Evaluating Stochastic Discount Factors from Term Structure Models

by

Heber K. Farnsworth

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

University of Washington

1997

Approved by Wayne Ferson
(Chairperson of Supervisory Committee)

Program Authorized
to Offer Degree Business Administration

Date June 25, 1997

UMI Number: 9806973

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

In presenting this dissertation in partial fulfillment of the requirments for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature Heber Farnsworth

Date 19 June 1997

University of Washington

Abstract

Evaluating Stochastic Discount Factors from Term Structure Models

by Heber K. Farnsworth

Chairperson of Supervisory Committee: *Professor Wayne Ferson*
Department of Finance

This paper introduces a new approach to testing continuous time models which can be applied to almost any pricing model and does not rely on the pricing equation having a closed form solution. It also allows use of more primitive assets then factors without having to specify an error distribution. This is achieved by examining the model in terms of the Stochastic Discount Factor (SDF) implied by the model. In this paper I examine the performance of asset pricing models from the term structure literature to illustrate the approach. I propose an observable proxy for the SDF from continuous time models and demonstrate via monte carlo methods that this proxy is an extremely good approximation to the true, unobservable SDF. I show how to use the SDF to value a series of returns. I find that the most popular one-factor parametric models can be rejected by the data. I also examine the most popular two-factor models which have been proposed in the literature.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This paper has benefitted from useful comments by my chairman, Wayne Ferson, as well as Richard Bass. Comments by seminar participants at the University of California at Berkeley, the University of Illinois at Urbana-Champagne, Ohio State University, Columbia University, Arizona State University, the University of Arizona, the University of Utah, Brigham Young University and the University of Southern California are also acknowledged. All remaining errors are my own. I also gratefully acknowledge financial support provided by the Thomas Foster as well as the Samuel Stroum fellowhips at the University of Washington.

Chapter 1

INTRODUCTION

1.1 *Stochastic Discount Factors*

A unifying concept in the modern asset pricing literature is the notion of a *stochastic discount factor* or *pricing kernel.* A stochastic discount factor (SDF) is a scalar random variable $m(t, T)$ having the property that:

$$E_t(m(t,T)R_{i,T}) = 1, \ \forall i \tag{1.1}$$

where $R_{i,T}$ is the gross return to be realized at time T on the ith asset purchased at time $t < T$, and the subscript on the expectation operator indicates that the expectation is conditional on information available to market participants at time t. Each asset pricing model posits a different form for $m(t, T)$. For example, consider the stochastic Euler equations that arise in inter-temporal optimization models. In a representative agent economy with additive time separable utility the Euler equation is of the form

$$p_t u_c(c_t, t) = E_t[u_c(c_{t+\tau}, t+\tau)p_{t+\tau}]$$

where $u_c(c_t, t)$ is the marginal utility of the optimal consumption flow c_t at date t and p_t is the value of an asset at date t. Dividing the right hand side by the left gives

$$1 = E_t[\frac{u_c(c_{t+\tau}, t+\tau)}{u_c(c_t, t)} R_{t+\tau}] \tag{1.2}$$

from which we see that the ratio of marginal utilities is a stochastic discount factor. Looking at asset pricing models in terms their stochastic discount factors is useful because the testable implications of a model (risk/return tradeoff, etc) are, in fact, statements about the SDF. Because of this, recent work in empirical asset pricing has been focused on examining the SDF directly. This move has come in part because it implies a very straightforward approach to testing, based on the Generalized Method of Moments (GMM) as in Hansen and Singleton [20]. The empirical contingent claims literature has lagged behind in adopting this approach to testing. This paper attempts to bridge this gap by showing how the stochastic discount factors implied by continuous time models can be identified and estimated. This method can be applied to a broad class of continuous-time models.

Contingent claims models provide an excellent framework for SDF methodology. These models allow us to write the SDF in terms of an observable state variable which is the price of a financial asset rather than a macroeconomic aggregate. This allows us to avoid the known measurement problems associated with variables such as aggregate consumption or wealth. In addition the problem of instrument selection is avoided because these models specify that the state variables generate all information which is available to traders. Hence only lagged values of the state variable are involved in the conditional expectation in (1.1). Previous studies using the stochastic discount factor approach have chosen instruments in a rather ad hoc manner.

1.2 Testing Term Structure Models

Although this paper deals with testing term structure models most of the discussion applies equally well to testing contingent claims models in general. The reason is that a bond, like an option, has a maturity date and the timing and amounts of future payments are stipulated in the bond contract. The difference is in what state variables the price of the bond is contingent upon. For an option this is obvious because the option is written on another claim. With bonds we assume that there are one or more state variables in the economy and that the bond can be viewed as a contingent claim on those state variables. A model of the term structure specifies three things

1. How many state variables does it take to adequately describe term structure dynamics?

2. What is the stochastic behavior of these state variables?

3. What is the reward (if any) which the market gives for bearing the risk inherent in each state variable?

The answers to these questions define a functional relationship between state variables and bond prices. This function is given by the solution to a particular differential equation which may or may not have a solution which we can write in closed form.

In some models the state variables are identified as variables which are directly observable (such as interest rates). In these cases much work has been done relating to the second question above. Papers like Chan, Karolyi, Longstaff and Sanders (CKLS) [11], and Aït-Sahalia [2] have looked at whether particular stochastic models for state prices are reallistic. This literature has become quite large recently. Such research serves a very useful purpose in model building but it is limited since it can only be used when state variables are directly observable. This paper instead focuses on testing the relationship between state variables and bond prices which each model implies [1].

Any attempt to test the pricing relation of a contingent claim model must first address the following question. *At any given time we observe more prices than state*

[1] Although we are using the term bond this discussion applies equally well to options on bonds, interest rate futures, and other assets.

variabes but how can we use them all in testing?. In order to illustrate this problem consider a two-state variable model for which the state variables are not observed. Now suppose the econmetrician has three bond prices available at every point in time. Using any two bond prices she could (in principal) use the pricing formula to "back out" the state variable values for that date. Doing this for every date gives a time series of the state variable. However the series she gets may depend on her choice of bonds to use. If this is the case then, strictly speaking, the model is not consistent with all observed prices and so must be false. There is no room for pricing errors in this approach. While logically correct this sets a higher standard for our models then we may wish to apply. There are at least two alternatives which have been adopted in the literature.

Add pricing errors to model If we assume that the observed prices are equal to model prices plus a mean zero error then we can "average" in some sense. An example is the work of Brown and Dybvig [10] who estimate parameters and the implied short rate level for each date from the one-factor model of Cox, Ingersoll, and Ross [14] using non-linear cross sectional regressions involving bonds of various maturities. As they state

> To estimate the parameters of the model we make the further assumption that the bond price quoted in period t, $V(t, c, d)$, deviates from the model price $V^*(t, c, d)$, by a mean zero error ... (which) is assumed to be independent and identically distributed as Normal in the cross section of bonds that cover the maturities traded at that point in time.

A similar assumption was made by Jegadeesh and Pennacchi [24] in order to apply Kalman filter techniques to a two factor model.

> We assume that the futures prices at which current transactions could be made are observed with error ... the error terms are assumed to be normal

and independent both serially and cross-sectionally ...

Of course how one chooses to model these error terms has profound implications for how the empirical work should be done. Normally distributed errors were convenient in these cases in light of the econometric approaches the authors were using. However there is little guidance as to how to choose an error distribution. Brown and Dybvig point out that the existence of these errors could be motivated by the fact that the data are the mean of bid and ask prices rather than actual transaction prices. However this motivation seems inconstistent with normally distributed errors unless we believe that spreads could be infinitely wide in some state of the world.

One could attribute pricing errors to model mispecification. However Brown and Dybvig point out that if this is the case then it is unreasonable to assume that the errors are independent across assets.

The real difficulty here is that we are dealing with models which were derived under the assumption that riskless hedges could be formed and so there is no mispricing possible. Hence any assumed error distribution is in some sense inconsistent with the model. Such an approach replaces the model by an incomplete markets model.

Use only as many bonds as state variables Another approach is to throw away the exta bonds and examine the implied state variables directly to see if their distribution is the same as the hypothesized one. This was the approach taken by Pearson and Sun [30] who used likelihood methods to examine a two-factor model using two bonds. Likelihood methods are, of course, very desirable from an estimation standpoint. However in order to test the model the authors had to estimate a more general model and see if the restrictions of the model they were examining held. The paper by Dai and Singleton [15] is similar in that they use three bills to estimate a three factor model. However they used simulated method of moments techniques which do allow for tests of fit. In either case the tests are based on the information in the prices of only two or three assets which limits our confidence in the inferences so obtained. It is not very convincing to say that we have failed to reject a model of the entire yield curve based on observations at two or three points on the curve.

A paper which avoids this issue is Gibbons and Ramaswamy [18] who test the CIR model by comparing moments of bond prices with those implied by the model. This in the spirit of this paper but our approach is simpler in that we need not be able to solve the pricing PDE and find theoretical moments of prices. We only need compute the joint moments of returns and state-variables implied by the SDF relation (1.1). In cases where the state variables are observed we can interpret this as a test of whether the relationship between the state variables and prices is correct. In cases where state variables are not observed we shall choose a number of reference assets equal to the number of state-variables and "back out" the state variables to form the SDF. Then we apply the SDF to other assets and interpret the results as a test of whether the values of all the assets used (including the test assets) is consistent with the model.

Previous tests of term structure models could only be applied to very simple claims (like bonds or call options on discount bonds). When viewed as a SDF the model can be applied to more general returns such as portfolios of these assets where the exact makeup of the portfolio is unknown. In addition to expanding the set of assets that can be used in testing this means that the previously almost untouched area of performance evaluation of bond funds can be examined using modern models[2].

In summary this paper presents a method for testing contingent claims pricing models which is very simple, can be applied to any type of claim and allows us to use all the data available to us without having to hypothesize the existence (or the distribution of) pricing errors.

[2] Some work on performance evaluation of bond funds does exist. I am aware of three papers: Blake, Elton, and Gruber [5, 6] and Kang [25]. However none of these papers use term structure models in evaluating performance

Chapter 2

TERM STRUCTURE MODELS

2.1 Mathematical Framework

The term structure models we shall examine are cast in a continuous-time framework. The underlying uncertainty in the model comes from state variables that are modeled as solutions to stochastic differential equations (SDE's) of the form

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \tag{2.1}$$

where $\{W_t,\ t \geq 0\}$ is a Weiner process. For the time being we will restrict our attention to the case of one state variable.

By employing arbitrage arguments it can be shown [1] that prices of claims C must satisfy a certain parabolic partial differential equation (PDE)

$$-\frac{\partial C}{\partial t} + rC = \mathcal{A}C;\ in\ [0,T) \times \Re \tag{2.2}$$

with terminal condition [2] equal to the terminal value of the claim.

$$C(T,x) = g(x);\ x \in \Re \tag{2.3}$$

where $g(x)$ is the time T payoff of the claim as a function of the state variable, r is the instantaneous risk-free rate of interest, and $\mathcal{A}$ is a second-order differential operator.

[1] c.f. Hull [23] or the appendix of this paper

[2] Sometimes C is expressed as a function of τ, the time to maturity $(T-t)$ rather than t. In this case the only difference is that the partial derivative on the left becomes a partial derivative with respect to τ and the minus sign is dropped. In this case the terminal condition becomes the initial condition of the PDE

In the one state-variable framework $\mathcal{A}$ is the operator which when applied to a twice differentiable function C, gives

$$\mathcal{A}C = \frac{1}{2}\sigma^2(x)\frac{\partial^2 C}{\partial x^2} + (\mu(x) - q(x)\sigma(x))\frac{\partial C}{\partial x}.$$

where q is a function that does not depend on the nature of the claims and is sometimes called the *market price of risk.*

The argument by arbitrage used in deriving the above shows that a PDE like this one must hold *for some* q. However there is no guidance as to how to choose a functional form for q. Models which simply assume a specific form are called *Arbitrage Models.* By contrast a *General Equilibrium Model* specifies q as part of the equilibrium. We will look at examples of both types of models.

2.1.1 Testing term structure models

Intuitively the question we would like to address in testing a model of the type described above is "Are predicted prices as given by the model consistent with observed prices?". We are interested in whether model prices are *close to* observed prices in some sense. We would like the model prices to be be correct on average, so that the model has no bias. Also we would like any deviations to be unrelated to the state variable so that the model is using all the information in the state variable. This leads one to consider tests based on

$$E\{p_t - C(X_t, t)|X_t\} = 0 \tag{2.4}$$

where p_t denotes the observed price of a claim at time t. While very intuitive there are several problems with this approach.

First, in order to implement this approach we must solve the PDE above for each asset we examine and for each time period in our sample. In cases where the PDE has no closed form solution this is a daunting task. In order to estimate parameters we must repeat this process for every function evaluation in whatever optimization algorithm we are using. The situation becomes worse if we wish to use data on American style claims which can be exercised early.

Second, although (2.4) is intuitively motivated it is statistically incorrect. Note that the null hypothesis we wish to test is

$$H_0 : p_t = C(X_t, t).$$

In words the null hypothesis is that model prices are exactly equal to observed prices. The distribution of any test statistic based on (2.4) under this null must be degenerate. This is because the framework of arbitrage pricing does not allow for any pricing errors. However this type of testing *requires* the existence of such errors.

The next section presents a methodology which retains the intuitive appeal of (2.4) while avoiding the problems noted above.

2.1.2 SDF's from term structure models

In the absence of arbitrage, it can be shown (see Harrison and Kreps [22]) that prices of contingent claims which make no payments until expiration are given by the "risk neutral" expectation of the value of the claim at expiration discounted at the risk-free rate. This means that the expectation is not taken with respect to the true distribution of the claim but with respect to an alternative probability distribution.

To make the connection between this probabilistic interpretation of the price and the PDE above we note that the solution to the pricing equation (2.2) is

$$C(t, x) = \int_{-\infty}^{\infty} \exp\left(-\int_t^T r_s ds\right) g(\xi)\Gamma(x, t, \xi, T) d\xi \tag{2.5}$$

where Γ is the *fundamental solution*[3] of the differential equation. It is well known that fundamental solutions to parabolic PDEs are transition probability density functions

[3] It has become common in finance to refer to the Green's function of a PDE. The fundamental solution is a more basic construct and has the same interpretation in terms of Arrow-Debreu

(conditional pdf's) of Ito processes. This means that (2.5) is a conditional expectation. However the process for which Γ is the conditional pdf is not the process which solves (2.1). Instead it is the diffusion process which solves the following SDE[4]

$$dX_t = (\mu(X_t) - q(X_t)\sigma(X_t))dt + \sigma(X_t)dW_t;\ 0 \leq t < \infty$$

Equation (2.5) states that the expectation using this conditional distribution of the terminal value of the claim, $g(X)$, times the discount function, $\exp\{-\int_t^T r_s ds\}$, gives the price of the claim. So Γ must correspond to the "risk neutral" probability measure. So we can rewrite (2.5) as[5]

$$C(t,x) = \tilde{E}[\exp\{-\int_t^T r_s ds\}g(X_T)|X_t = x]. \tag{2.6}$$

The tilde over the expectation operator indicates that the expectation is with respect to the risk neutral probability measure[6].

This equation has little empirical content since the data we observe are assumed to be generated by the process (2.1) and so we cannot estimate the risk neutral moment above. However Harrison and Kreps proved there exists a non-negative function, $\psi(X)$, which serves as a *density*[7] between the true probability distribution and the

---

securities. The Green's function also has the property that the integral of the terminal condition times the Green's function gives the solution to the PDE. The difference is that the term Green's function only applies to PDEs that have a boundary condition in addition to the terminal condition. In this case the Green's function is related to the fundamental solution in a simple way (see Friedman [17]).

[4] This process, sometimes called the risk-neutral process, is easy to identify from (2.2). It is the process which has $\mathcal{A}$ as its infinitesimal operator.

[5] The astute reader will note that if r_t is stochastic then it must be a function of X_t in order for this expectation to make sense. In fact in the singe state variable models we will examine we will have $r_t = X_t\ \forall\ t$.

[6] This equation is called the Feynman-Kac representation of the solution to (2.2).

[7] The term density in probability theory refers to a Radon-Nikodym derivative between two mea-

risk neutral one. This means that using $\psi(X)$ we can rewrite (2.6) as an expectation with respect to the true probability measure:

$$C(t,x) = E[\psi(X_T)\exp\{-\int_t^T r_s ds\}g(X_T)|X_t = x] \tag{2.7}$$

We call $\psi(X)$ a *State-Price Density* (SPD hereafter) [8]

The difference between the process under the risk neutral probability measure and the hypothesized process for the state variable is that the drift of the process changes from $\mu(X_t)$ to $\mu(X_t) - q(X_t)\sigma(X_t)$. The SPD $\psi(X_t)$ which changes the drift in this way can be easily found via Girsanov's theorem and is given by[9]

$$\psi(X)_T = \exp(-\int_t^T q(X_s)dW_s - \frac{1}{2}\int_t^T q(X_s)^2 ds) \tag{2.8}$$

Substituting this into (2.7) we get an expression similar to equation 18 of [32] (in that paper it was assumed that $X = r$ and $g = 1$).

An interesting and important property of the SPD is that it is a mean one martingale when indexed by T. From (2.7) we obtain a general expression for the stochastic discount factor. It is given by the discount function times the SPD, i.e.

$$m(t,T) = \exp(-\int_t^T r_s ds - \int_t^T q(X_s)dW_s - \frac{1}{2}\int_t^T q(X_s)^2 ds) \tag{2.9}$$

In a one state variable general equilibrium model where the state variable is an Ito process as in (2.1) then (2.9) follows immediately from (1.2). Note that optimal

sures. The probability density function $f(X)$ of a random variable X, for instance, is the Radon-Nikodym derivative of the measure which X induces on the real line with respect to Lebesgue measure on the real line so that we can compute the expectation of $g(X)$ by computing the integral of $g(X)f(X)$ with respect to Lebesgue measure.

[8] There is not perfect agreement in the literature on the usage of this term. Some use it to refer to the SDF.

[9] See the appendix for details of how to apply Girsanov's theorem to change the drift of a diffusion.

consumption for the representative agent will be a function of the state variable and so will be an Ito process.

$$dc^* = \mu_c(X_t)dt + \sigma_c(X_t)dW_t$$

If we write the ratio of the marginal utility of the representative agent in terms of the state variable we get (2.9) with [10]

$$q(X) = \frac{u_{cc}\sigma_c}{u_c}$$

Now consider rewriting (2.4) using (2.7).

$$E\{p_t - E[m(t,T)g(X_T)|X_t]|X_t\} = 0.$$

But this is the same as

$$E\{p_t - m(t,T)g(X_T)|X_t\} = 0.$$

There are several things to point out about this as an alternative to (2.4). First, in order to perform tests based on this equation we do not need to solve a PDE. We need only to construct m which is equally easy for all models, even when they have no closed form solution.

We can also handle American style claims. For an American claim the above should be changed to

$$E\{p_t - m(t,\tau)g(X_\tau)|X_t\} = 0$$

where τ is the date of exercise. Let t' be a fixed time which is before τ. Then we have

$$\begin{aligned} &E\{p_t - E[m(t,\tau)g(X_\tau)|X_{t'}]|X_t\} \\ &= E\{p_t - m(t,t')E[m(t',\tau)g(X_\tau)|X_t\} \\ &= E\{p_t - m(t,t')C(X_{t'},t')|X_t\} \end{aligned}$$

[10] See the appendix for a short proof, see also Theorem 4 of [13].

so we can use any future value $C(X_{t'}, t')$, not just the terminal payoff.

Note also that this equation does not require the existence of pricing errors. This is because even under the null hypothesis the quantity inside the expectation is not degenerate. However it is not the case that we will reject any model in which we observe mispricings. For instance, in the above equation we can replace $C(X_{t'}, t')$ by the observed price $p_{t'}$. Let $\eta = C(X_{t'}, t') - p_{t'}$. As long as we have that $E(m(t, t')\eta | X_t) = 0$ we will not reject the model. This has a nice economic interpretation as discussed in Bossaerts and Hillion [7]. Consider a position which is short the claim at time t and is long in a hedging strategy (based on the model) in the amount p_t. If the model is not exactly correct then the payoff to this "hedged" position will be η. But, recalling the definition of m in term of the ratio of marginal utilities, if $m(t, t')\eta$ has conditional mean zero then the agent is indifferent to this risk. In other words the agent would agree that the model is appropriate to use for hedging. So it is advantageous that this method would not reject the model because of such mispricings.

Note that if the market price of risk is identically zero the SPD degenerates to the trivial martingale $\psi(X)_T = 1, \ \forall T$ so that the risk neutral measure and the true measure are the same. Another way to say this is that the Local Expectations Hypothesis (LEH) as described in [12] holds. The SDF in this case is simply

$$m(t, T) = \exp(-\int_t^T r_s ds)$$

We will designate this as the LEH-SDF. It is a special case of any other model [11]. It will serve as a benchmark for comparison later on.

2.2 An observable proxy for the SDF

In order to test the stochastic discount factor for a particular model we must be able to construct $m(t, T)$ from observed data. As noted in (2.9) $m(t, T)$ is an exponen-

[11] In a model with several state variable LEH-SDF will be the SDF if all the relevant market prices of risk are zero

tial of an expression involving integrals with respect to time and stochastic integrals with respect to Brownian motion. The difficulty is that these continuous time processes are not observed continuously so some discretization is necessary. There is a large literature on approximating diffusions via discrete processes using normally distributed random numbers as proxies for Brownian increments. Our aim is somewhat the reverse. We already observe discrete points of the process X_t and we wish to approximate the terms in the SDF expression using these observations. The notion of convergence we are interested in is called *weak convergence*. An approximation Y_t^Δ based on a time step of size Δ converges in the weak sense with order β to a process X_t as $\Delta \to 0$ if

$$|E(f(X_T)) - E(f(Y_T^\Delta))| \leq K\Delta^\beta \tag{2.10}$$

for some K which depends on f but not on Δ. To see that this is the appropriate notion of convergence define the following system of stochastic equations

$$\begin{aligned} X_t^1 &= \int_0^t r_s ds \\ X_t^2 &= \int_0^t q(X_s^1) dW_s \\ X_t^3 &= \int_0^t q(X_s^1)^2 ds \end{aligned}$$

and consider the random vector $\mathbf{X}_t = [X_t\ X_t^1\ X_t^2\ X_t^3]$ where X_t is the state variable process. The quantity $f(\mathbf{X}_T) = m(t,T)g(X_T)$ is a function of $\mathbf{X}_t$. So if we have approximations $Y^{\Delta,1}$, $Y^{\Delta,2}$, and $Y^{\Delta,3}$ such that $\mathbf{Y}_t^\Delta = [X_t\ Y_t^{\Delta,1}\ Y_t^{\Delta,2}\ Y_t^{\Delta,3}]$ converges weakly to $\mathbf{X}_t$ then we would have that

$$|E(m(t,T)g(X_T)) - E(exp(-Y^{\Delta,1} - Y^{\Delta,2} - \frac{1}{2}Y^{\Delta,3})g(X_T))| \leq K\Delta^\beta \tag{2.11}$$

which is what we want.

For $Y^{\Delta,1}$ and $Y^{\Delta,3}$ we can use the Riemann sum approximations

$$Y^{\Delta,1} = \sum_{i=1}^{N} r_{t_i}\Delta$$

$$Y^{\Delta,3} = \sum_{i=1}^{N} q(X_{t_{i-1}})^2\Delta$$

where we suppose that in the interval $[t, T]$ there are available N observations at times $t = t_1 < t_2 < \cdots < t_N = T$ where $t_{i+1} - t_i = \Delta$. These are known to be weak approximations of order 1[12].

$Y^{\Delta,2}$ is more difficult because it is supposed to approximate an Ito integral and we do not observe the process W_t directly. If we did observe W_t then the simplest approximation scheme to use would be the *stochastic Euler approximation.*[13] Applying this scheme to the SDE of equation (2.1) gives:

$$X_{t+\Delta} - X_t \approx \mu(X_t)\Delta + \sigma(X_t)(W_{t+\Delta} - W_t)$$

However, as mentioned above, our purpose is not to approximate X_t using $(W_{t+\Delta} - W_t)$ but to use observations of X_t to construct the approximation $Y^{\Delta,2}$.

Define the *Backward Euler Scheme* by the equation

$$\epsilon_{t+\Delta} \equiv (X_{t+\Delta} - X_t - \mu(X_t)\Delta)/\sigma(X_t)$$

then $\epsilon_{t+\Delta}$ approximates a Brownian increment so we would expect that

$$Y^{\Delta,2} = \sum_{i=1}^{N} q(X_{t_{i-1}})\epsilon_{t_i}$$

might do well. There are general conditions under which this scheme will converge weakly to X^2 but these are difficult to verify[14]. Instead we shall examine the closeness of the approximation numerically in a later section for particular models of interest.

[12] See for example [26]

[13] Not to be confused with the stochastic Euler equation of (1.2).

[14] See Theorem 14.5.2 in [26]

In summary, the approximation to the SDF of (2.9) which we shall use in this paper is

$$\hat{m}(t,T) = \exp(-\sum_{i=1}^{N} r_{t_i}\Delta - \sum_{i} q(X_{t_{i-1}})\epsilon_{t_i} - \frac{1}{2}\sum_{i} q(X_{t_{i-1}})^2\Delta) \tag{2.12}$$

This representation is useful in empirical work in cases where we have frequent observations on state variables. For example, if daily observations on the state variable are available and if we use monthly data on bond returns then $T-t$ is one month in the above equation and Δ is one day.

Below I show some examples for some well-known asset pricing models.

2.3 Single state variable models

In the analysis to this point the identity of the state variable has been left unspecified. Cox, Ingersoll, and Ross [14] showed in a general equilibrium framework that under certain conditions it is possible to "invert the term structure" and use interest rates as instrumental variables in place of the unnamed state variables. For instance, in a one state variable model the stochastic properties of the short rate of interest are completely determined by the stochastic properties of the state variable and so prices of interest rate derivative securities (including all government bonds) can be expressed in terms of the short rate rather than the state variable [15]. CIR use this fact to build a single state-variable model. They model the process for the short rate as

$$dr_t = \kappa(\theta - r_t)dt + \sqrt{r_t}\sigma dW_t$$

and the market price of risk as

$$q(X) = \frac{\lambda}{\sigma}\sqrt{r_t}$$

[15] Actually CIR point out that this is only true in a one state variable model in which (a) the representative agent has constant relative risk aversion and (b) the short rate is a monotone function of the state variable. Their one-factor model is an example of such a model

where λ is a market risk parameter.

The arbitrage model of Vasicek [32] is similar. In his model the short rate process solves

$$dr_t = \kappa(\theta - r)dt + \sigma dW_t$$

and the market price of risk q is assumed to be constant [16].

The drift term in these models captures the apparent mean reverting nature of the short rate. Since the drift of the short rate in these models is the same any difference in the performance of the models must be due to the diffusion function or the market price of risk. It is generally held that CIR is a better model because it incorporates a level effect in the diffusion function which seems to be in evidence in the data. CKLS [11] find evidence for a level effect in the diffusion but of a different form than that proposed by CIR. They estimate a process called a constant elasticity of variance (CEV) which nests both of the above models. The drift is identical to the CIR and Vasicek models but the diffusion is of the form σr_t^γ. CKLS reject the CIR formulation which restricts γ to be .5 and estimate γ to be nearer to 1.5[17]. However, Aït-Sahalia [2] has shown that the entire class of CEV processes can be rejected.

In spite of the evidence against the assumptions of both these models it may be that one of them provides an adequate model for an SDF. That is to say, some misspecification of the short rate process may not be important for pricing. This is possible because two processes may be very dissimilar and yet imply SDF's which are much the same. Therefore my objective is to test whether the associated SDF's are valid for pricing a set of returns on primitive assets.

---

[16] Although Vasicek developed this model as an arbitrage model Goldstein and Zapatero [19] have shown that this short rate process and market price of risk can be supported in a general equilibrium framework.

[17] Brenner, et al [9], and Andersen and Lund [3] have shown that the point estimate of .5 is acceptable if σ is taken to follow a GARCH or E-GARCH process. However this takes us out of the one-factor framework of this section.

2.4 *Monte Carlo Evidence*

What is the extent of the bias introduced by the discretization scheme proposed above? The answer will depend on the functions μ and σ as well as the size of Δ. We can examine this bias for a particular model by means of monte carlo simulation.

In order to examine the consequences of discretization we must be able to generate a sequence which has the same (finite dimensional) distribution as though it had been sampled from a continuous path. For the Vasicek model we can easily generate a sequence which has *exactly* the right distribution since the transition densities of the process are normal. For the CIR model some approximation is necessary so a high order approximation was used on a mesh of points between each sampled value (see the appendix for details of how each series was simulated).

Now we can construct the ϵ_{t_i} as above[18] and construct $\hat{m}(t,T)$ using formula (2.13). Our purpose is to see if $\hat{m}(t,T)R_T$ has the same mean[19] as $m(t,T)R_T$ where R_T is the return on an asset or portfolio. Recall that any portfolio of bonds may be considered as a portfolio of pure discount bonds whose weights in various maturities are selected to match the timing and quantity of the coupon payment and maturity payments of the original bonds[20]. Therefore it suffices to examine the prices of pure discount bonds. As a first step we therefore examine whether $\hat{m}(t,T)$ has the same mean as $m(t,T)$ for all T.

Consider a bond paying \$ 1 at time $T_0 > T$. If we let p_t denote the time t price of this bond then clearly

$$p_t = E_t[m(t,T_0)].$$

Now examine the intermediate time T. By the law of iterated expectations we can

[18] This introduces the bias we wish to examine

[19] This mean will be unity if the model is correct.

[20] As long as portfolio does not contain any callable bonds, variable coupon bonds, etc

rewrite the above as

$$p_t = E_t[m(t,T)E_T[m(T,T_0)]]$$

where $m(T,T_0)$ is given by the same SDF formula but with the integrals taken over $[T,T_0]$. This amounts to pricing the strategy of buying the bond at t for p_t and selling it at time T for a price p_T.

$$p_t = E_t[m(t,T)p_T].$$

Dividing through by the left hand side gives

$$1 = E_t[m(t,T)R_T]$$

which is what we wish to examine. So our strategy will be to calculate p_t using $\hat{m}(t,T_0)$ and compare it to the theoretical value which is given by the bond pricing formula.

Since the data we have are daily observations and we have on average 252 trading days in a year we use $\Delta = 1/252$. Using the algorithm described above 1000 paths of length corresponding to 1 year were simulated. Then we use the formula to compute $\hat{m}(t,T_0)$ for each path and then average over all paths. Table 2.1 presents the results for the CIR and Vasicek models. The parameter values used were $\kappa = 2$, $\theta = .06$, $\sigma = .03$, $q = -.01$ (for the Vasicek model) or $\lambda = -.01$ (for the CIR model).

Table 2.1: Comparison of theoretical price with Monte Carlo mean

	CIR	Vasicek
Price	0.9457	0.9460
Mean of m	0.9460	0.9459
Mean of $\hat{m}$	0.9456	0.9458

This suggests that $\hat{m}$ has the right mean and so will price strips correctly and hence should price bond and bond portfolios correctly. In order to check whether $\hat{m}$ will price other claims correctly we need to ask whether $E_t(\hat{m}(t,T)R(r_T)) = E_t(m(t,T)R(r_T))$ where $R(r_T)$ is some function of r_T which could be the gross return on a security. Since we can't examine all such functions $R(r_T)$ we will try to answer the question of whether the joint distribution of $(\hat{m}(t,T), r_T)$ equals the joint distribution of $(m(t,T), r_T)$. We shall do this by simulating paths for the joint process (m_t, r_t) and then using sampled values of r_t (at daily intervals) from these paths we will form $\hat{m}(t,T)$. For this exercise we will take $T - t$ to be one month[21]. Plots of 10,000 realizations of these pairs for the CIR model model are presented in figures 2.4 and 2.4.

This Suggests that for these models the approach outlined in this paper is very good. It is reasonable to assume that for other models the bias will also be small and good tests can be obtained using daily data.

2.5 *Estimation and Testing*

In order to estimate parameters of the CIR and Vasicek models we will use the GMM. The moment conditions we will use are based on the definition of a stochastic discount factor, namely

$$E_t(m(t,T)R_T) = 1$$

where R_T is the gross return, realized at time T, on an interest rate contingent security (or portfolio). This equation says that the present value of a dollar invested in a risky asset is one dollar. If for some traded asset we have that this conditional expectation is less than one then we can say that the model under-prices the asset.

[21] The same parameter values are used here as above except that the value of λ is taken to be -0.1 rather than -0.01 to more clearly show the nonlinear relationship between r and m.

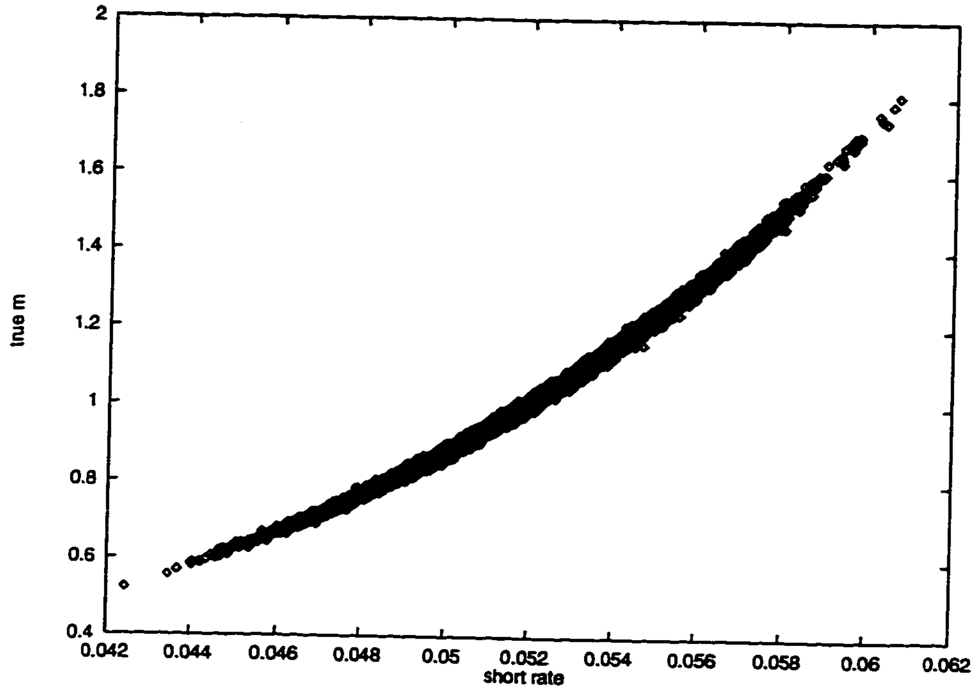

Figure 2.1: Short rate and Stochastic Discount Factor - Cox, Ingersoll and Ross Model

Similarly if the conditional expectation is greater than one then the model overprices the asset. Implicit in the formulation of these models is the assumption that the state variables convey all relevant information in the economy. This means that "admissible" trading strategies depend only on the state variables. It also means that the conditional expectation above becomes an expectation conditional on the current value of the state variables. We now convert these conditional moments into the following set of orthogonality conditions for GMM estimation

$$E\begin{bmatrix} \hat{m}(t,T)\mathbf{R}_T - \mathbf{1} \\ \hat{m}(t,T)\mathbf{R}_T \otimes z_t - \mathbf{1} \otimes z_t \end{bmatrix} = \mathbf{0}$$

where z_t is a vector of instruments which contains r_t and may also contain any functions of r_t [22].

[22] The results reported in this paper use only r_t as the instrument because it was found that adding

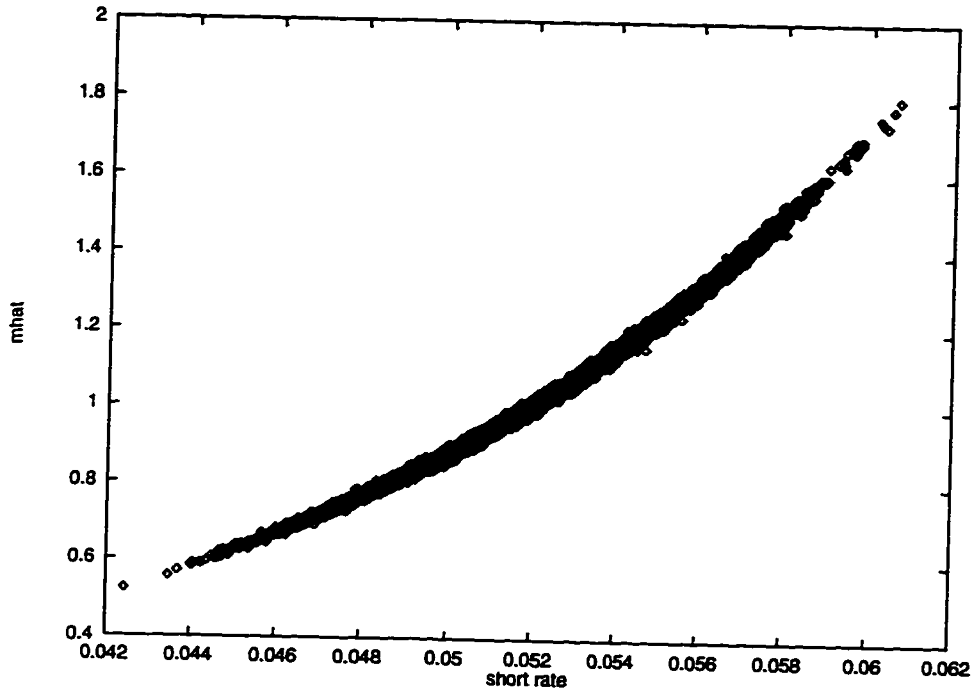

Figure 2.2: Short rate and Stochastic Discount Factor Proxy – Cox, Ingersoll and Ross Model

To form the SDF for the Vasicek model we define $\epsilon_{t_i}^{Vas} \equiv (r_{t_i} - r_{t_{i-1}} - \kappa(\theta - r_{t_{i-1}}))/\sigma$. Plugging this in to (2.12) along with the constant market price of risk we obtain

$$\hat{m}(t,T)^{Vas} = \exp(-\sum_i r_{t_i}\Delta - q\sum_i \epsilon_{t_i}^{Vas} - \frac{1}{2}q^2 N\Delta) \tag{2.13}$$

For the CIR model define $\epsilon_{t_i}^{CIR} \equiv (r_{t_i} - r_{t_{i-1}} - \kappa(\theta - r_{t_{i-1}}))/\sigma\sqrt{r_{t_{i-1}}}$. If we plug this into(2.12), recalling that the market price of risk in the CIR model depends on the parameter σ we obtain

$$\hat{m}(t,T)^{CIR} = \exp(-\sum_i r_{t_i}\Delta - \frac{\lambda}{\sigma}\sum_i \sqrt{r_{t_{i-1}}}\epsilon_{t_i}^{CIR} - \sum_i \frac{\lambda^2}{2\sigma^2} r_{t_{i-1}}\Delta)$$

Table 2.2: Descriptive Statistics of Primitive Asset Returns

Variable	N	Mean	Std Dev	Minimum	Maximum
RF	258	1.005729	0.002362	1.002040	1.013546
PORT6	258	1.006510	0.003239	1.001160	1.024230
PORT12	258	1.006755	0.005609	0.991280	1.045070
PORT48	258	1.007444	0.016273	0.939850	1.093150
PORT120	258	1.007525	0.022116	0.926460	1.109230
LTGOV	258	1.007680	0.032512	0.915900	1.152300

2.6 Data

The primitive assets which will be used used in estimation and testing all the models in this paper are the one-month risk-free return from the CRSP risk-free rates file; monthly returns on bond portfolios with target maturities of 6 months, 24 months, 4 years, and 10 years from the Fama-Bliss bond portfolios file; and the Long-Term Government Bonds: Total Returns series from Ibbotson Associates which is the holding period return on a bond which has approximately 20 years to maturity. This data covers the period from July 1973 to December 1994. Table 2.2 gives summary statistics of the primitive asset data.

For the Vasicek and CIR models the state variable is the instantaneous short rate which we take to be observable. The proxy we shall use is the bid-ask midpoint of the 7-day Eurodollar rate [23] expressed as an annualized rate. This time series covers the period from June 1973 to February 1995 comprising 5505 daily observations. This series is plotted in figure 2.3.

Since the interest rate data is expressed as a yearly rate the value of Δ is 1 divided

additional functions of r_t had little effect.

[23] I thank Yacine Aït-Sahalia for providing me with this data.

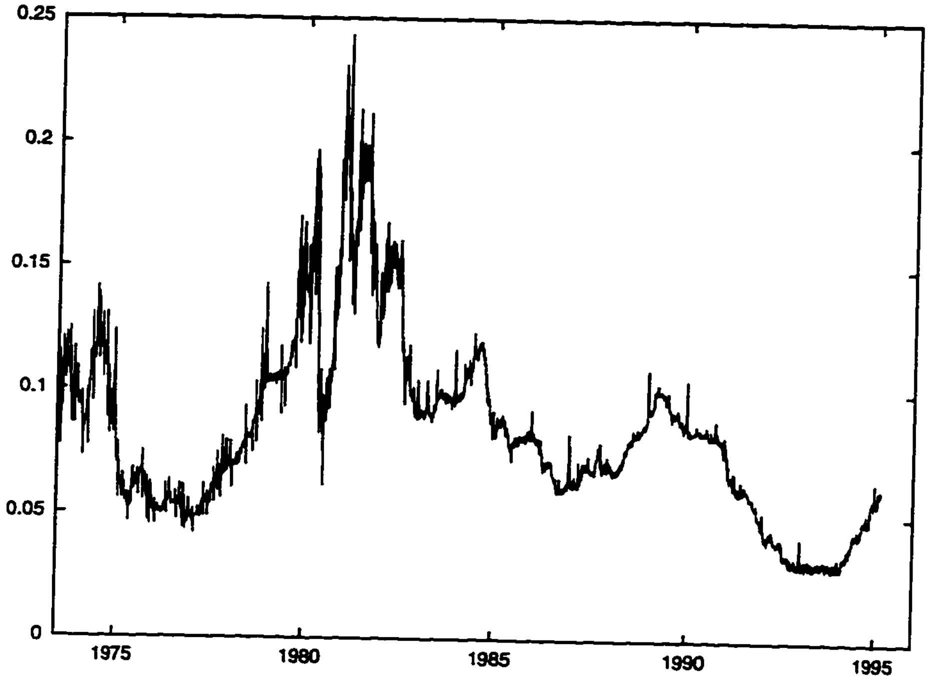

Figure 2.3: 7-day Eurodollar rate

by the number of trading days in a year (252 is the number we use). This does not allow for any "weekend effect" (i.e. Monday is considered to be the next day after Friday) but no weekend effect seems to be in evidence in this data[24]. Since the returns on the primitive assets are monthly returns the sums in the $\hat{m}$ formula are sums over all the days in a particular month.

2.7 Results for One-Factor Models

Parameter estimates for these two models are presented in table 2.3. Qualitatively speaking these parameter values do not seem unreasonable. Note that the market price of interest rate risk is negative. This confirms our intuition that if the values of most assets fall when interest rates rise then an asset which is positively correlated

[24] The variance of day to day changes is approximately the same for consecutive weekdays as it is for Friday to Monday

to interest rates will tend to reduce the risk of a portfolio it is added to. It may seem peculiar that the estimate of σ is so much lower in the Vasicek model. The is because the parameter has a different interpretation in each model. Recall that the diffusion function for the Vasicek model is σ^2 while it is $\sigma^2 r_t$ for the CIR model.

There are no t-statistics reported for either model. For the Vasicek model this is because the GMM weighting matrix was ill-conditioned so that the standard errors were enormous. For the CIR model the problem was different. Note that the point estimated of λ was very nearly zero. Now recall that any model for which the market price of risk is zero reduces to the LEH model which has no parameters. Hence at this point the gradient of the GMM objective function could not be calculated so the standard errors could not either.

However this does not prevent us from examining the performance of these two models. Our purpose is not to make inferences about the parameter values but to test whether the models are "good" in the sense that the moment conditions hold. For this reason J-statistics are reported for each model. The J-statistic is a test of the over-identifying restrictions of the GMM system and in this case it is asymptotically distributed as Chi-square with 8 degrees of freedom under the null hypothesis that the model is true. So these models are both strongly rejected.

Table 2.3: Estimates for Vasicek and Cox, Ingersoll and Ross models

	CIR	Vasicek
κ	1.2857	0.3816
θ	0.1215	0.0980
σ	0.1055	0.0156
q or λ	-1.55e-9	-0.1566
J-stat	68.6766	64.4979

It is instructive to compare these parameter estimates with estimates obtained using just the short rate data. Of course in this way we cannot estimate market price of risk parameters, only the parameters of the process.

Although we do not observe the short rate process continually we can still make estimates of the parameters of the process which are not subject to discretization bias. Aït-Sahalia [1] shows that for both of these models $E[(r_{t+\Delta} - r_t)|r_t] = (1 - e^{-\kappa\Delta})(\theta - r_t)$. This is the exact conditional expected change in the rate over an interval of length Δ. The exact conditional variance of interest rate changes is different for each model. First define $u_{t+\Delta} \equiv (r_{t+\Delta} - r_t) - E[(r_{t+\Delta} - r_t)|r_t]$. Then we have

$$E[u_{t+\Delta}^2|r_t] = \begin{cases} (\sigma^2/2\kappa)(1 - e^{-2\kappa\Delta}) & \text{Vasicek} \\ (\sigma^2/2\kappa)[2(e^{-\kappa\Delta} - e^{-2\kappa\Delta})r_t + (1 - e^{-\kappa\Delta})^2\theta] & \text{CIR} \end{cases}$$

These moments will serve to identify the parameters of the interest rate process.

$$E\begin{bmatrix} u_{t+\Delta} \\ u_{t+\Delta} \otimes z_t \\ u_{t+\Delta}^2 - E[u_{t+\Delta}^2|r_t] \\ (u_{t+\Delta}^2 - E[u_{t+\Delta}^2|r_t]) \otimes z_t \end{bmatrix} = 0$$

Table 2.7 presents such estimates. Notice that for both models the parameter θ is somewhat higher but stays reasonably close to the values dictated by the short rate. The parameter κ gets smaller for the Vasicek model. Recall that this parameter represents a "speed of adjustment" while θ is the mean of the process toward which the process reverts. This decline in κ is interesting in light of recent work by Ball and Torous [4] which points out that since the short rate is close to a unit root process then looking only at the time series behavior of the rate using GMM will give an upwardly biased estimate of this parameter.

In order to get a better feel for why these models are rejected we can examine the moment conditions which are presented in table 2.5 with standard errors and t-statistics for each model, including the LEH model. The first six rows of each table show the average pricing errors (APEs), defined as the sample mean of $m(t,T)R_T - 1$

Table 2.4: Parameter Estimates based on time series

	CIR	Vasicek
κ	0.89218	0.85837
	(0.27036)	(0.26250)
θ	0.090495	0.089102
	(0.00940)	(0.01041)
σ	0.180948	0.04678
	(0.00164)	(0.00011)

for each portfolio. This information is presented graphically in figure 2.4. Since a the CIR price of risk is essentially zero a plot of APEs for the LEH model would plot right on top of that for the CIR model and so it is not shown. This means that in terms of average pricing performance the CIR model does no better than the LEH model. Surprisingly it seems that the Vasicek model does even worse than the LEH model. This is largely a result of the poorly conditioned weighting matrix. The APEs from the first stage looked very similar to that for the CIR model but in the second stage the APEs got much worse while the other moment conditions got slightly better. The third and fourth colums of table 2.5 give the sample mean of $(m(t,T)R_T - 1)r_t$ which would be the covariance of pricing errors with the instrument if $m(t,T)R_T - 1$ was mean zero[25].

To aid our intuition for the exceptionally poor performance of the Vasicek model on average pricing error consider the following example. Suppose for simplicity that we have a SDF with only one parameter, θ. For testing we use only one return but two moment conditions

[25] The standard errors reported in the table are the square roots of the diagonal elements corresponding to these moment conditions of the inverse of the GMM weighting matrix .

Table 2.5: Moment Conditions from Vasicek and Cox, Ingersoll and Ross models

	Average Pricing Errors		Orthogonality Conditions	
	CIR	Vasicek	CIR	Vasicek
rf	-0.001305	-0.002184	-0.000137	-0.000040
	(0.000107)	(0.006639)	(0.000014)	(0.000919)
port6	-0.000530	-0.001573	-0.000059	0.000014
	(0.000146)	(0.006545)	(0.000020)	(0.000906)
port12	-0.000287	-0.001540	-0.000038	0.000008
	(0.000324)	(0.006422)	(0.000045)	(0.000891)
port48	0.000399	-0.001340	0.000019	0.000006
	(0.00100)	(0.006198)	(0.000125)	(0.000869)
port120	0.000482	-0.001478	0.000014	-0.000026
	(0.001367)	(0.006123)	(0.000167)	(0.000861)
ltgov	0.000638	-0.001739	0.000018	-0.000077
	(0.002010)	(0.005999)	(0.000234)	(0.000837)

$$g(\theta) = \begin{bmatrix} g_1(\theta) \\ g_2(\theta) \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\sum_{t=1}^{N} m_{t+1}(\theta)R_{t+1} - 1 \\ \frac{1}{N}\sum_{t=1}^{N}(m_{t+1}(\theta)R_{t+1} - 1)z_t \end{bmatrix}$$

Suppose that after the first stage we estimate the covariance matrix of the moment conditions to be

$$S = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

and we find that σ_{12} is nearly equal to σ_{22} but that σ_{11} is large relative to the other elements so that the matrix is not singular (this is roughly what happened in the Vasicek model). In the second stage we solve

$$\min_{\theta} g(\theta)' S^{-1} g(\theta)$$

which leads (after re-arranging) to the first order condition

$$g_1 = g_2 \frac{g_1' \sigma_{12} - g_2' \sigma_{11}}{g_1' \sigma_{22} - g_2' \sigma_{12}}$$

Now recalling the definitions of g_1 and g_2 we have that to a first approximation

$$g_2' \approx g_1' \sum z_t$$

Substituting this into the first order condition above gives

$$g_1 = g_2 \frac{\sigma_{12} - \sigma_{11} \sum z_t}{\sigma_{22} - \sigma_{12} \sum z_t}$$

Now recall that $\sigma_{12} \approx \sigma_{22}$ and that $\sigma_{11} > \sigma_{12}$ and assume that $\sum z_t > 1$ (as it is in this case). This means that the second stage estimate will make

$$g_1 > g_2$$

or in other words the unconditional moment will not be as small as the conditional moment. The larger variance of the unconditional moment condition has caused it to be sacrificed for a better fit of the conditional moment condition. This same phenomenon appears in other models as well.

One possible reason for the poor performance of these models is that there appear to be volatility regimes in the short rate data which is inconsistent with the one-factor framework of these models. One possible source of such behavior of the short rate is the monetary experiment conducted by the Federal Reserve which concluded at the end of 1982. A rather natural question to ask is whether in the post 1982 period these one-factor models might do a better job. Parameter estimates from this subperiod are presented for both models in table 2.6. Sample moment conditions for this estimation are presented in table 2.7 and the APE's are presented in figure 2.5. Note that the estimates and performance of the CIR model doesn't change much compared to the full sample period. However the Vasicek model changes drastically

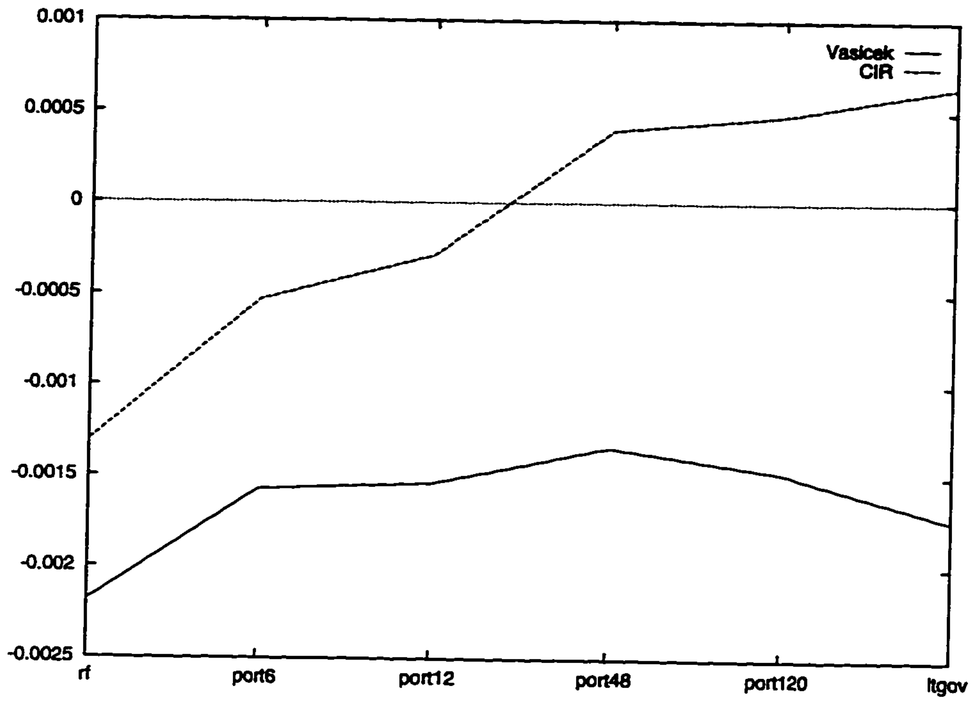

Figure 2.4: Average Pricing Errors

and in this subperiod seems to do rather well. This is what might be expected. The 1979 through 1982 period is characterized by both higher rates and higher volatility of rates. The CIR model can explicitly handle this but the Vasicek model cannot. On the other hand in the post 1982 period it appears that volatility is fairly constant which would tend to favor the Vasicek model. However it also appears that all the mean reversion comes from the 1979 through 1982 period and so the Vasicek model has a hard time fitting a linear drift to something that is essentially a random walk.

The two models we have examined in this chapter are not the only two that have ever been proposed. There are many others and we could apply the SDF methodology just as easily to these other models. However there is a growing consensus in the literature that a one state-variable model is insufficient to describe the dynamics of the term structure. For this reason several two and three state-variable models have been proposed. In the next chapter we shall examine them empirically.

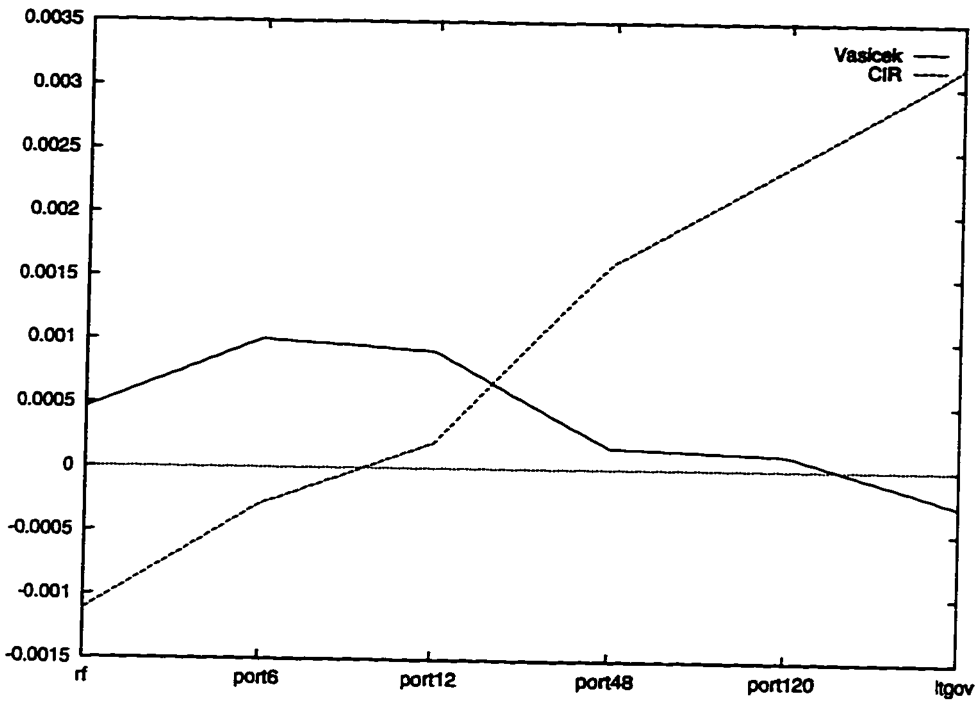

Figure 2.5: Average Pricing Errors – post 1982 subperiod

Table 2.6: Parameter Estimates – post 1982 subperiod

	CIR	Vasicek
κ	1.139734	0.023740
θ	0.117813	0.500495
σ	0.085087	0.000026
q or λ	-0.002751	-0.003441

Table 2.7: Moment Conditions – post 1982 subperiod

	Average Pricing Errors		Orthogonality Conditions	
	CIR	Vasicek	CIR	Vasicek
rf	-0.001113	0.000463	-0.000087	-0.000326
	(0.000109)	(0.066548)	(0.000009)	(0.004597)
port6	-0.000285	0.000999	-0.000024	-0.000287
	(0.000089)	(0.066537)	(0.000008)	(0.004596)
port12	0.000197	0.000913	0.000014	-0.000292
	(0.000226)	(0.066433)	(0.000019)	(0.004588)
port48	0.001611	0.000167	0.000125	-0.000347
	(0.001019)	(0.066121)	(0.000081)	(0.004563)
port120	0.002383	0.000116	0.000190	-0.000351
	(0.001556)	(0.066110)	(0.000126)	(0.004559)
ltgov	0.003165	-0.000281	0.000255	-0.000364
	(0.002486)	(0.065818)	(0.000195)	(0.004543)

Chapter 3

MULTIPLE STATE-VARIABLE MODELS

The analysis above is only slightly modified in the case of several state variables. Consider the case where X is a d-dimensional vector of state variables and W is a k-dimensional Brownian motion such that X satisfies the following SDE written component-wise as

$$dX_t^{(i)} = \mu_i(X_t)dt + \sum_{j=1}^{k} \sigma_{ij}(X_t)dW_t^j;\ 1 \le i \le d.$$

The vector $\mu(x)$ with ith element $\mu_i(t,x)$ is the *drift vector* of X and the matrix $\sigma(x)$ with entry i,j given by $\sigma_{ij}(x)$ is called the *dispersion matrix* of the process. As before we can employ an arbitrage argument to derive a PDE that claim prices must satisfy. This PDE is identical to (2.2) except that the second-order differential operator $\mathcal{A}$ applied to C gives

$$\mathcal{A}C = \frac{1}{2}\sum_{i=1}^{d}\sum_{k=1}^{d} a_{ik}(x)\frac{\partial^2 C}{\partial x_i \partial x_k} + \sum_{i=1}^{d}(\mu_i - q_i\sigma_i)\frac{\partial C}{\partial x_i}$$

where $a_{ik}(x) \equiv \sum_{j=1}^{d} \sigma_{ij}(x)\sigma_{kj}(x)$ for $1 \le i,k \le d$. The matrix $a(x)$ with entries $a_{ik}(x)$ is called the *diffusion matrix*.

Since there are d state variables each has a market price, q_i. The multidimensional version of the Girsanov theorem implies that the stochastic discount factor associated with such a model is

$$m(t,T) = \exp(-\int_t^T r_s ds - \sum_{i=1}^{d}[\int_t^T q_i dW_s^i - \frac{1}{2}\int_t^T q_i^2 ds])$$

3.1 Two Observable Factors

A well-known arbitrage model that falls into this class is the two state variable model of Brennan and Schwartz [8]. In their framework they hypothesize that the two state variables can be represented by the instantaneous rate of interest and the *long rate* defined to be the yield on a consol bond which pays coupons continuously. They begin by assuming that the short rate and long rate processes solve SDE's of the form

$$dr = \beta_1(r, l, t)dt + \eta_1(r, l, t)dW_t^{(1)}$$

$$dl = \beta_2(r, l, t)dt + \eta_2(r, l, t)dW_t^{(2)}$$

where the two Wiener processes have correlation ρ.

Brennan and Schwartz assume the following

$$\begin{aligned}
\beta_1 &= r[\alpha \ln(l/pr) + \frac{1}{2}\sigma_1^2] \\
\eta_1 &= r\sigma_1 \\
\eta_2 &= l\sigma_2 \\
\beta_2 &= l^2 - rl + l\sigma_2^2 + q_2 l\sigma_2
\end{aligned}$$

where α, p, σ_1, and σ_2 are constants. They also assume that both q_1 and q_2 are constant[1].

Brennan and Schwartz adopted a two-step approach to testing this model. First by linearizing the system in an approximate sense they estimated the parameters of the process. Then for any choice of market risk parameter the PDE can be solved numerically (there is no closed form solution to this model) and the result can be compared to observed prices of strips. Iterating this procedure, while very time

[1] Brennan and Schwartz point out that if a consol bond actually exists as a traded asset then this combination of drift and diffusion for the long rate must imply a constant price of risk.

consuming, can yield an estimate for this parameter. However this procedure is not really a test and furthermore the statistical properties of the market risk parameters estimated in the second step are uncertain.

We will estimate these parameters using the GMM as in the last section. The data for the long rate is the daily market yield on long term U.S. government securities available from the Federal Reserve (series DLTGS). Parameter estimates are given in table 3.1.

Table 3.1: Estimates for Brennan-Schwartz model

	Parameter Estimate
α	3.063664
p	6.847845
σ_1	3.363512
σ_2	-0.013633
q_1	0.180964
q_2	0.033503
J-stat	70.6146

It helps in interpretting these estimates if instead of looking at the SDE for the short rate in terms of r and we instead look at it in terms of $\ln(r)$. Ito's lemma gives that

$$d\ln(r) = \alpha(\ln(l) - ln(pr))dt + \sigma_1 dW^{(1)}$$

so that the natural log of the short rate is constantly trying to revert to the level of the log of the long rate as long as p is positive (which it is). The parameter α relates to the speed of mean reversion. Therefore it should not be surprising that the estimate of this parameter is large and positive. The interpretation of the other parameters is rather straightforward. One which might make one curious is the negative estimate for σ_2. Since the long rate is always positive this would seem to say that the diffusion function for the long rate is always negative. However this is incorrect. We can simply define another Brownian motion which is equal to $-W_t^{(2)}$ and we have a positive diffusion function again.

Moment conditions and standard errors for this model are presented in table 3.2 and are plotted in figure 3.1

Table 3.2: Moment conditions for Brennan-Schwartz model

	APE	Orth. to r	Orth. to l
rf	0.003511	0.000344	0.000039
	(0.005831)	(0.000614)	(0.000374)
port6	0.004167	0.000407	0.000079
	(0.005758)	(0.000602)	(0.000370)
port12	0.004187	0.000402	0.000084
	(0.005612)	(0.000581)	(0.000363)
port48	0.003886	0.000357	0.000081
	(0.004961)	(0.000503)	(0.000325)
port120	0.003366	0.000291	0.000065
	(0.004560)	(0.000458)	(0.000301)
ltgov	0.002542	0.000203	0.000052
	(0.003922)	(0.000396)	(0.000258)

An interesting thing to notice about this model is that although it is rejected the pattern of APEs is very different from what we saw in models which only use the short term rate as a factor. The pricing errors for longer maturity securities are very low. In some sense this is what we might expect would happen since the long rate is also a factor. Also we see in table 3.2 that the same phenomenon we observed in the Vasicek

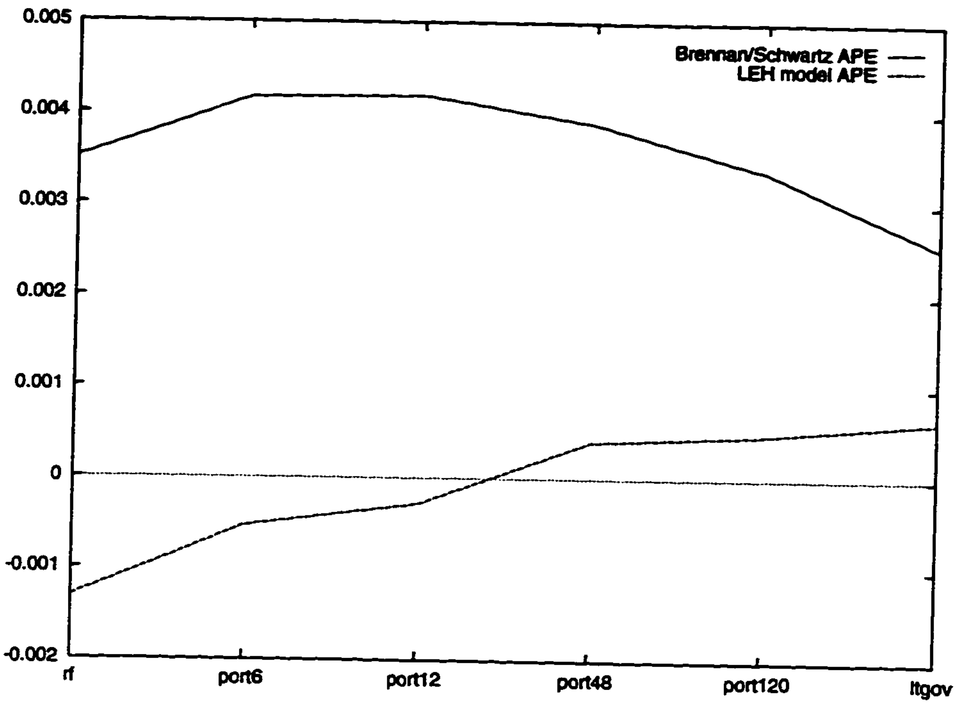

Figure 3.1: Average Pricing Errors for Brennan-Schwartz model

model seems to be in evidence here, namely that the sample moment conditions in the second and third columns (those relating to the conditional moments) are much smaller than the average pricing errors. The reason is probably the same.

3.2 Models with Unobserved Factors

In previous sections we have examined factor models in which the factors were directly observable at daily frequency. Many models which have been suggested in the literature involve factors which are either not easily observable or cannot be observed by the econometrician.

3.2.1 An Inflation Factor

Perhaps the best known model was suggested by Cox, Ingersoll and Ross [14] as a useful 2-factor extension of their one-factor model. In the two factor version they consider a security whose payoff at maturity is specified in nominal terms[2]. Its value in real terms will depend on the price level as well as the short rate and we will denote it by by $N(t, r, p)$ where p is the price level. If p is assumed to solve an SDE of the form

$$dp = \mu(p)dt + \sigma(p)dW_t^{(2)}$$

where $W^{(2)}$ is a Brownian motion independent of the Brownian motion driving the (real) short rate. CIR show that N satisfies

$$-N_t + rN = \mathcal{A}N$$

$$N(T, r, p) = \frac{g(r)}{p(T)}$$

where

$$\mathcal{A}N = \frac{1}{2}\sigma^2 r N_{rr} + \frac{1}{2}\sigma^2(p)N_{pp} + (\kappa(\theta - r) - \lambda r)N_r + \mu(p)N_p.$$

Note that the coefficient on the N_p term is the same as the drift of the price level process. This means that the drift of the price level process is unchanged under the risk neutral measure. In other words the market price of inflation risk is zero in this model. This is a direct consequence of assuming that inflation has no real

[2] The one-factor model gave the price of real claims. In a previous chapter we tested this model using both nominal short rate data and returns on nominal, rather than real claims. This was out of necessity. While one might be able to get monthly returns on primitives which have been adjusted for inflation there is no daily data series of the real short rate. Daily observations on price levels are simply not available.

effects. Since the price level process is uncorrelated with the state variable r then the relationship between the real values of nominal and real claims is simply

$$N(t,r,p) = C(t,r)E_t[1/p(T)] = E[m(t,T)g(r)|r_t]E_t[1/p(T)]$$

where $m(t,T)$ is the CIR SDF from the previous section. The nominal price of a nominal claim is

$$\hat{N}(t,r,p) = p(t)N(t,r,p) = C(t,r)E_t[p(t)/p(T)].$$

Let $\hat{R}_{t+1}$ be the gross nominal return on this claim, then

$$E_t[m(t,T)\hat{R}_T] = E_t[m(t,T)R_T]E_t(p(t)/p(T)) = E_t(p(t)/p(T)) \tag{3.1}$$

Hence this 2-factor model is really just the same 1-factor model where the returns on the primitive assets are adjusted for expected inflation. So testing the model would be the same as the tests of the one-factor version of a previous section except that we would have to know the expectation of one over the gross inflation rate.

Unfortunately this quantity is not observable. However consider a richer model in which the price level follows

$$dp = ypdt + \sigma_p p\sqrt{y}dW_t^{(2)}$$

where y is another state variable. CIR consider two different versions of this model. The one they call model 2 posits the following behavior for y

$$dy = \kappa_2(\theta_2 - y)dt + \sigma_2\sqrt{y}dW^{(3)}$$

where the correlation between $W^{(2)}$ and $W^{(3)}$ is ρ. In this case

$$E_t[p(t)/p(T)] = C(\tau)\exp\{-D(\tau)y_t\} \tag{3.2}$$

where $\tau = T - t$ and

$$C(\tau) \equiv \left[\frac{2\xi e^{(\xi+\kappa_2+\rho\sigma_2\sigma_p)(\tau)/2}}{(\xi+\kappa_2+\rho\sigma_2\sigma_p)(e^{\xi(\tau)}-1)+2\xi}\right]^{2\kappa_2\theta_2/\sigma_2^2}$$

$$D(\tau) \equiv \frac{2(e^{\xi(\tau)}-1)(1-\sigma_p^2)}{(\xi+\kappa_2+\rho\sigma_2\sigma_p)(e^{\xi(\tau)}-1)+2\xi}$$

$$\xi \equiv \sqrt{(\kappa_2+\rho\sigma_2\sigma_p)^2+2\sigma_2^2(1-\sigma_p^2)}$$

Recall that in the CIR 1-factor model the price of a pure discount bond which matures at time T is

$$C(t, r) = A(\tau)\exp\{-B(\tau)r_t\}$$

where

$$A(\tau) = \left[\frac{2\gamma e^{(\gamma+\lambda+\kappa)(\tau)/2}}{(\gamma+\lambda+\kappa)(e^{\gamma(\tau)}-1)+2\gamma}\right]^{2\kappa\theta/\sigma^2}$$

$$B(\tau) = \frac{2(e^{\gamma(\tau)}-1)}{(\gamma+\lambda+\kappa)(e^{\gamma(\tau)}-1)+2\gamma}$$

$$\gamma \equiv \sqrt{(\kappa+\lambda)^2+2\sigma^2}$$

So we have a closed form solution for the price of a nominal, pure discount bond which is given by

$$\hat{N}(t, r, y) = A(\tau)C(\tau)\exp\{-B(\tau)r_t - D(\tau)y_t\}.$$

This model was tested by Pearson and Sun [30] who used this closed form solution to solve for the "implied" levels of the state variables. In doing so they used two bonds since for each set of parameter values the above formula gives a one to one correspondence between a pair of bond prices at a given time and the level of the two state variables at that time. Although this is quite elegant it has the drawback that it can only use the information in those two bonds.

By using the framework developed in previous sections we can do better. We can use two bonds to get the state variables and then use the implied factors to construct the stochastic discount factor for the model and then apply it to other term structure assets.

For this model I have chosen the 182 day and 91 day bills to use for extracting implied state variables. Both of these series are available at daily frequency from the Federal Reserve. I also collected data on auction dates and issue dates in order to calculate the exact days to maturity for each bill [3].

The form of the stochastic discount factor in this case is identical to that used in the previous section when testing the one-factor model. This is because, as mentioned above, there is only one market price of risk in this model and it is associated with the short rate. The only difference between the two is that in this model we construct $\hat{m}$ from data on implied r_t. The moment conditions are slightly different because we are using an SDF which is appropriate for real assets and applying it to gross nominal returns on nominal assets. The modified moment conditions are (from (3.1) and (3.2)).

$$E\begin{bmatrix} \hat{m}_T\mathbf{R}_T - \exp\{D(\tau)y_t\}/C(\tau)\mathbf{1} \\ (\hat{m}_T\mathbf{R}_T - \exp\{D(\tau)y_t\}/C(\tau)\mathbf{1})r_t \\ (\hat{m}_T\mathbf{R}_T - \exp\{D(\tau)y_t\}/C(\tau)\mathbf{1})y_t \end{bmatrix} = \mathbf{0}$$

Parameter estimates are presented in table 3.3. The parameters κ,θ,σ, and λ have exactly the same interpretation as they did in the 1-factor case. The parameters

[3] Both bills are issued once a week. The Federal Reserve begins using a new bill in constructing the series the day after the auction. Although this is before the actual issue the bill still trades in the secondary market on a "when issued" basis.

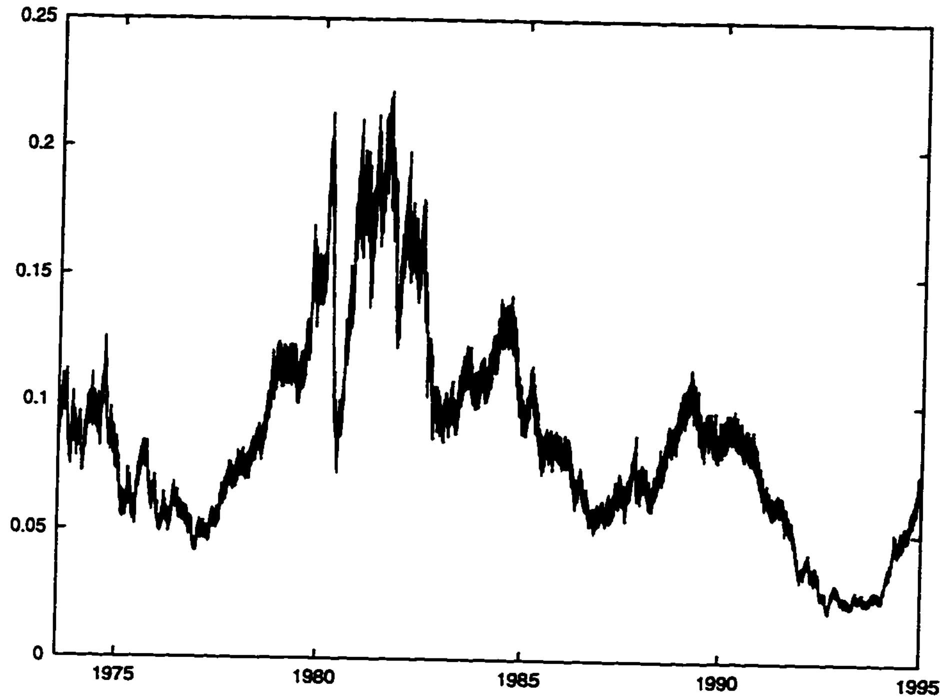

Figure 3.2: Implied Real Short rate

κ_2,θ_2, and σ_2 have analogous interpretations in terms of the dynamics of the second factor. With this in mind the estimates obtained do not seem unreasonable except in the case of the long run mean of the short rate which is estimated to be very large and negative. As a possible explanation for this we note that the factors here have been extracted from bills. In examining this same model Pearson and Sun [30] also found that by using bills they obtained a strange value for this parameter (in their case it was estimated to be 10.0301) while when using bonds to extract the factors more reasonable estimates were obtained. Theoretically of course there should be no difference if the model were true. The fact that parameter estimates depend on which assets are used to extract factors is very troubling and casts serious doubt on whether the model is correctly specified.

Another curious fact is that σ_p is negative. We can of course deal with this the same way as we did in the Brennan and Schwartz model by defining another

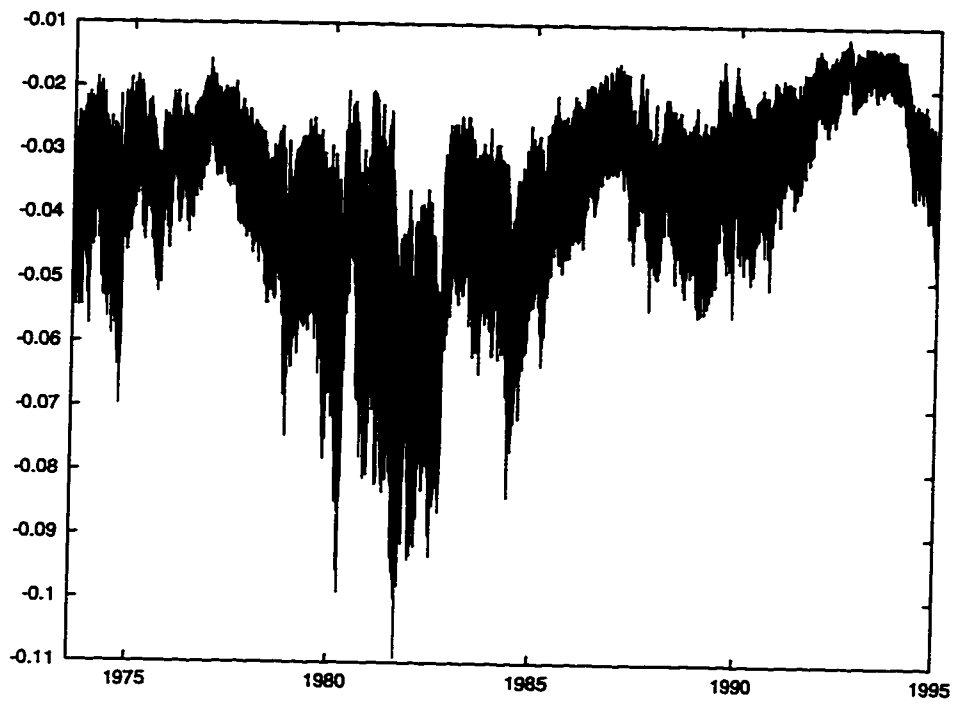

Figure 3.3: Implied Inflation Factor

Brownian motion equal to $-W^{(2)}$. This would also effect the parameter ρ which is the correlation between $W^{(2)}$ and $W^{(3)}$. Such a redifinition would mean that σ_p would be positive and ρ would be negative. Interestingly this is what Pearson and Sun found.

Once again it was impossible to compute standard errors for the parameters. On the basis of the J-statistic we will still reject this model. An examination of the moment conditions in table 3.4 indicates that the same increase in average pricing error which was apparent in the one factor model is still evident here. These APE's are plotted in figure 3.4. Also note that one again the APE's are much larger than the other moment conditions indicated that, once again, the procedure is trying too hard to fit the conditional moments at the expense of APE.

Among the possible reasons for rejection is an interesting pattern which emerged in the data. A close examination of figure 3.3 shows that the process exhibits very

Table 3.3: Parameter Estimates from 2-factor CIR Model

	Estimate
κ	0.375358
θ	-0.250115
σ	0.114001
λ	-0.001592
κ_2	3.465052
θ_2	0.060807
σ_2	0.648752
σ_p	-0.137574
ρ	0.183853
J-stat	63.3390

fast oscillation (this accounts for what appears to be a "thick" line). The periodicity of these oscillations is exactly 5 trading days. This indicates that y fluctuates with the time to maturity of the bills used to extract the factors. This suggests that the mispecification of the model relates to the functional dependence of prices on time to maturity.

3.2.2 Interest Rate Volatility as a Factor

Longstaff and Schwartz [27] introduced another two-factor model in a general equilibrium framework. The dynamics of the two state variables are governed by

$$dX_t = (\gamma - \delta X_t)dt + c\sqrt{X_t}dW_t^{(1)}$$

$$dY_t = (\eta - \xi Y_t)dt + f\sqrt{Y_t}dW_t^{(2)}$$

Table 3.4: Pricing Moment Conditions for CIR 2-Factor Model

	APE	Orth – short rate	Orth – y
rf	-0.000119	-0.000111	-0.000101
	(0.000119)	(0.000006)	(0.000016)
port6	0.000654	0.000646	0.000638
	(0.000106)	(0.000006)	(0.000016)
port12	0.000895	0.000883	0.000870
	(0.000277)	(0.000016)	(0.000040)
port48	0.001573	0.001550	0.001520
	(0.000959)	(0.000049)	(0.000122)
port120	0.001652	0.001629	0.001599
	(0.001326)	(0.000066)	(0.000164)
ltgov	0.001805	0.001782	0.001748
	(0.001977)	(0.000097)	(0.000237)

With no loss of generality we can rescale such that $f = c = 1$. The PDE for contingent claims is

$$-C_t + rC = \frac{x}{2}C_{xx} + \frac{y}{2}C_{yy} + (\gamma - \delta x)C_x + (\eta - \xi y - \lambda y)C_y$$

so that the market price of "X" risk is zero and the market price of "Y" risk is $\lambda\sqrt{Y_t}$. This means that the SDF can be written as

$$m(t,T) = \exp\{-\int_t^T r_s ds - \lambda \int_t^T \sqrt{Y_s}dW_s^{(2)} - \frac{\lambda^2}{2}\int_t^T Y_s ds\} \tag{3.3}$$

The short-term rate of interest in this model is

$$r_t = \alpha X_t + \beta Y_t$$

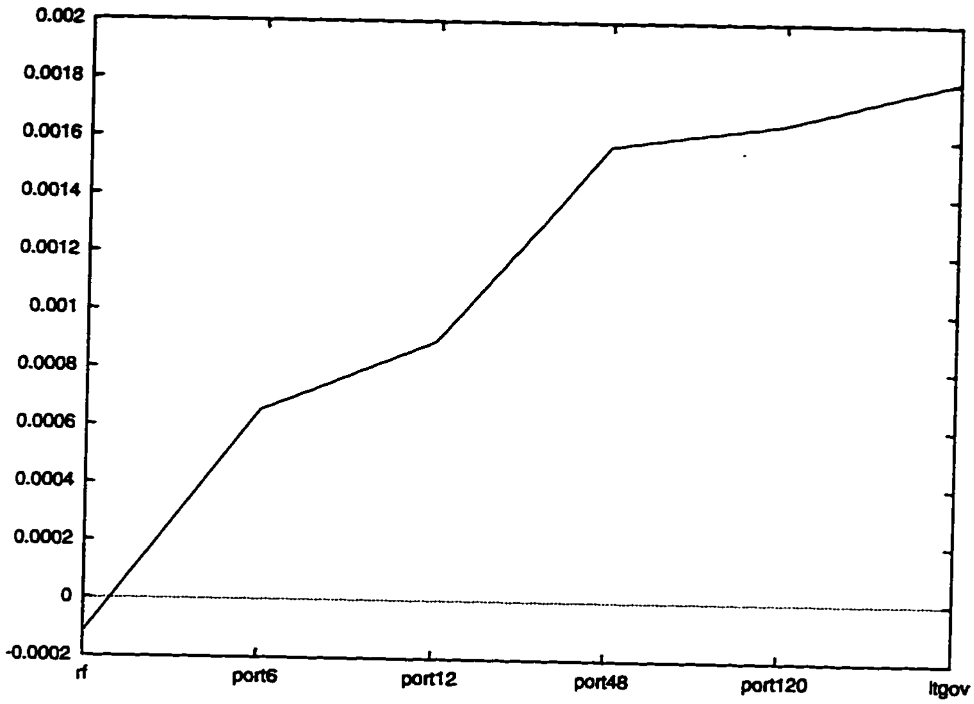

Figure 3.4: Average Pricing Errors for CIR 2-factor model

for $\alpha, \beta > 0$ while the variance of the short rate, V, is given by

$$V_t = \alpha^2 X_t + \beta^2 Y_t$$

Provided $\alpha \neq \beta$ this system is invertible and the short rate and the variance of the short rate can be used as proxies for the state variables, X and Y. In order to estimate the SDF we need only to solve for Y as follows

$$Y_t = \frac{V_t - \alpha r_t}{\beta(\beta - \alpha)}.$$

Applying Ito's lemma shows that the processes for r_t and V_t can be written as

$$\begin{aligned} dr &= \left(\alpha\gamma + \beta\eta - \frac{\beta\delta - \alpha\xi}{\beta - \alpha} r - \frac{\xi - \delta}{\beta - \alpha} V\right) dt \\ &\quad + \alpha\sqrt{\frac{\beta r - V}{\alpha(\beta - \alpha)}}\, dW^{(1)} + \beta\sqrt{\frac{V - \alpha r}{\beta(\beta - \alpha)}}\, dW^{(2)} \end{aligned} \tag{3.4}$$

$$
\begin{aligned}
dV &= \left(\alpha^2\gamma + \beta^2\eta - \frac{\alpha\beta(\delta-\xi)}{\beta-\alpha}\,r - \frac{\beta\xi-\alpha\delta}{\beta-\alpha}\,V\right)dt \\
&\quad + \alpha^2\sqrt{\frac{\beta r - V}{\alpha(\beta-\alpha)}}\,dW^{(1)} + \beta^2\sqrt{\frac{V-\alpha r}{\beta(\beta-\alpha)}}\,dW^{(2)} \qquad (3.5)
\end{aligned}
$$

So while the short rate by itself is not a Markov process the pair (r, V) form a vector Markov process.

While it can be argued that good short rate proxies are observable we have no way of observing V_t, the instantaneous volatility of r_t. In testing their model Longstaff and Schwartz used the one-month U.S. t-bill yield as the short rate and for V_t they use the conditional volatility at time t from a GARCH model. Of course the conditional volatility is not the same as the actual stochastic volatility although Nelson [28] has shown that the conditional volatility converges to the true volatility process as the time between observations goes to zero. Unfortunately the innovations of the conditional volatility do not converge to a Brownian motion (see footnote 15 of [28]). So this approach would not be suitable for forming a SDF since we must construct the SDF from the innovations of the factors.

Fortunately we do not have to address this issue in the same way. The relatively simple form of the pricing equation allows for a closed form solution for bill prices. Hence by utilizing two bills with known time to maturity we can obtain the level of r_t and V_t implied by the bill prices. This closed form solution is as follows

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

where τ is the time to maturity and

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

$$
B(\tau) = \frac{2\psi}{(\nu+\psi)(\exp(\psi\tau)-1)+2\psi}
$$

$$C(\tau) = \frac{\alpha\phi(\exp(\psi\tau) - 1)B(\tau)) - (\beta\psi(\exp(\phi\tau) - 1)A(\tau)}{\phi\psi(\beta - \alpha)}$$

$$D(\tau) = \frac{\psi(\exp(\phi\tau) - 1)A(\tau)) - (\phi(\exp(\psi\tau) - 1)B(\tau)}{\phi\psi(\beta - \alpha)}$$

and

$$\nu = \xi + \lambda$$

$$\phi = \sqrt{2\alpha + \delta^2}$$

$$\psi = \sqrt{2\beta + \nu^2}$$

$$\kappa = \gamma(\delta + \phi) + \eta(\nu + \psi)$$

Using this formula we back out the processes r_t and V_t from bill data as we did in the previous section. Once we have r_t and V_t we obtain Y_t by the relation above and using equation (3.3) form the SDF. The moment conditions we use for GMM estimation are

$$E\begin{bmatrix} \hat{m}(t,T)\mathbf{R}_T - \mathbf{1} \\ (\hat{m}(t,T)\mathbf{R}_T - \mathbf{1})r_t \\ (\hat{m}(t,T)\mathbf{R}_T - \mathbf{1})V_t \end{bmatrix} = \mathbf{0}$$

Estimating this model proved to be problematic because of the high degree of correlation between the moment conditions. In fact the weighting matrix was not invertible. For this reason only first stage estimates are presented in table 3.5. Longstaff and Schwartz were unable to estimate all the parameters of this model but interestingly the estimates they did obtain (α, β, and δ) are surprising close to

Table 3.5: First stage GMM estimates

	Estimate
α	-0.068571
β	0.194466
γ	0.010439
δ	0.370436
η	-0.006733
λ	-0.002558
ξ	35.778864

those presented here. As is clear from equations (3.4) and (3.5) the interpretations of the parameters in terms of the behavior of the short rate and it's volatility is not straightforward. More can be said be examing the implied processes for these variables themselves.

The implied r_t and V_t processes are presented in figures 3.5 and 3.6. The general pattern of the implied short rate is similar to the short rate proxy used in the previous chapter and similar to that of the CIR 2-factor model. However it seems to be translated in this model since all of the values are negative and very small in absolute value.

The most troubling aspect of this model is that the estimated volatility process doesn't seem to match the historical record. Figure 3.6 shows that just at the time when the short rate was most volatile the stochastic volatility factor reached it's lowest point. Also the stochastic volatility process seems to have the same 5 day oscillation pattern as the inflation factor in the previous sections although not as severe.

Despite these weakness the model seems to do rather well in terms of average pricing error as shown in figure 3.7 and table 3.6.

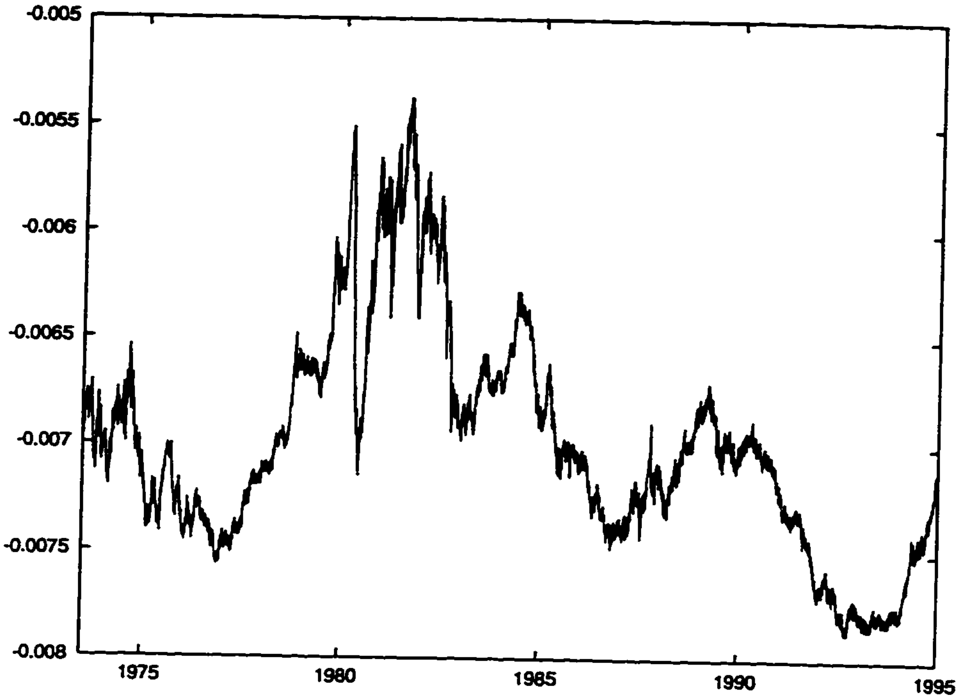

Figure 3.5: Implied Short Rate – Longstaff and Schwarz model

3.3 Conclusion

This paper has demonstrated a new technique for testing continuous time models. It attempts to bridge the gap between the empirical contingent claims literature and the literature of conditional asset pricing by formulating tests in terms of the stochastic discount factor implied by the model. Because the method does not rely on the pricing equation having a closed form solution it is very general. It also easily accomodates the complexity of models with multiple state variables. This approach also allows the use of rate of return data on arbitrary portfolios of assets rather than being restriced to simple claims. Here we have examined term structure models but the method is not restricted to this area. Any continuous time model where fequent (daily) observations on the state variable are available can be tested in this framework.

From the empirical work in this paper we learn the following about the models

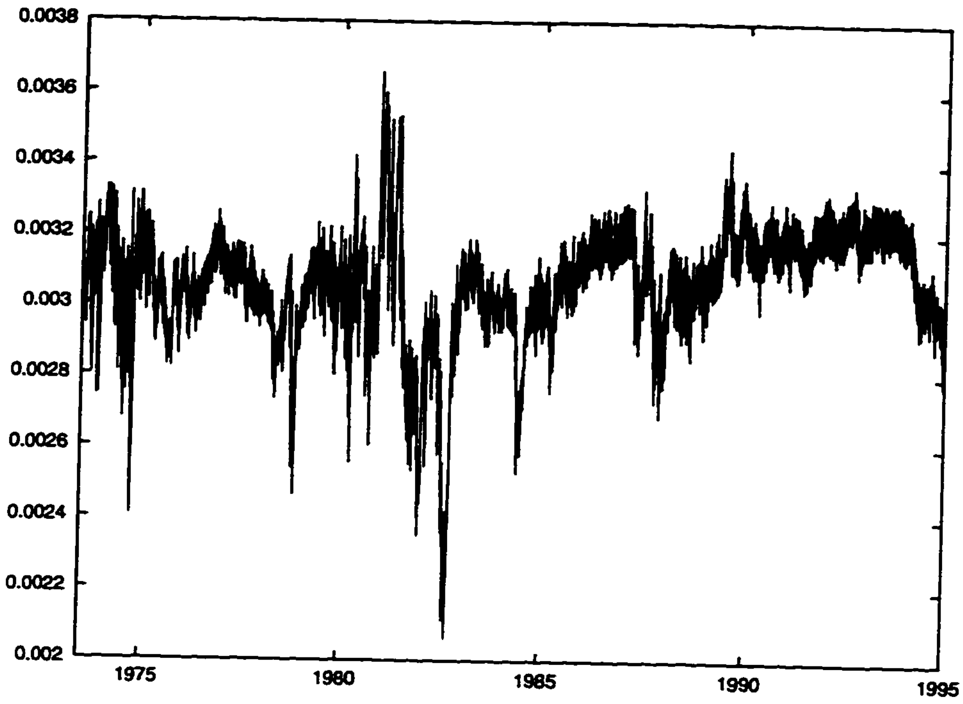

Figure 3.6: Implied Stochastic Volatility of Short Rate – Longstaff and Schwartz model

that have been proposed

It's hard to beat a LEH model One remarkable fact is that we can't do better in explaining cross sectional (across maturities) differences in expected return than a model in which we discount at the risk-free rate. This is especially troubling since there are clear rewards to holding longer term assets but none of our models seem to correctly account for this risk.

Subperiods matter It seems clear that the behavior of one-factor models is sensitive to time period which would indicate that the single factor is non-markovian. This does not come as much of a surpise to financial economists but it is interesting that it shows up in pricing as well as in the behavior of the short rate itself.

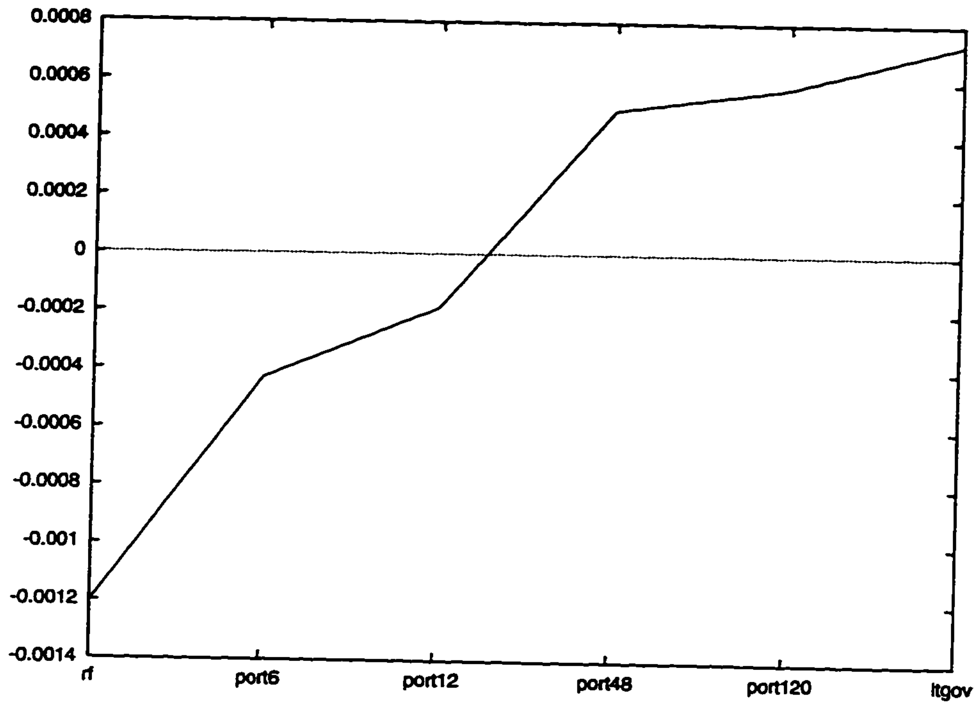

Figure 3.7: Average Pricing Errors from Longstaff and Schwartz model

Non-priced additional factors don't help Adding additional factors which have no price of risk seems to do nothing to account for differences in expected return. The only model which seems to account for these differences in any degree is the Brennan and Schwarz model which has two priced factors.

In addition this empirical excercize has taught us much about the SDF approach to contingent claims pricing

Daily is nearly continuous When using discretely sampled data to make inferences about continuous processes there is always a concern about the size of the discretization bias. In this paper I have demonstrated that for daily sampling the stochastic behavior of a SDF based on a simple approximation is very close to that which would obtain if the process could be observed continuously.

Table 3.6: Sample Moment Conditions

	APE	Orth – r	Orth – V
rf	-0.001204	-0.000004	0.000010
	(0.000164)	(0.000001)	(0.000001)
port6	-0.000428	-0.000002	0.000004
	(0.000203)	(0.000001)	(0.000001)
port12	-0.000185	-0.000001	0.000003
	(0.000348)	(0.000001)	(0.000002)
port48	0.000499	0.000001	-0.000002
	(0.001006)	(0.000003)	(0.000007)
port120	0.000580	0.000001	-0.000003
	(0.001367)	(0.000004)	(0.000009)
ltgov	0.000735	0.000001	-0.000004
	(0.002009)	(0.000006)	(0.000013)

Conditional moments are highly correlated with Unconditional A typical requirement in the conditional asset pricing literature is that $(m(t,T)R_T - 1)z_t$ have mean zero or in other words pricing errors should be uncorrelated with instruments. In this case this requirement is problematic because the sample mean of this random variable is highly correlated with the average pricing error. An approach that neglects instruments might be more desirable although perhaps less powerful.

Based on these observations I would make the following recommendations for future research.

1. Apply this approach to models with priced second and third factors to see if we can account for the risk premiums we observe.

2. In testing term structure models use constant maturity strips as primitives. While this is not necessary it allows one to solve for the elements of the GMM weighting matrix as a function of the parameters only, thus eleminating the problem of estimating the weighting matrix.

3. Apply this method to other contingent claims. Nothing we have done in this paper has required that the model be a term structure model or that the primitives be bonds. This approach could equally well be applied to options, futures, or any other contingent claim.

BIBLIOGRAPHY

[1] Yacine Aït-Sahalia. Nonparametric pricing of interest rate derivative securities. *Econometrica*, 1996.

[2] Yacine Aït-Sahalia. Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 1996.

[3] Torben G. Andersen and Jesper Lund. Estimating continuous time stochastic volatility models of the short term interest rate. Kellogg Graduate School of Management.

[4] Clifford A. Ball and Walter B. Torous. Unit roots and the esimation of interest rate dynamics. *Journal of Empirical Finance*, 1996.

[5] Christopher R. Blake, Edwin J. Elton, and Martin J. Gruber. The performance of bond mutual funds. *Journal of Business*, 1993.

[6] Christopher R. Blake, Edwin J. Elton, and Martin J. Gruber. Fundamental economic variables, expected returns, and bond fund performance. *Journal of Finance*, 1995.

[7] Peter Bossaerts and Pierre Hillion. A test of a general equilibrium stock option pricing model. *Mathematical Finance*, 1993.

[8] Michael J. Brennan and Eduardo S. Schwartz. A continuous time approach to the pricing of bonds. *Journal of Banking and Finance*, 1979.

[9] Robin J. Brenner, Richard H. Harjes, and Kenneth F. Kroner. Another look at models of the short-term interest rate. *Journal of Financial and Quantitative Analysis*, 1996.

[10] Stephen J. Brown and Philip H. Dybvig. The empirical implications of the cox, ingersoll, and ross theory of the term structure of interest rates. *Journal of Finance*, 1986.

[11] K. C. Chan, G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, 1992.

[12] John C. Cox, Jonathan E. Ingersoll, Jr., and Stephen A. Ross. A re-examination of traditional hypotheses about the term structure of interest rates. *Journal of Finance*, 1981.

[13] John C. Cox, Jonathan E. Ingersoll, Jr., and Stephen A. Ross. An intertemporal general equilibrium model of asset prices. *Econometrica*, 1985.

[14] John C. Cox, Jonathan E. Ingersoll, Jr., and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 1985.

[15] Qiang Dai and Kenneth J. Singleton. Specification analysis of affine term structure models. Stanford University Working Paper.

[16] Darrell Duffie. *Security markets : stochastic models*. Academic Press, 1988.

[17] A. Freidman. *Partial Differential Equations of Parabolic Type*, chapter 3. Prentice-Hall, 1964.

[18] Michael R. Gibbons and Krishna Ramaswamy. A test of the cox, ingersoll, and ross model of the term structure. *Review of Financial Studies*, 1993.

[19] Robert S. Goldstein and F. Zapatero. General equilibrium with constant relative risk aversion and vasicek interest rates. *Mathematical Finance*, 1996.

[20] Lars Peter Hansen and Kenneth J. Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, 1982.

[21] Wolfgang Hardle. *Applied nonparametric regression.* Cambridge University Press, 1990.

[22] J. Michael Harrison and David M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 1979.

[23] John C. Hull. *Options, Futures, and Other Derivative Securities.* Prentice-Hall, 1993.

[24] Narasimhan Jegadeesh and George G. Pennacchi. The behavior of interest rates implied by the term structure of eurodollar futures. *Journal of Money, Credit, and Banking*, 1996.

[25] Jangkoo Kang. Bond mutual fund performance evaluation: The numeraire portfolio approach. William E. Simon Graduate School of Business Administration.

[26] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, 1992.

[27] Francis A. Longstaff and Eduardo S. Schwartz. Interest rate volatility and the term stucture: A two-factor general equilibrium model. *Journal of Finance*, 1992.

[28] Daniel B. Nelson. Arch models as diffusion approximations. *Journal of Econometrics*, 1990.

[29] Jeffrey Nelson and Stephen Schaefer. The dynamics of the term structure and alternative portfolio immunization stategies. In *Innovations in Bond Portfolio Management:Duration Analysis and Immunization*. JAI Press, Inc, 1983.

[30] Neil D. Pearson and Tong sheng Sun. Exploiting the conditional density in estimating the term structure: An application to the cox, ingersoll, and ross model. *Journal of Finance*, 1994.

[31] Richard Stanton. A nonparametric model of term structure dynamics and the market price of interest rate risk. Haas School of Business.

[32] Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 1977.

Appendix A

PROOFS AND FURTHER RESULTS

A.1 Valuation Equation

Consider two claims $C^{(1)}$ and $C^{(2)}$ which are not identical. For simplicity in exposition assume neither claim makes any payment until maturity. The evolution of the prices of these claims is given by

$$dC^{(i)} = C^{(i)}\mu^{(i)}dt + C^{(i)}\sigma^{(i)}dW_t \; i = 1, 2$$

where the functions $\mu^{(i)}$ and $\sigma^{(i)}$ can be obtained via Ito's lemma. Since they both have the same source of risk (namely W_t) an instantaneously riskless portfolio can be formed by holding $C^{(2)}\sigma^{(2)}$ of $C^{(1)}$ and $-C^{(1)}\sigma^{(1)}$ of $C^{(2)}$. To prevent arbitrage this portfolio (call it Π) must earn the risk free rate, i.e.

$$\begin{aligned} d\Pi &= (\mu^{(1)}C^{(2)}\sigma^{(2)}C^{(1)} - \mu^{(2)}C^{(1)}\sigma^{(1)}C^{(2)})dt \\ &= r_t\Pi dt \end{aligned}$$

Rearranging shows that

$$\frac{\mu^{(1)} - r}{\sigma^{(1)}} = \frac{\mu^{(2)} - r}{\sigma^{(2)}} = q(t, X_t) \tag{A.1}$$

where q is a function that does not depend on the nature of the claims. The above equation states that the expected excess return on any asset is proportional to its volatility and the constant of proportionality is the same for all claims. Throughout this paper we will assume that q does not depend on time so we will write $q(X_t)$ in place of $q(t, X_t)$.

By Ito's lemma we have

$$\mu^{(i)}C = \frac{1}{2}\sigma^2(x)\frac{\partial^2 C}{\partial x^2} + \mu(x)\frac{\partial C}{\partial x} + \frac{\partial C}{\partial t}$$

and

$$\sigma^{(i)}C = \sigma(x)\frac{\partial C}{\partial x}$$

Substituting this into (A.1) and rearranging gives a PDE which the claim price must satisfy.

If there are d sources of risk then consider $d+1$ claims. The equation analogous to (A.1) is

$$\mu^{(i)} - r = \sum_{j=1}^{d+1} q^{(j)}\sigma_{ij}$$

A.2 Girsanov's Theorem

The basic idea of the Girsanov theorem is to begin with a Brownian motion under the original probability measure and construct a new measure under which a "translated" process is a Brownian motion.

Suppose that we are given a probability space $(\Omega, \mathcal{F}, P)$ and a process $\{W_t, \mathcal{F}_t; 0 \leq t \leq T\}$ such that W_t is a Brownian motion with respect to the filtration $\mathcal{F}_t$, under the measure P. Let another process $\{Y_t, \mathcal{F}_t; 0 \leq t \leq T\}$ be given which satisfies

$$P(\int_0^T Y_t^2 dt \ < \infty) = 1$$

so that the stochastic integral $\int_0^t Y_s dW_s$ is defined. If it is also true that

$$E[\exp(\int_0^T Y_t^2 dt)] \ < \infty$$

Then the process defined by

$$\psi_t(Y) \equiv \exp(\int_0^t Y_s dW_s - \frac{1}{2}\int_0^t Y_s^2 ds)$$

is a martingale with the property that $E\psi_t(Y) = 1$. In this case we can define a new measure $\tilde{P}_T$ on $\mathcal{F}_T$ by

$$\tilde{P}_T \equiv E[1_A \psi_T(Y)]; \;\; A \in \mathcal{F}_T$$

Girsanov's theorem says that under this new measure the process

$$\tilde{W}_t \equiv W_t - \int_0^t Y_s ds$$

is a Brownian motion under $\tilde{P}_T$. So suppose we are given a diffusion process

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$$

Then under the new measure we substitute for W_t from above

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)d\tilde{W}_t + \sigma(t, X_t)Y_t dt$$

$$= (\mu(t, X_t) + \sigma(t, X_t)Y_t)dt + \sigma(t, X_t)dW_t$$

If we wish to change the drift of X_t from $\mu(t, X_t)$ to $b(t, X_t)$ then we should select

$$Y_t = \frac{b(t, X_t) - \mu(t, X_t)}{\sigma(t, X_t)}$$

A.3 SDF's in General Equilibrium Models

If there is a single state variable in the economy which follows an Ito process then the optimal consumption for the representative agent will be a function of the state variable and so will be an Ito process.

$$dc^* = \mu_c(X_t)dt + \sigma_c(X_t)dW_t$$

Marginal utility of the representative agent will also be an Ito process which must solve

$$du_c(c_t^*, t) = \mu_p dt + u_{cc}(c_t^*, t)\sigma_c dW_t$$

where μ_p could be found explicitly using Ito's lemma. The stochastic discount factor in such an economy is given by

$$\frac{u_c(c_T^*, T)}{u_c(c_t^*, t)}$$

which we can rewrite as

$$\exp\{\log(u_c(c_T^*, T)) - \log(u_c(c_t^*, t))\}$$

but by Ito's lemma the expression inside the exponential can be written as

$$\int_t^T \frac{\mu_p}{u_c} ds + \int_t^T \frac{u_{cc}\sigma_c}{u_c} dW_s - \frac{1}{2}\int_t^T (\frac{u_{cc}\sigma_c}{u_c})^2 ds$$

Standard results from general equilibrium theory [1] state that the instantaneous risk-free rate of interest $r_t = -\frac{\mu_p}{u_c}$. Substituting this in gives us (2.9).

A.4 Generating data for Monte Carlo studies

Recall that the short rate process in the Vasicek model satisfies the SDE

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t.$$

The linear structure of this model allows us to solve the SDE to obtain

$$r_t = e^{-\kappa t} r_0 + \theta(1 - e^{-\kappa t}) + e^{-\kappa t}\sigma \int_0^t e^{\kappa s} dW_s.$$

Now consider the process $Y_t = \int_0^t e^{\kappa s} dW_s$. Note that this is a diffusion on natural scale and hence is a time change of brownian motion. Further note that the speed density

[1] See Duffie [16]

is non-stochastic and is given by $e^{-2\kappa t}$. So Y_t is normally distributed with mean zero and variance is $(e^{2\kappa t}-1)/2\kappa$. This suggests the following recursive algorithm for simulating a sequence of observations (sampled at interval Δ) from this process

$$r_{t_{i+1}} = e^{-\kappa\Delta} r_{t_i} + \theta(1-e^{-\kappa\Delta}) + e^{-\kappa\Delta}\sigma Z_{i+1}\sqrt{(e^{2\kappa\Delta}-1)/2\kappa}$$

where the Z_i are independent standard normal random variables. It should be emphasized that this algorith involves *no approximation.* So this sequence will have *exactly* the same finite dimensional distributions as a sample from the continuous process.

For the CIR model this is more difficult. I opt instead to use a Milstein scheme at a mesh of points between each observation I wish to sample. First take $\delta = \Delta/M$ for some positive integer M. Then given r_t define

$$r_{t+\delta} = r_t + \kappa(\theta - r_t)\delta + \sqrt{r_t}\sigma Z\sqrt{\delta} + \frac{1}{4}\sigma^2((Z\sqrt{\delta})^2 - \delta)$$

where Z is a standard normal random variable. Applying this algorithm recursively M times gives a value for $r_{t+\Delta}$ which will have approximately the right joint distribution with r_t. In fact the Milstein scheme converges with strong order 1 to the true process. Obviously larger M (smaller δ) give better approximations. I tried both $M=5$ and $M=15$ and found essentially no difference.

A.5 A Nonparametric SDF

In order to obtain the best estimates of the drift and diffusion of the short rate we can employ nonparametric methods. Aït-Sahalia [1] has shown how nonparametric estimates of the diffusion function can be obtained if a linear drift is assumed. However assuming a functional form for the drift is also undesirable so instead the methods of Stanton [31] are employed to obtain nonparametric estimates of both the drift and diffusion of the short rate.

Since we cannot calculate the exact conditional expected change or variance as before we are open to discretization bias. We note that a first order approximation to the drift and diffusion functions converges to the true drift and diffusion at rate Δ as $\Delta \to 0$

$$\mu(r_t) = \frac{1}{\Delta} E_t(r_{t+\Delta} - r_t) + O(\Delta)$$

$$\sigma(r_t) = \sqrt{\frac{1}{\Delta} E_t[(r_{t+\Delta} - r_t)^2] + O(\Delta)}$$

This discetization scheme was used by CKLS in estimating various parametric models. In order to minimize this bias Stanton (1996) suggests the second order approximations [2]

$$\mu(r_t) = \frac{1}{2\Delta}[4E_t(r_{t+\Delta} - r_t) - E_t(r_{t+2\Delta} - r_t)] + O(\Delta^2)$$

$$\sigma(r_t) = \sqrt{\frac{1}{2\Delta}[4E_t[(r_{t+\Delta} - r_t)^2] - E_t[(r_{t+2\Delta} - r_t)^2] + O(\Delta^2)}$$

which converges at rate Δ^2. To implement this in a nonparametric framework the conditional expectations can be approximated using nonparametric regression techniques such as kernel estimation [3].

The estimated[4] drift is shown in Figure A.5 along with the drift functions of the parametric models for comparison. The drift function is very similar to that found by Stanton (1996). Figure A.5 shows the nonparametric diffusion function as well as

[2] Using a time step of 3Δ in addition will lead to a third order approximation and so forth. See [31] for details

[3] For an excellent treatment of nonparametric regression techniques in general and kernel methods in particular see Hardle [21].

[4] Following Stanton a normal kernal was used with the window width $\hat{s}N^{-1/5}$ where $\hat{s}$ is the standard deviation of the observations and N is the number of observations.

Table A.1: Results for Nonparametric Model

q	0.011348
std. err.	0.037683
J-stat	100.7958
p-value	0

the diffusion functions of the parametric models. The diffusion function is noticable different from that estimated by Stanton. Stanton used the yield on the 3-month treasury bond as the short rate. Here we use the eurodollar rate following Aït-Sahalia [2] and find that both the drift and diffusion are similar to those he estimated using a flexible functional form.

In order to form an SDF we need to estimate the market price of risk. Stanton estimates this nonparametrically but his method allows for using only two bond return series. Aït-Sahalia [1] estimates a value for a constant market price of risk which corrosponds to his nonparametric model but his methodology allows for using only a single yield curve. If we assume a constant market price of risk then the $\hat{m}$ formula is identical to that for the Vasicek model. We can use a two step approach by taking the ϵ_t from the nonparametric model and then applying GMM to the pricing moment conditions. The results are presented in table A.1.

One must be careful in interpretting the J-statistic and the standard error reported in table A.1 because we have used ϵ_t as though it were data and haven't taken into account our uncertainty regarding the functions μ and σ.

Table A.2 presents the sample mean of the moment conditions for this model. Notice that this model seems to perform worse than the other one-factor models we have examined. This is due at least in part to the two step approach taken in estimating this model. The reason is that in the parametric models the parameters

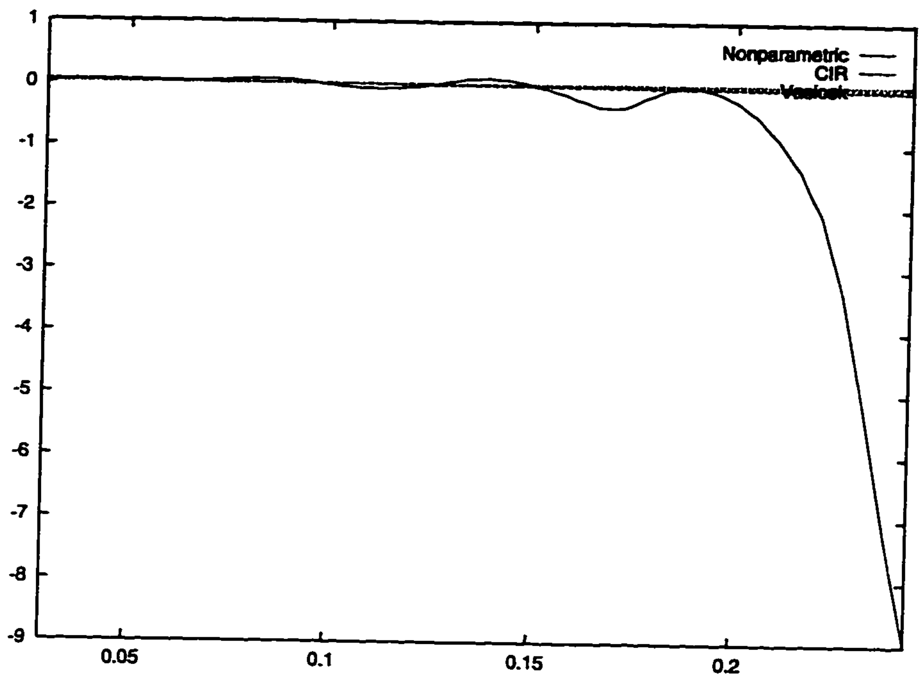

Figure A.1: Estimated Parametric and Nonparametric Drift Functions

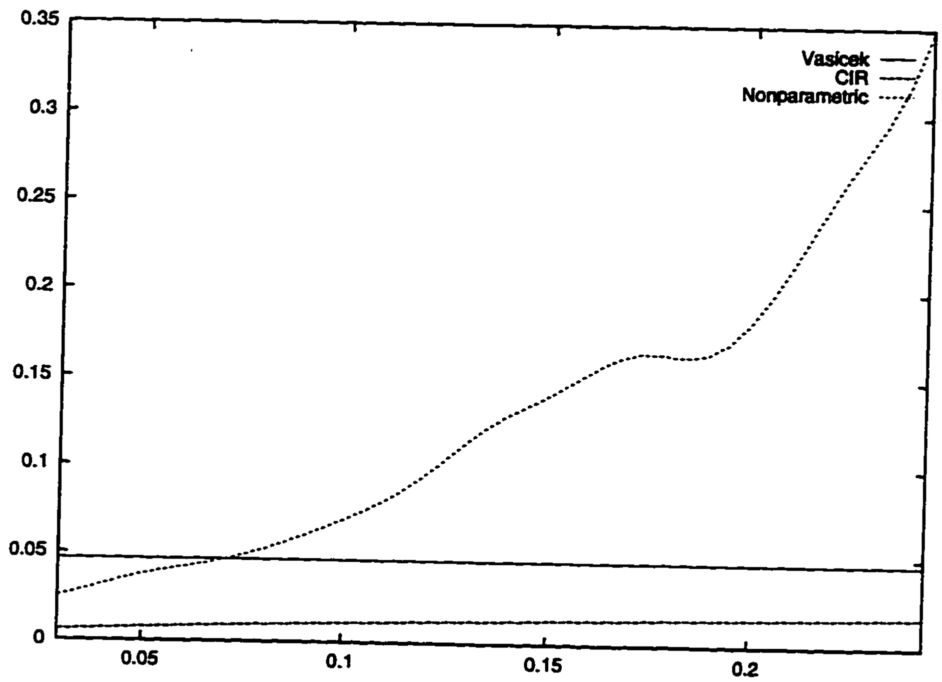

Figure A.2: Estimated Parametric and Nonparametric Diffusion Functions

Table A.2: Examination of Pricing Errors for Nonparametric Model

	APE	Orth
rf	-0.001204	-0.000154
	(0.000157)	(0.000023)
port6	-0.000429	-0.000076
	(0.000191)	(0.000025)
port12	-0.000185	-0.000055
	(0.000350)	(0.000047)
port48	0.000501	0.000002
	(0.001011)	(0.000125)
port120	0.000584	-0.000003
	(0.001375)	(0.000166)
ltgov	0.000741	0.000001
	(0.002017)	(0.000233)

of the drift and diffusion function were chosen to fit these moment conditions. In the nonparametric case the drift and diffusion functions were esimtated independently so only the price of risk could be adjusted to fit the moment conditions.

A.6 A Two-Factor Nonparametric Model

When Brennan and Scwartz tried to implement their two-factor model (which we have examined in a previous section) they had to use a two step approach. First they estimated the parameters of the long and short rates. Then they solved the pricing equations numerically for several possible values of the market price of risk. They choose as their estimate the market price of risk which made the predicted prices closest to actual prices of bonds which they had collected. We can maintain the spirit of this two step approach but with added generality and much lower time cost. First we discard the assumptions regarding functional form and choose a drift vector and diffusion matrix which seem to fit the data. We will use a multidimensional analogue of the nonparametric method of the previous section. For simplicity we will continue to assume that market prices of risk are constant.

It was found that neither component of the drift vector seemed to depend much on either variable except for very high values where the same nonlinear behaviour is apparent as was seen for the short rate itself. The diffusion matrix, on the other hand, exhibited quite a bit of nonlinearity (in each component). The vector $(\epsilon_t^r, \epsilon_t^l)$ was found to be nearly uncorrelated with correlation coefficient of 0.04. This is counter to the findings of Nelson and Schaefer [29] who found that innovations in the two rates were highly correlated but that the innovations in the short rate and the spread were nearly uncorrelated. Perhaps the nonlinear drift and diffusion have removed some spurious correlation.

Parameter estimates are presented in Table A.3. We denote the market price of "short rate risk" by q1 and that for "long rate risk" by q2.

Table A.3: Results for Nonparametric 2-Factor Model

q1	0.013350
	(0.083802)
q2	-0.016507
	(0.145925)
J-stat	92.719

Table A.4 presents an examination of the pricing errors for this model. Figure A.3 compares average pricing errors for both nonparametric models we have examined. The 2-Factor model performs even worse than the one-factor nonparametric model. However note that this model seems to do a better job than the Brennan and Schwartz model except that it does not do as good a job of removing the trend of increasing APE for longer maturity portfolios. Comparing APEs from these two models with the LEH model shows that neither model is a great improvement on the LEH in terms of APE. However both models do a better job in terms of orthogonality conditions.

Table A.4: Pricing Errors for 2-Factor Model

	APE	Orth − r	Orth − l
rf	-0.001126	-0.000132	-0.000117
	(0.000327)	(0.000032)	(0.000031)
port6	-0.000355	-0.000055	-0.000039
	(0.000300)	(0.000028)	(0.000028)
port12	-0.000119	-0.000034	-0.000012
	(0.000339)	(0.000040)	(0.000035)
port48	0.000525	0.000019	0.000069
	(0.000827)	(0.000109)	(0.000092)
port120	0.000579	0.000011	0.000086
	(0.001155)	(0.000148)	(0.000129)
ltgov	0.000688	0.000012	0.000115
	(0.001768)	(0.000213)	(0.000192)

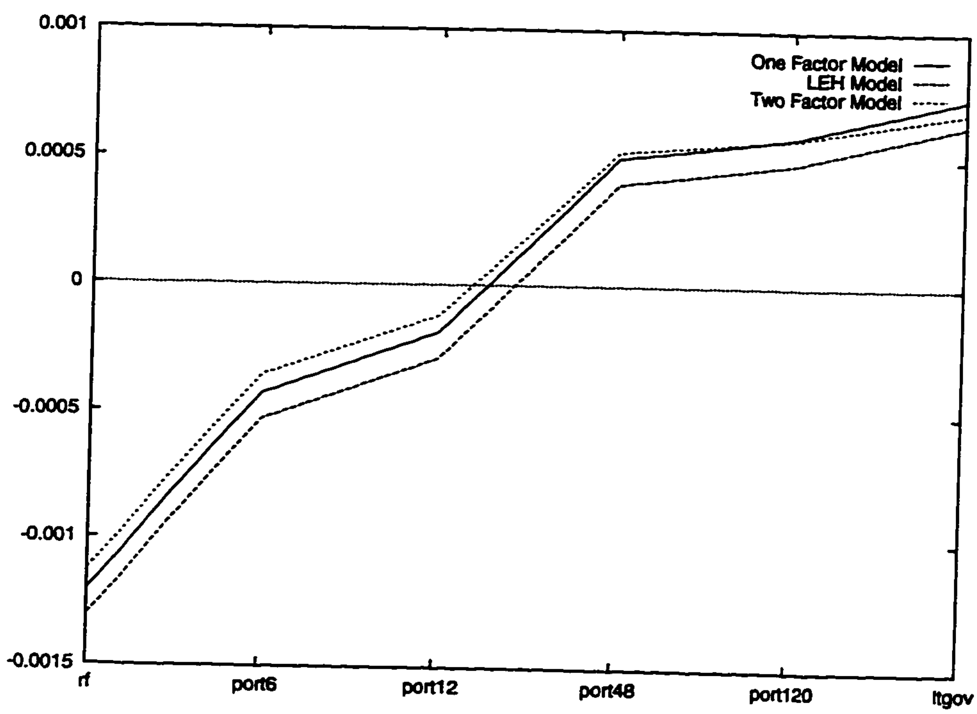

Figure A.3: APEs for Non-parametric models

VITA

Education

PhD., Finance, University of Washington, 1997.

Topic: Evaluating Stochastic Discount Factors from Term Structure Models

Advisor: Wayne Ferson

B.A., Economics, Brigham Young University, 1992

Birth Date

January 15, 1968