate dependent derivatives in either a stochastic interest rate or a stochastic volatility setting.

Among the forward rate models, we consider three-factor, two-factor, and one-factor variants of the HJM model, with constant as well as linear proportional volatility functions in each case. The single factor constant volatility model is a continuous time version of Ho and Lee (1986), in which forward rates are Guassian, while the proportional volatility variant involves diffusion functions that depend upon time-to-maturity only. In the multiple factor variants, we employ principal component analysis to empirically determine the unobservable factors, and then their volatilities.

This chapter is organized as follows. In section 2, we review the existing literature on interest rate derivative valuation models. We review the theory behind interest rate option models, stochastic processes and the empirical evidence regarding these models. In section 3, we outline the contribution of this study. This is classified in three areas: the non-parametric techniques, a broadening of the factor structure, and an empirical analysis of the new valuation techniques.

2. A REVIEW OF THE TERM STRUCTURE LITERATURE

In this section, we review the existing literature on term structure. On the theoretical side, we consider models of the term structure, as well as the pricing of bond options under interest rate uncertainty. The empirical part consists of estimation of continuous time stochastic processes, the testing of fixed income derivative pricing models, as well as numerical and non-parametric techniques for empirical testing and pricing. First, we consider general equilibrium models that rely on the spot rate as the input to producing an endogenous term structure that is used to price contingent claims, as exemplified by CIR (1985). Next, we review the no-arbitrage approach, or the forward rate models as illustrated by HJM (1992).

2.1 Theoretical Models

The traditional literature on yield curve movements and pricing securities takes the stochastic process for the short rate as given. It matches the parameters to market data as closely as possible and in the process deduces yield curve movements and derivative prices. The alternative approach of no-arbitrage valuation starts with identifying the process followed by instantaneous forward rates (or equivalently, the spot term structure), and derives results that hold for all arbitrage-free yield curve models.

2.1.1 The Equilibrium Approach and Spot Rate Models

This approach is seen in the work of Merton (1973), Black and Scholes (1973), Vasicek (1977), Dothan (1978), Courtadon (1982), Ball and Torous (1983), and the seminal paper by CIR (1985). Bond and option prices are derived in representative agent equilibrium, having an advantage of endogenous determination of asset prices that yields natural restrictions on the dynamics of the term structure with testable implications.

Merton (1973) presents a stochastic term structure in the valuation of option prices and derives a version of the Black-Scholes model in a setting where the discount bond and stock price process follows a bivariate geometric Brownian motion. Black (1976) derives a formula for options on default-free discount bond futures, assuming a constant short interest rate and constant volatility. Despite the fact that this is not a stochastic term structure model, it is an important benchmark, and is widely used in interest rate option modeling. Vasicek (1977) uses the stochastic interest rate approach in the context of fixed income contingent claims valuation. He derives a general form of the term structure of interest rates in a continuous time setting, under the assumption that the instantaneous (spot) rate of interest follows an Ohrnstein-Uhlenbeck mean reverting type of diffusion. When this is coupled with market efficiency, the price of a discount bond depends only upon the spot rates prevailing during its term. Under these assumptions, a no-arbitrage argument implies that the instantaneous term risk premia on bonds is proportional to the instantaneous standard deviation of return, which is reminiscent of the traditional capital asset pricing results. CIR (1985) develop an intertemporal continuous time equilibrium model to study the term structure of interest rates in a complete market setting. Here, all asset prices and stochastic processes are determined endogenously, the principal result being a partial differential equation that must be satisfied by all claims, giving rise to an equation for the equilibrium price of any asset in terms of the underlying real variables in the economy. Several factors traditionally mentioned as influencing the term structure are included in a way, which is consistent with utility maximizing behavior and rational expectations. The model leads to specific formulae for bond prices, which are well suited for empirical testing. Longstaff and Schwartz (1992; henceforth LS) extend the CIR model to two factors and develop a general equilibrium model of the term structure. The factors considered are the short-term riskless rate of interest and its volatility. They develop closed-form expressions for discount bonds and analyze the properties of the implied term structure, using Hansen's GMM framework to test the cross-sectional restrictions imposed by the model.

2.1.2 The No-Arbitrage Approach and Forward Rate Models

As an alternative to equilibrium models, Ho and Lee (1986) present a no-arbitrage valuation model in a discrete time binomial framework that allows for an exact fit to the current spot term structure. Black et al (1990) and Hull et al (1990) extend this with models capable of matching the current volatility structure. HJM (1990) extend the Ho and Lee model with a process followed by instantaneous forward rates and derive results that hold for all arbitrage-free yield curve models. They derive a no-arbitrage theoretical model of contingent claims valuation, in which a stochastic term structure for interest rates is outlined using the methodology of the equivalent martingale measure technique. This is a generalization of Ho & Lee (1986) model to a continuous time and continuous state space setting. Inputs to the model include the initial term structure of forward rates and a class of stochastic processes for its evolution. To be consistent with financial market equilibrium, no arbitrage conditions restrict the family of processes.

2.2 Empirical and Econometric Studies

Next, we review the empirical literature relating to the term structure and interest rate derivatives. The first set of results includes econometric tests of the term structure. These include attempts to estimate the parameters of the stochastic process for the short rate, as in Chan et al (1992) and the non-parametric model by Ait-Sahalia (1996). The other part includes direct tests of interest rate option pricing models by various numerical and non-parametric techniques.

2.2.1 Empirical Studies of the Term Structure

Several papers have attempted to test the empirical validity of popular models of the term structure. Chan et al (1992) compare various models of the short-term riskless rate using the Generalized Method of Moments (Hansen, 1982; GMM). They find that the most successful models are those that allow the volatility of interest rate changes to be highly sensitive to the levels of interest rates. Several models perform poorly in this comparison due to the implicit restrictions on the term structure volatility. This has important implications for the use of different term structure models in the valuation of interest rate contingent claims, as well as in the hedging of interest rate risk. Gibbons and Ramaswamy (1993) test a theory of the term structure of indexed bond prices based on CIR (1985). They utilize GMM to exploit the conditional probability distribution of the single state variable in CIR's model, thereby avoiding the use of aggregate consumption data, since it is prone to severe measurement error. They estimate a continuous-time model based on discretely sampled data, thereby avoiding temporal aggregation bias associated with discretization procedures. They find that the CIR model performs reasonably well when examining short-term U.S. Treasury bill returns and provides evidence of positive term premia and varied possible shapes for the yield curve. However, the fitted model is deficient in explaining the serial correlation structure in real Treasury-bill returns. Nowman (1997) presents a Gaussian estimation of continuous time dynamic models.[1] This accounts for exact discrete model to estimate the parameters of open continuous time systems from discrete stock and flow data in the manner of Bergstrom (1983)[2]. It also accounts for exact restrictions on the distribution of the

---

1

This is not to be confused with estimation that assumes a Guassian or normally distributed error term. Stronger technical conditions are imposed on the error terms, so that a Guassian likelihood function can be optimized to give (asymptotically) efficient estimates of the model parameters, without assuming normality. This is a "semi-parametric" technique, sometimes called Quasi-Maximum Likelihood Estimation (or QMLE), as opposed to ordinary parametric MLE.

2

Bergstrom (1983) studies the properties of efficient estimators of the structural parameters in closed lin- ear systems of higher order stochastic differential equations (SDEs), when the data are in discrete form, and the model includes variables of both the stock and flow variety.

discrete data, and does not rely on discretization procedures that depend on shortening the sampling interval to achieve convergence, in order to reduce temporal aggregation bias. He estimates several one-factor continuous time models of the short-term interest rate using a discrete time model and compares them to an approximation used by CKLS (1992). The volatility of the short rate is found to be sensitive to the level of interest rates in U.S.

A recent study in the empirical literature has been the non-parametric estimation of the structural parameters of underlying diffusion process. Pearson et al (1994) propose an empirical method that utilizes the conditional density of the state variables to estimate and test a term structure model, using data on both discount and coupon bonds. The method is applied to an extension of a two-factor model based on CIR (1985). Their results show that estimates based on only bills imply unreasonably large pricing errors for longer maturities and the original CIR model is rejected using a likelihood ratio test. They also find that the extended CIR model fails to provide an adequate description of the Treasury market. Ait-Sahalia (1996 a) employs a non-parametric estimation procedure for continuous-time stochastic models. In that prices of derivative securities depend crucially on the form of the instantaneous volatility of the underlying process, the volatility function is left unrestricted and is estimated non-parametrically. Although only discrete data are used, the estimation procedure does not rely on replacing the continuous time model by a discrete approximation. Instead, the drift and volatility functions are forced to match the densities of the process. He computes the SDE followed by the short-term interest rate, as well as non-parametric prices for bonds and bond options. In a related paper, Ait-Sahalia (1996 b) examines different continuous time models of the interest rate, testing parametric models by comparing their implied parametric densities to the densities computed non-parametrically. Even though the data are recorded at discrete intervals, the continuous time model is not replaced with a discrete approximation. It is found that the principal source of rejection with respect to existing models is the strong non-linearity of the drift. When it is close to its mean, the drift is virtually zero, and the interest rate behaves like a random walk. However, when far from its mean, the interest rate exhibits strong mean reversion. The volatility is found to be higher when the rate deviates from its long-run mean.

Stanton (1997) uses an alternative non-parametric technique for estimating continuous-time diffusion processes, which are observed at discrete intervals. He applies the methodology to three and six month Treasury Bill data from 1/65 to 7/95, for the estimation of the drift and diffusion of the short rate, as well as the price of interest rate risk. The estimated diffusion is similar to CKLS (1992), and there is strong evidence of non-linearity in the drift. It is close to zero for low to medium interest rates, with increasing mean reversion for higher interest rates. Jiang (1998) develops another non-parametric model of the term structure, which allows for maximal flexibility in fitting to the data. This is based only upon a spot rate process that admits only non-negative interest rates and a market price of risk that precludes arbitrage opportunities. The marginal density of the short rate, as well as the historical path of the term structure, are utilized to allow for robust estimation of the term structure. The model is estimated using U.S. government bond data, to provide comparability with existing literature. His results suggest that most traditional spot rate models are mis-specified and that the non-parametric model generates significantly different term structures and market prices of interest rate risk. Stutzer et al (1999) applies the *canonical valuation model*, a risk-neutral method that allows the specification of an individual assessment of the distribution of the underlying security at expiration, to CBOT bond futures for 21 randomly selected days from 10:96 to 01:97. Their model is found to outperform Black's (1976) model in absolute, but not percentage, terms.

2.2.2 Empirical Tests and Implementation of Interest Rate Option Models

Empirical tests of interest rate option valuation models are limited in comparison to the broader empirical and theoretical literature on interest rate models. The two factor model of Brennan and Schwartz (1982) is tested by Dietrich-Campbell et al (1986), using interest rate options on U.S. government bonds and treasury bills; the data fails to reject the cross-sectional restrictions of the model. Brown and Dybvig (1986) test the single factor version of CIR (1985), using a linear approximate estimation and find weak support for the model. Cakici (1989) tests single factor valuation models of European and American bond futures options. He estimates the models of Black (1976) and Barone-Adesi & Whaley (1987) assuming different stochastic processes for the short rate. Gibbons and Ramaswamy (1993) use GMM to test CIR (1985), finding that the cross-sectional restrictions of the model not rejected by the data. Flesaker (1994) tests a constant volatility version of the HJM model using GMM, for daily Eurodollar futures options on the CME, for the period 3/85-3/88. While the small sample properties of the GMM estimator are found to be significant in simulation analysis, the model performs poorly using actual data. Jordan et al (1995) analyze option pricing values implicit in callable Treasury bonds, in the framework of Longstaff (1992), who had reported the puzzle of negative values in these embedded derivatives. Employing an alternative empirical approach, the authors find implied option values to be generally positive and in line with theoretical predictions that do not allow arbitrage opportunities. Previous methodologies did not consider microstructural biases into consideration and this led to spurious negatively measured values.

Longstaff et al (1993) extend the two-factor model of Longstaff and Schwartz (1992). A simplified procedure, involving historical statistics on the short rate and GARCH volatility, is developed to estimate the stationary parameters of the model. Forecasts of future interest rates and volatilities are evaluated, using Monte Carlo simulation, to test the accuracy of the parameter estimates. The initial term structure parameter is fitted exactly by allowing a time varying term structure parameter, which is identified as the market price of risk. This modification of the model is applied to the pricing of interest rate caps and heir implied volatilities. Moraleda and Pelsser (1997) use market prices of daily caps and floors for 1993 and 1994 to determine whether spot ate (i.e., equilibrium style) or forward rate (i.e., no-arbitrage style) models of the term structure provide a better fit to market prices of options. The spot rate models of Hull and White (1994), Pelsser (1994), and Black and Karasinski (1991) are compared to the Gaussian, square root and proportional forward rate model developed by Ritchken and Sankarasubramanian (1995). They find that all spot rate models outperform their forward rate counterparts. A number of humped volatility models, obtained as extensions to the above models, are found to better fit the observed option prices in the spot interest rate setting, which is at variance with the findings of Bliss and Ritchken (1996), but consistent with the earlier literature. Buhler et al (1998) examine which interest rate valuation model best manages interest rate risk. They test the market for German warrants, using seven spot and forward-rate models with both one and two factors. They find that one forward-rate model (a single factor proportional volatility HJM model) and two spot rate models (a single factor extended CKLS and a two factor CIR model) outperform four other models.

Studies using numerical techniques go beyond econometric analysis to estimate the above theoretical models. In Ho et al (1986), the Geske-Johnson (1984) analytical approach is generalized to a stochastic interest rate economy. The method is implemented for options that can be exercised on one of a finite number of dates. It is shown that the value of an American option is an increasing function of interest-rate volatility. The magnitude of this effect depends on the extent to which the option is in the money, the volatilities of the underlying asset and the interest rates, as well as their degree of correlation. Barone-Adesi et al (1987) derive simple analytic approximations for pricing exchange-traded American call and put options written on commodities and commodity futures contracts. These approximations are accurate and computationally more efficient than finite-difference, binomial, or compound-option pricing methods. A paper that has important implications for numerical techniques is Carr et al (1992), which finds alternative representations of the McKean equation for the valuation of an American put. They estimate the value of an American put option as a combination of the corresponding European put price and the early exercise premium. Hull and White (1993) compare different approaches to developing arbitrage-free models of the term structure. They propose a numerical procedure capable of constructing a wide range of one-factor models of the short rate that have both the Markov property and are consistent with the initial term structure. They link the three main ways of constructing models in the no-arbitrage class: starting with discount bond prices, instantaneous forward rates and modeling the short rate. Their model involves the use of trinomial trees, which are found to be robust, efficient, and capable of implementing models, such as extensions of Vasicek (1977) and CIR (1985). The lognormal model (utilizing the explicit finite difference algorithm) of Black and Karasinski (1991) is also examined. The one-factor model of the short rate is fitted to the initial yield curve, and then extended to be consistent with initial volatilities.

3. THIS STUDY

This section outlines the details of our study. In testing interest rate valuation models, we first specify the basic characteristics of the models under consideration, which we term the process of *preselection*. We outline assumptions regarding the nature of the stochastic processes driving the term structure and the derivatives based on it. This is equivalent to the specification of the functional form for the underlying factors. We extend the analysis to non-parametric models, which rely on rather unrestrictive assumptions about the stochastic evolution of the term structure. Next, we specify the factors driving the term structure. We consider models that are driven by different number of factors, ranging from one to three. Furthermore, we consider the quality of the factors that must be considered: in particular, whether they are latent or observable, or whether they are statistically or theoretically motivated. To understand the term structure, we use a *principal component analysis* (Bliss (1997); PCA) of US Treasury instruments for the thirty-three year period, from 10:64 to 10:97. Our analysis includes new approaches. First, to check for stability over time, we extend our analysis for a more extensive period than done previously. Second, we employ a variant of the multivariate G.A.R.C.H. model to estimate a time varying variance-covariance matrix of yield changes, motivated by our finding that yield volatilities and their covariance's are time-varying.[3] Third, we use an improved computational approach to estimate the underlying

---

3

In particular, we estimate a *constant correlation multivariate generalized autoregressive-*

factors, which avoids questionable econometric assumptions (e.g., multivariate normality of residuals required in MLE estimation) and provides estimates that are more robust to specification error. Due to the consistently high degree of correlation among yields of various maturities, we find that three factors are sufficient to describe over 98% of the variation in the term structure. We compare the three-factor model to the existing two- and one-factor models. Our research also extends the universe of instruments considered. First, we analyze instruments of all three fundamental types: underlying short rates, zero-coupon bonds, and the derivatives on them.[4] Second, we extend the range of markets to include the spot interest rates in other countries (e.g., U.K. and Japan). In the context of bond pricing, we consider a longer historical period and a wider range of maturities than in the existing literature. Finally, with regard to derivatives, different interest rate options markets are examined.

---

moving average model (CCM-GARCH) model. This assumes that yield correlations do not change but yield volatilities do. This is motivated by the empirical finding that most of the variation in yield covariances can be attributed to changing yield volatilities (Rebonatto (1996)).

4

Future research would attempt to look at so-called *second generation* (e.g., swaptions and captions) and *third generation* (e.g., exotics such as barrier options) interest rate derivatives. Their building blocks are the Black (1976) model prices of bond options, rather than the underlying rates or bond prices themselves. However, valuing these instruments in even straightforward single factor parametric models is subject to both theoretical and practical difficulties. For example, since an equilibrium swap rate is a linear combination of forward rates, assuming that the forwards are log-normally distributed is inconsistent with assuming the same for swap rates, since linear combinations of log-normal random variables is not log-normal.

3.1 Non-parametric Alternatives

In this section, we focus on two types of non-parametric models. In contrast to the analytic and numerical approaches to valuing interest rate derivatives, which rely upon strong assumptions regarding the stochastic process governing the evolution of the term structure, these non-parametric models rely on weaker requirements. Such models typically impose only certain technical regularity conditions on the process of the state variable and the functional form of the true pricing formula.

3.1.1 Artificial Neural Networks

In this section, we consider a class of non-parametric models, *artificial neural networks*, which attempt to mimic the process of human cognition in developing pricing formulae, and have been shown to compete successfully with parametric counterparts in pricing a variety of derivatives. In general, derivative pricing functions are multivariate and non-linear. Hornik et al (1989) has shown that a broad class of neural networks, the *multilayer feed forward* variety (a variant of which we employ here), constitute so-called *universal approximators*. This means that under mild regularity conditions such networks are capable, to an arbitrary degree of accuracy, of representing any non-linear function.[5] White (1989, 1990) has shown that neural networks of this kind are amenable

---

5 A caveat in this regard is that the result is a theoretical one, holding precisely only under

to estimation by a non-linear regression, and can thus prove to be a very flexible tool in approximating non-linear functions.

In this approach, we make no assumptions about the functional form followed by the short rate, or any other appropriate term structure variable (e.g., the short rate volatility, which itself can be estimated non-parametrically by *kernel regression*). This approach has been used earlier to price equity derivatives by Malliaris et al (1993) and later extended by Hutchinson et al (1994). The first step in this procedure is the choice of input parameters and derivative prices that they influence. This set of factors *and* option prices observed both cross-sectionally and over time constitutes what is called a *pattern* in the artificial intelligence (AI) literature.[6] The data set is partitioned into three subsets: portions used for *training*, *cross-validation*, and *testing*. Parameter values are determined in the training set via non-linear optimization of a suitable criterion. The cross-validation portion is used to determine which specific network architecture performs best on the

---

the limiting case of an infinite number of network parameters. This leaves the question of determining the degree of precision attainable in computationally feasible applications to be answered by empirical experimentation.

6

There is a similarity in econometric methodology to a parametric approach that does not rely exclusively on historically estimated or cross-sectionally implied parameters. This combines both in order fit market prices at a point in time as well as predict option prices over time. An example of such modeling is *stochastic volatility analysis* (SVA), which uses extensions of GMM such as *simulated method of moments* (SMM-see Duffie et al (1993)) or *efficient method of moments* (EMM-see Gallant et al (2000)). These models can estimate unobservable factors, such as underlying asset volatility. We apply these estimation techniques to price discount bond options for several models, and compare the results to non-parametric pricing.

patterns in this set. Finally, the test set is used to assess the out-of-sample performance of the network selected.

The class of networks that we use is a version of the *multi-layer perceptron* artificial neural network (MLP-ANN), which is a network with a number of *hidden layers*.[7] While the number of inputs to the network is a fixed number, the number of hidden layers (called the *network architecture*) is chosen experimentally to achieve an optimal criterion (e.g., minimizing some distance measure between market and model prices). For each number of hidden units, we choose the free parameters of the network to minimize the root mean squared error (RMSE) of the pricing errors. Finally, we search over all network architectures for the optimal number of hidden units, to get the globally optimal vector of free-parameters.[8]

3.1.2 Multivariate Kernel Regression

Another non-parametric model is an extension of the *kernel regression* technique to several independent variables. Suppose that the price of a derivative depends on a vector of factors. In the context of bond future options, this is given by the short interest

---

7

These can be thought of as interaction terms in a multiple regression, in which the influence of the factors on each other are captured.

8

This is analogous to running different regressions with different numbers of independent variables until the best (adjusted) R-squared is found.

rate, the estimated volatility, the price of the underlying futures contract, the strike price, and the maturity of the option. We posit that the pricing relationship can be represented by any fixed and sufficiently smooth non-linear function. The idea behind this is to construct a "sophisticated" average of observed option prices around a fixed vector of characteristics. The averaging procedure puts more weight on observations that are paired with characteristics that are "closer", in the sense of Euclidean distance. The more local (global) the averaging, the more jagged (smooth) the estimated function. It is shown by Hardle (1990) that such a procedure, which utilizes a weighting function (or *kernel*) satisfying certain regularity conditions, results in an estimator that converges to the true pricing relationship asymptotically.[9]

The procedure can be summarized as follows. First, we define the multivariate weighting function, which is constructed from non-negative *smoothing kernels*, satisfying the conditions of integrating to unity. Second, we choose the *bandwidth parameter*, which controls the scale of the kernel function, much as a standard deviation parameter controls the dispersion of a probability distribution. Third, the *multivariate Nadaraya-Watson kernel estimator* (MNWKE) of the pricing function, which depends upon a vector of bandwidths, is chosen to minimize the weighted squared deviations from *cross-*

9

Basically, the weighting function should be positive and should integrate to one, like a probability density function. A normal density function satisfies these requirements. Other kernels could be used to give slightly more accurate results, but then one has to check other technical conditions, which in most cases the normal density function automatically satisfies.

validation function (CVF). The CVF for a particular observation is a MNWKE constructed by leaving that observation out.

3.2 Factor Structures

Another contribution to the literature involves the consideration of more factors than previously documented. First, we discuss the one- and two-factor models traditionally used by academicians and practitioners and then propose a 3-factor extension.

3.2.1 The Existing Single and Two-Factor Approaches

In the class of one factor models, the functional form of the stochastic process of the term structure is a feature that most discriminates between models. In the case of forward rate models, the parameterization of forward rate volatility is a key factor, while the specification of the process for the short rate is central to the implementation of spot rate models. In the latter case, the literature has focused primarily on one and two parameter variants, where it has been shown that the number of parameters is more important than the particular form of the model (Amin et al, 1994). The findings are that a two-factor forward rate model with linear proportional volatility best fits the data in and out of the sample. However, the one-parameter variant results in more stable estimates and consistently higher arbitrage profits from perceived mis-pricings over time.

Motivated by these findings, we follow Buhler et al (1999) in considering two single factor HJM models, a one-parameter constant volatility and a two-parameter linear proportional volatility variant.

In contrast to the forward rate models, the specification of the process for the short rate is critical in short-rate models. Motivated by empirical findings (Chan et al, 1992), we expect interest rate changes to be negative (positive) at relatively high (low) levels, which is termed mean reversion. There is also evidence of non-linearity in the drift of interest rate changes (Ait-Sahalia 1996), in the sense that the drift of the short rate has been found to increase gradually at low levels of the short rate and decrease sharply at very high levels. In order to account for this, we replace the linear drift function $\alpha + \beta r_t$[10], as used in Chan et al (1992), with the non-linear specification $\alpha + \beta r_t + v r_t \log(r_t)$. In the case of the diffusion function, several studies have verified that the magnitude of interest rates movements is directly related to the level of the short rate. In addition, motivated by theoretical considerations, we expect that the volatility of the short rate should not vanish at very low levels of the short rate. Hence, we nest the constant elasticity of variance (CEV) diffusion function of CKLS[11], σr_t^{γ}, with the displaced specification $(\delta + \sigma r_t)^{\gamma}$, where $\delta > 0$ is the parameter of displacement. An

---

10 In this context, we interpret $-\beta$ as the speed of adjustment parameter, and $-\alpha/\beta$ as the long run mean of the short rate. Therefore, we impose the restriction that $\alpha > 0$ and $\beta < 0$.

11 Technically, this is defined as $dt^{-1}\lim_{s \downarrow t} \mathrm{Var}_t[r_s]$

advantage of this generalized specification is that it allows us to nest continuous time versions of additional popular models, namely the continuous time versions of the Black and Karasinki (1991), Pearson and Sun (1994), and Black, Derman and Toy (1990).

The existing 2-factor models are of two types. One is the 2-factor CIR model of Longstaff and Schwartz (1992), which identifies the second pricing factor as random volatility of the short-rate. The second type is motivated by data analysis, which reveals that a factor related to the slope of the yield curve has explanatory power. Therefore, we introduce a 2-factor model, in which the level of the short rate and the spread between it and a long rate are the factors, similar to Brennan and Schwartz (1979, 1982) and Schaefer and Schwartz (1984). These models are special cases of the affine class of term structure models, as analyzed by Duffie and Kan (1996), in which the drift and volatility functions of the short rate are linear in the term structure factors.

3.2.2 Three-Factor Models and Principal Component Analysis

The study of more general models of the term structure are motivated by several other empirical findings. First, principal components and generalized multiple regression analysis reveals that most of the variation in the term structure can be attributed to a number of unobservable factors ranging between 2 and 3, depending upon the market and the time period. The first factor has similar impact (or "loading") across maturities, and as such can be interpreted as a "level factor". The second has a factor loading that varies directly with maturity, having opposite signs at the long and short ends of the term structure, and therefore having a natural interpretation as a "slope factor". Finally, the third factor has been described as having a loading profile that peaks at intermediate maturities, which results in twists in the term structure; this third factor can be considered a "curvature factor".[12]

In the context of forward rate models, the multifactor models are also of two varieties: constant and linear proportional volatility functions. In the cases of both two- and three-factor models, we empirically determine the functional forms of these, by calculating the volatility of factors as determined by *principal component analysis* (PCA). In the case of spot rate models, our choice is motivated by empirical studies of the spot rate, which reveal the non-normal features of volatility clustering and asymmetry in its distribution (Ait-Sahalia, 1996). Therefore, we examine a 3-factor affine term structure model (3F-ATSM). This is the most general family of models, nesting most of the widely studied models in the literature, which have closed-form solutions.[13]

12

The latter two factors, attributed to steepness and curvature of the yield curve, are related to the concepts of duration and convexity discussed in the traditional fixed income literature.

13

To be more specific, in ATSMs the short rate follows a *linear stochastic differential equation* , in which case a solution, should it exist, reduces to the solution of a system of ordinary differential equations (ODE's). It is only in this sense that we mean "closed form". In contrast, in non-linear SDE models, calibration to a discretized PDE or numerical integration of a discretized SDE is the only way to approximate a solution.

3.3 New Instruments Analyzed

Finally, in this study we examine a broader array of term structure instruments. The extension occurs along two dimensions. First, we take the comprehensive approach of building our empirical analysis from interest rates, then bonds and the term structure, and finally contingent claims on the latter. Second, we span multiple global markets over an extended period. This allows us to assess market efficiency regimes both spatially as well as temporally.

3.3.1 Short Term Interest Rates

Most of the existing literature on term structure has focused on the US Treasury Bill market. We extend our analysis, within a more generalized theoretical framework and using improved econometric methodology, to investigate if the stylized results of Chan et al (1992) hold in other global markets and over an extended period. We examine other international money markets (e.g., the Eurodollar, Eurosterling, and Euroyen), as well as additional domestic markets (e.g., Fed Funds and Bankers Acceptances rates.) We extend the CKLS methodology to data in the 1990's, which allows us an opportunity to test for structural changes.

3.3.2 Discount Bond Term Structure

Important studies of the term structure include the empirical test of the CIR model by Stambaugh (1988) and econometric estimation of a two-factor CIR model by Longstaff and Schwartz (1992). These studies concentrated on the U.S. market for government bills, notes, and bonds. This CRSP data set uses standard statistical techniques to infer the discount bond term structure from actual data on traded securities. We propose to investigate if the stylized results regarding the CIR hold with respect to this data over an extended period (10:64-10:97) and for different maturities.

3.3.3 Options on Interest Rate Futures

To date, the most comprehensive study of competing interest rate option models is Buhler et al (1999). They examine German interest rate warrants over the period 1990-1993. We extend their analysis CBOT options on Treasury bond futures for the first half of 1990, to compare our models with the traditional models that exist

4. THE REMAINDER OF THIS STUDY

This section summarizes the next two chapters of this study. In the upcoming sections, we implement the empirical tests of the term structure. In the next chapter, we present the term structure models to be tested empirically. In the third chapter, we test various global short-term interest rates, as well as U.S. government bonds. This involves a comparison of parametric and non-parametric term structure models. In the chapter following this, we extend the empirical tests to options on Treasury bond futures, comparing various types of parametric models to a non-parametric model for pricing interest rate derivatives by analyze pricing, hedging, and forecasting errors across models.

CHAPTER 2

THE SELECTION OF INTEREST RATE DERIVATIVE PRICING MODELS

1. INTRODUCTION AND DISCUSSION

In this section, we present the interest rate models to be estimated empirically. This involves two phases. First, we examine the scope of the universe of models under consideration, referred to as the process of *preselection.* This is supported by an extensive statistical analysis of the market for U.S. treasury bonds and their yields, as presented in Chapter 3, as well as reference to established term structure models. The second stage summarizes the main features of the models tested. To facilitate this, we view the universe of models along two dimensions-single versus multiple factor models on the one hand, and the spot versus forward rate models on the other. These are then compared to non-parametric models, which are independent of the stochastic dynamics of the term structure.

In testing interest rate valuation models, we first specify the basic characteristics of the models, which we term as the process of preselection. We specify, the factors driving the term structure models, which for the sake of tractability range between one and three. Furthermore, the quality of the factors is considered to examine if they are latent or observable, or whether they are motivated statistically or theoretically. Next, we examine the nature of the stochastic process driving the term structure.

To examine the term structure, we use *principal components analysis* (PCA) over a twenty-three year period, from 1964 to 1997. A variant of the multivariate G.A.R.C.H. model is used to estimate the time varying variance-covariance matrix of yield changes, motivated by the finding that yield volatilities and their covariances are time-varying. We use an improved computational approach to estimate the underlying factors, avoiding the assumption of multivariate normality of residuals in a maximum likelihood (ML) setting that provides estimates more robust to specification error. Due to the consistently high degree of correlation among yields of different maturities, we find that three factors are sufficient to describe over 98% of the variation in the term structure.

In one-factor models, the functional form of the stochastic process driving the term structure is a feature that distinguishes different models. In the case of forward rate models, the forward rate volatility is a key factor, while the specification of the process for the short rate is central to the implementation of spot rate models. In the latter case, the literature has focused primarily on one and two parameter variants, where it has been found that the number of parameters is more important than the particular form of the model (Amin et al, 1994). Most authors find that a two-factor forward rate model with linear proportional volatility best fits the data in and out of sample. However, the one-parameter variants result in more stable estimates and consistently higher arbitrage profits from perceived mispricings over time. We follow Buhler et al (1999) in considering two single factor HJM models, a one-parameter constant volatility[1] and a two-parameter linear proportional volatility variant.

In contrast to the forward rate models, the specification of the process is critical in short-rate models. Motivated by empirical findings such as Chan et al (1992; CKLS), we expect interest rate changes to be negative (positive) when rates are at relatively high (low) levels, which is termed *mean reversion*. There is evidence of non-linearity in the drift of interest rate changes (Ait-Sahalia, 1996), in the sense that the drift of the short rate has been found to increase gradually at low levels of the short rate and vice versa. In order to account for this, we replace the linear drift function $\alpha + \beta r_t$[2], as used in CKLS (1992), with the non-linear specification $\alpha + \beta r_t + v r_t \log(r_t)$. In the case of the diffusion function, it has been shown that the magnitude of interest rate movements is directly related to the level of the short rate. Motivated by theoretical models, we expect that the volatility of the short rate would not vanish at very low levels of the short rate. Hence, we nest the constant elasticity of variance (CEV) diffusion function of CKLS, σr_t^{γ}, within the displaced specification $(\delta + \sigma r_t)^{\gamma}$, where $\delta > 0$ is the parameter of displacement. Another advantage of this generalized specification is that it allows us to nest continuous time versions of other popular models, namely Black and Karasinki (1991), Pearson and

---

1 This can be recognized as the continuous time limit of the Ho and Lee (1986) model.

2 In this context, we interpret $-\beta$ as the speed of adjustment parameter, and $-\alpha/\beta$ as the long run mean of the short rate. Therefore, we impose the restriction that $\alpha > 0$ and $\beta < 0$.

Sun (1994), and Black, Derman and Toy (1990).

The study of more general models of the term structure are motivated by several well known empirical findings. First, principal components and generalized multiple regression analysis reveal that most of the variation in the term structure can be attributed to a number of unobservable factors ranging between 2 and 3, depending upon the market and time period. The first factor has similar impact (or "loading") across maturities, and as such can be interpreted as a "level factor". The second has a loading that varies directly with maturity, having influences of opposite signs at the long and short ends of the term structure, hence having a natural interpretation as a "slope factor". Finally, the third factor has usually been described as having a loading profile that peaks at intermediate maturities, resulting in twists in the term structure, and hence is referred to as a "curvature factor".[3]

In the context of forward rate models, the multifactor models that we consider are of two types: constant and linear proportional volatility functions. In both cases of two- and three-factor models, we empirically determine the functional forms of these models, through calculating the volatility of factors as determined by PCA. In the case of spot rate models, our choices are further motivated by empirical studies of the spot rate, that

---

3
The latter two factors, attributed to steepness and curvature of the yield curve, are related to the concepts of duration and convexity discussed in the traditional fixed income literature.

reveal the non-normal features of volatility clustering and asymmetry in its distribution (Ait-Sahalia, 1996). This motivates our choice of a *3-factor affine term structure model* (3F-ATSM). This is the most general family of models, nesting most of the models that have closed-form solutions.[4] Examples of these include the CIR (Cox et al, 1985) and Guassian term structure models. The 2-factor models are of two types. One is the 2-factor CIR model of Longstaff and Schwartz (1992; LS), which identifies the second pricing factor as random short rate volatility. The second type is motivated by data analysis, which reveals that a factor related to the slope of the yield curve has explanatory power. Hence, we implement a 2-factor model, in which the level of the short rate and the spread between it and the long rate are the two factors, an idea that is attributed to Brennan and Schwartz (1979) and Schaefer and Schwartz (1984). Both these models are special cases of the affine class of term structure models, as analyzed by Duffie and Kan (1996), in which the drift and volatility functions of the short rate are linear in the latent factors, and the parameters depend upon time.

---

[4] To be more specific, in ATSMs the short rate follows a linear *stochastic differential equation* (SDE), in which case a solution, should it exist, reduces it to a system of *ordinary differential equations* (SDEs). It is only in this sense that we mean "closed form". In contrast, in non-linear SDE models (such as the single factor general parametric model and the CKLS model), calibration to a discretized *partial differential equation* (PDE) or econometric analysis of discretized SDEs are the only way to approximate a solution.

2. THE SPOT RATE APPROACH

Spot-rate models of the term structure take the stochastic process of the underlying state variables as given, and produce as an output the spot term structure, as well as the equivalent (no-arbitrage) forward term structure. This endogenous term structure can be derived in a financial market equilibrium setting. These models are implemented by adaptation to current interest rates and their corresponding volatilities, and consistency is accomplished by allowing time dependency in the parameters of the state process. Due to various technical as well as theoretical reasons, we allow for a time dependent market price of risk (see CIR (1985)). These time dependent parameters of the stochastic factor processes are determined such that the endogenous term structure and volatility processes match their observable counterparts. We achieve this by inverting the valuation formula, finding the parameters that equate it to given bond prices, and then use the resulting formula to price options with no (pricing) bias at the underlying level.

This is contrasted to the forward-rate models, which adapt option prices to the current term structure through the specification of a stochastic forward rate curve. The difference is that the term structure is taken as an input, and market prices are determined as an output. Although the spot rate approach cannot insure consistency with the existing term structure, if one uses the correct underlying stochastic process, it is possible to exploit the mispicings over time in an internally consistent risk management program.

However, if the stochastic process driving the term structure is incorrect, this may lead to arbitrage.

2.1 Single Factor Models

In this section, we review the basic theoretical approach, in which the short rate is a sufficient statistic for determining the term structure. However, there are some limitations to this approach. First, we work in a continuous time framework, to be consistent with most derivative markets literature. Second, we implement the mathematical results of the existence of an equivalent martingale measure, which under technical conditions is equivalent to the absence of arbitrage. If we do this, the probabilities induced by this equivalent measure are such that the instantaneous expected return on any security is the short rate of interest. For mathematical details, see Duffie (1996) and Oksendahl (1995).

2.1.1 Mathematical Preliminaries for the Single-Factor Term Structure

We first present a *Standard Brownian Motion* (or SBM for short), $B_t \in \mathbb{R}$ and $t \in [0,T]$, or an independent continuous time random walk, to describe the source of uncertainty in the economy.[5] For simplicity, we stick to a finite time horizon $T < \infty$. This

---

[5] It is well-known that a SBM has $B_0 \underset{a.s.}{=} 0$ and is *independent, multivariate Gaussian* for

process is defined on a *complete probability space* $(\Omega, \Im, P)$, where Ω is a continuous sample space, $\Im$ is the σ-*algebra* generated by the SBM denoted $\sigma\left(B_s:\ s\in[0,T]\right)$[6], and P is the *probability measure* defined on $\Im$. We also fix a standard filtration $\mathscr{F}_t = \{\Im_t : t\geq 0\}$, defined as the σ-algebra formed from uniting the sigma-algebra generated by the Brownian motion, $\Im_t^B = \sigma\left(B_s:\ s\in[0,t]\right)$, and the null sets of $\Im$[7]. Furthermore, the probability measure is extended, such that $P(A) = 0 \quad \forall\ A = \varnothing$, which is to say that events of zero probability measure are considered as valid. The latter two qualifications are merely technical conditions necessary for the *completion* of the probability space. With this information structure in place, we proceed to our first assumption, which gives the basis of the model.

<u>Assumption 2.1.1</u>: There exists an adapted short-rate process $\{r_t : t\in[0,T]\}$ with $r:[0,T]\rightarrow\mathbb{R}_+\in\mathscr{L}^1[0,T]$.[8]

---

any non over- lapping times :
$\left(B_{t_1},..,B_{t_i}-B_{t_{i-1}},..,B_T-B_k\right)^T \sim N\left(0_{d\times k},\ (s\ -\ t)\mathbf{I}_k\right)\ \forall\ t_i\ \ni\ 0\leq t_1<\cdot\cdot t_i<\cdot\cdot<t_k\leq T.$

6
A sigma-algebra generated by a Brownian motion is a collection of all the outcomes that result from different possible paths of the process.

7See Protter (1990) for the case of a general information filtration.

8
This implies that the process r_t in the time interval [0,T] can be regarded a function from the domain set [0,T] to the range set of all non-negative real numbers, and this function is restricted to the set of all such functions with well-defined time integrals, for all sample paths that have positive probability.

This implies that r_t lives in the space of integrable functions from [0,T] to the real line, or $\int_{s=0}^{T} |r_t| dt \underset{a.s.}{<} \infty$.[9] This is the riskless, continuously compounded rate of interest on a money market account at time t. The assumption implies the existence of markets for instantaneous buying and selling at each point in time and provides a recipe for the calculation of arbitrage-free prices, assuming that the minimum technical conditions outlined are satisfied. This hypothesis is applicable throughout this study. We take the *numeraire* to be the *rolled-up money market account* $M_{0,t} \triangleq \exp\left[\int_{s=0}^{t} r_s ds\right]$, which is equivalent to discounting by the *unit discount bond* $\Lambda_{0,t} \triangleq \exp\left[-\int_{s=0}^{t} r_s ds\right]$. In the absence of arbitrage, there exists a probability measure Q, which has the property that the price of any security with a stochastic, finite variance time s > t payoff is given by

$$F(Z_t,t) = E_t^Q\left[\exp\left(-\int_{u=t}^{s} r_u du\right) Z_s\right] \tag{2.1.1.1}$$

where $E_t^Q(.)$ denotes expectation under risk neutralized measure conditional on $\mathfrak{F}_t$ and $F:\mathbb{R}\times[0,\tau]\to\mathbb{R}$ $\tau < T$ is a Borel measurable function. Consider the deflator, the pure discount bond itself, that pays one unit at some future time, $s \in (t,T]$. We assume that such a bond exists for each maturity and denote its price at time t by $\Lambda_{t,s}$. Equation (2.1.1.1), with Z=1, gives

$$\Lambda_{t,s} = E_t^Q\left[\exp\left(-\int_{u=t}^{s} r_u du\right)\right] \tag{2.1.1.2}$$

---

[9] This condition ensures that the expected value of the short rate is finite.

The function (2.1.1.2) Λ: $[0,T]\times[0,T] \to [0,1]$ is called the *term structure of interest rates* (or the yield curve). Given (2.1.1.2), we define the *continuously compounded yield* from t to s as

$$y_{t,\tau} = -\tau^{-1}\ln\left(\Lambda_{t,t+\tau}\right) \tag{2.1.1.3}$$

where $\tau = s - t$ is the holding period. The short rate is modeled in terms of the standard Brownian motion $\hat{B} \in \mathbb{R}$, under Q-measure that is obtained from $\mathbf{B}_t$ by the use of Girsanov's Theorem.[10] In this class of models, the short rate process is given by the *stochastic differential equation*

$$dr_t = \mu\left(r_t,t\right)dt + \sigma\left(r_t,t\right)d\hat{B}_t \tag{2.1.1.4}$$

where the functions $\mu:\mathbb{R}\times[0,T]\to\mathbb{R}$ and $\sigma:\mathbb{R}\times[0,T]\to\mathbb{R}$, the respective drift and diffusion, satisfy the standard technical conditions sufficient to guarantee the existence of (2.1.1.4). Given that the above equation is an *Ito process*, the short rate can be interpreted as

$$r_t = r_0 + \int_{s=0}^{t} \mu\left(r_s,s\right)ds + \int_{s=0}^{t} \sigma\left(r_s,s\right)d\hat{B}_s \tag{2.1.1.5}$$

In this model, r_t is the only factor governing the evolution of the yield curve (2.1.1.1). Mathematically, we can write $\Lambda_{t,s} = F\left(t,s,r_t\right)$ for some fixed, $\mathfrak{F}_t$-measurable function

10 For the Girsanov and Martingale Representation Theorems, see Chung and Williams (1990), Karatzas and Shreve (1988), and Revuz and Yor (1991).

$F:[0,t]\times[0,T]\times\mathbb{R}\to\mathbb{R}$.

2.1.2 The General Parametric Model, Affine Term Structures and the CIR Approach

Let the value of a contingent claim at time t, maturing at time T > t, where $T \in (0,T^{*}](3.7)$ for the longest maturity T^{*}, given the state of the economy X_t, be given by a sufficiently smooth function $F(X_t,t,T)$ with $0 < t \le T \le T^{*}$. Let the dynamics of the state variable, under the actual probability measure, follow the Ito stochastic differential

$$dX_t = \mu(X_t,t)dt + \sigma(X_t,t)dB_t \tag{2.1.2.1}$$

here the drift function is $\mu:\mathbb{R}_{+}\times[0,T] \to \mathbb{R}$ and the diffusion functions are $\sigma:\mathbb{R}_{+}\times[0,T] \to \mathbb{R}_{+}$.[11] CIR (1985) show that the function F(.) must satisfy the following partial differential equation (known as the *fundamental valuation equation*):

$$F_t + F_X(\mu - \Theta\sigma) + \frac{1}{2}F_{XX}\sigma^2 - rF = 0 \tag{2.1.2.2}$$

where $\Theta(\mathbf{X}_t,t)$ denotes the market price of risk for the state variable and r_t is the

---

11 There are certain regularity conditions that these functions must satisfy (so-called growth and Lipschitz conditions) that must be imposed for the stochastic process X_t to exist as a *strong solution* to an SDE. This means that the integral of the SDE will equal the value of the process at each time and for each path.

instantaneously risk-free interest rate. The absence of arbitrage implies that $\Theta:\mathbb{R}\times(0,T] \to \mathbb{R}$ is a function of only the state variables and time. The values for contingent claims can be found by solving the parabolic partial differential equation (2.1.2.2), subject to the appropriate initial and boundary conditions. Typically, the state variable is identified as the short-rate, r_t. The single factor models that we consider are all nested within the following SDE, which incorporates most of the parametric continuous-time models in the literature:

$$dr_t = \left[\alpha(\tau) + \beta(\tau)r_t + \nu(\tau)r_t\log(r_t)\right]dt + \left[\delta(\tau) + \sigma(\tau)r_t\right]^{\gamma(\tau)}d\hat{B}_t \qquad (2.1.2.3)$$

where $\tau = T - t$ is the time-to-maturity and the parameter vector $\boldsymbol{\Theta}^T = (\alpha,\beta,\gamma,\delta,\sigma,\nu):[0,T] \to \mathbb{R}^6$ depends at most on τ.[12] We refer to this as the *single factor general parametric model* (1F-GPM) and consider several single factor models that are nested within (2.1.2.3). The solution to this term structure is obtained numerically. An important specialization of this is the *affine single factor spot rate* model (1F-ASR), in which the drift function and instantaneous variance functions are linear in the short rate:

$$dr_t = (\alpha + \beta r_t)dt + (\delta + \sigma r_t)^{\frac{1}{2}}d\hat{B}_t \qquad (2.1.2.4)$$

12
This keeps the possibility of time dependent parameters open, and to correctly price bonds, the time dependency of at least one of these parameters must be assumed.

An appealing computational aspect of this model is that the yield-to-maturity on discount bonds are affine in the short rate as well, and parameter estimates to match market to model prices of discount bonds are obtained by solving a system of ODEs. Let the price at time t of a default-free, unit discount bond maturing at time T > t be given by:

$$F(r_t, T-t) = \exp\left[a(T-t) + b(T-t)r_t\right] \tag{2.1.2.5}$$

$a:[0,T] \to \mathbb{R}$, $b:[0,T] \to \mathbb{R}$ are functions of time-to-maturity. These functions must satisfy the following system of ODEs:

$$b(\tau)\left(\beta(\tau) + \frac{1}{2}\sigma(\tau)b(\tau)\right) - \left(1 + \frac{\partial b(\tau)}{\partial \tau}\right) = 0 \tag{2.1.2.6}$$

$$\frac{\partial(\tau)}{\partial \tau} - b(\tau)\left(\alpha_1(\tau) + b(\tau)\beta_1(\tau)\right) = 0 \tag{2.1.2.7}$$

where $\tau = T - t$ is the time-to-maturity.[13] Another popular family of models nested in (2.1.1.4) is the *Gaussian model.* The SDE for the short-rate dynamics of these models takes the form:

$$dr_t = \left[\alpha(\tau) + \beta(\tau)r_t\right]dt + \sigma(\tau)d\hat{B}_t \tag{2.1.2.8}$$

The important property of this class of models is that the short rate over any finite set of

---

13

In this model, F(•) is time homogenous, meaning that it is dependent upon calendar time only through the state variable and time to expiration T-t.

times is distributed multivariate normal under risk neutralized measure Q:

$$\begin{pmatrix} r(t_1) \\ . \\ . \\ .r(t_k) \end{pmatrix} \underset{Q}{\sim} N(\boldsymbol{\mu},\boldsymbol{\Sigma}), \tag{2.1.2.9}$$

for time $\{t_1,..,t_k\}\in[0,T]$ $\forall$ $k < \infty$, mean vector $\boldsymbol{\mu} \in \mathbb{R}^k$, and positive definite covariance matrix $\boldsymbol{\Sigma}\in\mathbb{R}^{k\times k}$. Equation (2.1.2.9) is a one-dimensional example of an important class of *linear* SDEs, which have the property of an explicit solution in terms of related ODEs. The closed form solution is given by:

$$b(t,T) = -e^{\alpha(T)}\left(\int_t^T e^{-\alpha(u)}\,\beta(u)\,du\right) + \frac{1}{2}\,e^{2\alpha(T)}\left(\int_t^T e^{-2\beta(u)}\,\sigma(u)\,du\right). \tag{2.1.2.10}$$

$$a(t) = -e^{\alpha(t)} \tag{2.1.2.11}$$

Another approach, which requires numerical calibration as in the 1F-GPM case, is the CKLS formulation:

$$dr_t = \kappa[\theta - r_t]dt + \sigma r_t^{\nu} d\hat{B}_t \tag{2.1.2.12}$$

where $\theta = \dfrac{-\alpha}{\beta}$, σ, and ν are positive constants and $\kappa = -\beta$. While this does not lead to a closed form expression for discount bond prices and neither is the solution time homogenous, it has the advantage of accommodating the high elasticity of variance as well as mean reversion, which matches empirical observations. Following Hull and White (1990), we allow a parameter to depend upon the time-to-maturity, in order that we can match an exogenously given schedule of discount bond prices. Since the elasticity parameter v is non-zero, there exists no closed form solution for the CKLS term structure, and so this calibration must be performed numerically. It is appropriate to let the market price of risk $\Theta(t)$ be time dependent[14], in that premia that varies with the time-to-maturity is consistent with several hypotheses of the term structure (such as the local expectations hypothesis). This gives rise to a natural economic interpretation of the calibration process.[15] The 1F-GPM and CKLS term structure models are numerically calibrated in several steps. First, estimates of the parameters of the short-rate drift and the volatility function are determined by an econometric technique that can accommodate the unobservability of a likelihood function. Examples of these include the *generalized method of moments* (Hansen, 1982; GMM), the *simulated method of moments* (Duffie and Singleton, 1993; SMM), and the *efficient method of moments* (Gallant et al, 2000;

14
Assuming a second parameter to exhibit time-dependency, it is possible to calibrate the model to the current volatility structure as well. However, there is evidence that such a procedure results in future volatility estimates that are not only unstable, but of unrealistic magnitudes. Therefore, a two-point calibration is achieved, by exactly matching the volatilities of the long- and short-rates, while interpolating the volatilities of intermediate maturities. This calibration is achieved by the mean reversion parameter, which governs the transmission of the short-rate volatility to the long rates.

15
There are technical reasons for this choice (see Heath et al (1992) or Uhrig et al (1996)).

EMM).[16] We then use a numerical algorithm to simultaneously solve for the time dependent function $\Theta(t)$ and the mean-reversion parameter κ, to calibrate the model to the current yield curve and long-rate volatility. This involves solving versions of the PDE equation (2.1.2.1) with r_t as the single state variable. In the case of the 1F-GPM model with constant coefficients, we solve:

$$F_t + F_r\left([\alpha + \beta r + vr\log(r)] - \Theta(t)[\delta + \sigma r]^{\gamma} \right) + \frac{1}{2}F_{rr}[\delta + \sigma r]^{2\gamma} - rF = 0 \quad (2.1.2.13)$$

where $\mu(r_t) = \alpha + \beta r_t + vr_t\log(r_t)$, $\sigma(r_t) = (\delta + \sigma r_t)^{\gamma}$, and $X_t = r_t$. In the case of the CKLS model, we solve:

$$F_t + F_r\left(\kappa[\theta - r] - \Theta(t)\ \sigma r^{v} \right) + \frac{1}{2}F_{rr}\ \sigma^2 r^{2v} - rF = 0 \quad (2.1.2.14)$$

where $\mu(r_t) = \kappa(\theta - r_t)$, $\sigma(r_t) = \sigma r_t^{v}$, and $X_t = r_t$. These are both solved subject to the maturity condition for zero coupon bonds:

$$F(r_T,T,T) = 1 \quad (2.1.2.15)$$

The fitting condition for bond prices, which matches endogenous (model) to observed

---

16 The latter two methods, SMM and EMM, involve estimation based upon a *hermite* expansion of a Gaussian transition density (called an *auxiliary model*) coupled with simulation of the underlying system based upon these parameters. This has the advantage of avoiding the *aggregation bias*, inherent in either ML or GMM estimation, using an Euler discretization of the SDE. An alternative that has no such bias is the Gausssian estimation technique (Nowman, 1997).

zero bond prices $\hat{F}(\bullet)$, is:

$$F(r_0,0,T) = \hat{F}(T) \quad T \in (0,T^*] \tag{2.1.2.16}$$

and the fitting condition for the long rate volatility:

$$-\frac{F_r(r_0,T^*,T^*)}{T^*\, F(r_0,T^*,T^*)} = \frac{\hat{\sigma}_{T^*}}{\hat{\sigma}_0} \tag{2.1.2.17}$$

where, $\hat{\sigma}_{T^*}$ is the historical volatility of the longest maturity (T^*) yield, $\hat{\sigma}_0$ is the sample estimator of the short-rate volatility, and the left-hand-side of (2.1.2.17) is the endogenous ratio of instantaneous long- to short-rate volatility.[17] The system of differential equations (2.1.2.14) to (2.1.1.17) is solved by the *inverted implicit finite difference* method (Uhrig et al (1996), Buhler et al (1999)).[18]

17

Since bond prices are decreasing in the short rate, the negative sign is needed to make the left- hand-side positive, since the right-hand-side is positive (as volatilities are positive). This can be recognized as the (average) absolute elasticity of the long bond price with respect to the short rate, evaluated at the initial value r_0.

18

This version of the implicit finite difference method has been shown superior to explicit schemes, with respect to convergence and stability, by Uhrig and Walter (1996). Finite difference techniques are related to the trinomial algorithm implemented by Hull and White (1993).

2.2 Multifactor Models

In this section, we sketch the analytical primitives of various multifactor models. The choice of 2 and 3 factor models are motivated by the empirical findings that up to three independent factors explain approximately 98% of all the variation in the term structure for U.S. Treasury securities. The inversion of the contingent claim pricing function is based upon the fundamental valuation equation derived by CIR (1985) in a general equilibrium setting. Let there be a k-vector of instantaneously independent factors, which under actual probability measure follows the multivariate Ito process :

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t,t)dt + \Sigma(\mathbf{X}_t,t)d\mathbf{B}_t \tag{2.2.1}$$

where is a vector of k instantaneously independent factors, $\mathbf{B}_t = (B_{1t},..,B_{dt})^T$ is a vector of independent d-dimensional Brownian motions, $\boldsymbol{\mu}(\mathbf{X}_t,t) = (\mu_1(\mathbf{X}_t,t),..,\mu_k(\mathbf{X}_t,t))^T$ is a vector of drift functions, and $\boldsymbol{\sigma}(\mathbf{X}_t,t) = \text{diag}(\sigma_1(\mathbf{X}_t,t),..,\sigma_k(\mathbf{X}_t,t)^T)$ is a diagonal matrix of diffusion coefficients. Let the time t value of a contingent claim maturing at time T > t, where $T \in (0,T^*]$ for the longest maturity T^*, given the state of the economy $\mathbf{X}_t$, be given by a sufficiently smooth function $F(\mathbf{X}_t,t,T)$ where $0 < t \leq T \leq T^*$. CIR (1985) show that this function must satisfy the following partial differential equation (known as the *fundamental valuation equation*):

$$F_t + \sum_{i=1}^{k}\left[F_{X_i}\left(\mu_i - \Theta_i\sigma_i\right) + \frac{1}{2}F_{X_iX_i}\sigma_i^2\right] - rF = 0 \tag{2.2.2}$$

where $\boldsymbol{\Theta}\left(\mathbf{X}_t,t\right) = \left(\Theta_1\left(\mathbf{X}_t,t\right),..,\Theta_k\left(\mathbf{X}_t,t\right)\right)^T$ denotes the vector of market prices of risk. The absence of arbitrage implies that $\boldsymbol{\Theta}:\mathbb{R}^k \rightarrow \mathbb{R}^k$ is a vector-valued function of the state variables but directly of time. The values for contingent claims can be found by solving the parabolic PDE (2.2.2), subject to the appropriate initial and boundary conditions, for the cases of k=1,2, and 3.

2.2.1 The 2-Factor Longstaff and Schwartz Model

We present a generalization of the 2-factor CIR model, originally developed by Longstaff and Schwartz (1992). In this equilibrium setting, two stochastically independent factors are proposed:

$$dX_{it} = \kappa_i\left[\varphi_i - X_{it}\right]dt + \sigma_i X_{it}^{\frac{1}{2}}d\tilde{B}_{it}, \tag{2.2.1.1}$$

for i =1,2. Assuming that the first factor influences production uncertainty, while both factors affect its expected return, it follows that this first factor is the only one that is priced:

$$\Lambda_{it} = \begin{cases} \sigma_{it}^{\frac{1}{2}}\lambda_t & \text{if } i = 1 \\ 0 & \text{if } i = 2 \end{cases} \tag{2.2.1.2}$$

Instantaneous returns on production are given by the following stochastic differential equation:

$$\frac{dQ_t}{Q_t} = \left(\mu X_{1t} + \theta X_{2t}\right)dt + \sigma_Q\sqrt{X_{1t}}\,d\tilde{B}_{Qt} \qquad (2.2.1.3)$$

The parameter vector $\left(\mu,\theta,\sigma_Q\right)^T$ is a constant and $\tilde{B}_{Qt}$ is a standard Brownian motion. Additionally, a *time additive preference* structure is assumed, which give rise to a logarithmic derived utility of wealth.[19] This implies that the equilibrium interest rate equals the instantaneous expected rate of return on physical production less its conditional variance, which is equivalent to the following:

$$r_t = \psi_1 X_{1t} + \psi_2 X_{2t} \qquad (2.2.1.4)$$

The constants of the short rate diffusion are related to the parameters of the production

---

19

Time additive preference, a standard assumption in multi-period equilibrium models, means that the utility function of the representative investor is the integral over utility at each point in time. This implies that the *optimal* discounted value of utility can be expressed as

$$V^*\left(W_t,t\right) = \max_{\{C_s\}_{s=t}^{T}}\left\{E_t\left[e^{r_s(T-s)}\int_{s=t}^{T} U\left(C_s,s\right)\right]\right\},$$

which is known as the *derived* (or *indirect*) *utility of wealth*. It has been shown (Merton, 1973) that in this setting, the derived utility of wealth can be written in the logarithmic form: $V^*\left(W_t,t\right) = G(t)\log\left(W_t\right)$. Applying Ito's Lemma to the first order conditions of the optimization problem leads to the derivation of the process for wealth and for the short rate.

process as $\psi_1 = \alpha\sigma_1^2$ and $\psi_2 = \left(\theta - \sigma_Q^2\right)\sigma_2^2$. Applying Ito's Lemma to (2.2.1.4), the instantaneous variance of the short rate is

$$V_t = \psi_1^2\sigma_1^2 X_{1t} + \psi_2^2\sigma_2^2 X_{2t} \tag{2.2.1.5}$$

This is a non-negative process for all sample paths given that the process (2.2.1.3) is non-negative as well. Provided that the state variables are not collinear, this system is invertible everywhere, and the state variables can be expressed in terms of the observable short rate and its volatility. Assuming that $\psi_2\sigma_2^2 \neq \psi_1\sigma_1^2$, the unobservable state vector is:

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \kappa_{1r} & \kappa_{1V} \\ \kappa_{2r} & \kappa_{2V} \end{pmatrix} \begin{pmatrix} r_t \\ V_t \end{pmatrix}, \tag{2.2.1.6}$$

where $\kappa_{1r} \triangleq \left(\psi_1\left(\psi_2 - \psi_1\left(\frac{\sigma_1}{\sigma_2}\right)^2\right)\right)^{-1}$, $\kappa_{1V} \triangleq \left(\psi_1\left(\psi_1\sigma_1^2 - \psi_2\sigma_2^2\right)\right)^{-1}$, $\kappa_{2r} \triangleq \left(\psi_2\left(1 - \psi_2\left(\frac{\sigma_2}{\sigma_1}\right)^2\right)\right)$, and and $\kappa_{2V} \triangleq \left(\psi_2\left(\psi_2\sigma_2^2 - \psi_1\sigma_1^2\right)\right)^{-1}$. We substitute the solution for the state variables in (2.1.1.6) into the following partial differential equation, which is satisfied (subject to the appropriate boundary conditions) by all interest rate derivatives in this model:

$$F_t + \left(\frac{\kappa_1\theta_1}{\sigma_1^2} - \left(\kappa_1 - \lambda_t\right)\sigma_1^2\right)F_1 + \left(\frac{\kappa_2\theta_2}{\sigma_2^2} - \kappa_2\sigma_2^2\right)F_2 +$$

$$+ \frac{1}{2}\left(\sigma_1^2 X_1^2 F_{11} + \sigma_2^2 X_2^2 F_{22}\right) - rF = 0 \tag{2.2.1.7}$$

The parameters of this model are estimated using the following procedure. It can be shown that

$$E[r_t] = \gamma\left(\frac{\alpha}{\delta}\right) + \eta\left(\frac{\beta}{\xi}\right) \tag{2.2.1.8}$$

$$Var[r_t] = \frac{1}{2}\left[\gamma\left(\frac{\alpha}{\delta}\right)^2 + \eta\left(\frac{\beta}{\xi}\right)^2\right] \tag{2.2.1.9}$$

$$E[V_t] = \gamma\left(\frac{\alpha^2}{\delta}\right) + \eta\left(\frac{\beta^2}{\xi}\right), \tag{2.2.1.10}$$

$$Var[V_t] = \frac{1}{2}\left[\gamma\left(\frac{\alpha^2}{\delta}\right)^2 + \eta\left(\frac{\beta^2}{\xi}\right)^2\right] \tag{2.2.1.11}$$

We set the historical means and standard deviations of both the short rate, as well as its estimated volatility[20], on the left-hand-side of equations (2.2.1.8)-(2.1.11). It can be shown that $\hat{\alpha} = \min_{t\in(0,1..,T]}\left\{\frac{\hat{V}_t}{r_t}\right\}$ and $\hat{\beta} = \max_{t\in(0,1..,T]}\left\{\frac{\hat{V}_t}{r_t}\right\}$ consistently estimate α nd β. We use a simple Gauss-Newton iterative procedure to solve for these auxiliary parameters and

20

The volatility is estimated non-parametrically, following Ait-Sahalia (1996) and Jiang (1998), in contrast to the approach of Longstaff and Schwartz (1992), who estimate interest rate volatility in a GARCH (1,1) frame- work. This is based on an Euler discretization of the continuous time processes for r_t and V_t, which is subject to aggregation bias, and we avoid that here. See the next chapter for a description of this kernel estimation procedure.

then use the same procedure in a second stage to recover the six underlying parameters. Finally, we substitute these solutions in (2.2.1.6), and then estimate the PDE (2.2.1.7). This allows us to solve a parabolic PDE in observable and estimable variables $\{r_t, \hat{V}_t\}$ by a multiple separation of variables

.

2.2.2 The 2-Factor Schaeffer and Schwartz Model

This is an extension of the 2-factor model of Schaefer and Schwartz (1984), in which the two factors are posited as the short rate, as well as the spread ($s_t = r_t - l_t$) between the short rate r_t and the long rate l_t. In this model, the short-rate is a function of the drift and two stochastically independent factors. Assume that the short rate process is the difference between the two factor processes:

$$r_t = l_t - s_t \tag{2.2.2.1}$$

The dynamics of the two instantaneously independent factors are given by:

$$dl_t = \kappa_l(\theta_l - l_t)dt + \sigma_l\sqrt{l_t}d\tilde{B}_{lt} \tag{2.2.2.2}$$

$$ds_t = \kappa_s(\theta_s - s_t)dt + \sigma_s\sqrt{l_t}d\tilde{B}_{st} \tag{2.2.2.3}$$

The PDE that must be solved to match the observed term structure and to price discount bond options is given by:

$$F_t + \left(\kappa_l(\gamma_l - l_t) - \varphi\sqrt{l_t}\sigma_l\right)F_l + \left(\kappa_s(\gamma_s - s_t) - \lambda_t\sigma_s\right)F_s +$$

$$+ \frac{1}{2}\left(\sigma_l^2 l_t^2 F_{ll} + \sigma_s^2 F_{22}\right) - \left(l_t - s_t\right)F = 0 \qquad (2.2.2.4)$$

The market price of risk is chosen to satisfy:

$$-\tau^{-1}\log\left(F\left(l_t, s_t, t, t+\tau\right)\right) = l_t \quad \forall\ t < T^* = t + \tau \qquad (2.2.2.5)$$

Equation (2.2.2.5) implies that we set the premium for long rate risk such that model yields on the long bond are consistent with observed long term rates. We use the market price of long-rate risk to overcome an internal inconsistency inherent in models where all the factors are market rates. In other words, the price (and therefore yield-to-maturity) on the longest-term bond depends on both l_t and s_t, where we label the long term factor as the "long-rate". However, since both rates determine its price, l_t does not have this property. Therefore, we determine the parameter φ (in the second term of equation (2.2.2.4)), such that equation (2.2.2.5) is satisfied.

2.2.3 Three Factor Affine Term Structure Models (ATSMs)

The models in this class used to value interest rate derivatives are three factor versions of the Guassian Term Structure Model (3F-GTSM) and the Cox-Ingersoll-Ross model (3F-CIRTSM). In all the 3-factor ATSMs, the short rate process is a linear combination of three state variables:

$$r_t = \psi_0 + \psi_1 X_1 + \psi_2 X_2 + \psi_3 X_3 \qquad (2.2.3.1)$$

In the 3F-GTSM model, we assume that the state process is 3-dimensional Guassian: here

$$dX_{1t} = -\kappa_{11} X_{1t} dt + d\tilde{B}_{1t} \qquad (2.2.3.2)$$

$$dX_{2t} = -\left[\kappa_{21} X_{1t} + \kappa_{22} X_{2t}\right] dt + d\tilde{B}_{2t} \qquad (2.2.3.3)$$

$$dX_{3t} = -\left[\kappa_{31} X_{1t} + \kappa_{32} X_{2t} + \kappa_{33} X_{3t}\right] dt + d\tilde{B}_{3t} \qquad (2.2.3.4)$$

the condition $\kappa_{ij} > 0$ for $i=j$ is imposed to assure admissibility. These models are studied theoretically by Vasicek (1977). An empirical estimation of a 2-factor Guassian term model is presented in Jegadesh and Pennachi (1996). The implications of the over-identifying restrictions imposed by this model, as compared to a general 3-factor ATSM, are examined by Dai et al (1998). The other 3-factor ATSM that we consider is characterized by a state vector that follows three correlated square-root diffusions (the 3F-CIRTSM model). In this case, the SDEs governing the state processes are given by:

$$dX_{it} = -\left[\kappa_{i1} X_{1t} + \kappa_{i2} X_{2t} + \kappa_{i3} X_{3t}\right] dt + X_{it}^{\frac{1}{2}} d\tilde{B}_{it} \text{ for } i = 1,..,3 \qquad (2.2.3.5)$$

The condition $\kappa_{ij} < 0$ for $i \neq j$ is imposed to assure admissibility. While the factors are instantaneously uncorrelated, they exhibit correlation over finite time periods. This is due to the relaxation of the over-identifying restriction that **K** is diagonal. Examples of such multi-factor CIR models are Chen and Scott (1993) and Pearson and Sun (1994).

3. FORWARD RATE MODELS

The starting point of the HJM model is the entire yield curve, from which the processes governing the short rate and its derivatives are derived. This is in contrast to the approach taken by specifying underlying state variables, and then deriving an endogenous term structure by imposing equilibrium conditions, as with the spot-rate models. In essence, there is an infinite state space at each point in time. The no-arbitrage requirement means that the yield curve can be presented in terms of forward rates prevailing at any point in time. Consider a continuous time economy with a market for zero coupon, default-free unit discount bonds. There exists a market for forward borrowing and lending, where the time t forward price of a loan with unit face value in the interval $[\tau,s]$ is denoted by $\Phi_{t,\tau,s}$. With no arbitrage, prices of forward contracts are related to the prices of discount bonds by

$$\Phi_{t,\tau,s} = \frac{\Lambda_{t,s}}{\Lambda_{t,\tau}} \tag{3.1}$$

Define the time t *forward rate* as

$$\varphi_{t,\tau,s} \triangleq -(s-\tau)^{-1}\ln\left(\Phi_{t,\tau,s}\right) = \frac{\ln\left(\Lambda_{t,\tau}\right) - \ln\left(\Lambda_{t,s}\right)}{s - \tau} \tag{3.2}$$

In the absence of arbitrage $\varphi_{t,\tau,s} \underset{a.s}{<} 1 \Leftrightarrow r_t \underset{a.s.}{>} 0 \ \forall \ t$. This is the continuously compounded rate of return on a bond purchased forward for the period [τ,s]. The *instantaneous forward rate* is defined in terms of the right-hand limit of (3.1) as s approaches τ.

$$f(t,\tau) = \lim_{s \downarrow \tau} \varphi_{t,\tau,s} \tag{3.3}$$

Partial differentiation and the definition of the instantaneous forward rate process (3.3) show that $f(t,\tau) = -\frac{\partial}{\partial \tau}\ln\left(\Lambda_{t,\tau}\right)$. The instantaneous forward rate process exists if and only if for each $t,\tau \in [0,T]$ the forward rate function defined in (3.2) is differentiable with respect to s, meaning that $\exists \ \frac{\partial \varphi_{t,\tau,s}}{\partial s} \ \forall \ s \in [\tau,T]$. Given this characterization of the forward rate curve, the entire zero-coupon yield curve is derived by noting from equation (3.3) that $f(t,\tau) = -\frac{\partial}{\partial \tau}\ln\left(\Lambda_{t,\tau}\right)$, which implies

$$\Lambda_{t,s} = \exp\left[-\int_t^s f(t,u)du\right] \tag{3.4}$$

This shows the instantaneous forward rate curve to be the building block of the spot term structure, in the sense that the price of a default-free, zero coupon bond has the integral representation (3.4). Equation (3.4) shows that the instantaneous forward rate schedule is a sufficient statistic for the determination of the spot term structure. This however has two problems in implementation. First, the forward rate curve is unobservable in the market. Second, although interest rate dependent contingent claims are priced consistently with respect to the initial term structure, there is no guarantee of such consistency with respect to spot rate dynamics determined endogenously in dynamic equilibria. In other words, there may be arbitrage opportunities if the existing term structure is not consistent with respect to the underlying structural model (Jiang, 1998). These complications are avoided by the assumption that there exists a well behaved short-rate process that is consistent with the requirement of the HJM formulation, as the limit of a sequence of instantaneous forward rate processes (see Heath et al, 1992).

Assumption 3.1: The spot rate process under the HJM framework is given by the right-hand limit of the process (3.3), as maturity approaches time t:

$$r_t = f(t,t) \triangleq \lim_{\tau \downarrow t} f(t,\tau). \tag{3.5}$$

In its "proper" stochastic integral representation, the forward rate process is:

$$f(t,s) = f(0,s) + \int_{u=0}^{t} \mu(u,s,f(.))ds + \int_{u=0}^{t} \boldsymbol{\sigma}^T(u,s,f(.))d\hat{\mathbf{B}}_u \quad 0 \le t \le s \le T \tag{3.6}$$

where and $\mu:\hat{T}\times\Omega \to \mathbb{R}$, for $\hat{T} \triangleq \left\{(t,s)\in\mathbb{R}^2 : 0\le t\le s\le T\right\}$, are measurable functions and adapted processes in this product space, in order to assure that (3.6) is a well-defined Ito

Process on the real line. $\hat{\mathbf{B}}_t \in \mathbb{R}^d$ is a d-dimensional standard Brownian motion (SBM)[21], under equivalent martingale measure[22], that arises in the application of Girsanov's Theorem. HJM show that, given the regularity conditions and the standard no-arbitrage restriction, the drift $\mu(t,T)$of the forward rate process under risk neutral measure is determined by the vector of volatility functions $\boldsymbol{\sigma}(t,T,f(t,T)) = \left(\sigma_1(t,T,f(.)),...,\sigma_d(t,T,f(.))\right)^T$.

Theorem 3.1: Consistency Restriction on the HJM Drift Function

Under the appropriate conditions, and in the no-arbitrage context of the HJM model, the instantaneous drift process of the forward rate (3.6) satisfies:

$$\mu(t,s) = \boldsymbol{\sigma}^T(t,s) \int_{u=t}^{s} \boldsymbol{\sigma}(t,u) du \qquad (3.7)$$

Proof:

See HJM (1992).

21

A SBM is a stochastic process, defined as a jointly measurable function $\mathbf{B}:\Omega\times\mathbb{R}^{+}\rightarrow\mathbb{R}^d$, given a complete probability space $(\Omega,\mathscr{F},P)$ and standard filtration of $F \triangleq \{\mathscr{F}_t : t=>0\}$

22

A probability measure Q defined on a (continuous) state space-tribe pair $(\Omega,\mathscr{F})$ is said to be equivalent to probability measure P if and only if $Q(A) > 0 \Leftrightarrow P(A) > 0 \ \forall \ A \in \mathscr{F}$, for P defined on the same space.

3.1 Single Factor Forward Rate Models

In a single factor forward rate model, the uncertainty in the forward rate process is described by a one-dimensional Brownian motion:

$$f(t,s) = f(0,s) + \int_{u=0}^{t} \mu(u,s,f(.))ds + \int_{u=0}^{t} \sigma^T(u,s,f(.))d\hat{B}_u \quad 0 \le t \le s \le T, \qquad (3.1.1)$$

where $\sigma:\hat{T}\times\Omega \to \mathbb{R}$ and $\mu:\hat{T}\times\Omega \to \mathbb{R}$, for $\hat{T} \triangleq \{(t,s)\in\mathbb{R}^2 : 0\le t\le s\le T\}$, are measurable functions and adapted processes on this product space, to assure that (3.1.1) is a well-defined Ito Process on the real line.

3.1.1 The Constant Volatility HJM Model

The basic single factor forward rate model, called Absolute Volatility I (HJM-AI) model, suggests a constant volatility function:

$$\sigma_{AI}(t,s,f(.)) = \sigma \qquad (3.1.1.1)$$

where $\sigma > 0$ for $0 < t < s < T$ is a strictly positive parameter. Integrating (3.1.1.1) implies that the drift function is linear in time-to-maturity:

$$\mu_{AI}(t,s,f(.)) = \sigma^2(s - t) \tag{3.1.1.2}$$

This implies a call option pricing formula that is similar to the standard Black-Merton-Scholes formulation, in a stochastic interest rate setting. Integration of the SDE of the forward rate function, given by $df(t,T) = \sigma^2(T - t)dt + \sigma d\hat{B}_t$ yields

$$f(t,T) = f(0,T) + \sigma^2 t(T - t)dt + \sigma\hat{B}_t \tag{3.1.1.3}$$

Taking the limit of (3.1.1.3) as $T \rightarrow t$ shows the short rate to be the following linear function (for given time and initial forward rate) of the Brownian motion:

$$r_t \triangleq f(t,t) = f(0,t) + \frac{\sigma^2 t^2}{2} + \sigma\hat{B}_t \tag{3.1.1.4}$$

Equation (3.1.1.4) shows that the short rate can take a negative value with positive probability, which is undesirable. By taking the exponent and integrating, unit bond prices at time t for maturity $T > t$, are given by:

$$\Lambda_{t,T} = \frac{\Lambda_{0,T}}{\Lambda_{0,t}}\left\{\exp\left[-\sigma(T - t)\left(\frac{\sigma t}{2} + \hat{B}_t\right)\right]\right\} \tag{3.1.1.5}$$

The price at time t of a European call option $F\left(t,\Lambda_{t,T}\middle|T^*,K\right)$ on a unit discount bond $\Lambda_{t,T}$, with option expiration $T^* < T$ and strike price K, is given by the risk neutral expected present value of the option at expiration:

$$F\left(t,\Lambda_{t,T}\middle|T^{*},K\right) = E_t^Q\left[\exp\left(-\int_{s=t}^{T^{*}} r_s ds\right)\left(\Lambda_{T^{*},T} - K,0\right)^{+}\right] \tag{3.1.1.6}$$

Given that the short rate is normally distributed, we have:

$$F\left(t,\Lambda_{t,T}\middle|T^{*},K\right) = \Lambda_{t,T}\Phi\left(d_1\right) - K\Lambda_{t,T^{*}}.\Phi\left(d_2\right) \tag{3.1.1.7}$$

where $\Phi(x) = \frac{1}{2\sqrt{\pi}}\int_{u=-\infty}^{x} e^{-u^2}du$ is the standard normal density function and

$$d_{1,2} \triangleq \frac{\log\left(\frac{\Lambda_{t,T}}{K\Lambda_{t,T^{*}}}\right) \pm \frac{1}{2}\sigma^2\left(T - T^{*}\right)^2\left(T^{*} - t\right)}{\sigma\left(T - T^{*}\right)\sqrt{\left(T^{*} - t\right)}} \tag{3.1.1.8}$$

3.1.2 The Linear Proportional Volatility HJM Model

The second single factor forward-rate model is the Proportional Volatility I model (HJM-VI), in which volatility is affine in time-to-maturity:

$$\sigma_{PI}(t,s,f(.)) = \left(\sigma_0 + \sigma_1(s - t)\right)\times\min(f(.),M) \quad \forall\ 0 < t < s < T, \tag{3.1.2.1}$$

where $0 < M \ll \infty$ is a large positive constant that bounds volatility. Integrating the drift function shows it to be a cubic function of the maturity time s:

$$\mu_{PI}(t,s,f(.)) = \sigma_1\left(\frac{3}{2}t^2\sigma_0\right) + \left\{\sigma_1\left[\sigma_1\left(1-\frac{3}{2}t^2\right)3t\sigma_0\right]\right\}s +$$

$$+\left[\frac{3}{2}\sigma_1\left(\sigma_0-\sigma_1 t\right)\right]s^2 + \left(\frac{1}{2}\sigma_1^2\right)s^3 \tag{3.1.2.2}$$

As in the HJM-AI model, closed form solutions for discount bond and bond option prices are straightforward (see Heath et al (1992))..

3.2 Multi-Factor Forward Rate Models

In the multi-factor forward rate models, the uncertainty in the forward rate process (3.6) can be described by two or three independent Brownian motions:

$$f(t,s) = f(0,s) + \int_{u=0}^{t}\mu(u,s,f(.))ds +$$

$$+\sum_{i=1}^{k}\left[\int_{u=0}^{t}\sigma_i(u,s,f(.))d\hat{B}_{iu}\right] \qquad 0 \le t \le s \le T,\ k = 2,3 \tag{3.2.1}$$

where $\sigma:\hat{T}\times\Omega \rightarrow \mathbb{R}$ and $\mu:\hat{T}\times\Omega \rightarrow \mathbb{R}$, for $\hat{T} \triangleq \{(t,s)\in\mathbb{R}^2:0\le t\le s\le T\}$, are measurable functions and adapted processes on this product space, to assure that (3.2.1) is a well-defined Ito Process on the real line and $\hat{\mathbf{B}}_{it}$ are standard Brownian motions (SBMs) under equivalent martingale measure, for i = 1,2,3. To implement these models, the first step involves estimating the current yield curve, in the form of initial forward rate curve. To estimate the volatility parameters in the single-factor cases, we use the volatilities of the estimated forward rate changes. In the multi-factor case, we apply principal components analysis to the estimated forward rates, and calculate the volatilities of the independent factors derived from the estimated factor loadings. The second step involves computing option prices from a binomial tree. We use the backward induction procedure, which accommodates the American feature of the bond futures options, taking at each node the maximum of the intrinsic and model value of the option. The tree is non-recombining in the case of proportional volatility models. We find that seven time steps are sufficient to achieve accurate option prices, in agreement with Amin et al (1994) and Buhler et al (1999). Such a tree contains 68 final nodes for the one factor case, 253 final nodes for the two-factor case, and 847 nodes for the three-factor case.

3.2.1 The 2-Factor HJM Model

The 2-factor versions of the HJM-AI and HJM-PI models are straightforward generalizations:

$$\boldsymbol{\sigma}_{AII}(t,s,f(.)) = \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix}, \qquad (3.2.1.1)$$

$$\boldsymbol{\sigma}_{PII}(t,s,f(.)) = \begin{bmatrix} \sigma_1 \min(f(.),M_1) \\ \sigma_2 \min(f(.),M_2) \end{bmatrix} \tag{3.2.1.2}$$

where (3.2.1.1) and (3.2.1.2) are the HJM-AII and HJM-PII models, respectively.

3.2.2 The 3-Factor HJM Model

The 3-factor versions of the HJM absolute and proportional volatility models are also straightforward generalizations:

$$\boldsymbol{\sigma}_{AIII}(t,s,f(.)) = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{bmatrix} \tag{3.2.2.1}$$

$$\boldsymbol{\sigma}_{PIII}(t,s,f(.)) = \begin{bmatrix} \sigma_1 \min(f(.),M_1) \\ \sigma_2 \min(f(.),M_2) \\ \sigma_3 \min(f(.),M_3) \end{bmatrix} \tag{3.2.2.2}$$

where (3.2.2.1) and (3.2.2.2) are the HJM-AIII and HJM-PIII models, respectively.

4. NON-PARAMETRIC TERM STRUCTURE MODELS

We next consider models that are not based on restricted mathematical structures, which arise out of assumptions on the underlying stochastic processes and where the choice of factors may be motivated by either theoretical or empirical considerations. The two types of non-parametric models that we consider are *multivariate kernel estimation* models (henceforth MKE) and *artificial neural networks* (henceforth ANN) The particular case of MKE models that we consider is the *multivariate Nadaraya-Watson kernel estimator* (henceforth MNWKE), while the ANN model we consider is the *multi-layer perceptron* (MLP) pricing procedure. We first introduce the MLP architecture for neural networks. The second section summarizes the kernel regression and the cross-validation process for choosing the optimal bandwidth, both of which can be fruitfully applied to the pricing of interest rate derivatives.

4.1 The Multilayer Perceptron Network Architecture

In this section, we consider a class of non-parametric models, *artificial neural networks*, which are models that try to mimic the process of human cognition in developing pricing formulae, and have been shown to compete successfully with parametric counterparts in the pricing of a variety of derivatives. In general, derivative pricing functions are multivariate and non-linear. Hornik et al (1989) has shown that a broad class of neural networks, the *multilayer feedforward* variety, constitute the so-called *universal approximators*. White (1989, 1990) has shown that neural networks of this kind are amenable to estimation by non-linear regression. This method prices interest rate derivatives by estimating an empirical valuation function, thereby making no assumptions regarding the specific functional form followed by the "underlying" asset, which in this context is the short rate and other appropriate term structure variables (see Malliaris et al (1993), Hutchinson et al (1994)). The estimation procedure involves partitioning the data set into subsets used for training, cross-validation, and testing. Parameter values are determined in the training set via non-linear optimization of a suitable criterion (see Chapter 3 for details). The particular class of networks that we use to price US Treasury notes and bonds is given by a version of the *multi-layer perceptron artificial neural network* (MLP-ANN):

$$F^{MLP}\left(\mathbf{X}_{it} \middle| \theta, J\right) = \beta_i + \sum_{j=1}^{J} \left(\frac{\omega_{ij}}{1 + \exp\left(\sum_{k=1}^{K} \omega_{ikj} X_{k,it} + \beta_{ij} \right)} \right) \tag{4.1.1}$$

where $F^{MLP}\left(\mathbf{X}_{it} \middle| \theta, J\right)$ denotes the i^{th} option price at time t as approximated by the network with J hidden layers. The 5-dimensional vector $\mathbf{X}_{it} = \left(X_{1,it},..,X_{k,it}\right) = \left(r_t, \hat{V}_t, F_{it}, K_i, \tau_{it}\right)^T$ represent the inputs to the network where: k=5 is the fixed number of inputs, r_t is the short rate of interest at time t, $\hat{V}_t$ is the (non-parametrically) estimated volatility of the short rate, F_{it} is the time t price of the i^{th} underlying futures contract, K_i is the strike price of the i^{th} option, and $\tau_{it} = T_i - t$ is the time-to-maturity. The free parameters of the network are $\theta = \left(\{\beta_i\}_{i=1,..,I}, \{\beta_j, \omega_j\}_{j=1,..,J}, \{\omega_{ikj}\}_{i,k,j=1,..I,K,J} \right)^T$ and J is chosen experimentally to achieve the best criterion. We choose the free parameters, for the J^{th} number of hidden units, to minimize the root mean squared error (RMSE) of the pricing errors:

$$\mathrm{RMSE}_{J,\theta} \triangleq \sqrt{\frac{1}{T\times I}\sum_{t=1}^{T}\sum_{i=1}^{I}\left(F^{MLP}\left(\mathbf{X}_{it}\middle|\theta,J\right) - Y_{it}\right)^2} \tag{4.1.2}$$

The optimal parameter vector for the MLP with J hidden units is given by

$$\hat{\theta}(J) = \underset{\theta}{\operatorname{argmin}}\left\{\mathrm{RMSE}_{j,\theta}\right\} \tag{4.1.3}$$

Finally, we search over all network architectures for the optimal number of hidden units, to give the globally optimal vector of free-parameters:

$$\hat{\theta}^{*} = \underset{J}{\operatorname{argmin}}\left\{\mathrm{RMSE}_{j,\hat{\theta}(J)}\right\} \tag{4.1.4}$$

4.2 The Nadarya-Watson Kernel Estimator

Next, we consider an extension of the technique of kernel regression, to the case where we know that the dependent variable depends upon several independent variables. Suppose that the price of the i^{th} derivative, Y_{it}, depends upon a k-vector of factors $\mathbf{X}_{it} = \left(X_{i_1 t},..,X_{i_k kt}\right)^T$ for $t = 1,..,T;\ i = 1,..,I$. In the application to options on bond futures, this is given by the 5-dimensional vector $\mathbf{X}_{it} = \left(r_t,\hat{V}_t,F_{it},K_i,\tau_{iT_i}\right)^T$, where the five variables are defined earlier. We posit the pricing relationship as

$$Y_{it} = F(\mathbf{X}_{it}) + \varepsilon_{it} \tag{4.2.1}$$

where $F:\mathbb{R}^k \to \mathbb{R}$ is an arbitrary, but fixed and sufficiently smooth, non-linear function and $\varepsilon_{it} \sim \text{i.i.d.}(0,\sigma_\varepsilon^2)$. The idea is to construct a sophisticated average of Y_{it}'s around a fixed $x_{i_o t_o}$ as the estimator of F (.). This procedure puts more weight on observations Y_{it} corresponding to $\mathbf{X}_{it}$'s that is "closer" to x_{i_o,t_o} in the sense of Euclidean distance. The more local (global) the averaging, the more jagged (smoother) the estimated function. It is shown by Hardle (1990) that such a procedure, utilizing a weight function that satisfies certain regularity conditions, results in an estimator that converges to F (.) asymptotically in several ways. Define the multivariate weighting function as

$$\omega_{j,it}^{h} = \frac{\prod_{j=1}^{K} K_{j,h_j}(X_{j,it})}{\sum_{i,t}\prod_{j=1}^{K} K_{j,h_j}(X_{j,it})} \tag{4.2.2}$$

where $K_{j,h_j}(X_{i,it}) \geq 0 \ \forall$ i,j,t is the non-negative *smoothing kernel* for the j^{th} variable, which satisfies

$$\int_{\forall x} K_{j,h_j}(x)dx = 1 \qquad \forall\, j \tag{4.2.3}$$

The *bandwidth parameter* h_j controls the scale of the j^{th} kernel as follows:

$$K_{j,h_j}(x)dx = h_j^{-1}K\left(\frac{x}{h_j}\right) \tag{4.2.4}$$

where K (.) is some suitable function.[23] The *multivariate Nadaraya-Watson kernel estimator* (MNWKE) of the function F (.) is

$$\hat{F}(\mathbf{x}|\mathbf{h}) = \frac{\sum\limits_{\forall i,t}\prod\limits_{j=1}^{K} K_{j,h_j}\left(x_j - X_{j,it}\right)Y_{j,it}}{\sum\limits_{\forall i,t}\prod\limits_{j=1}^{K} K_{j,h_j}\left(x_j - X_{j,it}\right)}, \tag{4.2.5}$$

where $\mathbf{x} = \left(x_i,..,x_k\right)^T \in \mathbb{R}^k$ is a fixed vector of characteristics. The procedure for choosing the optimal vector of bandwidths, denoted by $\mathbf{h}^* = \left(h_1^*,...,h_k^*\right)^T$, involves minimizing the cross-validation function. Let the MNWKE, with observation Y_{j,i^*t^*} deleted, be defined by

$$\hat{F}_{-i^*t^*}(\mathbf{x}|\mathbf{h}) = \frac{\sum\limits_{\forall i,t \neq i^*j^*}\prod\limits_{j=1}^{K} K_{j,h_j}\left(x_j - X_{j,it}\right)Y_{j,it}}{\sum\limits_{\forall \neq i^*j^*}\prod\limits_{j=1}^{K} K_{j,h_j}\left(x_j - X_{j,it}\right)} = \sum_{\forall i,t \neq i^*,t^*} \omega_{j,it}^{h} Y_{j,it} \tag{4.2.6}$$

---

[23]A popular choice which we implement is the *Gaussian kernel* :

$$K_h(x) = (2\pi h)^{-\frac{1}{2}}\exp\left(-\frac{x^2}{2h^2}\right).$$

This has nothing to do with a normality assumption on the error term!

The *cross-validation function* for bandwidth $\mathbf{h}$ (denoted $CVF_{\mathbf{h}}$) is a weighted average of the squared differences between (4.2.6) evaluated at $\mathbf{x} = X_{j,it}$ and the $Y_{j,it}$:

$$CVF_{\mathbf{h}} = \sum_{\forall i,t} \left(F_{-it}\left(X_{j,it} \middle| \mathbf{h} \right) - Y_{j,it} \right)^2 \delta\left(X_{j,it} \right) \tag{4.2.7}$$

The function δ (.) is some appropriate weighting that mitigates boundary effects. The estimation of the optimal MNWKE, for a given data-set, reduces to minimizing (4.2.7) over all $\mathbf{h}$:

$$\hat{F}^{*}(\mathbf{x}|\mathbf{h}) = \hat{F}(\mathbf{x}|\mathbf{h}^{*}) \tag{4.2.8}$$

The optimization in (4.2.8) is subject to the following:

$$\mathbf{h}^{*} = \underset{\forall \mathbf{h} \in \Theta(\mathbb{R}^k)}{\operatorname{argmin}} \left\{ CVF_{\mathbf{h}} \right\} \tag{4.2.9}$$

5. CONCLUSION

In this chapter, we review the theoretical literature for the empirical tests of stochastic interest rate term structure models. First, we compare the fundamentally different models chosen according to well-accepted criteria. These criteria are realism in the stochastic process governing the term structure, agreement with finance theory and empirical evidence, and usefulness in financial practice. Second, we discuss the different approaches to the term structure. These include spot rate (or equilibrium) models (e.g., CIR (1985)), forward rate (or no-arbitrage) models (e.g., HJM (1992)), and non-parametric alternatives such as kernel estimators (Hanke (1999) and neural networks (Hutchingson et al (1994)). We also discuss variants of these models, depending upon the number of factors determining the short rate for the equilibrium models, or depending upon the number of volatilities of the forward rates for the no-arbitrage models. In the case of the spot rate models, we present the 3 factor affine term structure model (Duffee and Kan (1996); 3F-ATSM), the 2 factor CIR model of Longstaff et al (1992; 2F-LSTSM), the 2 factor short rate-spread model of Brennan et al (1984; 2F-BSTSM), and the single factor general parametric model of Duffie (1996; 1F-GPM). In the case of the forward rate models, we present the HJM constant and proportional volatility models for one through three factors.

In Chapters 3 and 4, we analyze of the term structure and interest rate derivatives based upon these models. In Chapter 3 we test international short-term interest rates in the framework of Chan et al (1992) as well as the US Treasury bond term structure using principal components analysis, neural networks, and an implementation of the 2F-LSTSM. In Chapter 4 we test the full range of spot and forward rate models on CBOT treasury bond options in the manner of Buhler et al (1998). We extend this by considering a three factor Guassian jump diffusion model (3F-GJDTSM) adaptation of Duffie et al (1998), the non-nested quadratic term structure model (Bolye et al 1999, Ahn et al 2000; QTSM), and the fully non-parametric multivariate kernel estimation model developed in this chapter.

CHAPTER 3

TERM STRUCTURE OF INTEREST RATE MODELS: INTERNATIONAL EMPIRICAL EVIDENCE

1. INTRODUCTION AND DISCUSSION

The purpose of this chapter is to perform an empirical analysis of the term structure that will contribute to two different aspects of the literature. One focuses upon the underlying theories of the term structure, using econometrics as a tool to draw market efficiency implications of them (e.g., Stambaugh (1987)). The second aspect centers its methodology on the design and evaluation of models that are capable of accurately pricing of interest rate dependent derivatives (e.g., Jamshidian (1989)). The link between these is the characterization of the stochastic process governing the evolution of the yield curve.

In the design and improvement of pricing and risk management models, we must consider both these traditional approaches. We can then rank models based on well-accepted criteria: financial market equilibrium (or the non-existence of arbitrage), economic theory, financial practice, computational efficiency, simplicity, and empirical facts. Consistency across these dimensions gives us more confidence in implementation, in case the assumptions underlying a particular model do not hold in a particular application. Such considerations also further our efforts to understand how market participants process information and formulate forecasts. It also helps us refine theoretical precepts from financial data. To do this, we must formulate a null hypothesis and make propositions amenable to empirical verification.

The relevance of the process followed by the short rate, and its implication for the term structure of interest rates, is that it has a direct bearing on the pricing of interest rate derivatives. Pricing models that incorporate *stochastic interest rates* or *stochastic volatility* include, but are not limited to, the works of Merton (1973), Vasicek (1977), Dothan (1978), Courtadon (1982), and Ball and Torous (1983). We may add to this the equilibrium approach of Cox et al (1985) as well as the no-arbitrage approach of Heath et al (1992). In this chapter, we extend this literature in several directions. First, we address unresolved issues in both the econometric and pricing literature by estimating the stochastic process followed by the short interest rate, and then estimating and applying term structure models, in a stochastic interest rate and stochastic volatility environment. We extend this analysis Japanese, UK, and Euro markets. Previous models are extended and compared to non-parametric alternatives. Finally, we analyze the term structure through hedging and forecasting analyses of interest rates and interest rate derivatives..

This chapter is organized as follows. Section 2 reviews some of the statistical tests used in this study, both before econometric estimation as well as to model errors. These include tests for normality, unit root/stability, autocorrelation, random volatility/heteroscedasticity, and long-memory. We also present diagnostic tests for short-term interest rates and US treasury bond yields. Section 3 concentrates on the estimation of various diffusion models of the short rate, utilizing popular econometric methodologies, and presents results on a cross section of these variables. Various parametric models are then compared to a non-parametric model of the short interest rate, in terms of explaining as well as forecasting interest rate movements. We extend the analysis to other global markets. Section 4 focuses on characterizing the term structure, under various assumptions both about the process followed by the short rate, as well as the modeling approach taken to price bonds in a stochastic interest rate/volatility environment. This involves examination of the pricing accuracy, the utility of the model in hedging unanticipated interest rate and volatility shifts, and the ability of the models to dynamically forecast yield changes. It also involves a comparison of parametric spot and forward rate models to their non-parametric counterparts.

2. SUMMARY STATISTICS AND DIAGNOSTIC TESTS

The purpose of this section is to statistically characterize the various data used in this study. Before presenting the results, we briefly summarize the statistical tests that we use. These fall into two categories: tests of the generating distribution and tests of the process properties. The first examines whether the sample is likely to come from a normal population. The second looks at more general characteristics of the data-stationarity/stability, serial correlation, long memory, and patterns in volatility, etc. Taken together, they give a comprehensive evaluation of the data, in terms of whether they are suitable for use in our econometric models. Even if the data does not follow all the standard statistical assumptions in a strict sense, we would still be comforted by knowing that there are no detectable systematic patterns that would bias our estimation.

The basic test of a normal distribution focuses on the first four sample moments of the data. Let $\mathbf{X}_T = (X_1,...,X_T)^T$ denote a time series sample of market data. Under the null hypothesis, these are independent realizations from a normal distribution with mean $E(X_t) = \mu$ and variance $Var(X_t) = E(X_t - E(X_t))^2 = \sigma^2$ (both constant), or

$$H_0: X_t \sim NID(\mu,\sigma) \quad t = 1,..,T \tag{2.1}$$

While it is possible to base tests of normality upon the mean and variance[1], due to their unobservability, such test statistics are seldom used in finance applications. However, the third and fourth central moments can be calculated, and used in conjunction with the

---

[1] Under this null, the normalized sample mean $\hat{\mu} \triangleq T^{-1}\sum_{t=1}^{T} X_t$ and sample standard deviation $\hat{\sigma} \triangleq \left[(T - 1)^{-1}\sum_{t=1}^{T}(X_t - \hat{\mu}^2)^2\right]^{\frac{1}{2}}$ have the following distributions:

$$Z_\mu \triangleq T\times\left(\frac{\hat{\mu} - \mu}{\sigma}\right) \underset{H_0}{\sim} N(0,1)$$

$$Z_\sigma \triangleq \left(\frac{\hat{\sigma} - \sigma}{\sqrt{T}}\right) \underset{H_0}{\sim} N(0,1)$$

mean and standard deviation, to produce a statistic with a known distribution. These moments are estimated by the sample skewness $\hat{s} \triangleq (T - 1)^{-1}\sum_{t=1}^{T}(X_t - \hat{\mu})^3$ and sample kurtosis $\hat{\kappa} \triangleq (T - 1)^{-1}\sum_{t=1}^{T}(X_t - \hat{\mu})^4$.[2] The Berra and Jarque *J-statistic*, a basic tests of normality, is given by the following chi-squared random variable with 2 degrees of freedom.[3] :

$$J \triangleq T\left[\frac{\hat{s}^2}{6} + \frac{(\hat{\kappa} - 3)^2}{24}\right] \underset{H_0}{\sim} X^2(2) \tag{2.2}$$

This is not the most powerful test, as it often fails to detect similar alternative distributions. We compute the Kolmogorov and Smirnov *D-Statistic*, from the sample cumulative distribution of the pre-whitened series, which is distributed i.i.d. with a zero mean under the null hypothesis of normality. We can then consider more general tests of

---

2 It can be shown that for a normal random variable:

$$\hat{\kappa} \underset{H_0}{\sim} N\left(3, \frac{24}{T}\right)$$

$$\hat{s} \underset{H_0}{\sim} N\left(0, \frac{6}{T}\right),$$

3 The sums-of-squares of independent standard normal random variables has a chi-squared distribution with degrees of freedom equal to the number of terms in the sum: $Z_t \sim N(0,1) \Rightarrow \sum_{t=1} Z_t^2 \sim X^2(T)$

random patterns in the data that require neither a normal distribution, restrictions on moments higher than the second, nor whether the series is stationary or trending (*unit root*). This test of first order dependence, or autocorrelation, imposes the *white noise* hypothesis:

$$H_0: \quad X_t \sim ID\left(0,\sigma_X^2\right)$$

$$E\left(X_t X_{t-k}\right) = 0 \quad \forall\, k \neq 0 \tag{2.3}$$

Tests of this hypothesis are based upon the sums-of-squares of the L^{th} order sample auto-correlations of the series r_L. A commonly used test statistic used is the *Q-statistic* of Box and Pierce (1970)[4], which under the null hypothesis (2.3) is distributed as a chi-squared random variable, with degrees of freedom equal to the number of lags considered. A refinement due to Ljung and Box (1979), which has better small sample properties, is given by a declining weighted average of squared autocorrelations.[5] While these tests are

---

[4] This is given by

$$Q_{BP}(L) \triangleq T\sum_{j=1}^{L} r_j^2 \underset{H_0}{\sim} \chi^2(L) \qquad L = 1,\ldots,T-1.$$

[5] This is given by

$$Q_{LB}(L) \triangleq T(T-2)\sum_{j=1}^{L} \frac{r_j^2}{T - j} \qquad L = 1,\ldots,T-1.$$

less restrictive than the more traditional Durbin-Watson (1950) statistic, it can be criticized for the arbitrary specification of L. Furthermore, simulation studies have found this statistic to have low power against several hard-to-detect alternatives to the white noise hypothesis. Another characteristic of interest is stationarity, which has the property of a series having the same distributional characteristics over time. This is based on a simple first order autoregressive representation:

$$X_t = \gamma X_t + \varepsilon_t \tag{2.4}$$

In standard regression analysis, if $|\gamma|<1$, then the OLS estimator[6] has the property that $\sqrt{T}(\gamma - \hat{\gamma}) \xrightarrow{d} N(0,1 - \gamma^2)$. Under the unit root hypothesis $H_0{:}\gamma=1$, Dickey et al (1979) show that the estimator is *super-consistent*[7] for a random variable C, such that $T\hat{\gamma} \xrightarrow{d} (\mu_C,\sigma_C^2)$, where $E(C) = \mu_c < 1$ and $Var(C) = \sigma_c < \infty$. This has two implications:

---

[6] This is calculated as

$$\hat{\gamma} = \frac{\sum_{t=2}^{T} X_t X_{t-1}}{\sum_{t=1}^{T} X_t^2}$$

[7] That is, the speed of convergence is $O(T^{-2})$ vs. $O(T^{-1})$ in OLS, where $O(f(x))$ is the *asymptotic order* of a variable y with respect to x, and f(.) is some function. This means that as X gets very large, the variable in question is proportionate to f(x), or the ratio is bounded by a constant, $\frac{y}{O(f(x))} = c < \infty$.

the OLS estimator is biased below one, since it approaches a random variable with such an expectation, and the calculated standard errors are small, given that the speed of convergence is fast. This means that we are more likely to reject the unit root hypothesis. Through Monte Carlo simulation, Dickey et al (1979) develop a set of correct standard errors. The Dickey-Fuller (DF) *Z-statistic* is given by:

$$Z_{DF} = \frac{1 - \hat{\gamma}}{\sigma_{DF}} \tag{2.5}$$

where σ_{DF} is the standard error calculated from simulating a large number of unit-root regressions of varying sample sizes.. We use the Phillips and Perron (1982) statistic, Z_{PP}, which is robust to a general form of heteroscedasticity. A related statistical property that we examine is *long memory.* Some processes may seem uncorrelated upon casual analysis, but tend to have subtle patterns of dependence that represent deviations from randomness. This is termed "long-wave" (or "low-frequency") dependence, in that one may have to observe the data for longer periods in order to detect this. The basis for this model is Granger and Joyeaux (1980), who introduce the Autoregressive Fractionally-Integrated Autoregressive Process (ARFIMA) of order (p,d,q):

$$\Phi_p(L)(1 - L)^d X_t = \Theta_q(L)\varepsilon_t \tag{2.6}$$

Where $d \in (0,1)$ is the differencing parameter, $L^d X_t = X_{t-d}$ is the d^{th} order lag operator,

$\varepsilon_t \sim iid(0,\sigma_\varepsilon^2)$ is a random shock and Θ_p,Φ_q are lag polynomials of orders p and q[8]. If the stationary stochastic process defined by $\{X_t\}_{t=0}^{\infty}$ is invertible, then it has the infinite order moving representation:

$$X_t = H_\infty(L)\varepsilon_t \tag{2.7}$$

Where $H_\infty(L) = (1 - L)^{-d}\Phi_p^{-1}(L)\Theta_q(L) \triangleq \sum_{i=0}^{\infty} h_i L^i$. We consider a simple log price process (p=q=1), so that differencing d times leads to stationarity. This is known as a fractional noise process (Mandelbrot and Van Nees, 1968). If d=1, y_t~I(1) (integrated of order 1: random walk/unit root); when d=0, y_t~I(0) (integrated of order 0: white noise). The autocorrelations of a white noise (random walk) are zero at all lags (i.e., they never die down), whereas for a fractionally differenced process, it can be shown that the autocorrelation coefficients are given by the infinite approximating expansion (Hosking, 1981),

$$(1 - L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k - d)L^d}{\Gamma(k + 1)\Gamma(- d)} \tag{2.8}$$

Where Γ(X)=(X-1)! is the standard Gamma function. For a large number of lags k, it can be shown that the autocorrelations behave approximately as k^{2d-1}, which is suggestive of a slow hyperbolic decay (Hosking, 1981). To further illustrate this point, note that

---

8 This is written as $A_r(L) = \sum_{i=0}^{r} \alpha_i L^i$.

Hamilton (1994) has shown that for a fractional white noise process (2.6) and large k the impulse response coefficients behave approximately like $h_j \cong (j+1)^{d-1}$, where h_j is the j^{th} lag coefficient in the MA (∞) representation of X_t (which exists by stationarity and invertibility). For d=1, the h_j=1 all j, X_t is a non-stationary random walk; for d=0, h_j = 1/1+j, and has autocorrelations with the geometric decay of a stationary process; for 0< d<1, these coefficients die out hyperbolically at a rate $\left(\frac{1}{j+1}\right)^{1-d}$, slower than the geometric decay of a stationary ARMA process. Hamilton (1994) shows that a fractionally integrated process is not covariance stationary unless $d < \frac{1}{2}$. The estimation procedure for the parameter d is motivated by an examination of the *spectral density* of X_t :

$$f_X(\omega) = \frac{1}{\pi}\sin^{-2d}\left(\frac{\omega}{2}\right)\sum_{k=-\infty}^{\infty}\gamma_k^u e^{-i\omega k} \tag{2.9}$$

where $u_t \triangleq (1 - d)^{-d}\Phi_p^{-1}(L)\Theta_q(L)\varepsilon_t$ is a stationary process, $i \triangleq \sqrt{-1}$ is the imaginary unit, and $\gamma_k^u = \mathrm{Cov}(u_t,u_{t+k})$. The meaning of this can be gleaned from a Taylor series expansion around $\omega = 0$, which reveals that $f_X(\omega) \cong \omega^{-2d}$ for small ω, showing that the spectral density to be unbounded at low frequencies, and hence the variation in the process is dominated by long wavelengths. Geweke and Porter-Hudak (1983; GPH) examine the estimation of a fractionally differenced process and estimate d from the frequency domain regression:

$$\log\left(\hat{f}_X(\omega_j)\right) = \beta_0 + \beta_1 \log\left(\sin^2\left(\frac{\omega_j}{2}\right)\right) + \eta_j \qquad j = 1,..,T-1 \tag{2.10}$$

where $\omega_j \triangleq \frac{2\pi}{j}$, β_1=-d, $\hat{f}_X(\omega) = \frac{1}{T\pi}\sum_{k=1}^{T-1} \hat{\rho}_k^X e^{-i\omega_j k}$ is the sample periodogram, an estimator of the population spectral density at frequency ω_j (which can be recognized as the exact finite Fourier transform of the series $\{X_1,...,X_T\}$ scaled by 2/T at each harmonic frequency), and $\hat{\rho}_k^X = \left[(T-1)\hat{\rho}_0\right]^{-1}\sum_{t=1}^{T-k} X_t X_{t+k}$ are the sample autocorrelations of the series. The integral of the population spectrum gives the variance of a series attributable to a particular sampling frequency. A random walk (white noise) has a linearly declining (flat) spectrum. Advantages of the GPH approach are that it does not require time series identification of the orders of the process and robustness to heteroscedasticity, autocorrelation, and regime shifts in the process.

The final statistic calculated is a measure of changes in the variance of the process, a phenomenon previously termed heteroscedasticity, random volatility, or "GARCH effects". The underlying *generalized autoregressive conditional heteroscedasticity model* (GARCH) is based on Bollerslev (1987). The model is summarized as follows:

$$\varepsilon_t = \sqrt{h_t}u_t \tag{2.11}$$

$$u_t \sim NID(0,1) \tag{2.12}$$

$$h_t = h_0 + \pi(L)u_t^2 \tag{2.13}$$

$$\pi(L) = \frac{\sum_{i=0}^{m} \alpha_i L^i}{1 - \sum_{i=0}^{r} \delta_i L^i} \tag{2.14}$$

where ε_t is the heteroscedastic error term, h_t is the GARCH conditional variance, $\pi(L)$ is a rational lag polynomial, L is a lag operator such that $L^k X_t = X_{t-k.}$ and u_t is a standard normal error term. This implies that the variance is:

$$h_t = h_0 + \sum_{i=0}^{r} \delta_i h_{t-i} + \sum_{i=0}^{m} \alpha_i u_{t-i}^2 \tag{2.15}$$

Equation (2.15) shows the conditional variance to have an unconditional autoregressive component (to capture persistence or mean reversion in variance), and a moving average component (to measure the impact of mean equation innovations on variance). Since it can be shown that a low order GARCH(r,m) process can be reasonably approximated by a high order GARCH(0,m) process, we use the test statistic of Engle (1982). The first step involves calculating an OLS regression and collecting residuals which proxy for the u_t process. Then we run the second stage regression:

$$\hat{u}_t^2 = a_0 + \sum_{i=0}^{m} a_i \hat{u}_{t-i}^2 + e_t \qquad (2.16)$$

The Engle's T×R² test statistic, which under the null hypothesis of homoscedasticity has an asymptotic chi-squared distribution with m degrees of freedom, is given by

$$T \times R_U^2 \sim \chi^2(m) \qquad (2.17)$$

where the uncentered coefficient of determination is given by

$$R_U^2 = \frac{\sum_{t=1}^{T} \left(\hat{\alpha}_0 + \sum_{i=1}^{m} \hat{a}_i \hat{u}_{t-i}^2 \right)^2}{\sum_{t=1}^{T} \hat{u}_t^2} \qquad (2.18)$$

2.1 Short Term Interest Rates

As a proxy for the instantaneously compounded interest rate, we take the yield-to-maturity on pure short-term money market instruments. These are the one month T-bill rates (1MTB), the weekly federal funds rates (1WFF), the three month Libor Dollar rates (3MLD), the one month Japanese government bond yields (1MJGB), and the three month Euro-sterling rate (3MES). In the case of 1MTB, we utilize an extended and updated version of the dataset used by CKLS (1992). These are annualized yields based on the average of bid and ask spreads for Treasury bills. The sample period for the 1MTB short rate process runs from December 1964 to October1997, for a total of 391 observations. Table 2.1.1 presents the distributional statistics for various short rates. The mean-of the 1MTB rate over the period 12/64-10/1997 is 6.25% (0.041 bps for the first difference), with a standard deviation of 2.5% (75.1 bps for the first difference). This series exhibits marked non-normality, with respective excess skewness and kurtosis of 1.2803 and 1.8112, significant at the 1% level. This is reflected in a large J-statistic of 160.27, far above the 1% critical value of 9.21 for a $\chi^2(2)$ random variable. The results are similar and more pronounced in the case of the first difference, with a J-statistic of 1.99×10^3, the only difference being negative skewness. This is supported by the D-statistic in the case of the level, although not for the differenced series. There is high autocorrelation for both the levels and the differences of the 1MTB rate, with values of 4160.3 and 50.52, where the critical value for the first twenty autocorrelations is 37.6. The level and difference of the short rate exhibits neither a unit root nor long memory. Finally, rates exhibit statistically significant GARCH effects based on the Engle test, where the calculated $T \times R^2$ value of 98.6 far exceeds the 1% critical value of 15.1. The results other short rates, 1WFF rate and the 3MLD rate, from 66:5 to 96:20 and 90:1 to 99:12 respectively, are remarkably similar. While the 1WFF is about 100 bps higher in both mean and standard deviation, it displays similar skewness and kurtosis, and rejects as normality as well. However, based on the unit root tests, we reject stationarity in the undifferenced series in 1WFF. The 3MLD, sampled at daily frequencies, is 30 bps higher than the mean for 1MTB but has similar qualitative results. The 1MJGB exhibits only slightly different behavior from the U.S. market short rates. We find significant autocorrelation in both levels and differences, a unit root (stationarity) in the levels (differences), no long memory and significant GARCH effects. The Japanese rates have a lower mean of 4.71% as well as substantially lower skewness and kurtosis. We fail to reject normality by the J statistic of 1.92 for the 1MJGB, although the KS statistic rejects this null. The first differences, however, strongly reject normality by

TABLE 2.1.1: Distributional Statistics and Diagnostic Tests on Short Rates

Series	1 Month Treasury Bills	1 Week Federal Funds	3 Month Libor Dollar	1 Month Japan Govt.	3 Month Bond Euro-Sterling
Period	64:12-97:10	66:05-96:20	90:1-99:12	78:11-99:02	75:Q1-99:Q6
Count	390	1576	2480	244	97
Mean					
Level	0.06253^{c}	0.07453^{c}	0.0658^{c}	0.04716^{c}	0.09816^{c}
Δ	4.10×10^{-6}	3.98×10^{-6}	7.55×10^{-5}	1.76×10^{-4}	7.86×10^{-4}
Standard Deviation					
Level	0.02502	0.03257	0.02658	0.02710	0.02935
Δ	7.51×10^{-3}	3.97×10^{-3}	4.22×10^{-4}	0.00371	0.01109
Skewness					
Level	1.28031^{c}	1.25206^{c}	0.28813^{c}	0.01954	0.09844
Δ	-1.07782^{c}	0.27731	-3.02427^{c}	0.33203^{c}	-2.52940
Kurtosis					
Level	1.81122^{c}	1.90097^{b}	-1.3900^{c}	-0.19055	-0.06894
Δ	10.8762^{c}	9.22495^{c}	76.290^{c}	13.727^{c}	-0.09327
Kolmogorov-Smirnov					
Level	0.9886^{c}	0.9745^{c}	0.9473^{c}	0.9409^{c}	0.0450
Δ	0.0157	0.0871^{c}	0.2921^{c}	0.1902^{c}	-0.0293
Berra-JarqueWald Test					
Level	160.27^{c}	649.07^{c}	125.52^{c}	1.9225	6.22^{b}
Δ	1997.23^{c}	5604.87^{c}	233.96^{c}	1.98×10^{3c}	3.6728^{c}
Ljung-Box Q					
Level	4160.25^{c}	8710.24^{c}	8710.25^{c}	3.73×10^{3c}	588.74^{c}
Δ	50.52^{c}	225.81	225.80^{c}	71.172	41.62^{c}
Dickey-Fuller Z					
Level	-16.3891^{b}	-11.427	-4.63	-2.4317	-5.09
Δ	-413.36^{c}	-1994.9^{c}	-11.61	-215.12^{c}	-8.63
Phillips-Perron Z					
Level	-14.6769^{b}	-12.408	-10.95	-14.904	-9.08
Δ	-415.44^{c}	-1865.57^{c}	-335.818^{c}	314.93^{c}	-78.59^{c}
Gewke-Porter-Hudak					
Level	1.12734	0.90320	1.02735	1.08220	0.69398
Δ	0.10299	-0.09609	-0.04927	0.21966	-0.61065^{b}
Engle's $T\times R^2$					
	9.86×10^{4c}	3.53×10^{5c}	8.57×10^{5c}	28.542^{c}	6720.0

a, b, and c represents statistical significance at the 1%, 5%, and 10% levels, respectively.

both tests. Both tests, however, reject normality for the first differences. That the differences exhibit significant positive rather than negative skewness, possibly due to the lower level of this rate.

The final short-rate series, the 3MES, exhibits substantially different behavior. In this case, we fail to reject normality by the non-parametric KS statistic and the less powerful J statistic, and clearly reject stationarity, in both means and differences. Furthermore, the GPH statistics are indicative of possible long memory, and characteristics of autocorrelation as well as GARCH are present. These variations from the general trend may be driven by the quarterly sampling for this series.

We summarize these results as follows:

1. Generally, we reject unconditional normality for the short rate across different markets, time periods, and sampling frequencies. This is a supporting factor in the use of a continuous time framework, which does not impose normality in the error structure.
2. Short rates exhibit unit roots in levels and stationarity in differences, autocorrelation in both, as well as GARCH effects. This motivates us to use generalized econometric approaches such as generalized method of moments (GMM) and kernel regression, as opposed to linear structural models.

2.2 Default Free Discount Bonds

Distributional statistics and diagnostic tests for the U.S. government bond discount term structure for the period 6/1964 to 10/1997 are presented in Table 2.2.1. Rates for 3,6, and 9 months and 1 to 5 years are expressed in annualized, continuously compounded form. The mean level ranges from 6.25% for 3 months to 7.58% for 5 years, so the average yield curve is consistent with theories such as the liquidity preference and local expectations hypothesis. However, the yield volatilities range from 2.38% to 2.63% and peak at the 1-year maturity. All maturities exhibit excess skewness and kurtosis in level, which results in strong rejection of normality for all maturities by both the B-J and K-S statistics. Q-statistics exhibit autocorrelation in all cases, which strengthens with maturity. The Z and GPH statistics support stationarity and no long memory in all cases. Yield changes have qualitatively similar results, except that skewness is negative instead of positive, as is the case with the levels. All maturities exhibit evidence of GARCH effects. Figure 2.2.1 graphically represents shifts in the term structure over the time period and the stochastic nature and co-dependence of the yield curve are evident. Table 2.2.2 presents the unconditional correlation matrix. As expected, there is decreasing correlation with increased difference in maturity, with the decrease being largest for closest maturity, a phenomenon called *exponential decorrelation* in the term structure literature (Rebonato, 1996).

TABLE 2.2.1
Distributional Statistics and Diagnostic Tests on U.S. T-Bond Annualized Continuously Compounded Yields
3 Months to 5 Years Maturities (6:64-10:97)

	3 Months	6 Months	9 Months	1 Year	2 Years	3 Years	4 Years	5 Years
Mean								
Level	0.0626[a]	0.0660[a]	0.0683[a]	0.0697[a]	0.0708[a]	0.0730[a]	0.0745[a]	0.0758[a]
Δ	4.3×10^{-5}	4.1×10^{-5}	4.4×10^{-5}	3.8×10^{-5}	4.08×10^{-5}	4.4×10^{-5}	4.4×10^{-5}	4.5×10^{-4}
Standard Deviation								
Level	0.0258	0.0266	0.0265	0.0263	0.0258	0.0250	0.0242	0.0238
Δ	0.0075	0.0062	0.0069	0.0059	0.0059	0.0051	0.0051	0.0045
Skewness								
Level	1.2764[a]	1.2569[a]	1.4567[a]	1.1426[a]	1.0778[a]	1.0235[a]	1.0235[a]	0.9712[a]
Δ	-1.0775[a]	-1.6983[a]	-1.6983[a]	-0.9730[a]	-1.0075[a]	-0.6922[a]	-0.6922[a]	-0.17687
Kurtosis								
Level	1.7982[a]	1.6653[a]	1.6653[a]	1.2805[a]	1.04132[a]	0.78345[a]	0.74216[a]	0.61903[a]
Δ	10.843[a]	12.697[a]	1.457[a]	10.517[a]	11.164[a]	8.237[a]	8.2369[a]	3.7834[a]
Kolmogorov-Smirnov								
Level	0.9440[a]	0.9476[a]	0.9490[a]	0.9502[a]	0.9516[a]	0.9570[a]	0.9606[a]	0.9627[a]
Δ	0.1063[a]	0.1478[a]	0.1478[a]	0.1639[a]	0.1223[a]	0.1469[a]	0.1198[a]	0.0855[a]
Berra-JarqueWald Test								
Levell	58.44[a]	147.75[a]	128.04[a]	111.51[a]	93.285[a]	78.062[a]	73.999[a]	67.539[a]
Δ	1981.04[a]	2800.2[a]	2887.82[a]	1854.00[a]	2086.10[a]	1130.73[a]	536.28[a]	234.04[a]
Ljung-Box Q								
Level	4147.9[a]	4414.2[a]	4527.0[a]	4534.0[a]	4572.7[a]	4871.5[a]	5057.7[a]	5146.2[a]
Δ	50.31[a]	79.56[a]	81.69[a]	73.54[a]	68.94[a]	55.39[a]	37.79[a]	40.67[a]
Phillips-Perron Z								
Level	-14.667[a]	-11.753[a]	-11.184[a]	-11.09[a]	-11.05[a]	-9.375[a]	-8.296[a]	-7.695[a]
Δ	-414.35[a]	-316.15[a]	-292.55[a]	-285.40[a]	-303.69[a]	-289.25[a]	-304.58[a]	-329.4[a]
Gewke-Porter-Hudak								
Level	1.1205	1.3027	1.2300	1.1776	1.1720	1.1832	1.1581	1.1292
Δ	0.0944	0.2310	0.2071	0.1548	0.1372	0.1386	0.1598	0.1852
Engle's $T\times R^2$								
	6.9×10^{4}[c]	3.5×10^{5}[c]	7.5×10^{5}[c]	9.6×10^{5}[c]	3.5×10^{5}[c]	7.8×10^{5}[c]	9.9×10^{5}[c]	9.6×10^{5}[c]

a, b, and c represents statistical significance at the 1%, 5%, and 10% levels, respectively.

Figure 2.2.1: Treasury Yield Curves

(64:6-97:10)

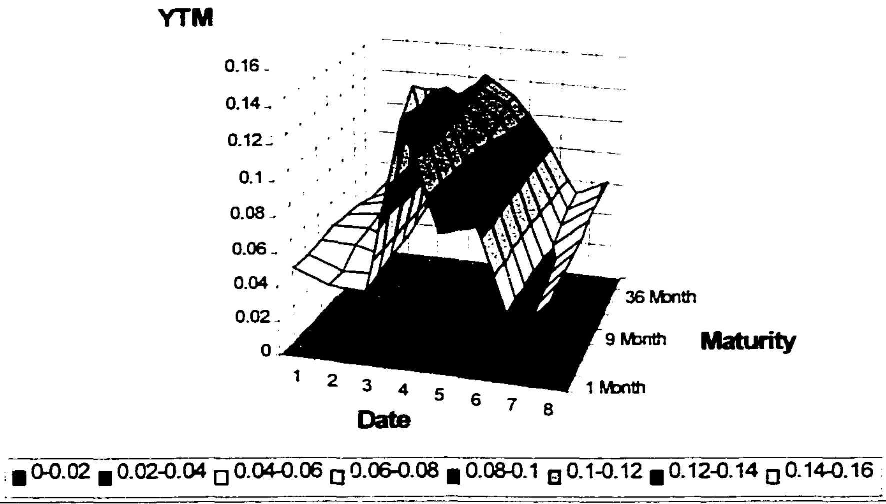


Table 2.2.2: Yield Change Correlation Matrix (US T-Bond Term Structure 6/64-10/97)

Month	1	3	6	9	12	24	36	48	60
1	1.0000	0.7680	0.6755	0.6174	0.6244	0.5331	0.5056	0.4463	0.4421
3	0.7680	1.0000	0.9466	0.8997	0.8702	0.7901	0.7410	0.6634	0.6616
6	0.6755	0.9466	1.0000	0.9672	0.9398	0.8826	0.8275	0.7577	0.7550
9	0.6174	0.8997	0.9672	1.0000	0.9684	0.9350	0.8886	0.8279	0.8223
12	0.6244	0.8702	0.9398	0.9684	1.0000	0.9447	0.9107	0.8590	0.8600
24	0.5331	0.7901	0.8826	0.9350	0.9447	1.0000	0.9616	0.9319	0.9297
36	0.5056	0.7410	0.8275	0.8886	0.9107	0.9616	1.0000	0.9573	0.9543
48	0.4463	0.6634	0.7577	0.8279	0.8590	0.9319	0.9573	1.0000	0.9637
60	0.4421	0.6616	0.7550	0.8223	0.8600	0.9297	0.9543	0.9637	1.0000

3. TESTS OF ALTERNATIVE SHORT RATE PROCESSES

3.1 Alternative Econometric Methodologies

This section focuses on strategies for consistent and efficient estimation of the structural parameters of the parametric term structure models to be reviewed in Section 3.2. This is of importance for option pricing and hedging, in that different models as well as estimation techniques lead to different conclusions about which stochastic process is most likely to characterize the term structure. We formalize this estimation algorithm by stating the null hypothesis that the parametric restrictions to be made in the subsequent sections are true, which is expressed as:

$$\begin{aligned} &H_0 : \exists \boldsymbol{\theta}_0 \in \Theta / \left\{ \mu_0(.,\boldsymbol{\theta}), \sigma_0^2(.,\boldsymbol{\theta}) \right\} \in P(\Theta) \\ &H_1 : \left\{ \mu_0(.,\boldsymbol{\theta}), \sigma_0^2(.,\boldsymbol{\theta}) \right\} \notin P(\Theta) \end{aligned} \tag{3.1.1}$$

where $\boldsymbol{\theta} \in \Theta \subset \bar{\mathbb{R}}^k / \theta < \infty$ is the parameter vector, assumed to exist in a compact sub-space of k-dimensional reals, and $P(\Theta)$ is a joint parametric family of functions. The null hypothesis proposes that there exist parameter values such that the parametric model is a reasonable representation of the process. For example, in our general parametric model, the parametric class is given by the set of continuous drift and diffusion functions $\mathbf{P} \equiv \left\{ \left| \mu_0(.,\boldsymbol{\theta}), \sigma_0^2(.,\boldsymbol{\theta}) \right| = \left[\alpha_1 + \alpha_2 r_t + \alpha_3 r_t \log(r_t), (\beta_1 + \beta_2 r_t)^{\nu} \right] \right\} / (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2 \nu) \in \Theta$. To state that the true parameter vector resides in this space, making it in principle an estimable quantity, is to say that the true drift and diffusion functions reside in the parametric space

P(.). Although this may be a conceptually straightforward exercise, testing this hypothesis encounters several difficulties. Estimation of these functions is difficult because they are in reality continuous mappings on the state space and time, and most estimation techniques rely on discretization of the continuous process. The true first and second moments of the data as calculated over discrete intervals are not given by these $\mu(.,\boldsymbol{\theta}),\sigma^2(.,\boldsymbol{\theta})$ functions, and as an approximation, this procedure is valid only as the length of the measurement interval vanishes. For instance, it can be shown that the conditional mean over the observation period depends on *both* the drift and diffusion functions, even in the simple case of a linear drift. This is known as the problem of *aggregation bias*.[9] This cannot be ameliorated by simply collecting data more frequently, in that as we approach the continuous time limit with transactions data, microstructural biases become an issue. These include the problems of price discreteness, bid-ask spreads, as well as non-synchronous trading.

The approaches to this problem fall into three broad categories. The most general approach involves moment estimation of a difference equation approximation to the underlying SDE (Chan et al (1992)). A more efficient algorithm is the estimation of discretely sampled data, implemented by maximizing a Gaussian likelihood function, even when the stochastic process is not itself Gaussian (Bergstrom (1983), Nowman

---

9

See Melino (1990) for the empirical finance literature on the estimation of continuous time models.

(1997)). The semi-parametric approach minimizes a distance criterion that depends on parametric and non-parametric marginal densities, thereby deriving consistent estimators even when the underlying stochastic process is mis-specified and where the diffusion function is estimated non-parametrically for a parametrically specified drift (Ait-Sahalia (1996)). Finally, Jiang (1998) has developed a purely non-parametric approach that relies on the conditional density of interest rate.

3.1.1 The Parametric GMM Approach

Here we will describe a simple approach to estimate the parameters of the term structure SDEs. We start with the unrestricted model. Following Brennan and Schwartz (1982), Dietrich-Campbell and Schwartz (1986) and Chan et al (1992), we approximate the SDE of the Single Factor General Parametric Model (1F-GPM) by a discrete system of difference equations, known as an Euler discretization. The econometric model is:

$$r_t - r_{t-1} = \alpha_1 + \alpha_2 r_{t-1} + \alpha_3 r_{t-1}\log(r_{t-1}) + \varepsilon_t \qquad (3.1.1.1)$$

$$E[\varepsilon_t] = E[\varepsilon_t | r_{t-1}] = 0 \qquad (3.1.1.2)$$

$$E[\varepsilon_t^2] = (\beta_1 + \beta_2 r_t)^{\gamma} \qquad (3.1.1.3)$$

$$E\left[\varepsilon_t^2 - \left(\beta_1 + \beta_2 r_t\right)^\gamma \Big| r_{t-1}\right] = E\left[\left(\varepsilon_t^2 - \left(\beta_1 + \beta_2 r_t\right)^\gamma\right) r_{t-1}\right] = 0 \qquad (3.1.1.4)$$

Equation (3.1.1.1) is a stochastic difference equation (or an *Euler discretization*) approximation to the drift function of the SDE followed by the short rate in the continuous time model, the right-hand-side terms being the expected rate change given information at time t-1, or $E\left[\Delta r_t \middle| r_{t-1}\right] = \alpha_1 + \alpha_2 r_{t-1} + \alpha_3 r_{t-1}\log\left(r_{t-1}\right)$. We interpret α_2 as the coefficient of mean reversion, measuring the speed with which the interest rate approaches $-\frac{\alpha_1}{\alpha_2}$, the long-run unconditional mean. The logarithmic term is intended to capture any non-linearities in the drift function. Equation (3.1.1.2) defines the innovation process, which proxies for small increments to a standard Weiner process in continuous time. In a discrete model, we only make the distributional requirement that the error terms have an unconditional mean of zero and are orthogonal to lags of the interest rate process. This implies that we cannot use information contained in past r_t to forecast the future, consistent with market efficiency. Equation (3.1.1.3) is the approximation to the diffusion function, constructed in the manner of a generalization of both the CEV and the displaced CIR specification. This specification admits conditional heteroscedasticity, an empirical as well as theoretical fact in interest rate processes of all kinds. The orthogonality condition (3.1.1.4) requires the error in forecasting the second moment of the process at time s > t-1 to be unpredictable, given the interest rate at time t-1, in analogy to the condition (3.1.1.2) for the first moment.

The econometric technique involves the Generalized Method of Moments (Hansen (1982)), which tests a set of overidentifying restrictions on the specification equations (3.1.1.1)-(3.1.1.4). The advantage in using this technique is that the distribution of interest rate changes is left unspecified, given that the true distribution may not be normal. For example, even though this distribution is normal in the Vasicek and Merton-Ho-Lee models, under CIR rate changes are distributed proportionately to a non-central χ^2 random variable. Even if we correctly specify the continuous time distribution, the temporal aggregation phenomenon makes the distribution of the discretely sampled data unknown. The GMM estimation is implemented by collecting the moment conditions into a k-vector with zero expectation:

$$\mathbf{f}_t(\boldsymbol{\theta}|r_{t-1}) = \begin{pmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+1}^2 - E\left[\varepsilon_{t+1}^2 | r_t\right] \end{pmatrix} \otimes \boldsymbol{\Pi}(r_t) \quad t=1,..,T \tag{3.1.1.5}$$

where $\otimes$ denotes the Kronecker product of two vectors, $\boldsymbol{\Pi}$ is a vector of k+1 transformations of r_t (with $\pi_0(r_t) \triangleq 1$ and $\pi_1(r_t) \triangleq r_t$), and $\boldsymbol{\theta} = (\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2,\nu)^T \in \Theta(\boldsymbol{\theta}) \subset \bar{C}_0(\mathbb{R}^6)$ is the parameter vector (that lives in a compact Euclidean subspace). The GMM estimator is given by solution to the following quadratic form:

$$\hat{\boldsymbol{\theta}}_T = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\left[\mathbf{g}_T(\boldsymbol{\theta}|\mathfrak{R}_T)^T \mathbf{S}^{-1} \mathbf{g}_T(\boldsymbol{\theta}|\mathfrak{R}_T)\right] \tag{3.1.1.6}$$

The vector $\mathbf{g}_T(\boldsymbol{\theta}|\mathfrak{R}_T) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{f}_t(\boldsymbol{\theta}|r_{t-1})$ is the sample average of the moment conditions, which should converge in probability to zero, and $\mathbf{S} = \lim_{T\to\infty} T\left[\mathbf{g}_T\left(\boldsymbol{\theta}_0\middle|\mathfrak{R}_T\right)\mathbf{g}_T\left(\boldsymbol{\theta}_0\middle|\mathfrak{R}_T\right)^T\right]$ is a weighting matrix, which is the asymptotic variance of the moment vector.[10]

3.1.2 The Non-parametric Kernel Approach

In this section, we outline the general theoretical framework for a single factor term-structure and then show the econometric methodology that is applied to estimate drift and diffusion functions, with no parametric restrictions on the SDE governing the evolution of the short rate.

Assumption 3.1.2.1: The short rate process is given by a *time-homogenous diffusion process*:

$$dr_t = \mu(r_t)dt + \sigma(r_t)dB_t \tag{3.1.2.1}$$

We impose the initial condition $r_t\big|_{t=t_0} = r_0 \in \mathbb{R}^{++} < \infty$, where $B_s - B_t \sim NID(0, s-t)\ \forall\ 0 \le t < s < T < \infty$ is a standard Brownian motion on the real line. The respective drift and diffusion functions $\mu:\mathbb{R}^{++}\to\mathbb{R}$, $\sigma:\mathbb{R}^{++}\to\mathbb{R}^{+}$ satisfy regularity conditions such that there exists a *strong solution* r_t by applying Ito's formula to the SDE (3.1.2.1). The underlying process is a *regular Markov process*. The requirement of time-homogeneity means that the drift and

---

[10] In order to derive the weighting matrix $\mathbf{S}^{-1}$, we may use the sample variance-covariance matrix of the T observations on the r innovations $\hat{\mathbf{S}}_T = \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{h}\left(\tilde{\boldsymbol{\theta}}_T, r_t\right)\mathbf{h}\left(\tilde{\boldsymbol{\theta}}_T, r_t\right)^T\right]$, evaluated at any consistent estimator $\tilde{\boldsymbol{\theta}}_T$ estimator.

diffusion functions in (3.1.2.1) do not depend upon calendar time, so that the process can be translated by any interval without changing its local stochastic properties, namely its instantaneous expected return and volatility. This is analogous to the assumption of constant first and second moments in the case of discrete time and state random processes. Karlin and Taylor (1981) have shown that, under these conditions, the transition density function is the fundamental solution to *Kolmogorov backward equation*:

$$\frac{1}{2}\sigma^2(y)\frac{\partial^2 p\left(r_t = x \middle| r_\tau = y\right)}{\partial y^2} + \mu(y)\frac{\partial p\left(r_t = x \middle| r_\tau = y\right)}{\partial y} = -\frac{\partial p\left(r_t = x \middle| r_\tau = y\right)}{\partial \tau} \quad (3.1.2.2)$$

Where $p\left(r_t = x \middle| r_\tau = y\right)$ is the conditional probability density function for r_t at time $t > \tau$, conditional on taking on the value y at time τ, subject to $p\left(r_\tau = x \middle| r_t = y\right) = \delta(x - y)$ (the *Dirac delta function*).[11] Equation (3.1.2.2) provides the dynamics of the conditional density, stating that its rate of change is proportional to its derivatives with respect to the

---

11 The Dirac delta is a unit point density, a continuous function of x (parameterized by y), with the property

$$\int_{-\infty}^{\infty} f(y)\delta(x - y)dy = f(x).$$

state variable, the constants of proportionality being the drift and diffusion coefficients of the state. The transition density also satisfies the Kolmogorov forward (or Fokker-Planck) equation[12]:

$$\frac{1}{2}\frac{\partial^2\left[\sigma^2(x)p\left(r_t = x \mid r_\tau = y\right)\right]}{\partial x^2} + \frac{\partial\left[\mu(y)p\left(r_t = x \mid r_\tau = y\right)\right]}{\partial x} = \frac{\partial p\left(r_t = x \mid r_\tau = y\right)}{\partial t} \quad (3.1.2.3)$$

The implications of (3.1.2.2) and (3.1.2.3) are that the conditional or dynamic properties of the Markov underlying process are completely characterized by the coefficient functions of the SDE (3.1.2.1).

<u>Definition 3.1.2.1</u>

A stochastic process is *strictly stationary* if and only if there exists an initial probability density $p\left(r_{t_0}\right)$ with the property that for any $0 \le t_0 < t < T < \infty$:

$$p\left(r_t = x\right) = \int_{\mathbb{R}^{++}} p\left(r_t = x \mid r_{t_0} = u\right) p\left(r_{t_0} = u\right) du = p\left(r_{t_0} = x\right) \quad (3.1.2.4)$$

There exists an unconditional density that does not depend on calender time, which is analogous to the time homogeneity of the underlying stochastic process. This implies that the drift and diffusion functions uniquely determine this stationary density. If a time homogeneous diffusion of the type (3.1.2.1) is strictly stationary, in the sense of (3.1.2.4),

[12] See Karlin and Taylor (1981) for an exposition of these facts.

then its stationary density is

$$p(r_t) = \frac{K}{\sigma^2(r_t)} \exp\left[2 \int_{u=r_{t_0}}^{r_t} \frac{\mu(u)}{\sigma^2(u)} du \right] \tag{3.1.2.5}$$

where K is a normalizing constant solving

$$K = \int_{v=r_{t_0}}^{\infty} \frac{1}{\sigma^2(r_t)} \exp\left[2 \int_{u=r_{t_0}}^{v} \frac{\mu(u)}{\sigma^2(u)} du \right] dv \tag{3.1.2.6}$$

Under the assumption of strict stationarity (3.1.2.4), the left hand side of the Kolmogorov forward equation (3.1.2.3) is zero. Then multiplication of both sides of the equation with the marginal density $p(r_t)$ and integration with respect to r_t, $t \in (0,+\infty)$, yields (3.1.2.5).[13] This illustrates that the marginal density, or the static properties of the underlying Markov process (3.1.2.1), are fully characterized by its drift and diffusion functions. It can be shown that under these assumptions there are restrictions on these coefficients, such that each depends functionally on the other, as well as the marginal density of the underlying stochastic process. From the Kolmogorov forward equation (3.1.2.3), the respective drift and diffusion of the stochastic process defined by equation (3.1.2.1) are

---

[13] This is subject to boundary conditions:

$$\lim_{r_t \uparrow \infty} p(r_t) = \lim_{r_t \uparrow \infty} \frac{d}{dr_t} p(r_t) = 0 \quad \lim_{r_t \uparrow \infty} \sigma^2(r_t) = \lim_{r_t \uparrow \infty} \frac{d}{dr_t} \sigma^2(r_t) = 0, r_{t_0} \in (0,\infty)$$

given by

$$\mu(r_t) = \frac{1}{2p(r_t)}\left\{\frac{d}{dr_t}\left[\sigma^2(r_t)p(r_t)\right]\right\} \tag{3.1.2.7}$$

$$\sigma^2(r_t) = \frac{2}{p(r_t)}\int_0^{r_t} \mu(r_u)p(r_u)du \tag{3.1.2.8}$$

Thus, given the marginal density of the short rate, the drift and diffusion functions mutually determine one another. This suggests that if we find robust estimators of either of these two, then the third is easily determined, and we have then fully characterized the process. Technical regularity conditions insure that a consistent *non-parametric kernel estimator* of the diffusion function is given by

$$\hat{\sigma}^2(r_t) = \frac{\sum_{i=1}^{n-1} K\left(\frac{r_{t_i} - r_t}{h_n}\right)\left(r_{t_{i+1}} - r_{t_i}\right)^2}{\Delta_n \sum_{i=1}^{n} K\left(\frac{r_{t_i} - r_t}{h_n}\right)} \tag{3.1.2.9}$$

Where $\left\{r_{t_i}: i=1,\ldots,n\right\}$ is a collection of n equi-spaced observations in the interval [0,T], such that $0 < t_1 < \cdots < t_i < t_{i+1} < \cdots < t_n < T, \Delta_n \triangleq \frac{T}{n}$ is the sampling interval. K(.) Is a positive kernel density function satisfying the requisite technical regularity conditions, and h_n is the window of the non-parametric estimator[14]. The conditions on the kernel and window are given by (Jiang 1998):

1. The kernel $K(x) \in \mathcal{L}^2(\mathbb{R})$ of order r, meaning that $\int_{x=-\infty}^{+\infty} x^j K(x)dx \begin{cases} = 0 & j = 1,..,r-1 \\ \in (0,\infty) & j = r \end{cases}$, is continuously differentiable to the rth degree on $\mathbb{R}$, $\int_{x=-\infty}^{+\infty} K(x)dx = 1$, & K (.) is symmetric around 0.
2. For diffusion estimation, the following limits hold with respect to the window h_n: $h_n \underset{\uparrow n}{\rightarrow} 0$, $nh_n \underset{\uparrow n}{\rightarrow} \infty$, $nh_n^5 \underset{\uparrow n}{\rightarrow} 0$, and $nh_n^3 \underset{\uparrow n}{\rightarrow} 0(\infty)$.

In practice, it is common to choose the Gaussian kernel, $K(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ for r = 2. The conditions above are sufficient to ensure that the bias in the diffusion function estimator is asymptotically negligible, as well as that the variance of this estimator goes to zero in large samples. It can be shown that the variance of $\hat{\sigma}(r_t)$ can be consistently estimated by

$$\hat{V}\left[\hat{\sigma}^2(r_t)\right] = \frac{\hat{\sigma}^2(r_t)}{\sum_{i=1}^{n} K\left(\frac{(r_{t_i} - r_t)}{h_n}\right)}. \tag{3.1.2.10}$$

Next, we consider the consistent estimation of the marginal density function of the short

[14] The larger (smaller) the window, the smoother (more jagged) is the non-parametric estimator of the diffusion function. See Jiang (1998) for results on the choice of this window.

rate. Obviously, since observed nominal rates are always positive (i.e., $r_t \underset{a.s.}{\geq} 0$), it follows that $p(r_t)$ must be estimated with this restriction on r_t. This can be stated mathematically as

$$p(r_t)\begin{cases} > 0 & r_t \in (0,\infty) \\ = 0 & r_t \leq 0 \end{cases} \tag{3.1.2.11}$$

The approach that we choose is estimation based on the augmented time series $\Re_n \triangleq \left(-r_{t_n}, -r_{t_{n-1}}, \ldots, -r_{t_1}, r_{t_1}, \ldots, r_{t_{n-1}}, r_{t_n}\right)$. This incorporates the restriction (3.1.2.11), without resorting to the use of boundary kernels, which can result in serious difficulties in implementation (Scott (1992)). Jiang (1998) shows that a consistent estimator of the short rate marginal density is given by

$$\hat{p}(r) = \frac{1}{nh_n}\sum_{i=1}^{n}\left[K\left(\frac{r_{t_i} - r}{h_n}\right) + K\left(-\frac{r_{t_i} - r}{h_n}\right)\right] \tag{3.1.2.12}$$

This is subject to the appropriate regularity conditions on the kernel function K(.). Given these consistent estimators of the diffusion function, it follows from equation (3.1.2.7) that the drift may be consistently estimated by

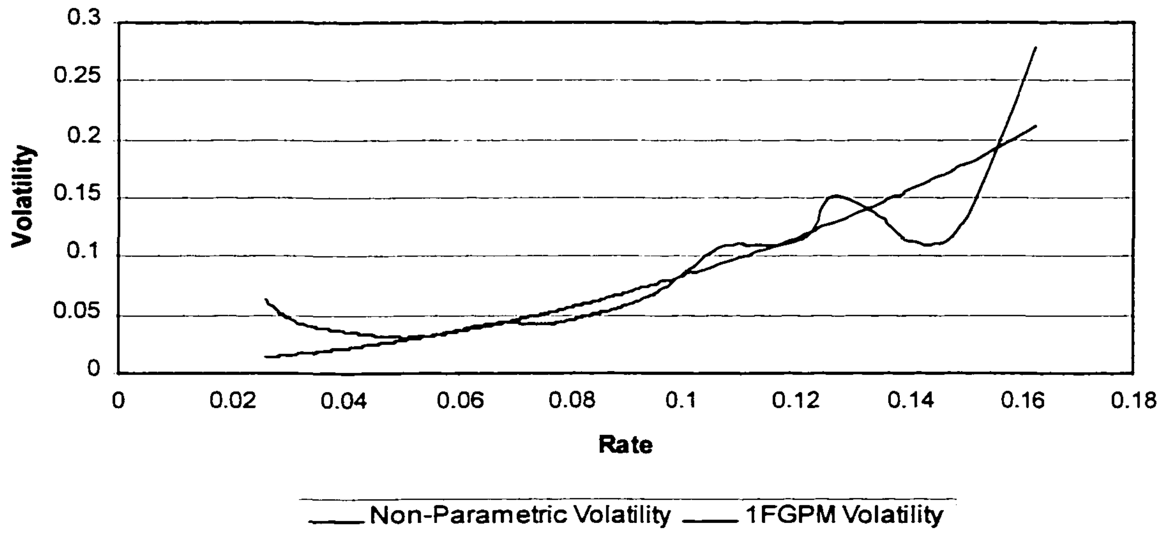
$$\hat{\mu}(r) = \frac{1}{2}\left[\frac{d\hat{\sigma}^2(r)}{dr} + \frac{\hat{\sigma}^2(r)}{\hat{p}(r)}\frac{d\hat{p}(r)}{dr}\right] \tag{3.1.2.13}$$

Figure (3.1.2.1) on the next page compares the estimated non-parametric diffusion function to a fitted parametric diffusion function for the 30 day T-bill rate in the period

12/64 to 10/97. The non-parametric diffusion is estimated from equation (3.1.2.9) with an optimally chosen bandwidth parameter.[15] The parametric diffusion function is estimated by applying the GMM procedure to the general parametric model (1FGPM), the unrestricted version of equations (3.1.1.1)-(3.1.1.4). Note that while both are increasing in the short rate, the non-parametric diffusion is more non-linear than the parametric diffusion. Figure (3.1.2.2) on the next page compares the estimated non-parametric marginal density function (NPMDF) to a fitted parametric (log-normal) density function (PDF) for this short rate data. The NPMDF is estimated from equation (3.1.2.12) with an optimally chosen bandwidth parameter. The log-normal PDF is calibrated to have the same mean and variance as the non-parametric density. Note that the NPMDF has a positive skewness (i.e., a fatter right tail), bi-modular (i.e., has concentrations at two levels), and is leptokurtotic (i.e., peaked-thinner density in the middle and more in the tails) than the PDF. Finally, Figure (3.1.2.3) on the next page compares the estimated non-parametric drift function to a fitted parametric drift function, from GMM estimation of the 1FGPM, for this short rate data. The non-parametric drift is estimated from equation (3.1.2.13) with an optimally chosen bandwidth parameter. Note that while both increase and then decrease, which is expected under mean reversion, the

[15] It is chosen to optimize a criterion that governs the trade-off between smoothness and fit to the data.

Figure 3.1.2.1: Non-Parametric and 1FGPM Diffusion Functions Compared (3M T-Bills 12:64-10:97)



Figues 3.1.2.2: Non-Paramteric and Log-Normal Marginal Densities Compared (3M T-Bills 12:64-10:97)

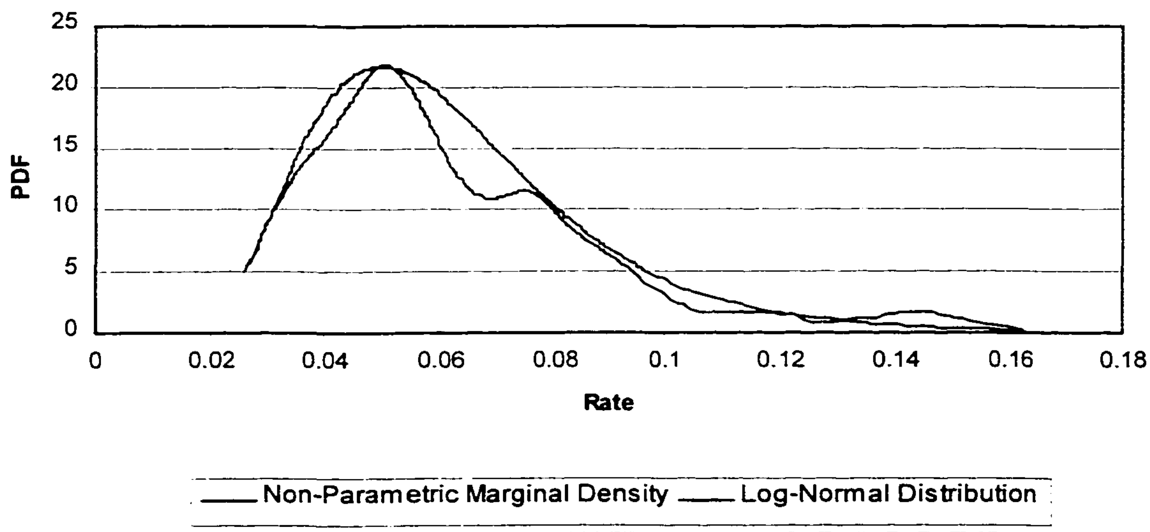
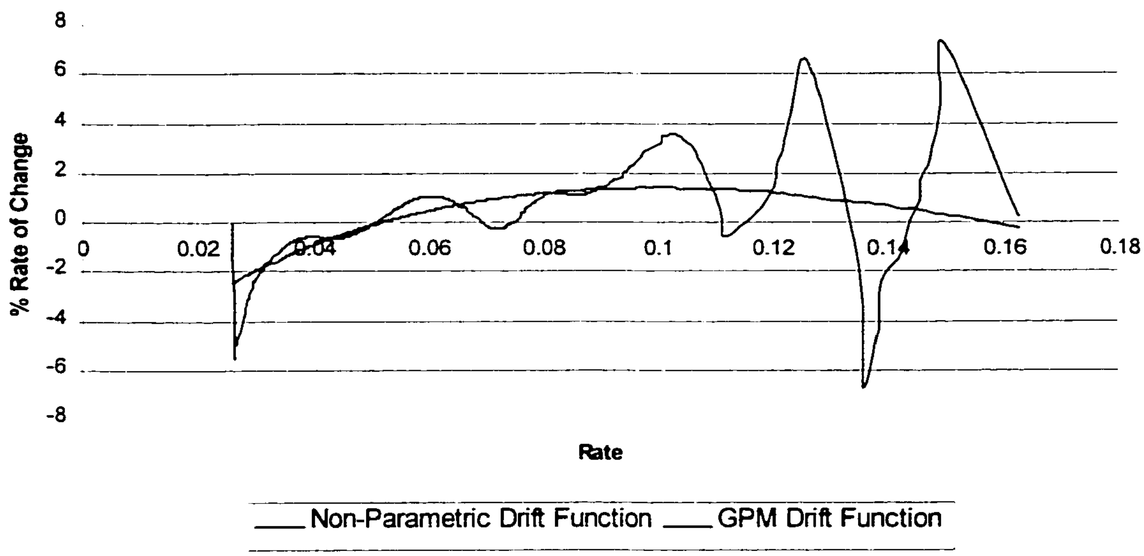


Figure 3.1.2.3: Non-Parametric and 1FGPM Drift Functions Compared (3M T-Bills 10:64-12:97)



non-parametric function is highly non-linear, and achieves higher rate-of-change levels at both very low and very high levels of the short rate.

3.2 A Comparison of Short Rate Models

We compare the relative performance of different models, many of which are nested within each other. We follow Chan et al (1992), in which an Euler approximation of a general SDE is implemented, and nested models are compared using various econometric procedures. These parametric models all impose dynamics for the instantaneous short rate that can be nested within the following SDE (see Duffie (1996)):

$$dr_t = \left[\alpha_1 + \alpha_2 r_t + \alpha_3 r_t \log(r_t)\right]dt + \left[\beta_1 + \beta_2 r_t\right]^{v} d\hat{B}_t \tag{3.2.1}$$

where the short rate process is defined by the Ito Integral
$r_t = r_0 + \int_{s=0}^{t}\left[\alpha_1 + \alpha_2 r_s + \alpha_3 r_s \log(r_s)\right]ds + \int_{s=0}^{t}\left[\beta_1 + \beta_2 r_s\right]^{v} d\hat{B}_s$, $\tilde{B}_t$ is a standard Brownian motion under risk neutralized probability measure Q, and the parameter vector $\boldsymbol{\theta}^T = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, v)$ is assumed to be constant. This is quite a flexible specification, in that it admits non-linearities in both the drift and diffusion functions. We call this the Single Factor General Parametric Model (1FGPM). The earliest specialization of this, developed by Merton (1973), is called the Arithmetic Brownian Motion term structure model[16] (ABM). This imposes the parameter restrictions $\alpha_2 = \alpha_3 = \beta_2 = v = 0$ and is represented by:

$$dr_t = \alpha_1 dt + \beta_1 d\hat{B}_t \tag{3.2.2}$$

This implies a normal distribution for the instantaneous short rate. The strong solution to (3.2.2) is given by $r_t = \alpha_1 t + \beta_1 \hat{B}_t$, which is a Brownian motion with drift. This model is nested within several subsequent generalizations. Among the best known of these is the geometric Brownian motion (GBM) model of Black and Scholes (1973):

---

16 Note that with time varying coefficients, this is often called the *Ho and Lee Model*, developed to price interest rate contingent assets in Ho et al (JF 1986).

$$\frac{dr_t}{r_t} = \alpha_2 dt + \beta_2 d\hat{B}_t \tag{3.2.3}$$

Equation (3.2.3) makes the parameter restrictions $\alpha_1 = \alpha_3 = \beta_1 = \nu = 0$ in (3.2.1). It can be easily shown by Ito's Lemma that the solution to (3.2.3) is given by $r_t = r_0 \exp\left[\left(\alpha_2 - \frac{1}{2}\beta_2^2\right)t + \beta_2 \hat{B}_t\right]$. This is called the log-normal model and since it follows that, the conditional distribution of continuously compounded returns is found to be $\ln(r_s) - \ln(r_t) \sim N\left(\left(\alpha_2 - \frac{1}{2}\beta_2^2\right)(s-t), \beta_2^2(s-t)\right) \quad \forall\ s>t$. This model is used by Rendleman and Barter (1980) and Marsh et al (1983) in pricing options on debt securities. It has the advantage of not admitting negative interest rates, but shares some of the disadvantages of other models that impose a linear form for the drift equation, $\mu(r_t,t) = \alpha_2 r_t$. The weakness lies in not modeling the reversion to the mean that is observed in time series short rate data. Furthermore, this process has an *absorbing barrier* at $r_t = 0$, which is counterintuitive to justify economically, although one might argue that the probability of reaching this point is negligible. A generalization of the GBM model that attempts to model heteroscedasticty observed in interest rate data is Constant Elasticity of Variance Model (CEVM) of Cox et al (1976) :

$$\frac{dr_t}{r_t} = \alpha_2 dt + \beta_2 r_t^{\nu-1} d\hat{B}_t \tag{3.2.4}$$

This specification allows the volatility of the interest rate to increase either less, greater, or in equal proportion to its level. The parameter γ is the sensitivity of the diffusion function to proportional changes in the interest rate, which follows from $\frac{d\ln\left|\sigma^2(r_t,t)\right|}{dr_t}r_t = \gamma$, which is constant. While this model allows for greater flexibility in capturing interest rate dynamics, in general it does not admit a closed form solution for bond and interest rate option prices. In a more analytically tractable model, Dothan (JFE 1978) restricted the GBM process to have zero drift in his discount bond valuation model:

$$\frac{dr_t}{r_t} = \beta_2 d\hat{B}_t \tag{3.2.5}$$

We call this the Dothan Model (DM). This has the advantage of precluding negative interest rates. Brennan and Schwartz (1977) use this SDE in developing numerical models of savings, retractable, and callable bonds for which analytical solution do not obtain, since the simplicity of (3.2.5) facilitates the finite difference approach. Another parameterization without a drift is Constantinides and Ingersoll (1992):

$$\frac{dr_t}{r_t} = \beta_2 r_t^{\frac{1}{2}} d\hat{B}_t \tag{3.2.6}$$

This has the advantage of having a constant proportional sensitivity of the diffusion with respect to interest rate, that fits well empirically, and like the GBM precludes negative interest rates. However, the Constantinides-Ingersoll model is non-Gaussian, which is less tractable.

The class of mean reverting models has in common the property of an *affine* drift function, which can be expressed as $\mu(r_t,t) = \alpha_1 + \alpha_2 r_t$, the interpretation being that $\bar{r} \triangleq -\frac{\alpha_1}{\alpha_2}$ is the long run unconditional mean interest rate (a constant by stationarity) and $\frac{\Delta dr_t^2}{\Delta r_{t-1}} \triangleq -\alpha_2$ is a speed of adjustment parameter. Vasicek (1977) imposes the parameter restrictions $\alpha_3 = \beta_2 = \nu = 0$ in equation (3.2.1) to this to give the Ohrenstein-Uhlenbeck or mean-reverting (henceforth VM) process:

$$dr_t = \left[\alpha_1 + \alpha_2 r_t\right] dt + \beta_1 d\hat{B}_t \tag{3.2.7}$$

This is a Gaussian model for the term structure, with closed form solutions used extensively in pricing discount bonds, bond futures and options on them, as well as varied types of contingent claims. The common property of these models is the independence of the diffusion from the level of the short rate, which is questionable given the documentation of conditional heteroscedasticity in time series data. Applications of this model include Jamshidian (1989) and Gibson and Schwartz (1990). It is nested within the model of Breenan and Schwartz (1979; BS) model, which postulates that $\nu = 1$:

$$dr_t = \left[\alpha_1 + \alpha_2 r_t\right] dt + \beta_2 r_t d\hat{B}_t \tag{3.2.8}$$

This is a mean-reverting as well as heteroscedastic model. Brennan and Schwartz (1982) use this model to derive a numerical solution for convertible bond prices. Another popular process in this mean-reverting class is the "square-root model" of CIR (1985), which specifies that $\nu = \frac{1}{2}$ and is expressed as:

$$dr_t = \left[\alpha_1 + \alpha_2 r_t\right] dt + \beta_2 r_t^{\frac{1}{2}} d\hat{B}_t \tag{3.2.9}$$

In equation (3.2.9) the short rate is allowed to both exhibit mean reversion as well as heteroscedasticity. CIR (1985) show that short rates and bond yields follow non-central chi-squared distributions under these parametric assumptions, while option formulae involve integrals over the latter. This model is been used to price mortgage backed securities in a partial equilibrium framework by Dunn and McConnell (1981). Ramaswamy and Sundaresen (1985) apply to the pricing of futures and options on futures in a general equilibrium framework for pricing the floating rate. Later this process is used to model the short rate in the valuation swap instruments by Sundaresen (1989). Longstaff (1990) use this model of rates in a yield option valuation model. An extension of this model was proposed by Pearson and Sun (1994), called the *displaced CIR model*, which is written as following:

$$dr_t = \left[\alpha_1 + \alpha_2 r_t\right]dt + \left[\beta_1 + \beta_2\sqrt{r_t}d\hat{B}_t\right] \qquad (3.2.10)$$

A model that tries to capture non-linearities in the drift of the short rate of interest documented in the literature is suggested by Black and Karasinski (1991)[17] in the context of bond and bond option pricing. This is the only model that allows $\alpha_3 \neq 0$, and is written as:

$$\frac{dr_t}{r_t} = \left[\alpha_1 + \alpha_3 \ln\left(r_t\right)\right]dt + \beta_2 d\hat{B}_t \qquad (3.2.11)$$

17

A popular special case of this model, used extensively in various forms by options traders, is the Black-Derman-Toy (henceforth BDT) model (Black et al 1990).

This has the desirable property of a sharply increasing drift at very low r^t if $\alpha_3<0$ while allowing for a linear diffusion function. A model similar to equation (3.2.11) in its implications is derived in a CIR theoretical framework by Longstaff (1989) as the Double Square Root Model (DSRM). This is later generalized to an HJM arbitrage-free setting as the Quadratic Term Structure (Boyle et al 1999, Ahn et al 2000; henceforth the QTSM). This model proposes the following process for the short rate of interest[18]:

$$dr_t = \left[\alpha_1 + \alpha_2\sqrt{r_t}\right]dt + \beta_2\sqrt{r_t}\,d\hat{B}_t \tag{3.2.12}$$

This model has the advantage of capturing non-linearity in the drift as well as proportionality of the variance to the level of the short rate. As in the CIR square root model, the parameter of mean reversion is given by $-\alpha_2$, and the long run mean interest rate level by $-\frac{\alpha_1}{\alpha_2}$. However, the restoring force is proportional to $-\left(\frac{\alpha_1}{\alpha_2} + \sqrt{r_t}\right)$ rather than $-\left(\frac{\alpha_1}{\alpha_2} + r\right.$, which implies a downward stickiness in the mean reversion process (i.e., interest rates move more rapidly toward the long-run mean from below than from above).[19] $-\left(\frac{\alpha_1}{\alpha_2} + r_t\right)$

18

A reflected Brownian motion process is assumed for the single underlying state variable, which is essentially an ABM process that is restricted to be almost everywhere positive.

19

It can be seen that the origin is a regular (or attainable) boundary, from which the rate returns to positive values with probability one in the next instant, and infinity is a natural (or Feller) boundary, which cannot be reached in finite expected time.

3.2.1 Explaining the Short Rate and its Volatility

Table 3.2.1.1 summarizes the parametric restrictions imposed by the different models of the short rate. $r_t = N \times \ln(1 + R_t)$ is the annualized, continuously compounded holding yield computed from the short rate R_t. $\varepsilon_t \sim ID(0, \sigma^2_{\varepsilon t})$ the random disturbance to the conditional mean of the yield change, orthogonal to the time t-1 information set $\Re_{t-1} = \{r_{t-1}\}$, with conditional volatility $\sigma^2_{\varepsilon t} = (\beta_1 + \beta_2 r_t)^{\gamma}$. $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \gamma)^T$ is the vector of parameters of the econometric model. The estimation used is a version of the BHHH algorithm, as part of a procedure for simultaneous non-linear systems of equations, in the MATLAB™ Optimization Toolbox.

Tables 3.2.1.1 through 3.2.1.4 present results of our econometric tests for the 1-Month U.S. Treasury Bill rates (1MTB), 1 Month Japanese Government Bond yields (1MJGB), 1 Month Euro Sterling deposit rates (1MESD), and 3 Month Libor Dollar rates (3MLD), respectively. Estimates are reported in the rows of along with asymptotic p-values beneath. χ^2 denotes Hansen's statistic for the test of over-identifying restrictions, which is the minimized GMM criterion, distributed as an asymptotic chi-squared random variable with degrees of freedom equal to the excess of orthogonality conditions over number of parameters. The goodness-of-fit statistics R_1^2 (R_2^2) are ratios of variation in rate changes (non-parametric volatility) explained by the fitted errors (estimated volatility).

TABLE 3.2.1.1

Parametric Restrictions Imposed on the General Model of the Term Structure

$$r_t - r_{t-1} = \mu(r_{t-1})\Delta t + \sigma(r_{t-1})\varepsilon_t, \qquad \varepsilon_t \sim \mathrm{IID}(0,1)$$

$$E[\varepsilon_t] = E[\varepsilon_t | r_{t-1}] = 0, \qquad E[r_t - r_{t-1}] = \mu(r_t), \qquad E[\varepsilon_t^2] = \sigma(r_t),$$

Model	$\mu(r_t)$	$\sigma(r_t)$	Parameter Restrictions
Single Factor General Parametric (1FGPM)	$\alpha_0 + \alpha_1 r_t + \alpha_2 r_t \log(r_t)$	$(\beta_0 + \beta_1 r_t)^\gamma$	None.
Non-linear Drift (NLDM)	$\alpha_0 + \alpha_1 r_t + \alpha_2 r_t \log(r_t)$	$\beta_1 r_t^\gamma$	$\beta_0 = 0$
Black-Derman-Toy (BDTM)	$\alpha_1 r_t + \alpha_2 r_t \log(r_t)$	$\beta_1 r_t$	$\alpha_0 = \beta_0 = \gamma = 0$
Displaced Diffusion (DDM)	$\alpha_0 + \alpha_1 r_t$	$(\beta_0 + \beta_1 r_t)^\gamma$	$\alpha_2 = 0$
Pearson-Sun Model (PSM)	$\alpha_0 + \alpha_1 r_t$	$(\beta_0 + \beta_1 r_t)^{\frac{1}{2}}$	$\alpha_2 = 0, \gamma = \frac{1}{2}$
Chow-Karolyi-Longstaff-Schwartz (CKLS)	$\alpha_0 + \alpha_1 r_t$	$\beta_1 r_t^\gamma$	$\alpha_2 = \beta_0 = 0$
Cox-Ingersoll-Ross Square Root (CIRSR)	$\alpha_0 + \alpha_1 r_t$	$\beta_1 r_t^{\frac{1}{2}}$	$\alpha_2 = \beta_0 = 0, \gamma = \frac{1}{2}$
Constant Elasticity of Variance (CEV)	$\alpha_1 r_t$	$\beta_1 r_t^{\frac{\gamma}{2}}$	$\alpha_0 = \alpha_2 = \beta_0 = 0$
Vasicek (VM)	$\alpha_0 + \alpha_1 r_t$	$\beta_1 r_t$	$\alpha_2 = \beta_0 = 0, \gamma = 1$
Merton Arithmetic Brownian Motion (MABM)	$\alpha_0 + \alpha_1 r_t$	β_1	$\alpha_2 = \beta_0 = 0, \gamma = 0$
Dothan Martingale Model (DMM)	0	$\beta_1 r_t$	$\alpha_0 = \alpha_1,$ $\alpha_2 = \beta_0 = 0, \gamma = 1$
Cox-Ingersoll-Ross Martingale Model (CIRM)	0	$\beta_1 r_t^{\frac{1}{2}}$	$\alpha_0 = \alpha_1 = \alpha_2 = \beta_0 = 0, \gamma = \frac{1}{2}$
Non-parametric Model (NPM)	Unspecified	Unspecified	None

Across most data-sets, and in contrast to CKLS for the case of the 1MTB, the parameters are statistically significant. This could be attributed to the longer sample length, more efficient use of the data, and an improved estimation technique. In all cases the χ^2 statistics of the GMM estimation indicates that we reject the hypothesis that all the moment restrictions fail to hold, which means that no model alone is rejected by the data[20]. In terms of parameter estimates, they tend to be similar across datasets for the more general models, but there are differences in some of the nested models. For instance, in the 1FGPM, we obtain elasticities of variance (EVs) of 2.36, 2.16, 2.63, and 2.93 for the US, Japanese, UK, and Eurodollar data, respectively. However, for the CKLS model, we obtain corresponding values of values of 1.96, 0.77, 1.68, and 0.52. This implies that the more general models tend to be more consistent (i.e., yield similar results) across different datasets. Our results differ from CKLS for the other data sets-we find much higher EV's in the US and UK markets, and much lower EV's in the Japanese and Eurodollar markets. This implies that interest rate volatility is more sensitive to the level of the rate in the US and UK, and less so in Japanese and the Eurodollar market.

Following CKLS, we conduct a simple test of model fit, by comparing the percent variation explained by the fitted drift and diffusion functions relative to their non-

20 This is analogous to a an F-statistic indicating that all the parameters of a linear regression are not simultaneously zero.

parametric counterparts. I define the R_1^2 (R_2^2) statistics to be the sample variance of the estimated drift (estimated volatility) divided by the sample variance of the non-parametric drift (non-parametric volatility). In general, across different data sets, these models explain a relatively small proportion of the variation in the non-parametrically estimated drift and diffusion functions, yet there is significant variation in relative magnitudes as well as rankings. These models explain a higher proportion of non-parametrically estimated variation in the drift and diffusion in the Japanese data, by about a factor of 10. These models explain drift as compared to diffusion differently across data-sets. In the case of the US data, the parametric models do a poor job of capturing variation in the drift as compared to volatility, with percentages explained in the drift ranging from 0.01% to 8.6%, while those in the volatility range from 0.1% to 70%. The situation is reversed in the case of the Japanese data, where the parametric models do better at capturing variation in the drift as compared to the volatility, with percentages explained in the drift ranging from 4% to 14%, while those in the volatility range from 0.1% to 16%. The results are similar for Euro and Japan with respect to explaining the drift much better than the volatility, but with magnitudes higher in the drift and lower in the volatility, with percentages explained in the drift ranging from 0.9% to 73.3%, while those in the volatility range from 0.0001% to 0.03%. For the UK drift is explained about as weli volatility, with percentages explained in the drift ranging from .01% to 6%, while those in the volatility range from 0.02% to 4%. These results differ with empirical findings based on US data, in which drift is estimated with much greater difficulty than the diffusion function.

Across data sets, the more general models tend to perform best in estimating drift and volatility, capturing the highest percentages of non-parametric variation in these explained. In the 1MTB, the 1F-GPM performs best, followed by the Non-Linear Drift and Black-Derman-Toy models. In terms of R_2^2, the CKLS and CEV models explain about half what the above models do, but outperform the remaining models. The CIR-SR model performs the weakest, with 2% of the variation explained. Surprisingly, the non-nested QTSM is very close to the CIR model by these measures, in spite of the fact that the QTSM is capable of accommodating a degree of non-linearity in the drift. In contrast, for the 1MJGB, the QTSM best describes the drift, with the 1FGPM a runner-up, while the PSM best describes volatility, with the QTSM the runner-up. The 1FGPM does a poor job of explaining the diffusion for the Japanese data. The QTSM is most consistent within data sets for the case of the Japanese. In the 1MESD data, models are generally less successful in capturing both the variation in the drift as well as the diffusion. The 1FGPM best describes the drift, while the nested BDTM best describes volatility, although the differences with other models is much less with the drift as opposed to the volatility. Finally, these results all differ in the case of the 3MLD, where the VM and BDTM best explain the drift and diffusion, respectively.

TABLE 3.2.1.1

Generalized Method of Moments Estimates of Autoregressive Approximation to Stochastic Differential Equations for the Short Rate Change Processes: 1 Month U.S. T-Bill Rates (4:64-10:97)

Coefficients and Statistics (P-Values in Parentheses)

	α_1	α_2	α_3	β_1	β_2	γ	χ^2	R_1^2	R_2^2
1FGPM:	1.6566	-0.8223	-0.2890	0.0076	0.0972	2.3643	729.29	0.0859	0.7022
	(0.00)[c]	(0.02)[b]	(0.00)[c]	(0.01)[c]	(0.00)[c]	(0.00)[c]	(0.00)[c]	(1)[1]	(1)
NLDM:	1.4509	-0.7444	-0.2649	(0)[2]	0.0045	2.3218	737.17	0.0828	0.6907
	(0.00)[c]	(0.00)[c]	(0.00)[c]		(0.07)[a]	(0.00)[c]	(0.00)[c]	(2)	(2)
BDTM:	(0)	0.0531	-0.0259	(0)	0.0472	(0)	829.68	0.0783	0.6734
	(0.00)[c]	(0.00)[c]	(0.00)[c]		(0.00)[c]	(3)	(3)		
DDM:	0.1370	-0.0271	-0.2115	(0)	0.1090	1.4351	796.95	0.0086	0.3668
	(0.00)[c]	(0.00)[c]	(0.79)		(0.06)[b]	(0.00)[c]	(0.00)[c]	(5)	(4)
PSM:	0.1130	-0.0228	(0)	-0.3371	0.1141	(½)	799.57	0.0062	0.1698
	(0.00)[c]	(0.00)[c]		(0.02)[b]	(0.00)[c]		(0.00)[c]	(6)	(7)
CKLS:	0.1150	-0.02862	(0)	(0)	0.0057	1.9579	797.96	0.0096	0.3600
	(0.00)[c]	(0.00)[c]			(0.44)	(0.00)[c]	(0.00)[c]	(5)	(5)
CIRSR:	0.1086	-0.0152	(0)	(0)	0.0474	(½)	824.88	0.0030	0.0018
	(0.00)[c]	(0.00)[c]			(0.00)[c]		(0.00)[c]	(7)	(11)
CEVM:	(0)	0.0048	(0)	(0)	-6.1×10^{-8}	-4.1×10^{-9}	947.44	0.0003	0.3330
		(0.08)[a]			(0.00)[c]	(0.00)[c]	(0.00)[c]	(10)	(6)
VM:	0.08601	-0.0075	(0)	(0)	0.1848	(0)	863.15	0.0007	0.0011
	(0.00)[c]	(0.00)[c]			(0.000)[c]		(0.00)[c]	(9)	(12)
MABM:	(0)	0.0282	(0)	(0)	0.2232	(0)	414.01	N.A.	N.A.
	(0.06)[a]	(0.00)[c]			(0.00)[c]				
DM:	(0)	(0)	(0)	(0)	0.0107	(1)	280.72	N.A.	0.0075
					(0.00)[c]		(0.00)[c]		(10)
CIRM	(0)	(0)	(0)	(0)	0.02232	(½)	632.91	N.A.	0.0135
					(0.03)[b]		(0.00)[c]		(8)
QTSM	0.1472	-0.0778	(0)	(0)	0.04601	(½)	884.16	0.0025	0.0017
	(0.78)	(0.00)[c]			(0.00)[c]		(0.00)[c]	(9)	(8)

1-Indicates the relative ranking of performance (1=best, 12=worst), 2-(.) indicates that the parameter is restricted to 0, 1, or ½.

TABLE 3.2.1.2

Generalized Method of Moments Estimates of Autoregressive Approximation to Stochastic Differential Equations for the Short Rate Change Processes: 1 Month Japanese Government Bond Yields (78:11-99:02)

Coefficients and Statistics (P-Values in Parentheses)

	α_1	α_2	α_3	β_1	β_2	γ	χ^2	R_1^2	R_2^2
1FGPM:	0.0632	-0.0997	-0.0583	0.2176	0.0292	2.1580	331.51	0.1246	0.0598
	(0.20)	(0.00)^c	(0.00)^c	(0.78)	(0.00)^c	(0.00)^c	(0.00)^c	(4)[1]	(6)
NLDM:	0.0462	-0.0741	-0.0460	(0)[2]	0.0404	0.7563	322.50	0.0950	0.0721
	(0.31)	(0.034)	(0.00)^c		(0.01)^c	(0.00)^c	(0.00)^c	(7)	(5)
BDTM:	(0)	-0.0461	-0.0369	(0)	0.0249	(0)	346.38	0.1146	0.0276
		(0.07)^c	(0.00)^c		(0.00)^c		(0.00)^c	(6)	(7)
DDM:	0.1073	-0.0500	(0)	0.1947	0.03186	2.0068	371.49	0.1335	0.0782
	(0.01)^c	(0.00)^c		(0.73)	(0.00)^b	(0.55)	(0.00)^c	(2)	(4)
PSM:	0.0041	-0.0473	(0)	0.0263	0.02297	(½)	360.09	0.1197	0.1680
	(0.01)^b	(0.00)^c		(0.03)^b	(0.00)^c		(0.00)^c	(3)	(1)
CKLS:	0.0904	-0.0460	0.0434	(0)	(0)	0.7695	362.76	0.1133	0.0895
	(0.03)^b	(0.00)^c	(0.00)^b			(0.00)^c	(0.00)^c	(5)	(3)
CIRSR:	0.0102	-0.0202	0.0166	(0)	(0)	(½)	473.62	0.0218	0.0028
	(0.00)^c	(0.00)^c	(0.01)^c				(0.00)^c	(8)	(9)
CEVM:	(0)	-0.0150	(0)	(0)	0.0447	0.4566	454.42	0.0120	0.0155
	(0.00)^c				(0.00)^c	(0.07)^a	(0.00)^c	(9)	(8)
VM:	0.0252	-0.0131	(0)	(0)	0.0615	(0)	464.17	0.01	1.2×10^{-4}
	(0.03)^b	(0.01)^c			(0.00)^c		(0.00)^c	(10)	(12)
MABM:	(0)	-0.0155	(0)	(0)	0.0539	(0)	214.63	N.A	N.A.
		(0.05)^a			(0.00)^c		(0.00)^c		
DM:	(0)	(0)	(0)	(0)	0.0043	(1)	295.2	N.A.	0.0008
					(0.42)^c		(0.00)^c		(10)
CIRM:	(0)	(0)	(0)	(0)	0.0196	(½)	295.2	N.A.	1.5×10^{-4}
					(0.00)^c		(0.00)^c		(11)
QTSM:	-0.1191	0.0643	(0)	(0)	0.0181	(½)	165.5	0.142	0.125
	(0.69)	(0.60)			(0.00)^c		(0.00)^c	(1)	(2)

1-Indicates the relative ranking of performance (1=best, 12=worst), 2-(.) indicates that the parameter is restricted to 0, 1, or ½.

TABLE 3.2.1.3

Generalized Method of Moments Estimates of Autoregressive Approximation to Stochastic Differential Equations for the Short Rate Change Processes: 1 Month Eurosterling Deposit Rates (75:Q1-99:Q6)

Coefficients and Statistics (P-Values in Parentheses)

	α_1	α_2	α_3	β_1	β_2	γ	χ^2	R_1^2	R_2^2
1FGPM:	0.0116	-0.3077	-0.1837	0.0001	0.01575	2.6304	128.21	0.0326	0.0073
	(0.08)[b]	(0.01)[c]	(0.00)[c]	(0.99)	(0.00)[c]	(0.00)[c]	(0.00)[c]	(1)[1]	(7)
NLDM:	0.0096	-0.0256	-0.1547	(0)[2]	0.0002	2.6249	322.50	0.0211	0.0076
	(0.00)[c]	(0.03)[b]	(0.00)[c]		(0.00)[c]	(0.00)[c]	(0.00)[c]	(4)	(6)
BDT:	(0)	-0.0995	-0.0411	(0)	0.0006	(0)	128.71	0.0143	0.0429
		(0.07)[a]	(0.07)[a]		(0.00)[c]		(0.00)[c]	(6)	(1)
DDM:	0.0036	-0.0433	(0)	-0.0271	0.0074	2.7483	142.56	0.0129	0.0013
	(0.01)[b]	(0.02)[b]		(0.00)[c]	(0.01)[c]	(0.01)[c]	(0.00)[c]	(7)	(10)
PSM:	0.0041	-0.0496	(0)	1.4×10^{-5}	0.0043	(½)	129.98	0.0169	0.0005
	(0.07)[a]	(0.00)[c]		(0.00)[c]	(0.00)[c]		(0.00)[c]	(5)	(11)
CKLS:	0.0042	-0.0512	(0)	(0)	0.0003	1.6814	129.90	0.0181	0.0093
	(0.01)[c]	(0.01)[c]			(0.06)[b]	(0.00)[c]	(0.00)[c]	(4)	(5)
CIRSR:	0.0031	-0.0445	(0)	(0)	0.0006	(½)	220.62	0.0218	0.0028
	(0.01)[c]	(0.00)[c]			(0.01)[b]		(0.00)[c]	(3)	(8)
CEV:	(0)	0.0120	(0)	(0)	0.0001	1.2004	219.11	0.0276	0.0290
		(0.00)[b]			(0.00)[c]	(0.00)[c]	(0.00)[c]	(2)	(3)
VM:	0.0048	-0.0633	(0)	(0)	5.2×10^{-5}	(0)	454.42	9.9×10^{-4}	0.0210
	(0.01)[b]	(0.00)[c]			(0.09)[a]		(0.00)[c]	(9)	(4)
MABM:	(0)	0.0010	(0)	(0)	5.8×10^{-5}	(0)	122.48	N.A.	N.A.
		(0.00)[c]			(0.00)[c]		(0.00)[c]		
DM:	(0)	(0)	(0)	(0)	0.0020	(1)	62.770	N.A.	0.0016
					(0.05)[b]		(0.00)[c]		(9)
CIRM:	(0)	(0)	(0)	(0)	7.8×10^{-5}	(½)	165.14	N.A.	0.0377
					(0.00)[c]		(0.00)[c]		(2)
QTSM:	0.0121	-0.0338	(0)	(0)	-0.0001	(½)	115.17	0.0072	1.8×10^{-5}
	(0.08)[a]	(0.00)[c]			(0.00)[c]		(0.00)[c]	(8)	(12)

1-Indicates the relative ranking of performance (1=best, 12=worst), 2-(.) indicates that the parameter is restricted to 0, 1, or ½.

TABLE 3.2.1.4

Generalized Method of Moments Estimates of Autoregressive Approximation to Stochastic Differential Equations for the Short Rate Change Processes: 3 Month Libor Dollar Rates (1:90-12:99)

Coefficients and Statistics (P-Values in Parentheses)

	α_1	α_2	α_3	β_1	β_2	γ	χ^2	R_1^2	R_2^2
1FGPM:	9.0×10^{-5}	-0.0019	-0.0008	1.0×10^{-6}	1.5×10^{-6}	2.9368	3762.1	8.9×10^{-4}	5.9×10^{-5}
	(0.00)[c]	(0.01)[c]	(0.00)[c]	(0.00)[C]	(0.00)[C]	(0.01)[c]	(0.00)[c]	(10)[1]	(7)
NLDM:	2.9×10^{-4}	-0.0341	-0.0154	(0)[2]	1.8×10^{-6}	0.5659	4139.9	0.1935	0.0054
	(0.09)[a]	(0.00)[c]	(0.000)[c]		(0.00)[c]	(0.00)[c]	(0.00)[c]	(3)	(2)
BDTM:	(0)	-0.0271	-0.0104	(0)	4.5×10^{-5}	(0)	4290.4	0.1864	0.0322
		(0.00)[c]	(0.00)[c]		(0.00)[c]		(0.00)[c]	(4)	(1)
DDM:	0.0002	-0.0052	(0)	0.0062	0.0083	4.2220	4413.7	0.0312	0.0001
	(0.00)[c]	(0.02)[b]		(0.00)[c]	(0.00)[c]	(0.00)[c]	(0.00)[c]	(7)	(5)
PSM:	0.0005	-0.0103	(0)	1.6×10^{-6}	3.1×10^{-5}	(½)	4235.3	0.1280	0.0030
	(0.00)[c]	(0.00)[c]		(0.00)[b]	(0.00)[c]		(0.00)[c]	(5)	(3)
CKLS:	0.0004	-0.0101	(0)	(0)	1.7×10^{-5}	0.5280	4244.9	0.1227	0.0047
	(0.00)[c]	(0.00)[c]			(0.00)[c]	(0.00)[c]	(0.00)[c]	(6)	(4)
CIRSR:	6.5×10^{-4}	-0.0215	(0)	(0)	3.0×10^{-5}	(½)	9361.1	0.5557	0.0164
	(0.00)[c]	(0.00)[c]			(0.00)[c]		(0.00)[c]	(2)	(2)
CEVM:	(0)	-0.0043	(0)	(0)	0.0208	0.1643	9448.3	0.0224	2.08×10^{-7}
		(0.06)[a]			(0.00)[c]	(0.00)[a]	(0.00)[c]	(8)	(10)
VM:	9.0×10^{-4}	-0.0247	(0)	(0)	3.1×10^{-6}	(½)	9083.1	0.7333	1.27×10^{-7}
	(0.00)[c]	(0.00)[c]			(0.00)[c]		(0.00)[c]	(1)	(11)
MABM:	-2.0×10^{-4}	(0)	(0)	(0)	3.1×10^{-6}	(0)	3059.6	N.A.	N.A.
	(0.20)				(0.00)[c]		(0.00)[c]		
DMM:	(0)	(0)	(0)	(0)	0.0013	(1)	295.2	N.A.	6.9×10^{-5}
					(0.00)[c]		(0.00)[c]		(6)
CIRM:	(0)	(0)	(0)	(0)	1.5×10^{-5}	(½)	7631.5	N.A.	2.8×10^{-5}
					(0.00)[c]		(0.00)[c]		(8)
QTSM:	0.0022	-0.0122	(0)	(0)	1.2×10^{-5}	(½)	116.6	0.0039	1.66×10^{-5}
	(0.01)[b]	(0.01)[b]			(0.05)[b]		(0.00)[c]	(9)	(9)

1-Indicates the relative ranking of performance (1=best, 12=worst), 2-(.) indicates that the parameter is restricted to 0, 1, or ½.

TABLE 3.2.1.5

Newey-West Chi-Squared Tests of Nested Autoregressive Approximation to Stochastic Differential Equations for the Short Rate Change Processes: 1 Month U.S. T-Bill Rates (4:64-10:97)

P-Values Beneath

nest →	GPM	DDM	PSM	NLD	BDT	CKLS	CIR	CEV	VM	ABM	DM	CIRM
GPM	-	799.5	29.23	1024.5	1210.4	29.26	234.7	737.9	510.4	500.0	234.3	344.5
		0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c
NLD	-	-	14.25	-	-	14.29	219.8	763.0	495.4	486.3	220.8	358.1
			1.6×10^{-4c}	-	-	1.6×10^{-4c}	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c
BDT	-	-	-	-	-	-	205.5	-	-	-	-	371.4
							0.00^c					0.00^c
DDM	-	-	-	-	185.8	-	-	-	-	-	-	-
					0.00^c							
PSM	-	-	-	-	-	-	-	-	-	-	-	-
CKLS	-	-	-	-	-	-	205.5	748.7	481.1	473.5	208.8	145.5
							0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c
CIR	-	-	-	-	-	-	-	-	-	-	-	562.3
												0.00^c
CEV	-	-	-	-	-	-	-	-	-	245.6	509.6	-
										0.00^c	0.00^c	
VM	-	-	-	-	-	-	-	-	-	-	-	-
ABM	-	-	-	-	-	-	-	-	-	-	264.0	-
											0.00^c	
DM	-	-	-	-	-	-	-	-	-	-	-	-
CIRM	-	-	-	-	-	-	-	-	-	-	-	-

TABLE 3.2.1.6

Newey-West Chi-Squared Tests of Nested Autoregressive Approximation to Stochastic Differential Equations for the Short Rate Change Processes

1 Month Japanese Government Bond Yields (78:11-99:02)

P-values Beneath

nest →	GPM	DDM	PSM	NLD	BDT	CKLS	CIR	CEV	VM	ABM	DM	CIRM
GPM	-	1207.2	1212.5	1024.5	210.4	1210.7	1187.7	1103.8	1106.2	901.0	1029	1976.8
		0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c
NLD	-	-	5.36	-	-	3.55	19.48	103.4	101.0	270.0	142.0	805.8
			0.02^b			0.06^a	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c
BDT	-	-	-	-	-	-	24.84	-	-	-	-	799.5
							0.00^c					0.00^c
DDM	-	-	-	-	185.8		-	-	-	-	-	-
					0.00^c							
PSM	-	-	-	-	-	-	-	-	-	-	-	-
CKLS	-	-	-	-	-	-	23.03	107.0	104.5	273.5	145.5	802.3
							0.00^c	0.00^c	0.00^c	0.00^c	0.00^c	0.00^c
CIR	-	-	-	-	-	-	-	-	-	-	-	824.0
												0.00^c
CEV	-	-	-	-	-	-	-	-	-	169.3	41.4	-
										0.00^c	0.00^c	-
VM	-	-	-	-	-	-	-	-	-	-	-	-
ABM	-	-	-	-	-	-	-	-	-	-	127.9	-
											0.00^c	
DMM	-	-	-	-	-	-	-	-	-	-	-	-
CIRM	-	-	-	-	-	-	-	-	-	-	-	-

TABLE 3.2.1.7
Newey-West Chi-Squared Tests of Nested Autoregressive Approximation to Stochastic Differential Equations for the Short Rate Change Processes
1 Month Euro-Sterling Deposit Rates (75:Q1-99:Q6)
P-valuesd Beneath

nest ¯	GPM	DDM	PS	NLD	BDT	CKLS	CIR	CEV	VM	ABM	DM	CIRM
GPM	-	29.6	13.5	4.60	8.47	15.10	8.50	14.51	11.92	112.7	236.4	269.4
		0.00[a]	0.00[c]	0.03[b]	0.00[c]	0.00[c]	0.00[c]	0.01[c]	0.00[c]	0.00[c]	0.00[c]	0.00[c]
NLD	-	-	17.24	-	-	18.86	4.97	10.75	16.69	29.80	373.0	226.2
			0.00[b]			0.00[c]	0.00[c]	0.00[c]	0.01[c]	0.00[c]	0.00[c]	0.00[c]
BDT	-	-	-	-	-	-	6.59	-	-	-	-	336.1
							0.01[c]					0.00[c]
DDM	-	-	-	-	3.81	-	-	-	-	-	-	-
					0.05[b]							
PSM	-	-	-	-	-	-	-	-	-	-	-	-
CKLS	-	-	-	-	-	-	5.59	26.6	3.17	74.2	272.7	330.9
							0.01[c]	0.00[c]	0.01[c]	0.00[c]	0.00[c]	0.00[c]
CIR	-	-	-	-	-	-	-	-	-	-	-	393.3
												0.00[c]
CEV	-	-	-	-	-	-	-	-	-	130.3	213.7	-
										0.00[c]	0.00[c]	
VM	-	-	-	-	-	-	-	-	-	-	-	-
ABM	-	-	-	-	-	-	-	-	-	-	348.7	-
											0.00[c]	
DM	-	-	-	-	-	-	-	-	-	-	-	-
CIRM	-	-	-	-	-	-	-	-	-	-	-	-

TABLE 3.2.1.8

Newey-West Chi-Squared Tests of Nested Autoregressive Approximation to Stochastic Differential Equations for the Short Rate Change Processes

3 Month Libor Dollar Rates (1:90-12:99)

P-Values Beneath

nest	GPM	DDM	BDT	NLD	PSM	CKLS	CIR	CEV	VM	ABM	DM	CIRM
GPM	-	1295.8	1211.4	165.6	1632.6	1606.7	4831.2	339.6	6312.3	43.0	117.7	17.2
		0.00^{a}	0.0^{c}	0.00^{c}	0.0^{c}	0.00^{c}	0.0^{c}	0.00^{c}	0.00^{c}	0.000^{c}	0.00^{c}	0.00^{c}
NLD	-	-	84.6	-	-	1772.3	4665.7	174.0	6227.3	179.8	25.1	154.0
			0.00^{b}			0.00^{a}	0.00^{c}	0.00^{c}	0.00^{c}	0.00^{c}	0.00^{c}	0.00^{c}
BDT	-	-	-	-	-	-	6463.8	-	-	-	-	20.2
							0.00^{c}					0.00^{c}
DDM	-	-	-	-	1794.2	-	-	-	-	-	-	-
					0.00^{c}							
PSM	-	-	-	-	-	-	-	-	-	-	-	-
CKLS	-	-	-	-	-	-	6437.9	1946.3	7999.6	40.4	12.3	14.6
							0.00^{c}	0.00^{c}	0.00^{c}	0.00^{c}	0.00^{c}	0.00^{c}
CIR	-	-	-	-	-	-	-	-	-	-	-	163.0
												0.00^{c}
CEV	-	-	-	-	-	-	-	-	-	160.8	6.4	-
										0.00^{c}	0.01^{b}	
VM	-	-	-	-	-	-	-	-	-	-	-	-
ABM	-	-	-	-	-	-	-	-	-	-	169.4	-
											0.00^{c}	
DM	-	-	-	-	-	-	-	-	-	-	-	-
CIRM	-	-	-	-	-	-	-	-	-	-	-	-

Table 3.2.1.5 through 3.2.1.8 present Newey-West specification tests of the restrictions imposed by each nested model, for the four short rate series, respectively . We do this for all possible nested pairs. We have the consistent finding that in no case do we fail to reject the parameter restrictions imposed by a model on any model that it is nested in. This supports the conclusion of Tables 3.2.1.1 through 3.2.1.4, in which the more general models tended to have a better in-sample performance. In addition, looking at the proportionate drops in the GMM criteria, we see that they become more dramatic as we move away from the more general model, supporting the conclusions of the econometric results.

We summarize these results as follows:

1. The GMM estimator and the Newey-West tests of various models and across different short interest rates support the conclusion that most of the popular models are mis-specified (i.e., likely to be different from the true data generating model) .
2. The variation in results across different rates, in terms of explaining the non-parametrically estimated drift and diffusion, shows that the popularity of several models (e.g., BDT, PS, CKLS) may have been driven by US historical experience. These models do not seem to fit other short rates globally.
3. The conclusion in the existing literature that most models explain the drift function better than the volatility function is seen to not hold globally.
4. In agreement with CKLS (1990), and in contrast to literature that accommodates more

parametricized approaches (such as GARCH effects), I find that the effect of the level of interest rates on the volatility is more significant than previously believed, when we allow the diffusion function to be a linear transformation of the short rate.

5. In agreement with economic theory, I find that both mean reversion and non-linearity is a significant feature of short rate drift processes, in agreement with the non-parametric finance literature (Ait-Sahalia 1996).
6. The non-nested quadratic term structure model (Boyle et al 1999), a HJM style model first estimated and compared with these other popular models in this study, is found to generally under perform despite its ability to model a mean reverting and non-linear drift.

3.2.2 Forecasting Interest Rate Movements

In this section, we compare the models in terms of out-of-sample forecasting performance. To measure the efficiency of each model in forecasting rates, at each point in time we re-estimate the models based upon information available up to that point, and then forecast the short rate for the next month on this basis. In the case of the non-parametric model, we multiply the estimated drift function by the time step (1/250) to generate the expected change in the short rate, adding it to the previously realized level to obtain a forecast. We use the following performance measures. Classical goodness-of-fit statistics presented are the mean squared error (MSE) and the mean absolute error (MAE). Another measure of the percentage of correctly predicted directional changes, the *directional variation symmetry* (DVS), is given by

$$DVS = \frac{1}{T-1}\mathcal{H}\left[\left(\hat{r}_t - \hat{r}_{t-1}\right)\left(r_t - r_{t-1}\right)\right], \text{ where } \mathcal{H}(x) = \begin{cases} 1 \text{ if } x > 0 \\ 0 \text{ otherwise} \end{cases}$$

is the heavy-side function. A directional change predictor that is robust to trends is the *normalized directional symmetry*, defined by the ratio $NDS = \frac{DS_{put}}{DS_{rwp}}$, where

$$DS_{put} = \frac{100}{T}\sum_{t=1}^{T}\mathcal{H}\left[\left(\hat{r}_t - \hat{r}_{t-1}\right)\left(r_t - r_{t-1}\right)\right]$$

is the *directional symmetry under test* and

$$DS_{rwp} = \frac{100}{T}\sum_{t=1}^{T}\mathcal{H}\left[\left(r_t - r_{t-1}\right)\left(r_{t-1} - r_{t-2}\right)\right]$$

is the *directional symmetry under random walk prediction*. The *weighted directional symmetry* (WDS) is a version of the DVS, given by the absolute prediction error weighted ratio of incorrectly to correctly predicted changes, which one seeks to minimize.[21] Theil's *U-Statistic*, given by

$$TU = \sqrt{\frac{\sum_{t=1}^{T}\left[\left(r_t - \hat{r}_{t+1}\right)/\hat{r}_t\right]^2}{\sum_{t=1}^{T}\left[\left(\hat{r}_{t+1} - \hat{r}_t\right)/\hat{r}_t\right]^2}},$$

is the ratio of random walk ("naive" model) to model the percentage forecast errors. This is a measure that considers the cost of large errors relative to a naive random walk predictor, which one seeks to minimize.

The results of the forecasting ability of the various models and using the various measures are given in Tables 3.2.2.1 through 3.2.2..4. The U.S. and Japanese short rate models are presented in the former two tables and the U.K. and Libor are in the latter two tables. The 1FGPM performs the best for all the four global interest rates based on most

---

21 This is given by:

$$WDS = \frac{\sum_{t=1}^{T}\mathcal{H}\left[-\left(r_t - r_{t-1}\right)\left(\hat{r}_t - \hat{r}_{t-1}\right)\right]\left|\hat{r}_t - r_t\right|}{\sum_{t=1}^{T}\mathcal{H}\left[\left(r_t - r_{t-1}\right)\left(\hat{r}_t - \hat{r}_{t-1}\right)\right]\left|\hat{r}_t - r_t\right|}$$

TABLE 3.2.2.1

Dynamic One-Step Ahead Forecasts: Single Factor Spot Rate Models

US 1 Month T-Bills (4:64-10:97)

Statistics (Rank in Parentheses)

	MSE	MAE	DVS	NDS	WDS	TU
GPM	0.0555 (1)	0.0308 (1)	0.5125 (3)	1.1130 (3)	1.4387 (1)	1.2985 (2)
DDM	0.0586 (4)	0.0321 (2)	0.4987 (9)	1.0030 (8)	1.4481 (2)	1.2825 (1)
PSM	0.0597 (5)	0.0339 (4)	0.5089 (5)	1.1115 (4)	1.5158 (7)	1.3439 (10)
NLD	0.0572 (2)	0.0365 (8)	0.5013 (7)	1.0186 (7)	1.4779 (3)	1.3146 (5)
BDT	0.0573 (3)	0.0357 (6)	0.5070 (6)	1.1047 (6)	1.5027(5)	1.3168 (7)
CKLS	0.0601 (7)	0.0361 (7)	0.5213 (1)	1.1186 (1)	1.4996(4)	1.3161 (6)
CIR	0.0616 (9)	0.0417 (10)	0.4937 (10)	1.0017 (10)	2.1370(9)	1.3263 (9)
CEV	0.0603 (8)	0.0395 (9)	0.5187 (2)	1.1174 (2)	2.1682(10)	1.3202 (8)
VCK	0.0650 (11)	0.0433 (11)	0.4861 (11)	1.0848 (11)	2.2949(12)	1.3472 (11)
ABM	0.0664 (12)	0.0664 (12)	0.5002 (8)	1.0017 (9)	1.8698 (6)	1.3529 (12)
QTSM	0.0645 (10)	0.0426 (10)	0.4835 (12)	1.0011 (12)	2.3107(11)	1.3004 (4)
NPM	0.0599 (6)	0.0324 (3)	0.5090 (4)	1.1059 (5)	1.6389 (8)	1.2994 (3)

TABLE 3.2.2.2

Dynamic One-Step Ahead Forecasts: Single Factor Spot Rate Models

1 Month Japanese Govt Bonds (78:11-99:02)

Statistics (Rank in Parentheses)

	MSE	MAE	DVS	NDS	WDS	TU
GPM	0.0730 (1)	0.0567 (1)	0.5588 (2)	1.2231 (1)	1.4071 (1)	3.4649 (2)
DDM	0.0769 (3)	0.0578 (2)	0.5522 (3)	1.1322 (3)	1.4255 (4)	6.3351 (2)
PSM	0.0780 (6)	0.0682 (9)	0.4911 (4)	1.1520 (2)	1.6729 (9)	3.4649 (2)
NLD	0.0861 (5)	0.0579 (3)	0.5885 (1)	1.0237 (10)	1.4162 (3)	3.4649 (2)
BDT	0.0961 (13)	0.0580 (4)	0.4885 (8)	1.0361 (9)	1.4079 (2)	3.4649 (2)
CKLS	0.0898 (9)	0.0682 (9)	0.4907 (5)	1.0520 (6)	1.6650 (5)	3.4649 (2)
CIR	0.0865 (11)	0.0631 (5)	0.4902 (6)	1.0578 (7)	1.6326 (5)	3.4649 (2)
CEV	0.0866 (12)	0.0657 (8)	0.4822 (9)	1.0785 (4)	1.6446 (8)	3.4649 (2)
VM	0.0862 (10)	0.0639 (7)	0.4160 (12)	1.0418 (8)	1.6356 (7)	3.4649 (2)
ABM	0.0828 (9)	0.0735 (6)	0.4900 (7)	1.0005 (11)	1.7360 (10)	3.4650 (2)
QTSM	0.0737 (2)	0.0753 (8)	0.4720 (10)	1.0580 (6)	1.7362 (11)	3.4650 (2)
NPM	0.0772 (4)	0.0850 (10)	0.4550 (11)	1.0017 (12)	1.5305 (5)	1.5702 (2)

TABLE 3.2.2.3

Dynamic One-Step Ahead Forecasts: Single Factor Spot Rate Models

3 Month Euro-Sterling Deposits (75:Q1-99:Q6)

Statistics (Rank in Parentheses)

	MSE	MAE	DVS	NDS	WDS	TU
GPM	0.0601 (1)	0.0549 (1)	0.5567 (4)	0.8308 (3)	0.8506 (1)	1.6419 (1)
DDM	0.0602 (2)	0.0568 (5)	0.5876 (3)	0.8769 (2)	1.2130 (8)	1.6476 (2)
PSM	0.0612 (4)	0.0567 (4)	0.5876 (3)	0.8769 (2)	1.1812 (7)	1.6476 (2)
NLD	0.0624 (7)	0.0563 (3)	0.5979 (2)	0.8923 (1)	1.0216 (5)	1.6476 (2)
BDT	0.06119 (3)	0.0561 (2)	0.5979 (2)	0.8923 (1)	0.9921 (2)	1.6476 (2)
CKLS	0.0620 (4)	0.0568 (6)	0.5979 (2)	0.8923 (1)	1.0475 (6)	1.6476 (2)
CIR	0.0620 (4)	0.0569 (6)	0.5979 (2)	0.8923 (1)	1.0132 (4)	1.6476 (2)
CEV	0.0622 (6)	0.0571 (7)	0.5876 (3)	0.8923 (1)	1.0100 (3)	1.6476 (2)
VM	0.0621 (5)	0.0572 (8)	0.5980 (1)	0.8769 (2)	1.2365 (9)	1.6476 (2)
ABM	0.0623 (6)	0.0861 (9)	0.5479 (5)	0.8239 (5)	1.8475 (10)	1.6674 (3)
QTSM	0.0625 (8)	0.0865 (10)	0.5794 (6)	0.8401 (5)	1.8574 (11)	1.6746 (4)
NPM	0.0625 (8)	0.0865 (10)	0.5794 (6)	0.8401 (5)	1.8574 (11)	1.6746 (4)

TABLE 3.2.2.4

Dynamic One-Step Ahead Forecasts: Single Factor Spot Rate Models

1 Month Euro-Libor Deposits (1:90-12:99)

Statistics (Rank in Parentheses)

	MSE	MAE	DVS	NDS	WDS	TU
GPM	0.0285 (1)	0.0186 (3)	0.4760 (1)	1.0348 (1)	0.9937(1)	1.5515 (1)
DDM	0.0326 (4)	0.0184 (1)	0.4559 (4)	0.9911 (4)	1.6857(2)	1.5559 (2)
PSM	0.0316 (3)	0.0189 (4)	0.4542 (6)	0.9874 (6)	2.0318(5)	1.5559 (2)
NLD	0.0333 (6)	0.0190 (5)	0.4583 (3)	0.9963(3)	1.7512 (3)	1.5559 (2)
BDT	0.0331 (5)	0.0190 (5)	0.4661 (2)	1.0133(2)	1.7578 (4)	1.5559 (2)
CKLS	0.0314 (2)	0.0188 (3)	0.4545 (5)	0.9882(5)	2.0209 (7)	1.5559 (2)
CIR	0.0431 (7)	0.0258 (7)	0.4535 (7)	0.9859(7)	2.0209 (7)	1.9961 (3)
CEV	0.0361 (8)	0.0237 (6)	0.4545 (5)	0.9882(5)	2.0209 (6)	1.5559 (2)
VM	0.0480 (9)	0.0279 (8)	0.4491 (8)	0.9763(8)	2.2056 (8)	1.5559 (2)
ABM	0.0485 (11)	0.0034 (9)	0.0920 (10)	0.8580 (9)	3.1360 (9)	3.4650 (4)
QTSM	0.0479 (10)	0.0036 (10)	0.0925 (9)	0.8270 (10)	3.3650 (11)	3.4650 (4)
NPM	0.0854 (12)	0.0038(11)	0.0962 (11)	0.8180 (11)	3.1630 (10)	3.4650 (4)

of the forecasting measures. The non-parametric model (NPM) tends to perform in the middle for three of the four interest rates, except the 3MLD. The QTSM model seems to perform the weakest, except for the Japanese short term rates. Most traditional models (e.g., CIR, CEV, and CKLS) seem to underperform in forecasting all four interest rates by most measures. The displaced diffusion (DD), Pearson-Sun, and Black-Derman-Toy models seem to perform fairly well and better than the NPM. The general parametric model (1FGPM) seems to have the best forecasting performance for all four short-term interest rates.

This demonstrates the performance of several continuous time interest rate models that have been developed in the last fifteen years. Most of the models have only been tested for U.S. short-term rates. Stanton (1997) tests the performance of a model nested in our 1FGPM for U.K. short-term rates but did not use GMM estimators and relied on Guassian approximate maximum likelihood. My extension of all the existing models to other global interest rates provides us with a test of robustness of the models. We also test the effectiveness of the existing models to a non-parametric model and find that it performs better than half the existing models.

4. Tests of the Discount Bond Term Structure.

In this section, we implement GMM and out-of-sample forecasting performance tests of alternative approaches to modeling term structure. We build upon the model of Chan et al (1992), who compare various one-factor models of the short rate. We compare alternative models using the Generalized Method of Moments (Hansen, 1982) and evaluate the term structure dynamics on a cross-section of U.S. government bond yields. Later, we extend the analysis to multiple factors, as in Longstaff et al (1992). The authors develop a general equilibrium model of the term structure, incorporating two factors: the short-term riskless rate of interest and its volatility. Contingent claims valuation is examined by means of closed-form expressions for discount bond options. In Longstaff et al (1992), Bollerslev's (1986) G.A.R.C.H. is utilized as a proxy for the interest rate volatility in a manner consistent with a 2-factor model, and the GMM framework is utilized to test the cross-sectional restrictions imposed by the model. I extend this by incorporating a non-parametric model for the interest rate volatility and comparing the in and out-of-sample forecasting performance of this model to single-factor models of the term structure. This generates forecasts that can be compared by MSE and MAE statistics.

4.1 Principal Components Analysis

Principal components analysis (PCA) is a purely statistical approach to modeling the yield curve. I linearly relate the yields on US T-Bond and T-Bills to a set of independent unobservable factors. I follow the approach of Bliss (1997), in which estimates of the factors are extracted from the variance-covariance matrix of yield changes. This is in contrast to the maximum likelihood approach (Johnson et al (1982)), which assumes the joint normality of yields, a questionable assumption (see the data analysis of yields in section 2.2). The innovation that we make is the use a time varying variance-covariance matrix. This approach has the advantage of simultaneously modeling imperfect correlation between yields of various maturities and the time variation in their variances. This leads to superior pricing performance (see Rebonatto (1996)). We construct a time-varying variance-covariance matrix using a *constant-correlation multivariate G.A.R.C.H.* specification (CCM-GARCH), in which the variance of the i^{th} yield change is:

$$Y_{i,t} - Y_{i,t-1} = \alpha_{1i} + \alpha_{2i} Y_{i,t-1} + \alpha_{3i} \sigma_{i,t-1} + \eta_{i,t} \quad i = 1,..,k \tag{4.1.1}$$

$$\sigma^2_{ii,t} = \beta_{1i} + \beta_{2i} \sigma^2_{ii,t-1} + \beta_{3i} \eta^2_{t-1} \tag{4.1.2}$$

$$\sigma_{ij,t} = \gamma_{ij} \sigma_{ii,t} \sigma_{jj,t} \quad i \neq j \tag{4.1.3}$$

$$\boldsymbol{\eta}_t \sim N(\mathbf{0}_k, \Sigma_t) \tag{4.1.4}$$

In contrast to a time-varying correlation set-up, this saves significant degrees of freedom. This structure allows for yield changes to be autocorrelated, depend on the conditional yield volatility, and exhibit autocorrelation in its variance. With the CCM-GARCH estimator of the factor loadings $\hat{\mathbf{L}}^*_t$ in hand, we estimate the underlying factors from the following relation:

$$\hat{\mathbf{X}}_t = \left(\hat{\mathbf{L}}_t^{\cdot T}\hat{\boldsymbol{\Sigma}}_t^{-1}\hat{\mathbf{L}}_t^{\cdot T}\right)^{-1}\hat{\mathbf{L}}_t^{\cdot T}\hat{\boldsymbol{\Sigma}}_t^{-1}\left(\mathbf{Y}_t - \bar{\mathbf{Y}}\right) \tag{4.1.5}$$

Where $\bar{\mathbf{Y}} = \mathrm{T}^{-1}\sum_{t=1}^{T}\mathbf{Y}_t$ is the vector of sample average yield-changes.

The estimation results of the CCM-GARCH model are shown in Table 4.1.1. The parameter estimates are in general statistically significant. The signs of the coefficients in equation (4.1.1), the expected yield changes, indicate that yield changes are mean reverting, and positively related to the lagged standard deviation of yield changes. The estimates for equation (4.1.2), the conditional standard deviation of the yield changes, indicate that the variance is autocorrelated and increasing in the lagged yield change residual. Finally, the estimates of the correlations between yield changes are of similar magnitude to the simple estimates in Table 2.2.1. However, the exponential decorrelation pattern (i.e., the decrease in correlation between a maturities is faster between closer maturities more slowly for further maturities) is more pronounced as compared to the simple estimates. This is a consequence of accounting for the random nature of the yield volatilities in the estimation. Such a pattern has been observed for many different interest rate markets across time (Rebonato, 1996)

TABLE 4.1.1

Constant-Correlation Multivariate Generalized Autoregressive Conditionally Heteroscedastic Estimation (CCM-GARCH)

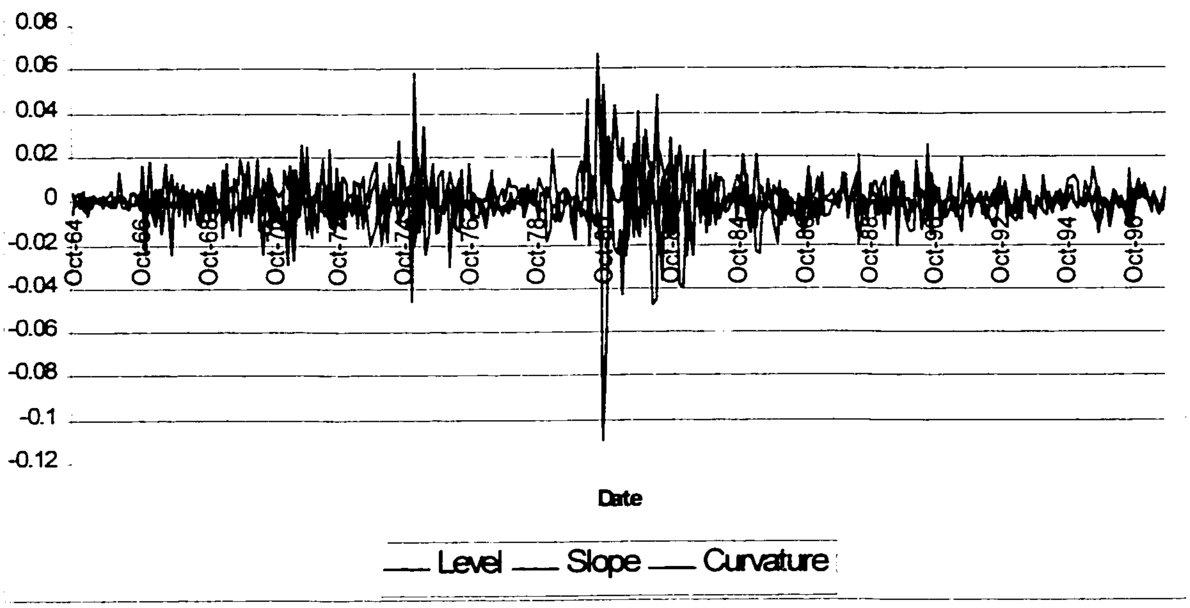
U.S. T-Bond Annualized Continuously Compounded Yield Changes

1 Month to 5 Years Maturities (6:64-10:97)

Month	1	3	6	9	12	24	36	48	60
α_1	3.1×10^{-5}	1.2×10^{-3}	1.5×10^{-3}	2.1×10^{-3}	2.3×10^{-3}	2.4×10^{-3}	2.1×10^{-3}	1.9×10^{-3}	1.5×10^{-3}
	(0.00)^{c}	(0.01)^{b}	(0.01)^{c}	(0.01)^{c}	(0.01)^{c}	(0.00)^{c}	(0.01)^{c}	(0.04)^{b}	(0.23)
α_2	-0.0803	-0.0170	-0.0181	-0.0259	-0.0310	-0.0278	-0.0199	-0.0214	-0.0151
	(0.00)^{c}	(0.06)^{a}	(0.12)	(0.04)^{b}	(0.03)^{b}	(0.05)^{b}	(0.19)	(0.29)	(0.62)
α_3	3.6×10^{-3}	-9.9900	-18.48	-18.757	-11.747	-18.858	-29.421	-11.392	-17.717
	(0.07)^{a}	(0.35)	(0.29)	(0.27)	(0.52)	(0.44)	(0.33)	(0.76)	(0.81)
β_1	0.1154	-3.4×10^{-6}	-4.0×10^{-6}	-4.2×10^{-6}	-3.6×10^{-6}	-4.8×10^{-6}	-7.0×10^{-6}	-1.0×10^{-5}	-8.9×10^{-6}
	(0.43)	(0.00)^{c}	(0.00)^{c}	(0.01)^{c}	(0.06)^{c}	(0.04)^{c}	(0.03)^{b}	(0.00)^{c}	(0.02)^{c}
β_2	0.2142	0.4437	0.5193	0.5393	0.5470	0.4553	0.3235	0.2043	0.3178
	(0.00)^{c}	(0.01)^{c}	(0.00)^{c}	(0.00)^{c}	(0.01)^{c}	(0.00)^{c}	(0.07)^{a}	(0.16)	(0.23)
β_3	7.6×10^{-6}	0.4437	0.2801	0.2622	0.2699	0.2309	0.2370	0.1945	0.1082
	(0.01)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.01)^{c}	(0.01)^{c}
$\gamma_{i,1}$	1	0.6136	0.5712	0.5336	0.5029	0.4724	0.4522	0.4591	0.6757
		(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}
$\gamma_{i,3}$	-	1	0.8959	0.8689	0.7877	0.7376	0.6938	0.6619	0.7621
			(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}
$\gamma_{i,6}$	-	-	1	0.9674	0.9194	0.8736	0.8371	0.7979	0.7733
				(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}
$\gamma_{i,9}$	-	-	-	1	0.9681	0.9161	0.8801	0.8584	0.8301
					(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}
$\gamma_{i,12}$	-	-	-	-	1	0.9463	0.9105	0.8618	0.8614
						(0.00)^{c}	(0.00)^{c}	(0.00)^{c}	(0.00)^{c}
$\gamma_{i,24}$	-	-	-	-	-	1	0.9615	0.9222	0.9081
							(0.00)^{c}	(0.00)^{c}	(0.00)^{c}
$\gamma_{i,36}$	-	-	-	-	-	-	1	0.9569	0.9227
								(0.00)^{c}	(0.00)^{c}
$\gamma_{i,48}$	-	-	-	-	-	-	-	1	0.9642

a,b, and c denotes statistical significance at the 10%, 5%, and 1% levels, respectively.

Figure 4.1.1: Term Structure PCA Factor Estimates (U.S. Treasuries 10:64-10:97)



4.1.1 Results of the Data Analysis

Principal components analysis (PCA) with time varying factor loadings, using a CCM-GARCH time varying variance-covariance matrix, is applied to yields-to-maturity on U.S. bills and bonds for the period 6/64 to 10/97. Rates for 3,6, and 9 months and 1 to 5 years are expressed in annualized, continuously compounded form. This data set is based upon a procedure due to Fama (1982), and subsequently updated by the Center for Research in Security Prices (CRSP), in which zero-coupon yields are extracted from the prices of traded government debt.

The results of the PCA analysis for the time-varying factor loadings are shown in Tables 4.1.1.1 (for various maturities under one year) and 4.1.1.2 (for maturities ranging from two to five years). The first Factor (i.e., "level"), with a mean loading varying between 0.256 (for five year maturity) and 0.377 (for five year months). Both the mean and the standard deviation are consistent across maturities. The second factor's (i.e., "slope") mean increases with maturity but the volatility is the same across all maturities. The mean factor loading is negative for short maturities up to six months and rises steadily to 2.70 for four years. The third factor (i.e., "curvature") is non-monotonic, with a peak mean for one-year maturity and a negative mean for nine-month maturity. The standard deviation for the third factor is much lower than the first two. The last panel in Table 4.1.1.2 presents the percentage of variation explained by the factors and is time-varying. The level factor explains over 80% of the variation, the second explains over 10%, and the third factors explain less than 5%. The variation (i.e., standard deviation) is highest for the first factor and lowest for the third factor.

Evidence in regard to higher order properties of the distribution of the factor loadings are gleaned by inspecting the Kolmogorov-Smirnov (KS), Ljung-Box Q, and Phillips-Peron Z statistics. These are shown in the lower three rows of each panel of Table 4.1.1. The KS statistics soundly reject normality for all factors and all maturities. The Q statistics are quite large and in all cases which leads us to accept the autocorrelation (i.e., if the slope factor loading has a relatively high value in one period, perhaps due to increased inflationary expectations, then it is expected that it will have a

Table 4.1.1.1
Principal Components Analysis of the U.S. Government Bond Term Structure
Annualized Continuously Compounded 1 Month to 1 Year Maturity Yields
(6:1964-10:1997)

Factor Loading Statistics

	Factor 1 (Level)	Factor 2 (Slope)	Factor 3 (Curvature)
1 Month			
Mean	0.366	-0.801	0.449
Standard Deviation	0.01348	0.01067	0.00582
K-S (normality)	0.1727	0.1589	0.1684
Q (autocorelation)	64.808	67.796	52.932
Z (stationarity)	-18.987	-18.991	-308.14
3 Months			
Mean	0.377	-0.215	0.442
Standard Deviation	0.01309	0.00934	0.00589
K-S (normality)	0.17979	0.14978	0.19647
Q (autocorelation)	65.264	65.378	49.491
Z (stationarity)	-18.422	-17.637	-308.14
6 Months			
Mean	0.374	-0.411	0.374
Standard Deviation	0.01252	0.00978	0.00445
K-S (normality)	0.18811	0.14290	0.18628
Q (autocorelation)	64.804	68.697	55.818
Z (stationarity)	-17.396	-19.992	-318.76
9 Months			
Mean	0.315	0.112	-0.270
Standard Deviation	0.01256	0.01012	0.00596
K-S (normality)	0.18641	0.14925	0.18766
Q (autocorelation)	66.657	65.489	44.979
Z (stationarity)	-18.539	-17.135	-286.40
1 Year			
Mean	0.379	0.134	0.517
Standard Deviation	0.01138	0.00949	0.00508
K-S (normality)	0.18349	0.16982	0.15585
Q (autocorelation)	64.214	65.394	44.208
Z (stationarity)	-17.912	-15.514	-303.18

Table 4.1.1.2
Principal Components Analysis of the U.S. Government Bond Term Structure
Annualized Continuously Compounded 2 to 5 Years Maturity Yields
(6:1964-10:1997)

Factor Loading Statistics	Factor 1 (Level)	Factor 2 (Slope)	Factor 3 (Curvature)
2 Years			
Mean	0.315	0.242	0.122
Standard Deviation	0.01233	0.01081	0.00485
K-S (normality)	0.17839	0.13973	0.18719
Q (autocorelation)	66.271	66.948	47.855
Z (stationarity)	-17.414	-15.293	-311.32
3 Years			
Mean	0.281	0.254	0.264
Standard Deviation	0.01200	0.01047	0.00505
K-S (normality)	0.16210	0.13685	0.18242
Q (autocorelation)	66.807	68.737	53.259
Z (stationarity)	-17.196	-15.065	-299.35
4 Years			
Mean	0.256	0.291	0.382
Standard Deviation	0.01123	0.01068	0.00439
K-S (normality)	0.16210	0.15077	0.15837
Q (autocorelation)	66.069	65.908	56.983
Z (stationarity)	-17.785	-19.051	-317.52
5 Years			
Mean	0.236	0.270	0.349
Standard Deviation	0.01331	0.01098	0.00409
K-S (normality)	0.15763	0.13129	0.16989
Q (autocorelation)	66.702	65.848	56.757
Z (stationarity)	-18.567	-17.317	-283.55
Percent Variation Explained			
Mean	81.11%	12.59%	4.03%
Standard Deviation	4.53%	2.49%	1.52%
K-S (normality)	0.2945	0.3729	0.3647
Q (autocorelation)	41.474	59.349	47.369
Z (stationarity)	-23.412	-28.633	-30.103

relatively high value in a subsequent period). Finally, the Z statistics reject that these series are stationary, which means that their distributions have changed significantly over time as the structure of the economy has changed.

4.2 The Pricing of Discount Bonds

In this section, we briefly review a benchmark theoretical model of the term structure and compare it to a non-parametric model. We use a continuous time framework to remain consistent with most of the derivative markets literature. Second, we implement the mathematical results of the existence of an equivalent martingale measure, which under technical conditions is equivalent to the absence of arbitrage. In doing so, the probabilities induced by this equivalent measure are such that the instantaneous expected return on any security is the short rate of interest. For mathematical details, see Duffie (1996) and Oksendahl (1995). In summary, we assume a complete probability space $(\Omega,\Im,P)$, where Ω is a continuous sample space, $\Im$ is the σ-algebra $\sigma(\mathbf{B}_s : s\in[0,T])$ generated by the standard Brownian motions (SBM) that describe the uncertainty in the economy, and P is probability measure defined on $\Im$. The general result is that, by the absence of arbitrage, there exists a probability measure Q, equivalent to P, such that the price of a stochastic unit discount bond $\Lambda_{t,s}$ with time $s > t$ payoff Z = 1 is given by

$$\Lambda_{t,s} = E_t^Q\left[\exp\left(-\int_{u=t}^{s} r_u du\right)\right] \quad (4.2.1)$$

where $E_t^Q(.)$ denote Λ: $[0,T]\times[0,T] \to [0,1]$ expectation under risk neutral measure and $\{r_t : t\in[0,T]\}$ is an adapted short-rate process. The function in equation (4.2.1) is called Λ: $[0,T]\times[0,T] \to [0,1]$ the *term structure of interest rates* (or the yield curve).

4.2.1 Alternative Pricing Models

At this point, we compare two types of term-structure models as well as their pricing given the setting outlined in Section 4.2 . Section 4.2.1.1 briefly summarizes the assumptions two factor Longstaff and Schwartz (1992) model, a variation of the CIR model (see Chapter 2 for details). Section 4.2.1.2 sketches an alternative non-parametric alternative model.

4.2.1.1 The 2-Factor Longstaff and Schwartz Model (2F-LS)

We implement a generalization of the 2-factor CIR model developed by Longstaff and Schwartz (1992). In this equilibrium setting, two stochastically independent factors are proposed:

$$dX_{it} = \kappa_i\left[\varphi_i - X_{it}\right]dt + \sigma_i X_{it}^{\frac{1}{2}} d\tilde{B}_{it} \quad i = 1,2 \quad (4.2.1.1.1)$$

It is assumed that only the first factor influences production uncertainty, while both affect both its expected return, from which it can be shown that the instantaneous short rate and its variance can be written in term of these factors. This leads to the following version of the CIR partial differential equation, which, subject to the appropriate boundary conditions, is satisfied by all interest rate dependent contingent claims:

$$F_t + \left(\frac{\kappa_1 \theta_1}{\sigma_1^2} - (\kappa_1 - \lambda_t)\sigma_1^2 \right) F_1 + \left(\frac{\kappa_2 \theta_2}{\sigma_2^2} - \kappa_2 \sigma_2^2 F_2 \right) F_2 +$$

$$+ \frac{1}{2}\left(\sigma_1^2 X_1^2 F_{11} + \sigma_2^2 X_2^2 F_{22}\right) - rF = 0 \qquad (4.2.1.1.2)$$

Where additional parameters θ_1 and θ_2 are from the production process and λ_t is the time dependent market price of risk. Following Uhrig (1996), we obtain these steady state estimates of these constants from the history of the 10-month T-Bill rate and its non-parametrically estimated volatility, using a simple Gauss-Newton iterative procedure. Finally, substitution of these solutions into (4.2.1.1.2) allows us to solve a parabolic PDE by a standard multiple separation of variables, subject to the appropriate boundary conditions.

4.2.1.2 The Multilayer Perceptron Artificial Neural Network Model (MLP-ANN)

In this section, we discuss a class of non-parametric models, *artificial neural networks*. These models that try to mimic the process of human cognition in developing pricing formulae. They have been shown to compete successfully with parametric counterparts in the pricing of a variety of derivatives, including bonds and options (Malliaris et al (1993), Hutchinson et al (1994)). Hornik et al (1989) has shown that a broad class of neural networks, the *multilayer feed forward* variety constitute so-called *universal approximators*, which means that under mild regularity conditions such networks are capable of representing any non-linear function. White (1989, 1990) has shown that neural networks are amenable to estimation by non-linear regression.

We price discount bonds by estimating an empirical pricing function, thereby making no assumptions regarding the specific functional form followed by the short rate and other term structure variables that are suitable. This involves partitioning the dataset into three subsets used for training, cross-validation, and testing. Parameter values are determined in the training set via non-linear optimization and the cross-validation portion is used to determine which specific network architecture performs best on the patterns in this set. The particular class of networks that we use to price bonds is given by the following version of the multi-layer perceptron artificial neural network (MLP-ANN):

$$F^{MLP}\left(\mathbf{X}_{it}\middle|\boldsymbol{\theta},J\right) = \beta_i + \sum_{j=1}^{J}\left(\frac{\omega_{ij}}{1 + \exp\left(\sum_{k=1}^{K}\omega_{ikj}X_{k,it} + \beta_{ij}\right)}\right) \qquad (4.2.1.2.1)$$

Where $F^{MLP}\left(\mathbf{X}_{it}\middle|\boldsymbol{\theta},J\right)$ denotes the i[th] bond price at time t as approximated by the network with J hidden layers. The 3-dimensional vector $\mathbf{X}_{it} = \left(X_{1,it},..,X_{k,it}\right) = \left(r_t,\hat{V}_t,\tau_{it}\right)^T$ represent the inputs to the network where: k=3 is the fixed number of inputs, r_t is the short rate of interest at time t, $\hat{V}_t$ is the (non-parametrically) estimated volatility of the short rate, and $\tau_{it} = T_i - t$ is the time-to-maturity of the i[th] bond, and $\boldsymbol{\theta} = \left(\{\beta_i\}_{i=1,..,I},\{\beta_j,\omega_j\}_{j=1,..,J},\{\omega_{ikj}\}_{i,k,j=1,..I,K,J}\right)^T$ are the free parameters of the networks. J is chosen experimentally to minimize the root mean squared error (RMSE) of the pricing errors.

4.2.2 Analysis of Pricing Errors

In this section, we compare the two term structure models by measuring their ability to capture the dynamics of zero coupon government bonds. In Tables 4.2.2.1 and 4.2.2.2, we compare the 2F-CIR model with the ANN-MLP model by a statistical examination of the bond pricing errors across various maturities. The results suggest that both models have effective pricing performance on average, as average errors are small and not statistically different from zero, and mean squared errors are approximately a few percentage points. The ANN-MLP model performs better, with slightly lower mean pricing errors (ranging from 0.08 to 13 bps as compared with 0.09 to 17 bps in the 2F-CIR model), as well as significantly lower MSEs (ranging from 18 bps to 4.32% as opposed to 24 bps to 5.78% in the 2F-CIR), at each maturity. The MAE statistics display a pattern similar to the MSE statistics.

Both models exhibit marked non-normality (excess skewness and excess kurtosis), but this is more severe in the 2F-CIR model, which can be seen in the higher Kolmogorov-Smirnov (KS) normality statistics in the later model. This implies that there may have been more instances of abnormally high or low pricing errors in the 2F-CIR model as opposed to the ANN-MLP model. Average pricing errors, MSE, MAE, as well as KS statistics deteriorate monotonically with maturity. This holds for both the ANN-MLP and the 2F-LS models. This suggests that it is more difficult to price longer maturities with either of these models. The explanation for this may be that more factors are needed to price long maturity bonds.

Finally, the tests of more general statistical properties-the Ljung-Box Q-statistics, Phillip-Perron Z-statistics and Geweke, Porter, & Hudack tests-suggest that the errors in both models are autocorrelated and possess long-memory, but are stationary. The magnitude of these test statistics are similar across all maturities and both models. Given that these errors are a by-product of models that exploit short and long term patterns in bond prices, the autocorrelation and long-memory patterns in the pricing errors are not surprising.

TABLE 4.2.2.1

Distributional Statistics and Diagnostic Tests

Longstaff & Schwartz 2 Factor CIR Model Bond Pricing Errors (78:2-97:10)

Months	3	6	9	12	24	36	48	60
Mean	-9.9×10^{-4}	-1.2×10^{-4}	-1.7×10^{-4}	-2.5×10^{-4}	0.004	30.00	0.011	0.017
Std. Err.	6.4×10^{-4}	1.8×10^{-4}	2.7×10^{-4}	4.0×10^{-4}	0.004	0.005	0.018	0.039
Skewness	-1.518[c]	-0.979[c]	-0.895[c]	-0.847[c]	-0.771	-0.783[c]	-1.27[c]	-0.192
Kurtosis	5.586[c]	5.376[c]	4.602[c]	4.375[c]	0.119	-0.294	1.113[c]	-0.413
KS[1]	0.449[c]	0.576[c]	0.586[c]	0.595[c]	0.902[c]	0.911[c]	0.879[c]	0.721[c]
LB-Q[2]	113.1[c]	119.8[c]	136.4[c]	205.3[c]	2449.3[c]	2993.9[c]	2569.2[c]	47.63[c]
PP-Z[3]	-184.2[c]	-166.19[c]	-143.84[c]	-104.11[c]	-9.921[b]	-4.0677	-7.929	-29.36[c]
GPH[4]	0.646[c]	0.499[c]	0.164[c]	0.234[c]	0.963c	1.048[c]	1.035[c]	0.451[c]
MSE	0.025	0.0704	0.0873	0.1958	1.576	1.447	4.876	5.779
MAE	0.0198	0.0451	0.0373	0.0666	1.291	1.353	3.732	4.474

a,b,c denote statistical significance at the 1%, 5%, and 10% levels, respectively.

1-Kolmogorov-Smirnov (normality), 2-Ljung-Box Q-statistics (autocorrelation), 3-Phillip-Perron Z-statistics (stationarity), 4-Geweke,Porter,&Hudack test (long-memory)

TABLE 4.2.2.2

Distributional Statistics and Diagnostic Tests

Multilayer Perceptron Neural Network Model Bond Pricing Errors (78:2-97:10)

Months	3	6	9	12	24	36	48	60
Mean	-8.8×10^{-5}	-1.0×10^{-4}	-1.03×10^{-4}	-1.9×10^{-4}	0.001	0.003	0.009	0.001
Std. Err.	4.6×10^{-4}	1.4×10^{-4}	2.1×10^{-4}	3.6×10^{-4}	0.003	0.004	0.011	0.021
Std. Err.	4.6×10^{-4}	1.4×10^{-4}	2.1×10^{-4}	3.6×10^{-4}	0.003	0.004	0.011	0.021
Skewness	-1.429[c]	-0.871[c]	-0.871[c]	-0.818[c]	-0.721	-0.758	-1.27[c]	-1.02[c]
Kurtosis	4.692[c]	5.692[c]	4.692[c]	3.914[c]	0.098	0.143	0.140	1.041
KS[1]	0.296[c]	0.409[c]	0.465[c]	0.499[c]	0.558[c]	0.637[c]	0.662[c]	0.735[c]
LB-Q[2]	112.69[c]	115.57[c]	138.84[c]	204.90[c]	273.8[c]	315.3[c]	474.1[c]	367.8[c]
PP-Z[3]	-102.25[c]	-178.28[c]	-198.62[c]	-150.55[c]	-132.09[c]	-172.8[c]	-137.4[c]	-157.0[c]
GPH[4]	0.585[c]	0.554[c]	0.184[c]	0.248[c]	1.032[c]	0.943[c]	0.974[c]	0.167[c]
MSE	0.0180	0.0532	0.0629	0.1400	1.0127	1.2132	3.4691	4.3244
MAE	0.01296	0.0451	0.0203	0.0273	1.0291	0.9049	3.0556	3.1247

See Table 4.2.2.1.

4.3 A Comparison of Model Performance

4.3.1 Hedging Interest Rate Risk and Forecasting the Term Structure

It is easy to derive hedging formulae in this setting. First, consider the 2F-CIR model. We assume that it is possible to trade in two short maturity bonds, one and three month maturities, such that the two sources of uncertainty-unanticipated changes in the short rate and its volatility-can be spanned. Define the estimated factor sensitivities for maturity τ by $\frac{dF(r_t, V_t, \tau)}{dr_t} \approx \hat{b}^t_\tau$ and $\frac{dF(r_t, V_t, \tau)}{dV_t} \approx \hat{c}^t_\tau$. Denote the two short maturities by $s_1 < s_2 < \tau$ and their sensitivities by $\frac{dF(r_t, V_t, s_1)}{dr_t} \approx \hat{b}^t_{s_1}$ and $\frac{dF(r_t, V_t, s_2)}{dr_t} \approx \hat{b}^t_{s_2}$. Let the hedging coefficients (estimated at time t), for the respective hedging of τ maturity bonds with s_1 and s_2 short bonds, be respectively denoted by $\hat{\Delta}_t(s_1, \tau)$ and $\hat{\Delta}_t(s_2, \tau)$. These are:

$$\hat{\Delta}_t(s_1, \tau) = \frac{\hat{b}_t^\tau \hat{a}_t^{s_2} - \hat{a}_t^\tau \hat{b}_t^{s_2}}{\hat{a}_t^{s_1} \hat{b}_t^{s_2} - \hat{a}_t^{s_2} \hat{b}_t^{s_1}} \tag{4.3.1.1}$$

$$\hat{\Delta}_t(s_2, \tau) = \frac{\hat{a}_t^\tau \hat{b}_t^{s_1} - \hat{b}_t^\tau \hat{a}_t^{s_1}}{\hat{a}_t^{s_1} \hat{b}_t^{s_2} - \hat{a}_t^{s_2} \hat{b}_t^{s_1}} \tag{4.3.1.2}$$

Similarly, we derive closed-form expressions for factor sensitivities, and ultimately hedge ratios, for the management of interest rate risk in an MLP-ANN setting. As in the case of the 2F-CIR model, assume that it is possible to trade in two short maturity bonds, denoted by $s_1 < s_2 < \tau$ (e.g., 1- and 3-month maturities), such that the two sources of uncertainty-unanticipated changes in the short rate and its volatility-can be spanned. Define the estimated neural network factor sensitivity of maturity τ to factor i (= r,v for the short rate and its volatility, respectively) as $\dfrac{dF^{net}(\mathbf{X}_{t,\tau})}{dX_{it}} \triangleq \hat{\beta}^{t}_{i,\tau}$. Differentiating:

$$\frac{dF^{net}\left(\mathbf{X}_{t,\tau}\middle|\boldsymbol{\theta}\right)}{dX_{it}} = -\omega_{ij}\sum_{j=1}^{J}\omega_j\left(\frac{1}{1 + \exp\left(\sum_{i=1}^{I}\omega_{ij}X_{it} + \beta_j\right)}\right)^2 \tag{4.3.1.3}$$

Let the neural network hedge coefficients (estimated at time t), for the hedging of τ maturity bonds with $s_1 < s_2 < \tau$ short bonds, be respectively denoted by $\hat{\Delta}^{net}_t(s_1,\tau)$ and $\hat{\Delta}^{net}_t(s_2,\tau)$. These are:

$$\hat{\Delta}^{net}_t(s_1,\tau) = \frac{\hat{\beta}^{\tau}_{t,v}\hat{\beta}^{s_2}_{t,r} - \hat{\beta}^{\tau}_{t,v}\hat{\beta}^{s_2}_{t,v}}{\hat{\beta}^{s_1}_{t,r}\hat{\beta}^{s_2}_{t,v} - \hat{\beta}^{s_2}_{t,r}\hat{\beta}^{s_1}_{t,v}} \tag{4.3.1.4}$$

$$\hat{\Delta}^{net}_t(s_2,\tau) = \frac{\hat{\beta}^{\tau}_{t,r}\hat{\beta}^{s_1}_{t,v} - \hat{\beta}^{\tau}_{t,v}\hat{a}^{s_1}_{t,v}}{\hat{\beta}^{s_1}_{t,r}\hat{\beta}^{s_2}_{t,v} - \hat{\beta}^{s_2}_{t,r}\hat{\beta}^{s_1}_{t,v}} \tag{4.3.1.5}$$

4.3.2 Analysis of Hedging and Forecast Errors

A comparison of the hedging positions between the 2F-CIR and ANN-MLP models is shown in Tables 4.3.2.1 and 4.3.2.2. The results suggest that both models have effective hedging performance on average, as mean pricing errors are small and not statistically different from zero, and mean squared errors are on the order of a few basis points. However, the MSE statistics are approximately 10 times larger in the ANN-MLP model (ranging from 3 to 86 bp) than the 2F-CIR model (ranging 1 to 9 bps). The comparison of MAE statistics is similar. This suggests that the 2F-CIR model yields superior hedging performance.

In both cases, the hedging positions exhibit marked non-normality (excess skewness and excess kurtosis), which can be seen in the high Kolmogorov-Smirnov (KS) normality statistics. However, this is more pronounced in the non-parametric model. Significant negative skewness in both models implies that large losses may have occurred in the sample period, with such extreme losses being larger in the non-parametric model as opposed to the parametric model.

Average hedging errors, MSE, MAE, as well as KS normality statistics deteriorate monotonically with maturity. This holds for both the ANN-MLP and the 2F-LS models. This suggests that it is more difficult to hedge longer maturities with either of these models. The explanation for this may be that more factors are needed to price long maturity bonds (e.g., they may be more sensitive to a "curvature" factor). We conclude that the parametric 2F-CIR model provides superior hedging performance as compared to the non-parametric ANN-MLP model. The reason for this is related to the nature of these models. Non-parametric models are designed primarily to fit prices in-sample, whereas parametric models are based upon long-term theoretical relationships which are expected to hold over longer time periods.

We also test the forecasting ability of various bond-pricing models. We generate 1-step ahead bond yield forecasts from four different bond-pricing models.[22] In the case of the single factor general parametric model (1F-GPM) the interest rate forecasts of the previous analysis are used with the appropriate PDE to numerically determine a forecasted prices and yields. For the ANN-MLP and 2F-CIR models, 1-step-ahead forecasts from the non-parametric kernel estimates of the drift and diffusion functions of the short are plugged into the re-estimated models at each point in time. Finally, for the *principal components analysis-constant correlation multivariate GARCH* (PCA-CCMGARCH) model, the time t estimated factors are forecast 1-step ahead in the PCA-CCMGARCH model, and then plugged back into the factor model to forecast the time t+1 yield.

We compare the forecasting performance of the models across maturities different in Table 4.3.2.3. We find that 2F-CIR does best at shorter maturities while the PCA-CCMGARCH is superior at longer maturities. This is consistent with our finding in the bond pricing and hedging analysis that the two factor models tend to do worse at longer

---

22

We did not analyze the pricing or hedging of the PCA-CCMGARCH model, because it is meant to fit bond yields and not prices, nor of the 1F-GPM, in that we defer comparing single to multiple factor models for Chapter 4.

maturities, and a third factor may be more significant for such maturities. The 1F-GPM is usually the worse performing model. The MLP-ANN is usually second best across all maturities and measures. In all cases, the MSE, MAE, and U increase (i.e., worsen) deteriorate as maturity increases, but the other directionally oriented measures (e.g., the direction variatinal symmetry or DVS) exhibit no such pattern. This suggests that there may be an omitted factor that is more effective with increased maturity.

TABLE 4.3.2.1

Distributional Statistics and Diagnostic Tests

Longstaff & Schwartz 2 Factor CIR Model Bond Hedging Errors (78:2-97:10)

Month	6	9	12	24	36	48	60
Mean	3.3×10^{-4}	1.1×10^{-3}	-2.0×10^{-3}	-3.4×10^{-3}	-2.2×10^{-3}	-9.2×10^{-3}	-2.1×10^{-4}
Std. Err.	4.0×10^{-4}	1.2×10^{-3}	2.7×10^{-3}	6.0×10^{-3}	4.8×10^{-3}	2.7×10^{-3}	1.6×10^{-3}
Skewness	-4.1689^{c}	0.2018^{c}	-3.7413^{c}	-2.7357^{c}	-2.2104^{c}	-1.9107^{c}	-0.9952^{c}
Kurtosis	51.981^{c}	33.819^{c}	56.650^{c}	52.872^{c}	50.454^{c}	50.454^{c}	22.910^{c}
KS	0.5758^{c}	0.5862^{c}	0.5953^{c}	0.5953^{c}	0.5953^{c}	0.5953^{c}	0.5953^{c}
LB-Q	119.82^{c}	136.38^{c}	205.34^{c}	205.34^{c}	205.34^{c}	205.34^{c}	205.34^{c}
PP-Z	-166.19^{c}	-143.84^{c}	-104.11^{c}	-104.11^{c}	-104.11^{c}	-104.11^{c}	-104.11^{c}
GPH	0.0049	0.0701	-0.9879	0.1940	0.7638^{c}	0.2618	-0.67811
BJ- J	2.8×10^{4c}	1.1×10^{5c}	3.2×10^{5c}	2.8×10^{5c}	2.5×10^{5c}	1.9×10^{5c}	5.2×10^{4c}
MSE	6.21×10^{-3}	0.0210	0.0411	0.0931	0.0732	0.0423	0.0142
MAE	2.2×10^{-3}	5.8×10^{-3}	0.0107	0.02446	0.0225	0.0173	0.0241

a,b,c denote statistical significance at the 1%, 5%, and 10% levels, respectively.

1-Kolmogorov-Smirnov (normality), 2-Ljung-Box Q-statistics (autocorrelation), 3-Phillip-Perron Z-statistics (stationarity), 4-Geweke,Porter,&Hudack test (long-memory)

TABLE 4.3.2.2

Distributional Statistics and Diagnostic Tests

Multilayer Perceptron Neural Network Model Bond Hedging Errors (78:2-97:10)

Month	6	9	12	24	36	48	60
Mean	2.8×10^{-4}	-1.7×10^{-2}	-1.1×10^{-2}	-8.1×10^{-3}	-1.0×10^{-2}	-1.9×10^{-2}	-1.2×10^{-3}
Std. Err.	2.0×10^{-4}	1.2×10^{-2}	8.2×10^{-2}	6.2×10^{-3}	3.2×10^{-2}	2.3×10^{-2}	1.9×10^{-3}
Skewness	9.693^{c}	-9.776^{c}	-9.753^{c}	-9.656^{c}	-0.335^{c}	-7.567^{c}	-5.994^{c}
Kurtosis	92.93^{c}	95.18^{c}	94.84^{c}	93.54^{c}	87.14^{c}	67.13^{c}	49.08^{c}
KS	0.3150^{c}	0.0208	0.0220	0.0257^{c}	0.0412	0.0683	0.0602
LB-Q	$47.98^{c}5$	1.595^{c}	51.60^{c}	49.787^{c}	47.39^{c}	41.27^{c}	32.46^{c}
PP-Z	-86.09^{c}	-190.2^{c}	-189.3^{c}	-189.4^{c}	-185.9^{c}	-200.6^{c}	-200.5^{c}
GPH	0.2425	0.2425	-0.0716	-0.0707	0.3514	-0.0586	-0.0438
BJ- J	7.2×10^{4c}	7.5×10^{5c}	7.5×10^{5c}	7.3×10^{5c}	6.1×10^{5c}	3.8×10^{5c}	2.0×10^{5c}
MSE	2.7×10^{-2}	0.17059	0.1138	0.8581	0.4426	0.3146	0.02576
MAE	2.8×10^{-3}	1.9×10^{-2}	0.0148	0.1464	0.1226	0.1289	0.1286

See Table 4.3.2.1.

TABLE 4.3.2.3
Dynamic One-Step Ahead Bond Yield Change Forecasts
Alternative Bond Pricing Models-3 Month to 1 Year Maturities
(relative ranking of models in parentheses)

	MSE	MAE	DVS[5]	NDS[6]	WDS[7]	Theil's U
3 Month						
2F-CIR[1]	9.3×10^{-7} (1)	5.6×10^{-7} (1)	0.4869 (1)	0.9490 (3)	0.9428 (1)	2.7131 (2)
MLP-ANN[2]	4.0×10^{-6} (2)	1.2×10^{-6} (2)	0.4513 (2)	0.9832 (2)	1.4642 (3)	0.1483 (3)
1F-GPM[3]	1.0×10^{-4} (3)	4.3×10^{-5} (3)	0.2276 (4)	1.0230 (1)	1.0111 (2)	0.0305 (4)
CCMGARCH[4]	8.4×10^{-2}(4)	8.4×10^{-2} (4)	0.3487 (3)	0.7640 (4)	1.9633 (4)	86.52 (1)
6 Month						
2F-CIR[1]	6.0×10^{-6} (2)	2.8×10^{-6} (1)	0.4974 (1)	1.0000 (1)	0.9423 (1)	4.0691 (2)
MLP-ANN[2]	5.3×10^{-6} (1)	4.8×10^{-6} (2)	0.4744 (2)	0.9686 (2)	1.0431 (2)	0.1711 (3)
1F-GPM[3]	2.6×10^{-4} (3)	1.1×10^{-4} (3)	0.2225 (3)	0.9560 (3)	1.2828 (4)	0.0503 (4)
CCMGARCH[4]	8.7×10^{-2} (4)	8.7×10^{-2} (4)	0.4949 (4)	0.8502 (4)	0.9467 (3)	165.9 (1)
9 Month						
2F-CIR[1]	1.7×10^{-5} (2)	7.5×10^{-6} (2)	0.4712 (3)	1.1605 (2)	0.9423 (3)	0.9161 (3)
MLP-ANN[2]	5.8×10^{-6} (1)	2.7×10^{-6} (1)	0.2046 (4)	0.7354 (4)	1.0431 (2)	6.5241 (2)
1F-GPM[3]	4.6×10^{-4} (3)	1.9×10^{-4} (3)	0.5564 (1)	1.3262 (1)	1.2828 (1)	0.0536 (4)
CCMGARCH[4]	0.0861 (4)	0.0861 (4)	0.5282 (2)	0.8957 (3)	0.8294 (4)	263.60 (1)
1 Year						
2F-CIR[1]	3.6×10^{-5} (3)	1.1×10^{-5} (3)	0.4712 (3)	0.9890 (2)	1.1699 (3)	0.1111(3)
MLP-ANN[2]	1.7×10^{-5} (2)	8.6×10^{-6} (2)	0.5641 (1)	0.9735 (3)	1.1545 (2)	34.79 (1)
1F-GPM[3]	1.6×10^{-3} (4)	7.1×10^{-4} (4)	0.2200 (4)	1.0238 (1)	1.1745 (4)	0.0220 (4)
CCMGARCH[4]	2.6×10^{-6} (1)	1.5×10^{-6} (1)	0.5000 (2)	0.8442 (4)	0.9939 (1)	1.7891 (2)

1-2 Factor Cox-Ingersoll-Ross Model of Longstaff & Schwartz, 2-Artificial Neural Network-Multilayer Perceptron Model , 3-Single Factor General Parametric Model 4-Principal Components Analysis Multivariate Constant Correlation GARCH , 5-Directional Variational Symmetry, 6-Normalized Directional Symmetry, 7-Weighted Directional Symmetry

TABLE 4.3.2.4

Dynamic One-Step Ahead Bond Yield Change Forecasts

Alternative Bond Pricing Models-2 to 5 Year Maturities

(relative ranking of models in parentheses)

	MSE	MAE	DVS[5]	NDS[6]	WDS[7]	Theil's U
2 Year						
2F-CIR[1]	4.3×10^{-4} (3)	2.1×10^{-4} (3)	0.5393 (3)	0.9717 (3)	0.7913 (3)	0.0183 (2)
MLP-ANN[2]	3.5×10^{-5} (2)	2.0×10^{-6} (2)	0.5769 (1)	0.9912 (1)	0.6976(1)	0.5812 (1)
1F-GPM[3]	3.0×10^{-3} (4)	1.3×10^{-3} (4)	0.2123 (4)	0.9765 (2)	1.2643 (4)	0.1070 (3)
CCMGARCH[4]	1.2×10^{-6} (1)	7.6×10^{-7} (1)	0.5539 (2)	0.9310 (4)	0.7224 (2)	0.4585 (2)
3 Year						
2F-CIR[1]	4.1×10^{-3} (3)	2.2×10^{-3} (4)	0.5602 (1)	0.9907 (1)	0.6979 (1)	0.0132 (3)
MLP-ANN[2]	2.0×10^{-4} (2)	8.2×10^{-5} (2)	0.5180 (2)	0.9528 (3)	0.9826 (2)	0.6194 (2)
1F-GPM[3]	4.4×10^{-3} (4)	2.0×10^{-3} (3)	0.2225 (4)	0.9775 (2)	1.1469 (3)	7.0×10^{-3}(4)
CCMGARCH[4]	9.3×10^{-7} (1)	5.9×10^{-7} (1)	0.4539 (3)	0.7763 (4)	1.1534 (4)	1.8602 (1)
4 Year						
2F-CIR[1]	5.6×10^{-2} (4)	3.4×10^{-2} (4)	0.5497 (1)	1.0096 (2)	0.7248 (1)	0.0030 (3)
MLP-ANN[2]	2.9×10^{-3} (2)	1.3×10^{-3} (2)	0.5282 (2)	1.0619 (1)	0.8493 (2)	0.0264 (2)
1F-GPM[3]	5.7×10^{-3} (3)	2.7×10^{-3} (3)	0.2276 (4)	0.9570 (3)	1.1431 (3)	4.6×10^{-3}(4)
CCMGARCH[4]	7.2×10^{-7} (1)	5.5×10^{-7} (1)	0.4641 (3)	0.8153 (4)	1.2129 (4)	2.0143 (1)
5 Year						
2F-CIR[1]	1.9×10^{-3} (4)	6.1×10^{-4} (4)	0.3927 (3)	0.9740 (3)	2.0320 (4)	0.0585 (4)
MLP-ANN[2]	1.9×10^{-4} (3)	5.7×10^{-4} (3)	0.4923 (1)	0.9846 (2)	1.2185 (3)	0.1260 (3)
1F-GPM[3]	9.7×10^{-5} (2)	3.9×10^{-5} (2)	0.4872 (2)	0.8597 (4)	1.1989 (2)	0.1470 (2)
CCMGARCH[4]	1.9×10^{-6} (1)	4.9×10^{-7} (1)	0.2251 (4)	1.0000 (1)	1.1625 (1)	3.0192 (1)

1-2 Factor Cox-Ingersoll-Ross Model of Longstaff & Schwartz, 2-Artificial Neural Network-Multilayer Perceptron Model , 3-Single Factor General Parametric Model 4-Principal Components Analysis Multivariate Constant Correlation GARCH , 5-Directional Variational Symmetry, 6-Normalized Directional Symmetry, 7-Weighted Directional Symmetry

5. SUMMARY AND CONCLUSION

In this chapter we have analyzed tested various term structure models, at the level of various international short rates of interest as well as the U.S. Treasury discount bond term structure. For short-term interest rates, we have contributed to the literature by considering a wider array of models (including non-parametric alternatives) than previously done, as well as by extending the tests to international markets. In the case of bond prices, we have added a comparison of non-parametric and statistical to traditional equilibrium term structure models, and have extended the analysis from pricing to hedging and yield forecasting.

For the short-term interest rates examined, GMM tests of various models and different international markets support the conclusion that most of the popularly used models are mis-specified. First, the variation in results across different rates, in terms of explaining the non-parametrically estimated drift and diffusion, show the several models do not seem to fit other short rates globally. Second, the conclusion in the existing literature that most models explain the drift function better than the volatility function is not universal. Third, in agreement with CKLS (1990), and in contrast to literature that accommodates more parametricized approaches, we find that the effect of the level of interest rates is more significant than previously believed, when we allow for a more general diffusion function. Fourth, in agreement with economic theory, we find that both mean reversion and non-linearity is a significant feature of short rate drift processes, in agreement with the non-parametric finance literature (Ait-Sahalia 1995). Finally, the non-nested quadratic term structure model (Boyle et al 1999) is found to generally under perform other models, despite its ability to model a mean reverting and non-linear drift.

In the analysis of the term structure, based upon US treasury bills and bonds, we find that fundamentally different approaches differ in their ability to fit bond prices as opposed to hedging and forecasting interest rate risk. We compare a traditional equilibrium approach (the Longstaff and Schwartz two factor CIR model: 2F-LS) to a non-parametric approach (the artificial neural network-multilayer perceptron model: ANN-MLP). We find that the non-parametric model to be superior in fitting bond prices, while the equilibrium model better hedges interest rate risk. In a comparison of interest rate forecasting performance of these two approaches to a statistical model (principal components analysis-constant correlation multivariate G.A.R.C.H.: PCA-CCMGARCH) and a short rate diffusion model (the single factor generalized parametric model:1F-GPM), we find several differences as well. The equilibrium model forecasts best at shorter maturities, the statistical model does so best longer maturities, while the single factor model tends to under perform across all maturities. However, we find that the two factor models, both equilibrium and non-parametric, perform worse by all measures and for all purposes (i.e., pricing as well hedging and forecasting) as maturity lengthens. This implies that additional factors may be necessary in order to model the term structure across a wide range of maturities.

In Chapter 4, we advance this study of the term structure to the level of derivatives interest rate dependent instruments. This is accomplished by testing single and multiple factor spot-rate and forward-rate models, as well as a non-parametric model, using bond options data. We attempt to what factors and models best explain the pricing of interest rate dependent derivatives and which ones perform best in hedging and forecasting. This is relevant to figuring out the best stochastic model for the term structure, to determine which type of model is best suited for interest rate risk management, and to determine which theories of the term structure are best supported by the data. As we will see, while many of these models may be related even mathematically equivalent, they tend to diverge widely in their practical implementation.

CHAPTER 4

A COMPARISON OF ALTERNATIVE INTEREST RATE DERIVATIVE PRICING MODELS

1. INTRODUCTION AND DISCUSSION

In this chapter, we test single and multiple factor spot-rate and forward-rate models, as well as a non-parametric model, using bond options data. We attempt to determine what factors and models best explain the pricing of interest rate dependent derivatives and which ones perform best in hedging and forecasting. This is relevant for several reasons. First, one needs to figure out the best stochastic model for the term structure. Second, it will help to determine which type of model is best suited for interest rate risk management. Third, there is a need to determine which theories of the term structure are best supported by the data. While many of these models may be related, if not mathematically equivalent, they tend to diverge widely in their practical implementation. The existing literature sheds little light on the factors that best describe the economic reality of the term structure. Very few studies have attempted to address this level of generality. One exception is Buhler et al (1999), in which competing models are compared using German warrants in the 1990-1993 period. We extend the existing literature to a different data set, the CBOT bond futures options.

Among the parametric models, we examine different stochastic structures.

Among the *spot rate models*, we distinguish between the single and multiple factor variants. The single factor spot rate models are driven by a generalization of Chan et al (1992; henceforth CKLS) stochastic differential equation, where we allow for non-linearity in the drift as well as a displaced diffusion function. This nests more term structure models than CKLS does, including the popular Black-Derman-Toy (1990; henceforth BDT) and Pearson-Sun (1994, henceforth PS) models. The multiple factor spot rate models include those with both two and three underlying factors. The most general model that we develop is a variant of the three-factor *affine term structure model* (ATSM) with a jump diffusion component. The number of factors is motivated by empirical evidence that factors associated with the level, the steepness and the curvature of the yield curve explain 98% of the term structure variation. The jump component is motivated by evidence that interest rates go through regime shifts (Bliss (1997)). We also consider two nested three factor models, one a correlated Gaussian model with unit variance log-normally distributed factors, and the other a multi-factor CIR model with hetereroscedastic factors following independent square-root processes. The other multiple factor models are restricted to having two factors. In the first of these, we generalize the Schaefer and Schwartz (1984) model, which identifies the two factors as the short rate and the difference between the long-rate and the short-rate. The second model postulates two unobservable factors that are linearly related to the short rate, which under equilibrium assumptions results in a term structure that depends upon the short rate and its volatility. This is an extension of the Longstaff and Schwartz (1992) model, which is a two-factor variant of the CIR model. The commonality among these models is that they price interest rate dependent derivatives in either a stochastic interest rate or a stochastic volatility setting. Among the forward rate models, we consider three-factor, two-factor, and one-factor variants of the HJM model, with constant as well as linear proportional volatility functions in each case. The single factor constant volatility model is a continuous time version of Ho and Lee (1986), in which forward rates are Gaussian, while the proportional volatility variant involves diffusion functions that depend upon time-to-maturity. In the multiple factor variants, we employ principal component analysis to empirically determine the unobservable factors, and their volatilities.

Since the focus of this study is risk measurement and control capabilities of alternative pricing models, we propose the following criteria for evaluating these models. First, since valuation models within a risk management system should be capable of forecasting future options prices, the criterion of *predictability* is of relevance. Second, there is the hedging performance of the model, or what we call its *replicability*. This refers to the practice of fitting model to market prices (e.g., by finding the implied volatility that matches the Black-Scholes price to a market option price). This is useful in pricing similar derivatives in another market or pricing the same derivative a short time later. While this is an important criterion for practitioners and has been the focus of traditional academic literature, it is not the only criterion for our purposes. Therefore, our tests are not just about market efficiency (i.e., assuming the model is true, how well does it price options?), but also for model quality (i.e., how well does the model manage risk in a broader context?). A third criterion, that we call *stability*, is the model's ability to value derivatives consistently across time. This involves the estimation of parameters in a time series setting, as opposed to the cross-sectional determination of parameters from derivative prices at each point in time. We evaluate these models for consistency, as opposed to whether the given model is valid. This avoids using parameter estimates that are biased toward a particular model and is in line with our focus on the predictability of options prices within a risk management system. By not using information contained in the derivatives markets in extracting these parameter values, we are able to accommodate large deviations between true and observed option prices. This approach constitutes a "global" test, as opposed to a "local" test, of competing derivative valuation models. Finally, we test the computational *cost* of implementing the model. This includes the computer resources needed to estimate model parameters, the ability of the model to the current term and volatility structure, and the complexity of the numerical algorithm used in valuing derivatives.

This chapter is organized as follows. In Section 2, we review the theoretical literature on interest rate derivative valuation models. Section 3 reviews the empirical models, while section 4 presents the empirical results of our study. Section 5 concludes this chapter.

2. REVIEW OF THE LITERATURE

2.1 Theoretical Models of the Term Structure

2.1.1 The General Equilibrium Approach and Spot Rate Models

Merton (1973) presents a stochastic term structure, a version of the Black-Scholes model in which the discount bond and stock price processes follow a bivariate geometric Brownian motion. Black (1976) derives a formula for options on default-free discount bond futures, assuming a constant short rate and volatility. Vasicek (1977) models a stochastic interest rate process (in an Ohrnstein-Uhlenbeck mean reverting diffusion) to derive closed form solutions for discount bonds and options on bonds. Cox, Ingersoll, and Ross (1985; henceforth CIR) present an intertemporal continuous time equilibrium model of the term structure under complete markets, from which they derive a square-root diffusion for the short rate and closed form solutions for interest rate derivatives. Longstaff and Schwartz (1992) extend the CIR model to two factors, the instantaneous short-term interest rate and its volatility, resulting in a parsimonious model that captures many empirically observed characteristics of the term structure.

A recent development in term structure research is the *affine term structure models* (ATSMs), variations of which are equivalent to several popular spot rate models, including Black (1976), Vasicek (1977), and CIR (1985). Brown and Schaeffer (1994) examine these and Duffie and Kan (1996) generalize them. The equilibrium short rate is assumed to be an affine (i.e. linear) function of state variables, from which it follows that yields on discount bonds are linearly related to the instantaneous short rate, with term structure parameters depending on time-to-maturity.[1] This set-up has been generalized to accommodate default risk, stochastic volatility, and jump components. However, the existing empirical evidence on ATSMs has not been encouraging. Duffee (1998) reports that affine models lack predictive power in explaining yield changes. Kimmel (1999) has similar findings in the context of stochastic volatility specifications of the term structure. Dai and Singleton (2000) present an exhaustive analysis of ATSMs, deriving conditions under which generalizations of popular models are *admissible* (i.e., the short rate is non-negative and discount bond prices are less than par). The authors find that the principle limitations of ATSMs lies in the trade-off between allowing time variation in volatility and imperfect correlation among state variables. This leads to an alternative research approach, *quadratic term structure models* (Gallant et al 2000; henceforth QTSMs), which posit that the short rate is a quadratic function of the state variables. While this guarantees a positive interest rate and a non-linear term premium, it is computationally inefficient.

More recent research in option pricing has centered on *stochastic volatility models*

---

[1] This can be derived by standard pricing techniques, such as expectation under equivalent martingale measure or solution of the Ricatti ordinary differential equations (ODEs).

(SVMs). Volatility is a key variable in pricing term structure derivatives. Chernov and Ghysels (1999) present a review of issues in SVM analysis, and point out that it is not feasible to estimate the risk-neutral parameters of SVM models from asset prices alone, since volatility is non-traded.[2] Recent attempts to estimate SVMs have focused on the Heston (1993) *affine-jump diffusion* class of models (AJDs). This is extended by Duffie et al (1998), using transform analysis and by Bakshi et al (1998) beyond the affine class of models. The parallel econometric developments are the *simulated method of moments* (SMM) of Duffie and Singleton (1993) and *efficient method of moments* of Gallant et al (1996, 2000). These techniques allow us to achieve maximum likelihood efficiency of parameter estimation in the presence of unobserved state variables (i.e., the volatility) and an unknown likelihood function. We apply the EMM technique to a different version of the AJD class of option models, and compare the results to alternative models.

2.1.2 The No-Arbitrage Approach and Forward Rate Models

The earlier models have been referred to in the literature as *arbitrage-free*, and we call this the "traditional arbitrage-free approach". These include the works of Vasicek (1977), Richard (1978), Dothan (1978), and Brennan and Schwartz (1979). As pointed out by several authors (CIR (1985), HJM (1992), and Back (1997)), there is a distinction

---

2

Risk neutral parameters can be estimated from asset prices alone if one assumes that volatility risk is idiosyncratic or unpriced, as in Hull and White (1987), which is at odds with empirical evidence (Bates, 1996).

between this approach and the absence of arbitrage opportunities, or what is called the "no-arbitrage approach". The possibility of arbitrage in the earlier models arises because of the exogenous specification of market price of risk. Although these models impose a no-arbitrage condition to price contingent claims, the risk premia may not be consistent with preferences in the market, and the model might in fact be susceptible to arbitrage. Equilibrium pricing (e.g., CIR (1985)), in which risk premia are endogenous, are not susceptible to arbitrage in this sense.[3] However, if one starts from the imposition of no-arbitrage and allows the risk premia to be endogenous, it is unnecessary to impose an equilibrium structure.

The no-arbitrage alternative to equilibrium models is first presented by Ho and Lee (1986), who derive a valuation model in a discrete time binomial framework, that allows for an exact fit to the current spot term structure. BDT (1990) and Hull and White (1990) extend this analysis with models that are capable of matching the current volatility structure. Heath et al (1990) extend the Ho and Lee model to a continuous time process followed by the instantaneous forward rates. HJM (1992) derive a generalized no-arbitrage theoretical model of contingent claims valuation and derive general results that hold for all arbitrage-free yield curve models. Starting with the initial term structure of forward rate and a class of stochastic processes for its evolution, and using the

---

3

This is equivalent to assuming the existence of a risk neutral probability measure, which the earlier models do not.

methodology of the equivalent martingale measure, prices of continuous claims can be derived. The conditions of Harrison & Kreps (1979) guarantee the existence and uniqueness of an equivalent martingale probability measure, which yields valuation formulae for interest sensitive contingent claims that do not depend explicitly upon the market price of risk.

2.2 Tests of Interest Derivative Models: Implementation and Estimation

Direct tests of interest rate option valuation models are limited in comparison to the broader empirical and theoretical literature on interest rate models. Dietrich-Campbell et al (1986) test the two-factor model of Brennan and Schwartz (1982; henceforth 2FBS) using options on U.S. Treasury bonds and Treasury bills. Brown and Dybvig (1986) test the single factor version of Cox et al (1985; henceforth 1FCIR), finding weak support using a linear model approximation. Cakici (1989) tests single factor valuation models of Black (1976) and Barone-Adesi & Whaley (1987), assuming different stochastic processes for the short rate, on CBOT bond futures options. Gibbons and Ramaswamy (1993) test the 1FCIR model using GMM, finding support for the model. Flesaker (1994) tests a constant volatility version of the HJM model using GMM, on CME Eurodollar futures options. While the simulation analysis shows the GMM results to be significant, the model performs poorly using actual data. Jordan et al (1995) analyze option-pricing values implicit in callable Treasury bonds, in the framework of Longstaff (1992). They employ an alternative empirical approach and resolve the puzzle of negative values in these embedded derivatives that Longstaff had found. Longstaff et al (1993) implement the 2FLS model. A simplified procedure is used to fit the initial term structure parameters exactly, allowing a time varying market price of risk, to price interest rate caps and estimate their implied volatilities. Moraleda and Pelsser (1997) use market prices of daily caps and floors for 1993 and 1994 to compare spot and forward rate models. The spot models of Hull and White (1994), Pelsser (1994), and Black and Karasinski (1991) are compared to the Gaussian, square root, and proportional forward rate model of Ritchken and Sankarasubramanian (1995). They find that all spot interest rate models outperform their forward rate counterparts. Buhler et al (1998) test the market for German warrants, using seven spot and forward-rate models with both one and two factors. They identify one forward-rate model (a single factor proportional volatility HJM model) and two spot rate models (a single factor extended CKLS and a two factor CIR model) that outperform the other four models.

3. REVIEW OF EMPIRICAL MODELS

In this section, we present the models to be estimated empirically. This involves two phases. We refer to the first process as *preselection.* This involves an extensive statistical analysis of the market for U.S. treasury bonds and their yields, as well as evaluates different theoretical models of the term structure. The second stage involves summarizing the main features of the various models: single versus multiple factor models on one hand and spot rate versus forward rate models on the other. These are then compared to non-parametric models, which offer a benchmark that is independent of the assumptions regarding the stochastic structure of the term structure.

3.1 Preselection of Models and Statistical Analysis of the Data

3.1.1 Data Analysis

We start by specifying the basic characteristics of the valuation models. First, the number of factors driving the term structure is specified. For the sake of tractability, this ranges between one and three. Second, we consider different types of factors: latent, observable, and statistically or theoretically motivated. Third, we make an assumption regarding the underlying stochastic process driving the term structure, which is equivalent to a specification of the functional forms of the underlying factors. This is determined by a comprehensive data analysis of U.S. Treasury bill and bond yields by *principal components analysis* (PCA) for the period 6:64 to 10:98, in which we employ a constant correlation multivariate G.A.R.C.H. (CCM-GARCH) model. Due to a high degree of correlation among yields, we find that three factors are sufficient to describe over 98% of the variation in the term structure.

3.1.2 Single Factor Models

In single factor forward rate models, the parameterization of forward rate volatility is a critical factor. The current literature has focused primarily on one and two parameter variants, where it has been shown that the number of parameters is more important than the particular form of the model (Amin et al, 1994). While two-parameter models with linear proportional volatility best fit the data in and out of the sample, the one-parameter model results in more stable estimates and better hedging performance. We follow Buhler et al (1999) in considering two single factor HJM models, a one-parameter constant volatility[4] and a two-parameter linear proportional volatility variant. In the case of spot rate models, the specification of the process for the short rate is critical. Motivated by the findings of Chan et al (1992), we expect interest rate changes to be directly related to their level. There is also evidence of non-linearity in the drift of interest rate changes (Ait-Sahalia (1996)). To account for this, we replace the linear drift function $\alpha + \beta r_t$[5] used in Chan et al (1992), by the non-linear specification $\alpha + \beta r_t + v r_t \log(r_t)$. In the case of the diffusion function, the magnitude of interest-rate movements are directly related to the level of the short rate. In addition, theoretical considerations dictate that the volatility of the short rate should not vanish at low levels of the short rate (Duffie (1996)). Hence, we nest the constant elasticity of variance (CEV) diffusion function of CKLS[6], σr_t^{γ}, with the displaced specification $(\delta + \sigma r_t)^{\gamma}$, where $\delta > 0$

4

This is the continuous time limit of the Ho and Lee (1986) model.

5

In this context, we interpret $-\beta$ as the speed of adjustment parameter, and $-\alpha/\beta$ as the long run mean of the short rate. Therefore, we must impose the restriction that $\alpha > 0$ and $\beta <$ 0.

6

echnically, this is defined as: $dt^{-1} \lim_{s \downarrow t} \mathrm{Var}_t[r_s]$

is the parameter of displacement. This generalized specification is such that it allows us to nest continuous time versions of the popular Black-Karasinki (1991), Pearson-Sun (1994), and Black- Derman-Toy (1990) models.

3.1.3 Multifactor Models

As shown earlier, there is evidence that most of the variation in the term structure can be attributed to 2 and 3 factors. The first factor has similar impact (or "loading") across maturities, and is interpreted as a "level factor". The second has a loading that varies directly with maturity, with influences of opposite signs at the long and short ends of the term structure, and is interpreted as a "slope factor". The third factor has a loading profile that peaks at intermediate maturities, resulting in twists of the yield curve, and can thus be interpreted as a "curvature factor".[7]

In the context of forward rate models, the multifactor models that we consider have constant and linear proportional volatility functions. We empirically determine the functional forms of these, by calculating the volatility of factors using principal components analysis. In the case of spot rate models, we develop a three-factor affine

---

7 The latter two factors, attributed to steepness and curvature of the yield curve, are related to the concepts of duration and convexity discussed in the traditional fixed income literature.

term structure model with a jump diffusion component (3F-AJDATSM). This family of models nests most of the widely studied models, which have closed-form solutions[8], i.e. the CIR and Guassian models. The two factor models that we consider include the CIR model of Longstaff and Schwartz (1992) and the Schaefer and Schwartz (1984) model. The former identifies the random volatility of the short-rate, while the latter identifies the slope of the yield curve as the second factor. Both these models are special cases of the affine class of term structure models, as analyzed by Duffee and Kan (1996), in which the drift and volatility functions of the short rate are linear in the factors, and parameters are time dependent.

3.2 Review of the Models

3.2.1 Forward Rate Models: The HJM Approach

The HJM model takes the entire yield curve as an input, from which the short rate and interest rate derivatives are derived, by no-arbitrage. In contrast, spot rate models specify the process of state variables and impose an equilibrium condition, from which an endogenous term structure is derived. The no-arbitrage requirement means that the yield

---

8

To be more specific, in ATSMs the short rate follows a *linear stochastic differential equation*, in where a solution (should it exist) reduces to a system of *ordinary differential equations* (ODEs). In contrast, non-linear SDE models require numerical calibration to a discretized *partial differential equation* (PDE).

curve can be described in terms of the forward rates prevailing at any point in time. We assume that the instantaneous forward rate follows an Ito process:

$$df(t,s) = \mu(f(t,s)t,s)dt + \boldsymbol{\sigma}^T(f(t,s)t,s)d\hat{\mathbf{B}}_t \quad \text{for} \quad 0 \le t \le s \le T \tag{3.2.1.1}$$

where the terms are defined in Chapter 2. Heath et al (1992) show that, in the absence of arbitrage restrictions, the drift $\mu(t,T)$ of the forward rate process under risk neutral measure is determined by the vector of volatility functions $\boldsymbol{\sigma}(t,T,f(t,T)) = \left(\sigma_1(t,T,f(.)),...,\sigma_d(t,T,f(.))\right)^T$:

$$\mu(t,s) = \boldsymbol{\sigma}^T(t,s)\int_{u=t}^{s} \boldsymbol{\sigma}(t,u)du \tag{3.2.1.2}$$

We consider various specializations of this. The Absolute Volatility (HJM-AVI) model proposes a constant volatility function. In the Proportional Volatility (HJM-PVI) model, volatility is linear in the time-to-maturity. The multifactor versions of these-the Absolute Volatility II (HJM-AVII), Proportional Volatility II (HJM-PVII), Absolute III (HJM-AVIII), and Proportional Volatility III (HJM-PVIII) models-are straightforward generalizations. The first step in implementing these models is to estimate the current yield curve by means of a cubic spline function, to produce the initial forward rate curve $f(0,T)$. To estimate the volatility parameters, we use principal component analysis and calculate the volatilities of the independent factors derived from the estimated factor loadings. The second step is computing option prices. We perform an Euler discretization of (3.2.1.1), construct a (non-recombining) binomial tree under risk-neutral measures, and apply backward induction. Similar to Amin et al (1994) and Buhler et al (1999), we find that seven time steps are necessary to achieve accurate option prices.

3.2.2 Spot Rate Models and the CIR Approach

The spot-rate models are based on the *fundamental valuation equation* derived by Cox et al (1985) in a general equilibrium setting. Let there be a k-vector of independent factors, which under actual probability measure follows the Ito process:

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t,t)dt + \boldsymbol{\Sigma}(\mathbf{X}_t,t)d\mathbf{B}_t, \tag{3.2.2.1}$$

where the terms are defined in Chapter 2. We assume that the value of a contingent claim, given the state of the economy, is a sufficiently well behaved function $F(\mathbf{X}_t,t,T)$ for $0 < t \leq T \leq T^*$ that must satisfy the following partial differential equation:

$$F_t + \sum_{i=1}^{k}\left[F_{X_i}(\mu_i - \Theta_i\sigma_i) + \frac{1}{2}F_{X_iX_i}\sigma_i^2\right] - rF = 0, \tag{3.2.2.2}$$

where $\boldsymbol{\Theta}(\mathbf{X}_t,t) = (\Theta_1(\mathbf{X}_t,t),..,\Theta_k(\mathbf{X}_t,t))^T$ denotes the market prices of risk for the state vector. The absence of arbitrage implies that $\boldsymbol{\Theta}:\mathbb{R}^k \rightarrow \mathbb{R}\times\mathbb{R}^k$ is a maturity independent vector valued function of only the state variables. The values for contingent claims can be found by solving equation (3.2.2.2), a parabolic PDE, subject to the appropriate initial and boundary conditions, for k=1,2, and 3.

3.2.2.1 Single Factor Spot Rate Models

The single factor models that we consider are all nested within the following stochastic differential equation, which incorporates most of the parametric continuous-time models in the literature:

$$dr_t = \left[\alpha(\tau) + \beta(\tau)r_t + \nu(\tau)r_t\log(r_t)\right]dt + \left[\delta(\tau) + \sigma(\tau)r_t\right]^{\gamma(\tau)}d\hat{B}_t \qquad (3.2.2.1.1)$$

where $\tau = T - t$ is the time-to-maturity and the parameter vector $\Theta^T = (\alpha,\beta,\gamma,\delta,\sigma,\nu):[0,T] \rightarrow \mathbb{R}^6$ depends at most on τ. This is referred to as the *single factor general parametric model* (1F-GPM). This is a generalization of the CKLS (1992) formulation. These do not have a closed form expression for discount bond prices, and hence must be numerically calibrated to a PDE. These models have the advantage of being able to model non-linearities in the drift or diffusion functions. This matches the empirical observation that term premia may be non-linear in state variables (Stambaugh (1988)). An alternative specialization of equation (3.2.2.1.1) is the *single factor affine spot rate* model (1F-ASR), in which the drift and diffusions functions are linear in the short rate. Here the yields-to-maturity on discount bonds are affine in the short rate. It can be shown that the solution is *time homogeneous*, which means that the parameters of the term structure may depend upon time-to-maturity, but not calendar time. A computationally appealing aspect of this model may be obtained by solving a system of ordinary differential equations.[9] Another important model nested in equation (3.2.2.1.1) is the *Gaussian model*, in which short rates over any finite set of times are distributed multivariate normal. The Black (1976) and Vasicek (1977) models are also nested within equation (3.2.2.1.1). As in Hull and White (1990), in order to match an exogenously given schedule of discount bond prices, we allow the market price of risk $\Theta(t)$ to be time dependent.[10] This is consistent with several hypotheses of the term structure, which predict that term premia may vary with both time and time-to-maturity. This gives a natural economic interpretation of the calibration process and overcomes some technical difficulties.[11] Estimates of parameters of the short-rate drift and volatility function for the spot rate models are determined by the *efficient method of moments* (Gallant (2000); henceforth EMM), which is based upon simulation and a Hermite expansion of a

---

9

Equation (3.2.2.1.1) is a one-dimensional example of an important class of SDEs known as *linear stochastic differential equations*, which have the property that an explicit solution can always be found in terms of a solution to a related system of *ordinary differential equations* (ODEs), which is much easier to solve than a PDE.

10

If we assume a second parameter to exhibit time-dependency, it is possible to calibrate the model to the current volatility structure as well. However, there is evidence that such a procedure results in future volatility estimates that are not only unstable, but are of unrealistic magnitudes (Uhrig et al (1996)). Therefore, we achieve a two-point calibration, by exactly matching the volatilities of the long- and short-rates, while interpolating the volatilities of intermediate maturities.

11

Technical reasons for this choice are related to the existence of a solution for the fitting function, their consequences of this for the endogenous volatility structure, and the relation between the risk-neutralized and original probability measure (Heath et al (1992), Uhrig et al (1996)).

Gaussian transition density.[12] We then use a numerical algorithm to solve versions of the PDE (3.2.2.1.1) by calibrating the model to the current yield curve and long-rate volatility.[13] These are solved subject to the maturity condition for zero coupon bonds:

$$F(r_T, T, T) = 1 \tag{3.2.2.1.2}$$

The fitting condition for bond prices, which matches endogenous to observed zero bond prices $\hat{F}(\bullet)$, conditional on the initial short rate r_0, is given by

$$F(r_0, 0, T) = \hat{F}(T) \quad T \in (0, T^*] \tag{3.2.2.1.3}$$

The final condition that we impose upon estimated bond prices to match the initial term structure is a fitting condition for long rate volatility:

$$-\frac{F_r(r_0, T^*, T^*)}{T^* \, F(r_0, T^*, T^*)} = \frac{\hat{\sigma}_{T^*}}{\hat{\sigma}_0}, \tag{3.2.2.1.4}$$

$\hat{\sigma}_{T^*}$ is the historical volatility of the longest maturity (T^*) yield and $\hat{\sigma}_0$ is the sample estimator of the short-rate volatility. The left-hand-side of equation (3.2.2.1.4) is the ratio

---

12

This has the advantage of avoiding the aggregation bias inherent in moment estimation of an Euler discretization of the SDE. See Ait-Sahalia (1999) for an approximate ML type estimation similar to EMM.

13

We use the the *inverted implicit finite difference* method (Uhrig et al (1996) and Buhler et al (1999)), a version of the implicit finite difference method that has been shown superior with respect to convergence and stability. These are related to tree techniques, such as the trinomial algorithm of Hull and White (1993), which are used to solve similar problems.

of instantaneous long- to short-rate volatility.

3.2.2.2 Multifactor Spot Rate Models

The most general multifactor spot-rate model that we consider is in the class of ATSMs. Duffie and Kan (1996) first studied this class of models, which yield essentially closed-form expressions for the prices of discount bonds. Dai and Singleton (1999) examine the technical conditions under which the latter yields admissible prices for discount bonds, and conduct an empirical analysis of such models, using the Simulated Method of Moments (SMM) of Duffie et al (1993).

3.2.2.2.1 Three Factor Spot-Rate Models

The model developed here is a generalization of various 3-factor models that have appeared in the recent literature, which are in the affine class and admit a jump-diffusion component. We call this the *three-factor general affine jump-diffusion term structure model* (3F-GAJDM), a version of Duffie et al (1998). We assume that the vector of term-structure factors $\mathbf{X}_t = \left(\log(r_t), X_{1t}, X_{2t}\right)^T$ follows:

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t, t)dt + \boldsymbol{\Sigma}(\mathbf{X}_t, t)d\mathbf{B}_t + d\mathbf{Z}_t \qquad (3.2.2.2.1.1)$$

The drift function is

$$\mu(\mathbf{X}_t,t) = \begin{pmatrix} r_0 \\ \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} 0 & \mu_1-1/2 & \mu_2-1/2 \\ 0 & -\kappa_1 & 0 \\ 0 & 0 & -\kappa_2 \end{pmatrix} \begin{pmatrix} r_t \\ X_{1t} \\ X_{2t} \end{pmatrix} \tag{3.2.2.2.1.2}$$

The diffusion function satisfies

$$\Sigma(\mathbf{X}_t,t)\Sigma^T(\mathbf{X}_t,t) = \begin{pmatrix} X_1+cX_2 & \rho_1\sigma_1X_1 & \rho_1\sigma_1X_1 \\ \rho_1\sigma_1X_1 & \sigma_1^2X_2 & \sigma_2^2X_2 \\ \rho_2\sigma_2X_2 & 0 & d\sigma_2^2x_2 \end{pmatrix} \tag{3.2.2.2.1.3}$$

These relations can be expressed equivalently in terms of univariate differentials as

$$dr_t = \left(\mu_0 + \left(\mu_1 - \frac{1}{2}\right)X_{1t} + \left(\mu_2 - \frac{1}{2}\right)X_{2t}\right)dt + \sqrt{X_{1t} + cX_{2t}}dB_{rt} \tag{3.2.2.2.1.4}$$

$$dX_{it} = \left(\theta_i - \kappa_iX_{it}\right)dt + \sigma_i\sqrt{X_{it}}dB_{it} \qquad i = 1,2 \tag{3.2.2.2.1.5}$$

$$dr_t^2 = \left(X_{1t} + cX_{2t}\right)dt \tag{3.2.2.2.1.6}$$

$$dX_{it}^2 = \sigma_i^2X_{it}dt \qquad i = 1,2 \tag{3.2.2.2.1.7}$$

$$dr_tdX_{it} = \rho_i\sigma_iX_{it}dt \qquad i = 1,2 \tag{3.2.2.2.1.8}$$

$$dX_{1t}dX_{2t} = 0 \tag{3.2.2.2.1.9}$$

Equation (3.2.2.2.1.4) states that the non-jump component of the short rate follows a type of mean reverting process in both the state variables. The unobservable state variables follow ordinary square root processes and are instantaneously uncorrelated. The univariate jump-diffusion component of $\mathbf{Z}_t = (Z_t \ 0 \ 0)^T$ is described by the following:

$$dZ_t = -\lambda(\mathbf{X}_t,t)\mu_J dt + \log(1+J_t)dq_t \tag{3.2.2.2.1.10}$$

$$dq_t \sim P(\lambda(\mathbf{X}_t)dt) \tag{3.2.2.2.1.11}$$

$$\lambda(\mathbf{X}_t,t) = \lambda_0 + \lambda_1 X_{1t} + \lambda_2 X_{2t} \tag{3.2.2.2.1.12}$$

$$\log(1+J_t) \sim N\left(\log(1+\mu_J)-\frac{1}{2}\sigma_J^2,\ \sigma_J^2\right) \tag{3.2.2.2.1.13}$$

Equation (3.2.2.1.1.11) says that the increments to q_t are distributed as Poison with instantaneous mean arrival rate of λ per unit time, which is a linear function of the latent state variables, and is also proportional to the drift of dZ_t. The jump size is determined by the log-normal variable $1+J_t$. The short rate is the following linear function of the latent factors:

$$r_t = r_0 + r_1 X_{1t} + r_2 X_{2t} \tag{3.2.2.2.1.14}$$

From equations (3.2.2.2.1.4),(3.2.2.2.1.5) and (3.2.2.2.1.14) the dynamics of r_t can be expressed as:

$$dr_t = \left(\left(r_1\theta_1 + r_2\theta_2\right) + r_1\kappa_1 X_{1t} + r_2\kappa_2 X_{2t}\right)dt + \left(r_1\sigma_1\sqrt{X_{1t}}\right)dB_{1t} + \left(r_2\sigma_2\sqrt{X_{2t}}\right)dB_{2t} \qquad (3.2.2.2.1.15)$$

Comparing this to equation (3.2.2.2.1.4) yields the following parameter restrictions:

$$r_1 = \frac{1}{\sigma_1} \qquad (3.2.2.2.1.16)$$

$$r_2 = \frac{c}{\sigma_2} \qquad (3.2.2.2.1.17)$$

$$r_0 = \mu_0 = \frac{\theta_1}{\sigma_1} + \frac{c\,\theta_2}{\sigma_2} \qquad (3.2.2.2.1.18)$$

$$\mu_1 = \frac{1}{2} - \frac{\kappa_1}{\sigma_1} \qquad (3.2.2.2.1.19)$$

$$\mu_2 = \frac{1}{2} - \frac{\kappa_2}{c\,\sigma_2} \qquad (3.2.2.2.1.20)$$

Nested models[14] in this class that are also used to value interest rate derivatives are three factor versions of the Guassian Term Structure Model (3F-GTSM) and the Cox-Ingersoll-Ross model (3F-CIRTSM). The mathematical details of these 3-factor models are

---

14 This model nests several of the stochastic volatility option pricing models that have appeared in the last 10 years, mostly in areas other than the pricing of interest rate options. This study is the first to apply this to term structure derivatives.

covered in Chapter 2.

3.2.2.2.2 Two Factor Spot-Rate Models

We consider 2-factor spot rate models of two types: the 2F-LS (2-factor CIR model of Longstaff and Schwartz (1992)) and 2F-SS (2-factor model of Schaefer and Schwartz (1984)). In the 2F-LS model, two stochastically independent factors are included:

$$dX_{it} = \kappa_i\left[\varphi_i - X_{it}\right]dt + \sigma_i X_{it}^{\frac{1}{2}} d\tilde{B}_{it} \quad i = 1,2 \tag{3.2.2.2.2.1}$$

Under the assumption that only the first factor influences production uncertainty and affect both its expected return in conjunction with time additive preferences, the state variables can be expressed in terms of the short rate and its volatility. These variables are then substituted into the CIR partial differential equation and bond options are priced. We follow Uhrig (1996) to estimate parameters from the unconditional mean and variance of the spot rate and its non-parametrically estimated volatility. In the 2F-LS model, the two factors are posited as the short rate and the spread $s_{t.} = r_t - l_t$ between the short rate r_t and the long rate l_t. The short rate process is the difference between the two factor processes:

$$r_t = l_t - s_t \tag{3.2.2.2.2.2}$$

The dynamics of the two instantaneously independent factors are derived from equation

(3.2.2.2.2.2) and the market price of risk is chosen such that the yield to maturity on the longest maturity bond is equal to l_t and independent of s_t.

3.2.3 Non-parametric Model

In this section, we consider a model that is not based on a restricted mathematical structure for the stochastic process driving the term structure. However, the choice of factors is still motivated by either theoretical or empirical considerations. The non-parametric model we use is the *multivariate Naturae-Watson kernel estimator* (henceforth MNWKE). Hardle (1990) shows that such a procedure, under certain regularity conditions, results in an estimator for a pricing function that converges asymptotically to the true one. We assume that the price of the i[th] derivative Y_{it}, depends on a k-vector of factors $\mathbf{X}_{it} = \left(X_{i_1 t},..,X_{i_k kt}\right)^T \quad t = 1,..,T;\ i = 1,..,I$. In our application to options on bond futures, this is given by the 5-dimensional vector $\mathbf{X}_{it} = \left(r_t, \hat{V}_t, F_{it}, K_i, \tau_{iT_i}\right)^T$, where r_t is the short rate of interest at time t, $\hat{V}_t$ is the (non-parametrically) estimated volatility of the short rate, F_{it} is the time t price of the i[th] underlying futures contract, K_i is the strike price of the i[th] option, and $\tau_{iT_i} = T_i - t$ is the maturity of the i[th] option. The MNWK estimator of the function F(.) is

$$\hat{F}(\mathbf{x}|\mathbf{h}) = \frac{\sum_{\forall i,t} \prod_{j=1}^{K} K_{j,h_j}\left(x_j - X_{j,it}\right) Y_{j,it}}{\sum_{\forall i,t} \prod_{j=1}^{K} K_{j,h_j}\left(x_j - X_{j,it}\right)} \tag{3.2.3.1}$$

where the function $K_{j,h_j}(x_j)$ is a weighting *kernel* for the j^{th} independent variable, $\mathbf{x} = (x_i,..,x_k)^T \in \mathbb{R}^k$ is a fixed vector of characteristics, and $\mathbf{h} = (h_1,...,h_k)^T$ is a vector of bandwidths that governs the range of the averaging. The procedure for choosing the optimal vector of bandwidths, involves minimizing the *cross-validation function* $CVF_{\mathbf{h}}$ over all bandwidths **h**. This is a weighted average of the squared differences between (3.2.2.1) estimated with the i^{th} observation left out and the $Y_{j,it}$, evaluated at $\mathbf{x} = X_{j,it}$, for some appropriate weighting function that mitigates boundary effects.

4. A COMPARISON OF PRICING RESULTS

Tables 4.1 and 4.2 summarize the spot and forward rate models, respectively, that are tested in this study. Table 4.1 reports the characteristics that distinguish the various spot rate models: acronym (as the names tend to be long in this set), factor description (short rate, spread, or latent), stochastic process for the factors, the endogenously determined market-price-of-risk, and the determination of the short rate (given directly or as a function of the factors). In Table 4.2 (the forward rate models), we list the model type (single versus multiple factors or linear versus proportional volatility), the stochastic process for the forward rate drift, and the stochastic process for the forward rate diffusion .

The estimation of the parameters of these models and the calibration to option prices can be summarized as follows. In the estimation step, we derive parameter values from the time series of government Treasury bond yields. In the case of the spot rate models, we implement the EMM (Gallant et al, 2000) procedure to estimate these parameters. This involves simulating the state processes for a given parameter setting, and then maximizing an approximation to the likelihood function of the data, by comparing the actual moments of the data to that of the simulated process. This is done for various parameter settings until an optimal criterion is achieved. In the case of the forward rate models, we use the output of a principal component analysis in order to estimate the term structure of the instantaneous forward rate processes. In both types of parametric models, and output of this is estimated volatility functions for various spot and forward rates, which are important inputs to the option pricing step. In the pricing step, we implement an alternating direction finite difference procedure to numerically solve the PDEs of the models, while for the forward rate models we price options by means of a lattice (which is non-recombining in the case of the proportional volatility models). Finally, in the case of the non-parametric model, we use a combined cross sectional-time series optimization algorithm to directly determine which vector of bandwidths best matches option prices.

Table 4.1:Characteristics of Various Spot-Rate Models

Model	Factors	Stochastic Process	Factor Risk[1]	Short Rate
1FGPTSM[2]	Short Rate	$dr_t = [\alpha + \beta r_t + v r_t \log(r_t)]dt + [\delta + \sigma r_t]^{\gamma} d\hat{B}_t$	$\theta = \lambda(t)$	Given
1FATSM[3]	Short Rate	$dr_t = (\alpha + \beta r_t)dt + (\delta + \sigma r_t)^{\frac{1}{2}} d\hat{B}_t$	$\theta = \lambda(t)$	Given
1FGTSM[3]	Short Rate	$dr_t = [\alpha(\tau) + \beta(\tau) r_t]dt + \sigma(\tau) d\hat{B}_t$	$\theta = \lambda(t)$	Given
CKLSTSM[4]	Short Rate	$dr_t = \kappa[\theta - r_t]dt + \sigma r_t^{\gamma} d\hat{B}_t$	$\theta = \lambda(t)$	Given
2FLSTSM[5]	2 Unspecified Factors X_1 & X_2	$dX_{it} = \kappa_i[\varphi_i - X_{it}]dt + \sigma_i X_{it}^{\frac{1}{2}} d\tilde{B}_{it}$, $i = 1,2$	$\theta_i = \begin{cases} \sigma_{it}^{\frac{1}{2}}\lambda_t & i=1 \\ 0 & i=2 \end{cases}$	$r_t = x_{1t} + x_{2t}$
2FSSTSM[5]	Spread between long and short rate s_t	$dl_t = \kappa_l(\theta_l - l_t)dt + \sigma_l\sqrt{l_t}\,d\tilde{B}_{lt}$ $ds_t = \kappa_s(\theta_s - s_t)dt + \sigma_s\sqrt{l_t}\,d\tilde{B}_{st}$	$\theta_i = \begin{cases} \theta\sqrt{l_t} & i=l \\ \lambda(t) & i=s \end{cases}$	$r_t = l_t - s_t$
3FGTSM[6]	3 Unspecified Multinormal Factors X_1, X_2, & X_3	$dX_{1t} = -[\kappa_{31}X_{1t}]dt + d\tilde{B}_{1t}$ $dX_{2t} = -[\kappa_{21}X_{1t} + \kappa_{22}X_{2t}]dt + d\tilde{B}_{2t}$ $dX_{3t} = -[\kappa_{31}X_{1t} + \kappa_{32}X_{2t} + \kappa_{33}X_{3t}]dt + d\tilde{B}_{3t}$	$\theta_i = \lambda_i(t)$ $i = 1,2,3$	$r_t = \psi_0 + \sum_{i=1}^{3} \psi_i X_{it}$
3FCIRTSM[6]	3 Unspecified Independent Factors X_1, X_2, & X_3	$dX_{it} = -[\kappa_{i1}X_{1t} + \kappa_{i2}X_{2t} + \kappa_{i3}X_{3t}]dt + X_{it}^{\frac{1}{2}} d\tilde{B}_{it}$ $i = 1,2,3$	$\theta_i = \lambda_i\sqrt{X_{it}}$ $i = 1,2,3$	$r_t = \sum_{i=1}^{3} X_{it}$
3FGJDATSM[7]	Short Rate and 2 Independent Factors X_1 & X_2	$dX_{it} = (\theta_i - \kappa_i X_{it})dt + \sigma_i\sqrt{X_{it}}\,dB_{it}$ $dr_t = \left(\mu_0 + \left(\mu_1 - \frac{1}{2}\right)X_{1t} + \left(\mu_2 - \frac{1}{2}\right)X_{2t}\right)dt + \sqrt{X_{1t} + cX_{2t}}\,dB_{rt}$	$\theta_i = \lambda(X_{it}, r_t, t)$ $i = 1,2$	$r_t = r_0 + \sum_{i=1}^{2} r_i X_{it}$

1-This is the functional form of the endogenously determined market price of risk.
2-1FGPM = Single Factor General Parametric Term Structure Model
3-1FA(G)TSM = Single Factor Affine (Gaussian) Term Structure Model
4-CKLSTSM = Chan, Karolyi, Longstaff, and Saunders (1992) Term Structure Model
5-2FLS(SS)TSM = 2-Factor Longstaff & Schwartz (Schaefer & Schwartz) Term Structure Model
6-3FG(CIR)TSM = 3-Factor Guassian (CIR) Term Structure Model
7-3FGJDATSM = 3-Factor Gaussian Jump-Diffusion Affine Term Structure Model (Duffie (1998))

Table 4.2
Forward-Rate Models Under Consideration

HJM Model Type[1]	Forward Rate Drift Process $\mu(t,s,f(t,s))$	Forward Rate Diffusion Process $\sigma(t,s,f(t,s))$
Panel A: Single Factor Models		
Absolute I (HJMAVI)	$\sigma^2(s-t)$	σ
Linear Proportional I (HJMLPVI)	$\left(\sigma_0(s-t)+\frac{1}{2}\sigma_1(s-t)^2\right)\times\left(\sigma_0+\sigma_1(s-t)\right)$	$\left(\sigma_0+\sigma_1(s-t)\right)$
Quadratic Term Structure (HJM1FQTSM)	$\left(\sigma_0(t,s)+\sigma_1(t,\tau)x_t\right)\times\left(\int_{\tau=t}^{s}\sigma_0(t,\tau)+x_t\int_{\tau=t}^{s}\sigma_1(t,\tau)\right)$	$\sigma_0(t,s)+\sigma_1(t,s)x_t$
Panel B: Two-Factor Models		
Absolute II (HJMAVII)	$(s-t)\sum_{i=1}^{2}\sigma_i^2$	$\sigma_i \quad i=1,2$
Linear Proportional II (HJMLPVII)	$\sum_{i=1}^{2}\left(\sigma_{0i}(s-t)+\frac{1}{2}\sigma_{1i}(s-t)^2\right)\times\left(\sigma_{0i}+\sigma_{1i}(s-t)\right)$	$\left(\sigma_{0i}+\sigma_{1i}(s-t)\right) \quad i=1,2$
Panel B: Three-Factor Models		
Absolute III (HJMAVIII)	$(s-t)\sum_{i=1}^{3}\sigma_i^2$	$\sigma_i \quad i=1,2,3$
Linear Proportional III (HJMLPVIII)	$\sum_{i=1}^{3}\left(\sigma_{0i}(s-t)+\frac{1}{2}\sigma_{1i}(s-t)^2\right)\times\left(\sigma_{0i}+\sigma_{1i}(s-t)\right)$	$\left(\sigma_{0i}+\sigma_{1i}(s-t)\right) \quad i=1,2,3$

1: HJM = Heath-Jarrow-Morton (1992) forward rate models.

Table 4.3 reports summary statistics for options on 30 year Treasury Bond futures traded daily on the Chicago Board of Trade for the first half of 1990 (1/1/90-5/31/90). There are 7979 observation in the six month period for the following variables: option premium (in points plus 1/32nd's expressed as a decimal), time-to-maturity (days to maturity as a percent of 250 business days), strike price and futures price (both as a percent of par), the short rate (the annualized continuously compounded 30 day T-Bill rate), and the short rate volatility (non-parametric kernel estimate). Call options constitute 46.9% of the sample. In-the-money options make up approximately 34% of the sample.

The average option price in this period is 3.9 points, ranging from an average of 19.4 for options greater than 10% in the money to 0.9 points for those greater than 10% out of the money. We find that option prices and the variability of prices increase with moneyness, and call (put) options are more valuable as interest rates decrease (increase). The average time-to-maturity is approximately 4 months and this average decreases with moneyness. The futures price is approximately 93% of par, the short rate of interest averages about 8%, and short rate volatility averages about 12 basis points.

Table 4.3: Summary Statistics
Daily CBOT Options on 30 Year Treasury Bond Futures (1/1/90-5/31/90)

	All	Calls	Puts	>10%ITM[1]	10-5%ITM	5-0%ITM	0-5%OTM[2]	5-10%OTM	>10%OTM
#	7979	3741	4238	662	872	1281	1443	1415	2306
%	100	46.9	53.1	8.3	10.9	16.1	18.1	17.7	28.9
OptionPrice[3]									
μ[4]	3.91	3.16	4.57	19.39	10.82	4.8	1.47	0.35	0.11
σ[5]	5.9	5.22	6.43	4.21	2.12	1.6	1.07	0.35	0.59
Time-to-maturity[6]									
μ	0.49	0.50	0.49	0.25	0.32	0.43	0.46	0.48	0.42
σ	0.28	0.28	0.28	0.39	0.36	0.37	0.38	0.35	0.36
Strike Price[7]									
μ	93.1	97.1	89.7	97.4	95.6	93.3	93.1	94.5	91.2
σ	9.83	8. 09	9.94	11.46	7.04	3.60	3.71	7.38	14.66
Futures Price[7]									
μ	92.5	92.4	92.6	92.1	92.9	92.9	92.8	92.7	92.7
σ	2.73	2. 71	2.75	4.21	2.49	2.43	2.38	2.36	3.27
Short Rate[8]									
μ	0.08	0.08	0.082	0.081	0.080	0.083	0.080	0.083	0.080
σ	.0.002	0.001	0.001	1.2×10^{-3}	1.1×10^{-3}	1.26×10^{-3}	1.27×10^{-3}	1.26×10^{-3}	1.25×10^{-3}
Short Rate Volatility[9]									
μ	0.001	0.001	0.001	1.21×10^{-3}	1.25×10^{-3}	1.26×10^{-3}	1.24×10^{-3}	1.24×10^{-3}	1.25×10^{-3}
σ	0.0002	0.0001	0.002	1.61×10^{-4}	1.77×10^{-4}	1.76×10^{-4}	1.75×10^{-4}	1.72×10^{-4}	1.71×10^{-4}

1-In the money.
2-Out of the money.
3-Premium as a percent of par.
4-Average.
5-Standard deviation.
6-Days to maturity as a percent of 250 business days
7-Percent of par
8-Annualized, continuously compounded 30 day Treasury bill rate
9-Nonparametrically estimated volatility of the short rate

Tables 4.4 and 4.5 report summary statistics of the estimated volatility functions, for spot and forward rates, as derived from the spot rate and forward rate models. The models are first fitted to the prices of bonds and the volatilities of various bonds are computed. This is the basic term structure input to the option pricing model and we present these estimates to be consistent with the existing literature (Buhler et al, 1999).

Since the volatility estimates from the various models are not directly comparable, we compute the volatilities for two selected rates, the spot (1 month) and the 10 year rate. In the case of the spot rate models, this long rate is the *10 year spot rate* (i.e., yield on a nine year zero coupon bond), while for the forward rate models, it is the *instantaneous forward* rate in 10 year's time.

The values obtained for the spot rate volatilities are broadly consistent across models, in the 2.6-2.7% range for spot rate and 2.2-2.8% range in the forward rate models. The 10 year spot volatilities in the spot rate models are lower in average and standard deviation, while the 10 year forward volatilities in the forward rate models are correspondingly higher. Common features include wide second order variation (standard deviations of volatility in the 7-8% range for both the short and forward rate), high positive skewness, autocorrelation, heteroscedasticity, as well as marked non-normality. As we go from one to three factors, computational time increases by about a factor of 1000, from about 30 minutes for the single factor model to close to 32,000 minutes for the three factor jump diffusion model.

Table 4.4: Estimated Absolute Short and Long Rate Volatilities for the Spot Rate Models

Model	Mean	σ	Skew	JB[1]	LBQ[2]	PPZ[3]	GPH[4]	Engle[5]	Time
1FGPTSM									
Short	0.027	0.007	12.79	22.33	68.80	-191.0	0.023	0.675	232 m.
Long	0.022	0.006	12.34	20.28	73.10	-190.9	0.101	0.580	229m
Relative	0.805	0.875	-1.535	7.951	69.80	-5.010	0.999	0.008	N/A
1FATSM									
Short	0.028	0.008	12.80	22.32	68.01	-190.9	0.028	0.674	187 m
Long	0.021	0.006	12.32	20.16	73.48	-191.0	0.106	0.521	73 m.
Relative	0.755	0.782	-1.464	7.944	69.80	-4.988	1.001	0.002	N/A
1FGTSM									
Short	0.027	0.007	12.79	22.64	61.10	-192.4	0.028	0.674	203 m.
Long	0.022	0.007	12.31	10.14	73.79	-199.2	0.103	0.567	197 m.
Relative	0.801	0.958	-1.535	7.946	69.80	-4.995	0.999	0.003	N/A
CKLSTSM									
Short	0.028	0.008	12.49	20.98	69.76	-191.1	-0.005	0.323	226 m
Long	0.022	0.005	13.04	19.48	32.98	-191.0	0.552	23.08	259 m
Relative	0.804	0.671	-1.526	78.42	69.76	-4.044	0.974	0.003	N/A
2FLSTSM									
Short	0.027	0.003	12.50	21.05	69.74	-191.4	-0.004	0.331	535 m
Long	0.022	0.002	3.046	19.90	33.45	-191.2	0.553	27.77	498 m
Relative	0.804	0.667	-1.526	27.85	69.74	-3.83	0.973	0.082	N/A
2FSSTSM									
Short	0.027	0.008	12.52	21.12	69.72	-191.2	-0.005	0.344	454 m
Long	0.022	0.005	33.29	27.60	31.30	-192.3	0.533	34.93	471 m
Relative	0.828	0.697	-1.526	7.841	69.71	-3.946	0.975	0.168	N/A
3FGTSM									
Short	0.026	0.007	12.54	21.20	69.68	-191.9	-0.004	0.356	11103m
Long	0.021	0.005	13.36	29.67	30.96	-191.3	0.511	15.19	1154m
Relative	0.804	0.708	-1.526	17.85	69.68	-3.872	0.974	70.18	N/A
3FCIRTSM									
Short	0.026	0.007	12.55	21.23	69.61	-191.6	-0.006	0.363	11194m
Long	0.021	0.007	3.33	28.94	31.35	-191.6	0.5245	24.23	1137m
Relative	0.804	0.918	-1.527	78.61	9.61	-4.011	0.976	0.113	N/A
3FGJDATSM									
Short	0.022	0.008	12.561	21.28	69.67	-191.4	-0.005	0.361	2843m
Long	0.016	0.005	73.371	30.40	30.61	-191.0	0.524	16.28	2831m
Relative	0.721	0.595	-1.527	27.87	9.67	-4.120	0.980	0.0808	N/A

1: Jarra-Barque J-statistic (normal distribution)
2: Ljung-Box Q-statistic (uncorrelated increments/white noise)
3: Phillips-Peron Z-statistic (stationarity/stability)
4: Geweke-Porter-Hudak statistic (long memory/fractional differencing)
5: Engle T×R2 statistic (GARCH Effects/random volatility)

Table 4.5
Estimated Absolute Volatilities of Spot and Forward Rates for the Forward Rate Models

Model	Mean	σ	Skew	JB[1]	LBQ[1]	PPZ[1]	GPH[1]	Engle[1]	Time
HJMAVITSM									
Spot	0.022	0.007	12.50	20.99	69.71	-191.4	-0.005	0.334	209 m
Forward	0.022	0.007	12.50	20.99	69.71	-191.4	-0.005	0.334	209 m
Relative	1	1	1	1	1	1	1	1	1
HJMLPVITSM									
Spot	0.028	0.008	12.80	22.32	68.01	-190.9	0.028	0.674	187 m
Forward	0.107	0.018	12.32	20.16	73.48	-191.0	0.106	0.521	173 m
Relative	1.242	2.3205	-1.464	7.944	69.80	-4.988	1.001	0.002	N/A
HJM1FQTSM									
Spot	0.027	0.007	12.79	22.64	61.10	-192.4	0.028	0.674	203 m
Forward	0.107	0.0183	12.31	10.14	73.79	-199.2	0.103	0.557	197 m
Relative	1.249	2.542	-1.54	7.946	69.80	-4.995	0.999	0.003	N/A
HJMAVIITSM									
Spot	0.028	0.008	12.49	20.97	69.8	-191.1	-0.005	0.323	226 m
Forward	0.108	0.018	3.039	19.48	32.98	-191.0	0.552	23.08	259 m
Relative	1.244	2.382	-1.526	78.42	69.76	-4.044	0.974	0.003	N/A
HJMLPVIITSM									
Spot	0.027	0.008	12.51	21.04	69.74	-191.4	-0.004	0.331	535 m
Forward	0.106	0.017	3.046	19.90	33.45	-191.2	0.553	27.77	498 m
Relative	1.244	2.312	-1.526	27.85	69.74	-3.836	0.973	0.082	N/A
HJMAVIIITSM									
Spot	0.027	0.007	12.52	21.12	69.72	-191.2	-0.005	0.344	454 m
Forward	0.102	0.017	53.291	27.60	31.30	-192.3	0.533	34.93	471 m
Relative	1.208	2.431	-1.525	57.84	69.71	-3.946	0.9751	0.168	N/A
HJMLPVIIITSM									
Spot	0.026	0.008	12.54	21.20	69.68	-191.9	-0.004	0.356	1103m
Forward	0.099	0.012	63.35	29.67	30.96	-191.3	0.5108	15.19	1154m
Relative	1.243	51.66	-1.526	7.85	69.68	-3.872	0.9747	0.184	N/A

1: Refer to Table 4.4 for definitions of these statistics.

Table 4.6 reports a comparative analysis of the option pricing errors using various models. We find that results vary greatly across model types. Average percentage error ranges from a tenth of a basis point (0.0014%) for the non-parametric model) to close to 5 basis points (0.0491%) for the single-factor spot rate models). However, in all cases they are statistically indistinguishable from zero . In general the models have no systematic bias, leading us to conclude that the numerical procedures used are effective on average.

In terms of MSE and MAE (as well as some higher moment measures such as skewness), there is great similarity between the one, two, and three factor spot rate models, as well as the HJM forward rate models (for all factors) and the non-parametric model. The best model in terms of mean squared error (1.4 %) and mean absolute error (0.99 %) is the non-parametric (MNWK) model. Among the parametric models, the 3-factor jump diffusion model (3FGJDATSM) performs the best with an MSE of 1.64%, while the HJM models perform the worst, the best MSE being 4.64% for both the 2 and 3 factor proportional volatility models. Among the single factor spot rate models, the general parametric model (1FGPTSM) outperforms affine (1FATSM), Guassian (3FGTSM) and generalized CKLS (CKLSTSM) models, though the results are close.

This is different than the previous literature, where the 1FATSM performed the best. Among the single factor HJM models, the quadratic term structure model (HJM1FQTSM) is superior, and more qualitatively similar to the better performing spot jump-diffusion model than the other HJM models. The two factor Longstaff et al (1992) version of the CIR model (2FLSTSM) model tends to outperform all the single factor models, though it is computationally expensive. Overall, we find that six models perform well in calibration to the market prices of options: the 1FGPTSM, 2FLSTS, and 3FGJDATSM spot rate models ; the HJM1FQTSM and 2-factor linear proportional (2FHJMLPII) HJM models; and the non-parametric MNWKTSM model.

Table 4.6

Statistical Analysis of Deviation Between Model and Market Values[1]

Model[2]	Mean	MSE	MAE	s[3]	KS[4]	LBQ[4]	PPZ[4]	GPH[4]	ETR2[4]	Time[4]
1FGPTSM*	0.049	3.331	2.591	1.07	0.391*	129.21*	-4.0e+3*	0.236	5.74*	2.85
1FATSM	-0.050	3.430	2.624	1.17	0.49*	99.48*	-6.4e+3*	0.193	6.17*	2.79
1FGTSM	-0.049	3.526	2.696	1.15	0.451*	94.35*	- 3.2e+3*	0.234	5.96*	2.91
CKLSTSM	0.052	3.368	2.682	1.37	0.425*	94.97*	-6.1e+3*	0.196	6.52*	2.81
2FLSTSM*	1.7e-3*	2.367	1.730	-1.73*	0.135	26.81	-113.92*	0.07*	4.19	6.57
2FSSTSM	1.8e-3	2.418	1.777	-1.69*	0.203	303.5*	-400.68*	0.08*	4.72	6.47
3FGTSM*	0.032	2.271	1.820	-0.24	0.460	841.3*	-101.01*	0.805	0.84	34.92
3FCIRTSM	-0.033	2.308	1.854	-0.32	0.465	803.9*	-116.6*	0.794	0.79	34.89
3FGJDATSM*	-0.028*	2.164	1.644	-0.33*	4.57*	785.0*	-1.4e+4	0.808	0.76	34.91
HJMAVITSM	-0.241	7.080	5.633	3.83	4.99	148.4	-3830.0	0.196	1.70	117.3
HJMLPVITSM	0.017	6.745	4.985	3.72	4.95	129.7	-3840.8	0.222	1.72	116.6
HJM1FQTSM*	0.004	2.185	1.700	-0.31	4.61	822.9	-553.4	0.752	0.78	64.86
HJMAVIITSM	-0.179	7.049	4.700	3.76	4.98	186.5	-3203.4	0.193	1.57	217.3
HJMLPVIITSM*	0.153*	6.594	4.636	3.45*	0.005*	12.9e+5*	-4.0e+3*	0.236*	0.95*	235.6
HJMAVIIITSM	0.019	7.297	5.213	3.64	4.91	130.4	-3304.6	0.214	1.89	418.9
HJMLPVIIITSM	-0.158	6.606	4.641	3.75	4.98	133.6	-3891.2	0.218	1.54	505.1
MNWKTSM[6]*	1.39e-3	1.40	0.999	0.37	0.21*	7778.1*	-90.66*	0.236*	1.7e+3*	872.3

1- A "*" denotes either a model outperforming in its class or a statistic that is significant at the 10% level or less.

2- Refer to Tables 4.1 and 4.2 for descriptions of the parametric models.

3- Symbol referred to excess skewness relative to normal.

4- Refer to Table 4.3 for a description of these statistics.

5- Measured in hours of CPU time.

6- Multivariate Nadarya-Watson Kernel term structure model.

Another important test the ability of these models to manage interest rate risk is an analysis of hedging errors, which is presented in Table 4.7. In each model, we calculate the sensitivity of model prices to the postulated factors for each trading day, and set up a hedge position that is held for one day. This is defined as a unit of the option minus a hedging instrument multiplied by the sensitivity of the option price with respect to it. In the single factor spot rate models, the sensitivity of the model price with respect to the 30 day T-bill rate (by recalculating the finite difference solution for a small change in the rate), and then an appropriate amount of this instrument is held. In the non-parametric model, the sensitivity with respect to the short rate is similarly calculated by re-optimizing the kernel estimator for a small change in the short rate, thereby giving us the appropriate quantity of the 30 day T-bill rate to hold along with the option. In the two and three factor spot rate models, other one-year or five-year Treasury notes are used as well. In the HJM models, the values of various forward rates are perturbed and the lattice price is recalculated, generating sensitivities that can be applied to these instruments.

Note that the same models still outperform within each group, as in the analysis of pricing errors, although the rankings are different across groups. In contrast to its ability to finely calibrate to market prices, the non-parametric model is no longer the best performing model in terms of hedging variability, having a mean-squared hedging error of 2.4%. The 2FLTSM is the best model in hedging options, with a MSE of 0.95%, while the worst performing model, 1FGPTSM, has a MSE of 4.07%. Furthermore, examining the higher order distributional properties of the hedging errors, the favored models consistently outperform their cohorts. For example, all models exhibit positive skewness, which means that there are a few extreme positive outliers in these series. However, this skewness is less pronounced in the case of the better performing models.[15] This illustrates that we may want to use different types of models to hedge dynamically rather than to calibrate to current market prices.

---

15 Since we are defining the error as model minus market, one can argue that these are instances of profit arising from exploiting arbitrage relative to the model; however, we will not pursue that line of reasoning here, and consider hedging error series closer to mean zero-normal as coming from superior models.

Table 4.7
Statistical Analysis of Model Hedging Errors[1]

Model[2]	Mean	MSE	MAE	s[3]	KS[4]	LBQ[4]	PPZ[4]	GPH[4]	ETR2[4]	Time[4]
1FGPTSM*	2.4e-5	4.069	1.453	11.51*	0.041	147.2	-8.6e+3	0.394*	45.6	2.85
1FATSM	-1.7e-5	4.104	1.483	11.58	0.057	150.2	-2.9e+3	0.345*	55.2	2.82
1FGTSM	-5.4e-5	4.202	1.618	11.59	0.049	155.4	-1.6e+4	0.396*	52.0	2.81
CKLSTSM	1.0e-5	4.165	1.458	11.56	0.134	147.3	-8.5e+3	0.444*	45.9	2.87
2FLSTSM*	0.003	0.591	0.366	-1.94	0.211	105.2	-5.6e+3	0.340	552.5	5.83
2FSSTSM	-0.002	0.595	0.407	-1.86	0.265	109.1	-5.2.e+3	0.340*	552.8*	5.52
3FGTSM*	0.108	0.669	0.588	6.841	0.167	1.9e+3	-79.62	2.188*	258.3	34.8
3FCIRTSM	-0.099	0.687	1.564	6.878	0.134	1.8e+3	-1.3e+5	2.076*	258.3	34.9
3FGJDATSM*	0.022	0.652	0.529	6.831	0.116	1.0e+3	-8.1+3	2.145	258.2	34.7
HJMAVITSM	-0.024	1.529	1.269	11.90	2.92	1.3e+3	-5.0e+3	0.268	61.79	58.9
HJMLPVITSM	0.029	1.549	1.254	11.48	2.89	1.1e+3	-5.4e+3	0.285	61.57	58.3
HJM1FQTSM*	-0.023	0.952	0.750	-11.26	2.91	1.4e+3	1.7e+3	0.256	61.29	119.0
HJMAVIITSM	0.027	1.561	1.270	11.68	2.95	1.8e+3	-5.5e+3	0.206	61.84	58.2
HJMLPVIITSM*	-0.023	1.522	1.233	11.25	2.87	1.1e+3	-5.6e+3*	0.241	61.28	119.0
HJMAVIIITSM	-0.026	1.709	1.563	11.62	2.89	1.4e+3	-4.9e+3	0.251	61.88	59.5
HJMLPVIIITSM	0.025	1.827	1.457	11.44	2.91	1203.9	-5.1e+3	0.260	61.59	356.7
MNWKTSM[6]*	1.1e-5	2.398	2.199	0.371	0.21	7.7e+3	-9.7e+3	0.236	1.7e+3	832.7

1- A "*" denotes either a model outperforming in its class or a statistic that is significant at the 10% level or less.
2- Refer to Tables 4.1 and 4.2 for descriptions of the parametric models.
3- Symbol referred to excess skewness relative to normal.
4- Refer to Table 4.3 for a description of these statistics.
5- Measured in hours of CPU time.
6- Multivariate Nadarya-Watson Kernel term structure model.

Finally we regress model pricing errors on a number of independent variables-the moneyness ratio (defined as the ratio of futures price to the exercise price), time-to-maturity (measured in days to option maturity as a percent of 250 business days), strike price (as a percent of par value), short rate (settlement of annualized, continuously compounded 1month T-Bill rate), as well as the volatility of the short rate (non-parametric kernel volatility). Regression results are presented in Table 4.8. In all models, the only significant coefficients are moneyness and time-to-maturity. We find that the pricing errors decrease with moneyness and increase with time-to-maturity. In terms of severity of the bias, the two-factor HJM model performs the worst and the non-parametric model performs the best, as judged by the magnitude of the coefficients. We can also look at the overall degree of mis-specification by the measure of R^2. Since these variables should have no explanatory power if the model is a "good fit", a lower the R^2 is considered the better. By this measure the 1FHJMQTSM performs the best and 1FGPTSM performs the worst.

Table 4.8
Least Squares Regression Analysis of Interest Rate Derivatives Model Pricing Errors
CBOT Options on 30 Year Treasury Bond Futures (1/1/90-5/31/90)

Model	β_0[1]	F/X[2]	τ[3]	X[4]	F[4]	r[5]	σ[6]	R^2*
1FGPTSM	-2.7327	-1.7740[a]	0.0925[a]	0.0019	0.0148	36.84	-279.10	0.1908
2FLSTSM	4.0937	-0.7075[a]	0.7279[a]	0.0094	0.0157	-24.56	177.75	0.0975
NWNPKTSM	-1.4506	-0.3385[a]	0.0853[a]	0.0015	0.0003	21.55	33.716	0.0079
3FGJDTSM	5.5567	-0.9215[a]	0.1777[a]	0.0099	-0.0131	-72.80	969.23	0.0121
2FLPHJTSM	-7.553	-37.914[a]	3.3886[a]	0.3392	-0.4721	66.15	-475.15	0.0950
1FHJMQTSM	-26.502	-0.9117[a]	0.6439[a]	0.0125	0.0442	327.10	-1488.7	0.0031

1-Intercept of the regression.
2-Moneyness: defined as the ratio of futures price to the exercise price.
3-Time-to-Maturity: days to option maturity as a percent of 250 business days
4-Strike Price: expressed as percent of par value
5-Short Rate: Annualized, continuously compounded yield derived from the daily settlement of price of the 1month Treasury bill
6-Short Rate Volatility: Non-parametric kernel volatility of the 1 month Treasury bill rate.
a,b.c denotes statistical significance at the 1%, 5%, and 10% levels, respectively.

5. SUMMARY AND CONCLUSION

In this chapter, we test single and multiple factor spot-rate and forward-rate models, as well as a non-parametric model, using bond options data. The idea is to determine which factors and derivative models best price and hedge interest rate risk. We test for the best stochastic model of term structure, which type of model is best suited for interest rate risk management, and which theories of term structure are best supported by the data. We find that while non-parametric models may be superior in pricing interest rate options in-sample, multi-factor equilibrium style models (e.g., a two factor Longstaff and Schwartz (1992) CIR model) or more parsimonious but highly flexible single factor models (e.g., the single factor general parametric model of Duffie (1996)) have superior hedging and forecasting performance.

The data consists of settlement prices for all options on 30 years Treasury Bond futures traded on the Chicago Board of Trade for the first half of 1990 (1/1/90-5/31/90). The values obtained for the spot rate volatilities are broadly consistent across models, while the 10 year spot volatilities in the spot rate models are lower in average and standard deviation, while the 10 year forward volatilities in the forward rate models are correspondingly higher. The best model in terms of average, mean squared, and mean absolute errors is the non-parametric (MNWKTSM) model. Among the parametric models, the 3-factor jump diffusion (3FGJDTSM) model is best, while the HJM models perform the worst. Among the single factor spot rate models, the general parametric model (1FGPTSM) outperforms affine (1FATSM), Guassian (3FGTSM) and generalized CKLS models (CKLSTSM). However, for the spot rate models the results are closer than for the other models, and the general parametric model is far more computationally intensive. In hedging analysis, while the better performing models still outperform within each grouping, the non-parametric model is no longer the best performing model, while the 2FLSM is the best model in hedging options in terms of MSE. In analyzing pricing errors by means of regressions on option pricing variables, in terms of severity of the bias, the two-factor HJM model performs worst and the nonparametric model the best.

We conclude that models motivated by term structure theory and having a realistic stochastic characterization of interest rates, such as the 2-factor CIR models, are found to have decent hedging and out-of-sample forecasting performance. On the other hand, models motivated by the design of risk management systems capable of accurately explaining the market prices of interest rate dependent derivatives, such as HJM variants or more recently developed non-parametric techniques, have been found to better explain prices in-sample, though often at an exorbitant computational cost.

CHAPTER 5

EMPIRICAL TESTS OF STOCHASTIC TERM STRUCTURE MODELS AND THE PRICING OF INTEREST RATE DERIVATIVES: PROLOGUE AND FUTURE DIRECTION

1. INTRODUCTION AND DISCUSSION

In this research, we empirically test various models of the term structure, in order to further the literature in several ways. First, we link two streams of research in financial economics, in the domain of term structure derivatives theory and interest rate risk management. The first focuses on term structure theory and stochastic characterization of interest rates. The second type concentrates upon the design and evaluation of models that are capable of accurately explaining the market prices of interest rate derivatives. Our second contribution is to compare fundamentally different modeling philosophies and theoretical perspectives. These include no-arbitrage pricing, general equilibrium theory, quantitative and statistical analysis, as well as non-parametric modeling of the term structure and interest rate derivatives. Finally, we expand the analysis to other global markets.

We also make another contribution, albeit indirect, to the practice of derivatives trading and interest rate risk management. In empirically analyzing the pricing, forecasting, and hedging capabilities of competing term structure models, we examine issues central to the measurement and control of interest rate risk exposure. Such modeling strategies are central to any system constructed to supervise interest rate risk, whether it is Value-at-Risk, sensitivity analysis, stress testing, or scenario analysis. Few studies in the current literature have evaluated the empirical performance of competing valuation models on the same data set.

This final chapter concludes the major findings of our research. Section 2 outlines the models we estimate, while Section 3 presents a conclusive summary and comparison of the various term structure models, including non-parametric ones. Section 4 presents avenues of future research.

2. VARIOUS TERM STRUCTURE MODELS

In this section, we highlight key areas of the term structure and derivatives literature that this research has built upon. We touch on three main areas in this study: econometric tests of global short term interest rates, analysis of U.S. default-free term structure, and the pricing of bond futures options.

In tests of short-term rates, we compare two fundamentally different approaches: estimation of the parameters of short rate stochastic differential equations versus non-parametric estimation of the drift and diffusion functions. For the parametric estimation, we follow Brennan and Schwartz (1982), Dietrich-Campbell and Schwartz (1986) and

Chan et al (1992; CKLS). We extend these studies by approximating the SDE of the Single Factor General Parametric Model (1F-GPM), as discussed in Duffie (1996), by a discrete system of difference equations (known as an *Euler discretization*). The econometric technique involves the Generalized Method of Moments (GMM: Hansen (1982)). We nest several more models than CKLS, including the *affine term structure model* (ATSM: Dai et al ((1999)), the Gaussian term structure model (GTSM: Jamshidian (1989), Gibson and Schwartz (1990)), the Black-Derman-Toy model (BDT: Black et al (1990)), the *non-linear drift model* (Black and Karasinski (1991)) and the *double square root model* (DSRM: Longstaff (1989)). The DSRM has recently been generalized to a HJM arbitrage-free setting as the *quadratic term structure model* (QTSM: Boyle et al (1999), Ahn et al (2000)). In the non-parametric case, we build upon Pearson's (1994) conditional density approach, Ait-Sahalia's (1996) comparison of implied parametric to non-parametric densities, Stanton's (1997) Gaussian estimation approach and Stutzer's (1996) canonical valuation model. Our contribution involves implementing and testing the model of Jiang (1998) in the sense of basing all estimation on a non-parametric kernel estimated density and volatility function for the short rate.

In our analysis of the discount bond term structure, the theoretical benchmark that we consider is an extension of Longstaff and Schwartz's (1992; henceforth LS) two-factor version of the CIR model. We take the non-parametric estimation of the interest volatility factor (Jiang (1998)), and combine it with the improved computational approach of Longstaff et al (1993), which allows us to hedge and forecast bond prices.

The non-parametric model that we use is based upon a neural network approach. We use a variant of Hornik et al (1989), the *multilayer feed forward multilayer perceptron* architecture, in conjunction with White's (1989, 1990) non-linear regression model. We compare these models to a purely statistical approach, *principal components analysis* (PCA), as discussed in Rebonato (1996) and implemented by Bliss (1997) for US treasury bonds.

Finally, our tests of interest rate derivatives are based upon several recent studies. Studies focusing on single factor models include the Gibbons and Ramaswamy (1993) test of the CIR model, Flesaker (1994) test of the constant volatility HJM model, and Jordan et al (1995) analysis of option pricing values implicit in callable Treasury bonds in the framework of Longstaff (1992). Moraleda and Pelsser (1997) compare the spot models of Hull and White (1994), Pelsser (1994), and Black and Karasinski (1991) to the Gaussian, square root, and proportional forward rate model of Ritchken and Sankarasubramanian (1995). Finally, Buhler et al (1999) test the market for German warrants, using seven spot and forward-rate models, each with both one and two factors. They identify one forward-rate model (a single factor proportional volatility HJM model) and two spot rate models (a single factor extended CKLS and a two factor CIR model) that outperform the other four models. In addition to considering three factor models, we also test an *affine term structure model* (ATSM) and *stochastic volatility model* (SVM). Variations of ATSMs, in which the equilibrium short rate is assumed to be an affine (i.e., linear) function of state variables, are equivalent to several popular spot rate models (namely Black (1976), Vasicek (1977), and CIR (1985)). Affine models are tested by Brown and Schaeffer (1994) and generalized by Duffie and Kan (1996). Duffee (1998) finds that affine models lack predictive power in explaining yield changes. Dai and Singleton (1998) present an exhaustive analysis of ATSMs, deriving conditions under which generalizations of popular models are *admissible*. Limitations of the ATSM approach have lead to the *quadratic term structure model* (Gallant et al 2000; QTSMs), which posits that the short rate is a quadratic function of state variables and thus guarantees a positive interest rate. We test a version of the QTSM, the *double square root model* of Longstaff (1989; DSR). SVM implementation has focused on the *affine-jump diffusion* class of models (Heston (1993); AJD). This is extended by Duffie et al (1998) using transform analysis and by Bakshi et al (1997) beyond the affine class of models. The parallel econometric developments are the *simulated method of moments* (SMM) of Duffie and Singleton (1993) and *efficient method of moments* of Gallant et al (1996, 2000). We contribute to this literature by applying the EMM technique to our version of the AJD class of option models, based upon the discussion of Chernov and Ghysels (1999), as adapted to term structure derivatives.

3. A SUMMARY AND COMPARISON OF EMPIRICAL RESULTS

Here we summarize the various empirical findings in this research. First, we discuss the findings with regard to the econometric estimation of international interest rates. Next, we discuss the results for the pricing of U.S. Treasury bills and bonds.

Finally, we present the conclusions of the CBOT bond future option analysis.

3.1 Econometric Tests of International Short Term Interest Rates

As a proxy for the instantaneously compounded interest rate, we take the yield-to-maturity on pure short-term money market instruments. These are the one month U.S. T-bill rates (1MTB), the weekly federal funds rates (1WFF), the three month $ Libor rates (3MLD), the one month Japanese government bond yields (1MJGB), and the three month Euro-sterling rates (3MES). In the case of 1MTB, we utilize an extended and updated version of the data set used by CKLS (1992). We first undertake a comprehensive statistical analysis of these rates and find that all the series exhibit marked non-normality, as evidenced by large and statistically significant Kolmogorov-Smirnov D-statistics. We find high autocorrelation for both rate levels and their differences based on the Q-statistics, and find significant GARCH effects based on the Engle (1982) $T \times R^2$ tests. On the other hand, long memory characteristics as measured by GPH statistics (Geweke et al, 1983), are not found in any instances. The main differences found are in stationarity, as the 1MTB does not exhibit a unit root in either the level or differences. However, the 3MES exhibits a unit root in both levels and differences, while 1WFF exhibits a unit root in levels and the 1MJGB only in differences. The conclusion of this data analysis, the rejection of unconditional normality for all short rates, is a supporting factor in the use of a continuous time framework, which does not impose normality in the error structure. This motivates the use of generalized econometric approaches, such as generalized method of moments (GMM) or kernel regression, as opposed to linear structural models.

The econometric tests involve estimating the parametric models by GMM and the non-parametric model by Jiang's (1998) procedure. First, we graphically compare the non-parametric model with a representative parametric model for the 1MTB (these qualitative results, unlike the econometric tests, are very similar across models and markets). We look at the diffusion function, marginal density function, and the drift function. In general, while the both the parametric and non-parametric diffusion functions are increasing in the short rate, the non-parametric diffusion is more non-linear than the parametric diffusion. As compared to parametric log-normal PDF's, calibrated to have the same mean and variance as the non-parametric density, the non-parametric PDF shows more extreme positive skewness, bi-modularity, as well as leptokurtosis. Comparing the estimated parametric and non-parametric drift functions, we find that while both first increase and then decrease with the short rate (which is expected under mean reversion), the non-parametric function is highly non-linear, and achieves higher rate-of-change levels at both very low and very high levels of the short rate.

We compare the various parametric models to each other and to the non-parametric model, using Hansen's χ^2 statistic for the test of overidentifying restrictions of the GMM and the goodness-of-fit statistics R_1^2 (R_2^2). The latter are ratios of the variation in non-parametric drift variability (non-parametric diffusion) explained by the variability in the fitted drift (parametrically estimated diffusion). Across most data sets, and in contrast to the results of CKLS (1992) for the case of the 1MTB, all the parameters are statistically significant. This could be attributed to the longer sample length, more efficient use of the data, and an improved estimation technique. In all cases the χ^2 statistics of the GMM estimation indicates that we reject the hypothesis that all the moment restrictions fail to hold, which means that no model alone is rejected by the data. In terms of parameter estimates, they tend to be similar across data sets for the more general models, but there are significant differences in some of the nested models. For instance, in the 1FGPM, the elasticities of variance (E'S) obtained are much closer across markets than for the CKLS model. This implies that the more general models tend to be more consistent across different data sets. Our results for the CKLS model differ from CKLS (1992) findings with respect to some of the other markets in that we find much higher E'S in the US and UK markets, and much lower EV's in the Japanese and Eurodollar markets. This implies that interest rate volatility is more sensitive to the level of the rate in the US and UK, and less so in Japan and the Euro market. Comparing the percent variation explained by the fitted drift and diffusion functions relative to the non-parametric counterparts, we find that across different data sets these models explain a relatively small proportion of the variation in the non-parametrically estimated drift and diffusion functions. However, there is significant variation in relative magnitudes as well as rankings across models and markets. These models explain a higher proportion of non-parametrically estimated variation in the drift and diffusion in the Japanese data, as compared to the other markets, by about a factor of 10. These models explain drift as compared to diffusion differently across data sets. In the case of the US data, the parametric models do a poor job of capturing variation in the drift as compared to volatility. However, the situation is reversed in the case of the Japanese data, where the parametric models do better at capturing variation in the drift as compared to the volatility. The results are similar for Euro and Japan with respect to explaining the drift much better than the volatility, but with magnitudes higher in the drift and lower in the volatility. For the UK, drift is explained about as well as volatility. These results differ with respect to various empirical findings based on US data, in which drift is estimated with much greater difficulty than the diffusion function.

Across data sets, the more general models tend to perform best in estimating drift and volatility, capturing the highest percentages of non-parametric variation explained. In the 1MTB, the 1FGPM performs best, followed by the Non-Linear Drift (OLD) and Black-Derman-Toy (BDT) models. In terms of R_2^2, the CKLS and CEV models explain about half what the above models do, but outperform the remaining models. The CIR model performs the weakest, with 2% of the variation explained. Surprisingly, the non-nested QTSM is very close to the CIR model by these measures, in spite of the fact that the QTSM is capable of accommodating a degree of non-linearity in the drift and the CIR model cannot. In contrast, for the 1MJGB, the QTSM best describes the drift, with the 1FGPM a runner-up, while the Pearson-Sun (TSM) best describes volatility, with the QTSM the runner-up. The 1FGPM does a poor job of explaining the diffusion for the Japanese data. In this case, the QTSM is most consistent within data sets for the case of the Japanese market. In the 1MESD data, models are generally less successful in capturing both the variation in the drift as well as the diffusion. The 1FGPM best describes the drift, while the nested BDT model best describes volatility, although the differences with other models is much less with the drift as compared to the volatility. Finally, these results all differ in the case of the 3MLD, where the SVM and BDT best explain the drift and diffusion, respectively. As a final in-sample test, we present Newel-West specification tests of the restrictions imposed by each nested model, for the four short rate series, for all possible nested pairs of models. The consistent finding that in no case do we fail to reject the parameter restrictions imposed by a model on any model that it is nested in. This supports the earlier observation, that in the context of a comparison to the non-parametric benchmark the more general models tended to have a better in-sample performance. Furthermore, the proportional drops in the GMM criteria become more dramatic as we move away from the more general model, supporting the conclusions of the econometric results.

As a test of the out-of-sample properties, we measure the forecasting ability of these models. The 1FGPM performs the best for all the four global interest rates based on most of the forecasting measures. The non-parametric model (GPM) tends to perform in the middle for three of the four interest rates, except the 3MLD. The QTSM model seems to perform the worst, except for the Japanese short-term rates. Most traditional models (e.g., CIR, CEV, and CKLS) seem to underperform in forecasting all four interest rates by most measures. The displaced diffusion (D.D.), Pearson-Sun, and BDT models seem to perform fairly well and better than the GPM.

In summary, we have find that the GMM estimator and the Newey-West tests of various models and across different short interest rates support the conclusion that most of the popular models are mis-specified (i.e., likely to be different from the true data generating model). First, the variation in results across different rates, in terms of explaining the non-parametrically estimated drift and diffusion, show that the popularity of several models (e.g., BDT, PS, CKLS) may have been driven by US historical experience. These models do not seem to fit other global short rates. Second, the conclusion in the existing literature that most models explain the drift function better than the volatility function does hold globally. In agreement with CKLS (1992), and in contrast to the literature that accommodates more parametricized approaches (such as GARCH effects), we find that the effect of the level of interest rates is more significant than previously believed, when we allow the diffusion function to be a linear transformation of the short rate. In agreement with economic theory and empirical evidence, we find that both mean reversion and non-linearity is a significant feature of short rate drift processes, in agreement with the non-parametric finance literature (Ait-Sahalia, 1996). The non-nested quadratic term structure model (Boyle et al 1999), a HJM style model that is estimated and compared with other popular models in this study, is found to generally underperform, despite its ability to model a mean reverting and non-linear drift.

3.2 Bond Pricing and Hedging Analysis of the U.S. Treasury Term Structure

As a first step in analyzing the default-free term structure, we perform a comprehensive statistical analysis of the U.S. government bond discount term structure, from 6/1964 to 10/1997. We analyze rates for 3, 6, and 9 months and 1 to 5 years, expressed in annualized, continuously compounded form. Mean levels are found to increase monotonically with maturity, while yield volatilities peak at one year, consistent with previous empirical evidence. However, all maturities exhibit excess skewness and kurtosis in level and result in strong rejection of normality. All series exhibit autocorrelation, which strengthens with maturity. We find evidence for stationarity and no long memory in all cases. All maturities exhibit evidence of GARCH effects. Most results for yield changes are qualitatively similar, except that skewness is negative instead of positive. Principal component analysis (PCA) of yield changes over the time period finds evidence of stochastic shifts in the relevance of various factors (e.g., level, slope, and curvature) that result in shifts of the term structure. An impulse response analysis of shocks to the yield curve reveals decreasing correlation with increased difference in maturity, with the decrease being largest for the closest maturity, a phenomenon called *exponential decorrelation* in the literature (Rebonato, 1996).

As a test of in-sample pricing performance of differing pricing approaches, we compare the equilibrium 2FCIR model with the non-parametric ANN-MLP model by a statistical examination of the bond pricing errors across various maturities. The results suggest that both models have effective pricing performance on average, as mean errors are small and not statistically different from zero, and mean squared errors are approximately a few percentage points. The ANN-MLP model performs better, with slightly lower mean pricing errors, as well as significantly lower MSEs. The MAE statistics display a pattern similar to the MSE statistics. Both models exhibit marked non-normality (excess skewness and excess kurtosis) in pricing errors, but this is more severe in the 2FCIR model, which implies that there may have been more instances of abnormally high or low pricing errors in the 2F-CIR model as compared to the ANN-MLP model. Average pricing errors, MSE, MAE, as well as KS statistics deteriorate monotonically (i.e., become larger) with maturity. This holds for both the ANN-MLP and the 2F-LS models. This suggests that it is more difficult to price longer maturities with either of these models. The explanation for this may be that more factors are needed to price long maturity bonds. Finally, the tests of more general statistical properties of the pricing errors suggest that pricing errors for both models are autocorrelated, stationary, and possess long-memory. The magnitude of these test statistics is similar across all maturities and both models. Given that these errors are by-products of models that exploit short and long term patterns in bond prices, the autocorrelation and long-memory patterns in the pricing errors are not surprising.

As a test of the out-of-sample properties of these models, a comparison of hedging positions between the 2F-CIR and ANN-MLP models is made. The results suggest that both models have effective hedging performance on average, as mean errors are small and not statistically different from zero, and MSEs are on the order of a few basis points. However, the MSEs are approximately 10 times larger in the ANN-MLP model than the 2FCIR model. The comparison of MAE statistics is similar. This suggests that the 2FCIR model yields superior hedging performance. However, in both cases the hedging positions exhibit marked non-normality, which is more pronounced in the non-parametric model. Significant negative skewness in both models implies that large losses may have occurred in the sample period, such losses being larger in the non-parametric model. Average hedging errors, MSE, MAE as well as normality test statistics deteriorate monotonically with maturity for both models. This suggests that it is more difficult to hedge longer maturities yield changes with either of these models. The explanation for this may be that more factors are needed to hedge shifts at the long end of the term structure (e.g., the "curvature" factor may have more importance there than at the short end). We conclude that the parametric 2FCIR model provides superior hedging performance as compared to the non-parametric ANN-MLP model. The reason for this is related to the nature of these models. Non-parametric models are designed primarily to fit prices in-sample, whereas parametric models are based upon long-term theoretical relationships which are expected to hold over time.

We also test the forecasting ability of various bond-pricing models by generating 1-step ahead bond yield forecasts from the two different such models. In addition, we compute these for the single factor general parametric (1FGPM) interest rate model and the *principal component analysis-constant correlation multivariate GARCH* (PCA-

CCMGARCH) yield change model. We find that 2FCIR does better at shorter maturities while the PCA-CCMGARCH does better at longer maturities. This is consistent with our finding in the bond pricing and hedging analysis that the two factor models tend to do worse at longer maturities, and a third factor may be more significant for such maturities. The 1FGPM is consistently the worst performing model. The MLP-ANN is usually second best across maturities and forecasting measures. In all cases, the MSE, MAE, and Theil-U statistics deteriorate (i.e., increase) as maturity increases, but the other directionally oriented measures exhibit no such pattern. This suggests that there may be an omitted factor that is more effective with increased maturity.

3.3 A Comparison of Stochastic Interest Rate Term Structure Models: The Pricing of Bond Options

In the final test of this research, we examine various parametric interest rate option pricing models and a non-parametric model on CBOT data. The spot rate models differ in factor description (e.g., short rate, spread, or latent), stochastic process for the factors (e.g., square root versus general Ito processes), endogenously determined market-price-of-risk (e.g., determined by the short rate or by unobservable factors), and the determination of the short rate (i.e., given directly or as a function of the factors). The forward rate models differ by the number of sources of uncertainty in the forward rate process (e.g., single vs. multiple factors, linear vs. proportional volatility), the stochastic process for the forward rate drift (e.g., constant, linear or non-linear in time-to-maturity), and the stochastic process for the forward rate diffusion (e.g., arithmetic, geometric or generalized Brownian motion).

The data set we test is options on 30 year Treasury bond futures traded on the Chicago Board of Trade (CBOT), settlement data for all traded instruments for the first half of 1990 (1/1/90-5/31/90). There are 7979 observations in this six-month period: option premium, time-to-maturity, strike and futures price, the short rate (the annualized continuously compounded 30 day T-Bill rate), and the short rate volatility (non-parametric kernel estimate). A statistical analysis of the data reveals option prices and the variability of prices increasing with moneyness, since call (put) options are more valuable as interest rates decrease (increase).

We estimate the volatility processes as derived from the spot rate and forward rate models. These are key inputs for the option pricing models. The models are first fit to the prices of bonds and volatilities of various bonds are computed. This is the basic term structure input to the option pricing models and we present these estimates to remain consistent with the existing literature (Buhler, 1998). Since the volatility estimates from the various models are not directly comparable, we compute the volatilities for two selected rates, the spot (1 month) and the 10-year rate. In the case of the spot rate models, this long rate is the *10-year spot rate* (i.e., yield on a ten year zero coupon bond), while for the forward-rate models, it is the *instantaneous forward* rate *in 10 years time*. We find values obtained for the spot rate volatilities to be broadly consistent across models, with the 10 year spot volatilities in the spot rate models lower in average and standard deviation, while the 10 year forward volatilities in the forward rate models are correspondingly higher. Common features of these estimates include high positive skewness, marked non-normality, autocorrelation, and heteroscedasticity. As we go from one to three factors, computational time to estimate volatilities increases by about a factor of 1000, from about 30 minutes for the single factor model to close to 32,000 minutes (i.e., about 22 days) for the three factor jump diffusion model.

A comparative analysis of the pricing errors reveals that results vary greatly across model types. Average percentage errors range from a tenth of a basis point (for the non-parametric model) to close to five basis points (for the single-factor spot rate models), although in all cases they are statistically indistinguishable from zero. In terms of MSE and MAE (as well as some higher moment measures such as skewness), there is great similarity between the single, double, and three factor spot rate models as well as the HJM (For all factors). The best model in terms of average, mean squared, and mean absolute errors is the non-parametric (MNWK) model. Among the parametric models, the 3-factor Guassian jump diffusion (3FGJJD) model is best, while the HJM models perform the worst. Among the single factor spot rate models, the 1FGPM outperforms the affine (1FATSM), Guassian (1FGTSM) and generalized CKLS models, though but the results are close. This is different from the previous literature, where the 1FATSM model has performed the best among the single factor HJM models, the QTSM is superior, and more qualitatively similar to the better performing spot rate 3FGJD models than the other HJM models. The two factor Longstaff et all (1992) CIR model (2FLS) tends to outperform all of the single factor models, though it is computationally expensive. Overall, we identify six models that perform best in calibration to market prices: the 1FGPM, 2FLS, and 3FGJDM spot rate models; the QTSM and 2-factor HJM linear proportional (2FHJMLPII); and the non-parametric MNWK model.

As a test of the comparative ability of these models to manage interest rate risk, we analyze hedging errors. In each model, we calculate the sensitivity of model prices to the postulated factors for each trading day, and set up a hedge position that is held for that day. This is defined as a unit of the option minus a hedging instrument multiplied by the sensitivity of the option price with respect to it. In the single factor spot rate models, the sensitivity of the model price with respect to the 30-day T-bill rate (by recalculating the finite difference solution for a small change in the rate), and then an appropriate amount of this instrument is held. In the non-parametric model, the sensitivity with respect to the short rate is similarly calculated by re-optimizing the kernel estimator for a small change in the short rate, thereby giving us the appropriate quantity of the 30- day T-bill rate to hold along with the option. In the two and three factor spot rate models, other bills are used as well. In the HJM models, the values of various forward rates are perturbed and the lattice price is recalculated, generating sensitivities that can be applied to these instruments.

Note that the same models still outperform within each group, as in the analysis of pricing errors, although the rankings are different across groups. In contrast to its ability to finely calibrate to market prices, the non-parametric MNWK model is no longer the best performing model in terms of hedging variability. The 2FLS model is the best model in hedging options, while the worst performing model is the 1FGPM. Furthermore, examining the higher order distributional properties of the hedging errors, the favored models consistently outperform their cohort. For example, while all models exhibit extreme positive skewness, which means that there are a few extreme positive outliers in these series, the chosen models are not as "bad" in this sense. This illustrates that we may want to use different types of models to dynamically hedge as opposed to calibration to current market prices.

Finally we regress model pricing errors on a number of independent variables-the moneyness index (defined as the ratio of futures price to the exercise price), time-to-maturity (measured in days to option maturity as a percent of 250 business days), strike price (as a percent of par value), short rate (settlement of annualized, continuously compounded 1month T-Bill rates), as well as the volatility of the short rate (non-parametric kernel volatility). In all models, the only significant coefficients are moneyness and time-to-maturity. We find that the pricing errors decrease with moneyness and increase with time-to-maturity. In terms of severity of the bias, the two-factor HJM model performs the worst and the non-parametric model performs the best, as judged by the magnitude of the coefficients. We can also look at the overall degree of misspecification by the size of the R-squared. Since these variables should have no explanatory power if the model is a "good fit", the lower the R-squared, the better. By this measure the 1FHJMQTSM performs the best and 1FGPTSM performs the worst.

4. SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

In this research, we empirically test different models of the term structure. First, we link two streams of research in financial economics, term structure derivatives theory and interest rate risk management. We show that various models which can be theoretically, empirically, or computationally preferred may perform differently depending upon the purpose to which they are put. Models motivated by term structure theory and having a realistic stochastic characterization of interest rates, such as the 2-factor CIR models, are found to have decent hedging and out-of-sample forecasting performance. On the other hand, models motivated by the design of risk management systems capable of accurately explaining the market prices of interest rate derivatives are found to better explain prices in-sample, although often at an exorbitant computational cost. However, significant variation in performance is found within these classes of models. Our second contribution is to compare fundamentally different modeling philosophies and theoretical perspectives (e.g., no-arbitrage, general equilibrium theory, statistical analysis, and non-parametric modeling). We show that the appropriateness of a particular approach depends upon what aspect of the term structure one we model. For example, for short-term rates, we show that diffusion function estimation works relatively better in the Japanese interest rates markets than elsewhere. Concerning bonds, a principal components approach is found to be more competitive to a neural networks approach in forecasting yield changes. Finally, for interest rate futures options, we show that while non-parametric models fit options in sample, they underperform in hedging.

This work may be extended in various directions. First, there are different classes of models that could be compared to the non-parametric alternatives examined in this research. For example, the recently developed Libor based models of the term structure (Jamshidian (1996) and Brace, Gatarek, and Musiela (1995)-BGM), which eschew with modeling a short or forward rate process for the term structure, could be profitable compared to other approaches. Second, different instruments or inter- international markets need to be analyzed, in the context of bond pricing and bond option pricing. For example, the Japanese government bond yield curve could be analyzed, or Eurodollar futures options could be priced by the techniques developed here. Third, this work could be extended to term structures models that contain credit risk premia, such as swap spreads, corporate bonds, or yield spread options. Fourth, the statistical and non-parametric approaches presented could be enhanced by considering macroeconomic variables or other predictors, to get better estimates of derivative prices. Fifth, a comparison of results for the option pricing analysis could be extended to more recent time periods or to different sampling frequencies. In regard to the former, we may compare these results on the CBOT options on Treasury bond futures to other selected years. With respect to the latter, we could look at intra-day or transaction data, in order to test these models in an analysis of the market microstructure of interest rate derivatives.

REFERENCES

Ait-Sahalia, Yacine, 1996a, Non-parametric pricing of interest rate derivative securities, *Econometrica* 64, 527-560.

________________, 1996b, Testing continuous time models of the spot interest rate, *Review of Financial Studies* 9, 385-426.

Ahn, H.H., Dittmar, R.F. and A.R. Gallant, 2000, Quadratic term structure models: theory and evidence, Manuscript, New York University.

Amin, K.I. and A. Morton, 1994, Implied volatility functions in arbitrage-free term structure models, *Journal of Financial Economics* 35, 141-180.

Bakshi, G., Cao, C. and Z. Chen, 1997, Empirical performance of alternative option pricing models, *Journal of Finance* 52, 2003-2049.

Ball, C.B. and W.N. Tourus, 1983, Bond price dynamics and options, *Journal of Financial and Quantitative Analysis* 13, 517-531.

Barone-Adesi, G. and R.E. Whaley, 1987, Efficient analytic approximation of American option values, *Journal of Finance* 42, 301-320.

Bates, D., 1996, Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche mark options, *Review of Financial Studies* 9, 69-107.

Bergstrom, A.R., 1983, Guassian estimation of structural parameters of higher-order continuous time dynamic models, *Econometrica* 51, 117-152.

Black, F., 1976, Pricing of commodity contracts, *Journal of Financial Economics* 3, 167-179.

_____ and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637-659.

______, Derman, E. E. and W. Toy, 1990, A one-factor model of interest rates and its application to treasury bond options, *Financial Analysts Journal* 46, 33-39.

______ and P. Karasinki, 1991, Bond and option pricing when short rates are log-normal, *Financial Analysts Journal* 47, 52-59.

Bliss, R. R. and P. Ritchken, 1996, Empirical tests of two state-variable Heath-Jarrow-Morton models, *Journal of Money Credit and Banking* 28, 452-481.

__________, 1997, Movements in the term structure of interest rates, *Economic Review*, Federal Reserve Bank of Atlanta, 4th Quarter, 16-33.

Bollerslev, 1987, A conditionally heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics* 69, 542-547.

Boyle, P.P. and W.D. Tian, 1999, Quadratic interest rate models as approximations to effective rate models, *Journal of Fixed Income* 9, 69-81.

Box, G. E. P. and D.A. Pierce, 1970, Distribution of residual correlations in autoregressive-integrated moving average time series models, *Journal of the American Statistical Association* 65, 1509–1526.

Brace, A., Gatarek, D., and M. Musiela, 1995, The market model of interest rate dynamics, *Mathematical Finance* 7, 127-154.

Brennan, M.J. and E.S. Schwartz, 1979, A continuous time approach to the pricing of bonds, *Journal of Banking and Finance* 3, 133-155.

______________, 1982, An equilibrium model of bond pricing and a test of market efficiency, *Journal of Financial and Quantitative Analysis* 17, 301-329.

Brown, S.J. and P.H. Dybvig, 1986, The empirical implications of the Cox, Ingersoll, Ross theory of the term structure of interest rates, *Journal of Finance* 41, 617-630.

Bühler, W., Uhrig-Homburg, M., Walter, U. and T. Weber, 1999, An empirical comparison of forward- and spot-rate models for valuing interest-rate options, *Journal of Finance* 54, 269-305.

Cakici, N., 1989, The pricing of interest rate dependent securities and the market price of risk, Unpublished Ph.D. Dissertation, The City University of New York.

Carr, P., Jarrow, R. and R. Myneni, 1992, Alternative characterizations of American put options, *Mathematical Finance* 2, 87-106.

Chan, K.C., Karolyi, G.A., Longstaff, F.A. and A.B. Saunders, 1992, An empirical comparison of various models of the short-term interest rate, *Journal of Finance* 47, 1209-1227.

Chen, R.R. and L. Scott, 1993, Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* 3, 14-31.

Chernov, M. and E. Ghysels, 2000, Estimation of stochastic volatility models for the purpose of option pricing, in Abu-Mostafa, Y.S., Lebaron, B., Lo, A., and A.S. Weigend, eds.: *Computational Finance 1999* (M.I.T. University Press, Cambridge).

Chung, K.L. and R.J. Williams, 1990, *An introduction to stochastic integration*, (Birkhauser, Boston).

Constantinides G.M. and J.E. Ingersoll, 1992, Optimal bond trading with personal taxes, *Journal of Financial Economics* 7, 85-116.

Cox, J.C., 1983, An equilibrium model of bond pricing and a test of market efficiency, *Journal of Financial and Quantitative Analysis* 17, 301-329.

______ and S.A. Ross, 1976, The valuation of options for various stochastic processes, *Journal of Financial Economics* 3, 145-166.

________ Ingersoll, J.E. and S.A. Ross, 1985a, An intertemporal general equilibrium model of asset prices, *Econometrica* 53, 363-384.

________________, 1985b, A theory of the term structure of interest rates, *Econometrica* 53, 385-407.

Courtadon, G.,.1982, The pricing of options on default free bonds, *Journal of Financial and Quantitative Analysis* 17, 75-100

Dai, Q., Liu, J. and K. Singleton, 1998, Admissibility and identification problems of affine asset pricing models, Manuscript, Stanford University.

Dickey, D.A. and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74, 427-31.

Dietrich-Campbell, B. and E.S. Schwartz, 1986, Valuing debt options: empirical evidence, *Journal of Financial Economics* 16, 321-343.

Dothan, U.L., 1978, On the term structure of interest rates, *Journal of Financial Economics* 6, 59-69.

Duffee, G., 1998, Treasury yields and corporate bond spreads, *Journal of Finance* 53, 2225–2242.

Duffie, D.and K. Singleton, 1993, Simulated moments estimation of Markov models of asset prices, *Econometrica* 61, 929-952.

______, 1996, *Dynamic asset pricing theory*, 2nd ed. (Princeton University Press, Princeton).

______ and R. Kan, 1996, A yield factor model of interest rates, *Mathematical Finance*

6, 379-406.

________, J. Pan and K. Singleton, 1998, Transform analysis and option pricing for affine jump diffusions, Manuscript, Stanford University.

Dunn, K.B. and J.J. McConnell, 1981, Valuation of GNMA mortgage-backed securities, *Journal of Finance* 36, 599-616.

Engle, R., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50, 987-1007.

Durbin, J. and G.S. Watson, 1950, Testing for serial correlation in least squares regression I, *Biometrika*, 37, 409-428.

Fama, E.F., 1982, Inflation, real returns and capital investment, *Journal of Monetary Economics* 17, 297-323.

Flesaker, B., 1993, Testing the Heath-Jarrow-Morton/Ho-Lee model of interest rate contingent claims pricing, *Journal of Financial and Quantitative Analysis* 28, 483-495.

Gallant, A.R. and G. Tauchen, 2000, The efficient method of moments, Manuscript, University of North Carolina.

Geweke, J. and S. Porter-Hudak, 1983, The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4, 221-237.

Geske, R. and H. Johnson, 1984, The American put option valued analytically, *Journal of Finance* 39, 1511-24.

Gibson, R. and E.S. Schwartz, 1990, Stochastic convenience yields and the pricing of oil contingent claims, *Journal of Finance* 45, 959-976.

Granger, C.W.J. and R. Joyeaux, 1980, An introduction to long range time series models and fractional differencing, *Journal of Time Series Analysis* 1, 15-30.

Hamilton, J.D., 1994, *Time series analysis* (Princeton U. Press, Princeton).

Hanke, M., 1999, Neural networks vs. Black-Scholes: an empirical comparison of the pricing accuracy of two fundamentally different option pricing methods, *The Journal of Computational Intelligence in Finance* 7, 26-35.

Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029-1054.

Hardle, 1990, *Applied Nonparametric Regression* (Cambridge University Press, London).

__________ and J. Scheinkman, 1995, Back to the future: Generating moment implications for continuous time Markov processes, *Econometrica* 63, 767-804.

Harrison, J.M. and D.M. Kreps, 1979, Martingales and arbitrage in multiperiod security markets, *Journal of Economic Theory* 20, 381-408.

__________ and S. Pliska, 1981, Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes and their Application* 11, 215-260.

Heath, D., Jarrow, R., and A. Morton, 1990, Bond pricing and the term structure of interest rates: a discrete time approach, *Journal of Financial and Quantitative Analysis* 25, 419-440.

____________, 1992, Bond pricing and the term structure of interest rates: a new method for contingent claims valuation, *Econometrica* 60, 77-105.

Heston, S, 1993, A closed form solution for options with stochastic volatility with applications to bond and currency options, *Econometrica* 60, 77-105.

Ho, T. S. and S. Lee, 1986, Term structure movements and pricing interest rate contingent claims, *Journal of Finance* 41, 1011-1029.

________, Stapleton, R. and M.G. Subrahmanyam, 1997, The valuation of American options with stochastic interest rates: a generalization of the Geske-Johnson technique, *Journal of Finance* 52, 827-840.

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

Hosking, J.R.M., 1981, Fractional differencing, *Biometrika* 68, 165-176.

Hull, J., and A. White, 1990, Pricing interest-rate-dependent securities, *Review of Financial Studies* 3, 573-592.

___________, 1993, One-factor interest rate models and valuation of interest rate derivative securities, *Journal of Financial and Quantitative Analysis* 28, 235-254.

Hutchinson, J.M., Lo, A.W. and T. Poggio, 1994, A nonparametric approach to pricing and hedging derivative securities via learning networks, *Journal of Finance* 69, 851-889.

Ingersoll, J., 1992, Optimal consumption and portfolio rules with intertemporally dependent utility of consumption, *Journal of Economic Dynamics and Control* 16, 681-712.

Jackwerth, J.C. and M. Rubinstein, 1996, Recovering probability distributions from option prices, *Journal of Finance* 51, 1611-1631.

Jegadeesh, N. and G. Pennachi, 1996, The behavior of interest rates implied by the term structure of Eurodollar futures, *Journal of Financial and Quantitative Analysis* 26, 426-446.

Jamshidian, F, 1989, An exact bond option formula, *Journal of Finance* 4, 205-209.

__________, 1991, Forward induction and construction of yield curve diffusion models, *Journal of Fixed Income* 1, 62-74.

__________, 1997, Libor and swap market models and measures, *Finance and Stochastics* 1, 291-328.

Jiang, G., 1998, Nonparametric modeling of U.S. interest rate term structure dynamics and implications on the prices of derivative securities, *Journal of Financial and Quantitative Analysis* 33, 465-497.

Johnson, H., Kotz, J. and M. Kemp, 1992, *Univariate discrete distributions, Second Edition* (John Wiley and Sons., New York).

Jordan, B.D., Jordan, S.D. and R.D. Jorgenson, A reexamination of option prices implicit in callable Treasury bonds, *Journal of Finance* 38, 141-162.

Karatzas, I. and S. Shreve, 1988, *Brownian motion and stochastic control* (Springer-Verlag, New York).

Karlin & Taylor, 1981, *A second course in stochastic processes* (Academic Press, New York).

Litterman R. and S. Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income* 1, 54-61.

Longstaff, F.A., 1989, A non-linear general equilibrium model of the term structure of interest rates, *Journal of Financial Economics* 23, 195-224.

__________, The valuation of options on yields, *Journal of Financial Economics* 26, 97-122.

__________, 1992a, The valuation of options on yields, *Journal of Financial Economics* 26, 97-122.

__________, 1992b, Are negative option prices possible?, *Journal of Business* 65, 571-592.

_________ and E.S. Schwartz, 1992, Interest rate volatility and the term structure: a two factor general equilibrium model, *Journal of Finance* 47, 1259-1282.

__________, 1993. Implementation of the Longstaff-Schwartz interest rate model, *Journal of Fixed Income* 3, 7-14.

Ljung, G. M. and G. E. P. Box, 1978, On a measure of lack of fit in time series models, *Biometrika* 65, 553–564.

Malliaris, M. and L. Salchenberger, 1993, A neural network model for estimating option prices, *Applied Intelligence* 3, 193-206.

Mandelbrot, B.B. and J.W. Van Nees, 1968, Fractional Brownian motions, fractional noises, and applications, *SIAM Review* 10, 422-437.

Marsh, T.A. and E.R. Rosenfeld, 1983, Stochastic processes for interest rates and equilibrium bond prices, *Journal of Finance* 38, 635-646.

Melino, A., 1990, Estimation of continuous time models in finance, in C. Sims, ed., *Advances in Econometrics, Sixth World Conference*, Vol. II.

Merton, R.C., 1973, Theory of rational option pricing, *Bell Journal of Economics and*

Management Science 4, 367-382.

Millerstein, K. Sandman, K. and D. Sonderman, 1997, Closed form solutions for term structure derivatives with log-normal interest rates, *Journal of Finance* 52, 409-430.

Moraleda, J.M. and A. Pelsser, 1997, Forward vs. spot interest rate models of the term structure: an empirical comparison, Manuscript, Tinbergen Institute of Erasmus University.

Nowman, K.B., 1997, Guassian estimation of single-factor continuous time models of the term structure of interest rates, Journal of Finance 52, 1395-1706.

Pearson, N.D. and T. Sun, 1994, Exploiting the conditional density in estimating the term structure: an application of the Cox, Ingersoll, and Ross model, *Journal of Finance* 49, 1279-1304.

Pelsser, A., 1994, A tractable yield-curve model: The direct approach to debt option pricing, *The Review of Futures Markets* 13, 461-514.

Phillips, P.C.B. and P., Perron, 1988, testing for a unit root in time series regression, *Biometrika* 75, 335-346.

Protter, P., 1990, *Stochastic integration and differential equations* (Springer, New York).

Ramaswamy, K. and M. Sundaresen, 1985, The valuation of floating-rate instruments: Theory and evidence, *Journal of Financial Economics* 17, 251-272.

Rebonato, R., 1996, *Interest rate option models* (John Wiley & Sons., Chichester, UK).

Rendleman, R. and B. Barter, 1980, The pricing of options and debt securities, *Journal Financial and Quantitative Analysis* 15, 11-24.

Revuz, D. and M. Yor, M., 1991, *Continuous martingales and Brownian motion* (Springer-Verlag, New York).

Ritchken and Sankarasubramanian, 1995, Volatility structures of forward rates and the dynamics of the term structure, *Mathematical Finance* 5, 55-72.

Rubinstein, M., 1994, Implied binomial trees, *Journal of Finance* 49, 771-818.

Schaeffer, S.M. and E.S. Schwartz, 1984, A two-factor model of the term structure: An approximate analytical solution, *Journal of Financial and Quantitative Analysis* 19, 413-424.

Sundaresan, S.M., 1989, Consumption and equilibrium interest rates in stochastic production economies, *Journal of Finance* 39, 77-92.

Stambaugh, R.F., 1987, The information in forward rates, *Journal of Financial Economics* 21, 41-70.

Stanton, R., 1997, A non-parametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* 52, 1973-2002.

Stutzer, M., 1996, A simple non-parametric approach to derivative security valuation, *Journal of Finance* 51, 1633-51.

________ and M. Chowdury, 1999, A simple non-parametric approach to bond option pricing, *Journal of Fixed Income* 8, 67-76.

Oksendahl, B., 1995, *Stochastic differential equations: An introduction with applications, 4th ed.* (Springer-Verlag, New York).

Uhrig W.M. and U. Walter, 1996, A new numerical approach to fitting the initial yield curve, *Journal of Fixed Income* 5, 82-90.

_______, 1996, An empirical examination of the Longstaff-Schwartz bond option pricing model, *Journal of Derivatives* 4, 41-54.

Vasicek, O., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177-188.

White, H., 1989, Some asymptotic results for learning in single layer feedforward networks, *Journal of the American Statistical Association* 11, 1008-1013.

_______, 1990, Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings, *Neural Networks* 3, 535-550.