Forthcoming:
Journal of Finance

A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk

*Richard Stanton**

Haas School of Business
U.C. Berkeley
545 Student Services Building #1900
Berkeley, CA 94720-1900
tel. (510) 642-7382
fax (510) 643-1420
email: stanton@haas.berkeley.edu

This Draft: January 30, 1997

ABSTRACT

This article presents a technique for nonparametrically estimating continuous-time diffusion processes which are observed at discrete intervals. We illustrate the methodology by using daily three and six month Treasury Bill data, from January 1965 to July 1995, to estimate the drift and diffusion of the short rate, and the market price of interest rate risk. While the estimated diffusion is similar to that estimated by Chan, Karolyi, Longstaff and Sanders (1992), there is evidence of substantial nonlinearity in the drift. This is close to zero for low and medium interest rates, but mean reversion increases sharply at higher interest rates.

*Haas School of Business, U.C. Berkeley. I gratefully acknowledge helpful comments and suggestions from Gurdip Bakshi, Pierluigi Balduzzi, Steve Grenadier, Dilip Madan, Matthew Richardson, René Stulz, an anonymous referee, seminar participants at Berkeley, Harvard, MIT, Michigan, NYU, LBS, Long Term Capital Management, Stanford, UCLA, U.C. Irvine, the 1996 Utah Winter Finance Conference, the 1996 meetings of the Western Finance Association and the Society for Economic Dynamics and Control, and the 1997 meetings of the American Finance Association. I am grateful for financial assistance from the Berkeley Program in Finance, from the U.C. Berkeley Committee on Research, and from the Fisher Center for Real Estate and Urban Economics.

Modern asset pricing theory allows us to value and hedge a wide array of contingent claims, given a continuous-time model for the dynamics of the underlying state variables.[1] Many such models have been developed to describe a range of economic variables, including stock prices (for pricing options and other derivatives) and real investment returns. Given the importance of valuing and hedging the huge institutional holdings of fixed income securities and derivatives, it is not surprising that one of the most common uses of continuous-time models has been in describing the dynamics of the short term riskless interest rate, r_t.

Unfortunately, while the theory tells us what to do once we have a model for the underlying variable, it gives us little or no guidance in choosing the right model in the first place. For example, researchers have proposed many different parametric models of short rate dynamics, each attempting to capture particular features of observed interest rate movements. Moreover, empirical tests of these models have yielded mixed results. As a result, several recent researchers have used *nonparametric* techniques to reduce the number of arbitrary parametric restrictions imposed on the underlying process. For example, Aït-Sahalia (1996a) estimates the diffusion, σ, nonparametrically, given a linear specification for the drift, μ.

This article avoids making parametric assumptions about either the drift or the diffusion. We develop a procedure for estimating both functions nonparametrically from data observed only at discrete time intervals. The procedure, which can be extended to a multivariate setting, involves first constructing a family of approximations to the true drift and diffusion. These approximations converge pointwise to μ and σ at a rate Δ^k, where Δ is the time between successive observations, and k is an arbitrary positive integer. We investigate the performance of these approximations for some commonly used parametric interest rate models, and find that with daily data, even the simplest, first order approximations are almost indistinguishable from the true functions over a wide range of values. As the sampling frequency decreases, the performance of all the approximations deteriorates. However, the higher order approximations remain indistinguishable from the true functions even with monthly data. This suggests that the approximation errors introduced should be small, as long as the series being studied is observed monthly or more often.

These approximations to the drift and diffusion of the underlying process can easily be estimated nonparametrically from discretely sampled data. We illustrate the methodology by estimating a diffusion process for the three month U.S. Treasury Bill rate, using daily data from 1965 to 1995. We find that the drift, μ, shows evidence of substantial nonlinearity. For low and medium interest rates, there is only very slight mean reversion. However, as interest rates continue to climb, the degree of mean reversion increases dramatically. The estimated volatility is similar to that estimated (parametrically) by Chan, Karolyi, Longstaff

[1] For an overview, see Merton (1992) or Duffie (1988).

and Sanders (1992).

To price assets which depend on some underlying state variable, we need to know not only the dynamics of that variable, but also the associated market price of risk, λ (the excess return required for an investor to bear each extra unit of risk). Most previous research assumes a specific parametric form for λ, often $\lambda \equiv 0$.[2] Our estimation procedure allows us to determine this price of risk nonparametrically. We illustrate this by explicitly estimating the functional relationship between the market price of interest rate risk and the level of interest rates, using daily excess returns on six month versus three month Treasury Bills over the period January 1965 to July 1995. Combining this with the estimated dynamics of the three month Treasury Bill rate to price interest rate dependent securities, we obtain results that differ substantially from those obtained assuming the price of risk to be identically zero.

While our methodology yields valuable insights when used to study interest rate dynamics, it is likely to prove even more useful for other, less studied, economic variables. As just one example, it is common to use large numbers of macroeconomic variables in predicting mortgage prepayment rates a few months in the future.[3] Since prepayment rates determine the cash flows received by mortgage-backed security (MBS) holders, any variable that affects prepayment should also affect the value of the MBS. However, valuation models typically ignore all but a few interest rate variables. This is in part because little is known about either the dynamics or the price of risk of the other explanatory variables. The estimation methodology described in this article offers a convenient way of estimating models for these variables.

The remainder of this article is organized as follows. Section I discusses the estimation of diffusion processes, and reviews existing parametric and nonparametric approaches to this problem. Section II uses Taylor series expansions to construct a family of approximations to the drift and diffusion of a diffusion process, and investigates the performance of these approximations for some common parametric interest rate models. Section III derives similar approximations to the market price of interest rate risk. Section IV shows how these approximations can be estimated nonparametrically from data observed only at discrete time intervals, and the results of this estimation using Treasury Bill data are given in Section V. Section VI concludes the article.

[2] i.e. that the *local expectations hypothesis* holds (see Cox, Ingersoll and Ross (1981)).

[3] These variables include housing starts, unemployment etc.

I. Estimating Diffusion Models

In valuing contingent claims, it is convenient to represent the underlying state variable(s) as a continuous-time diffusion process, satisfying a time-homogeneous stochastic differential equation,

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dZ_t. \tag{1}$$

Here, Z_t is a standard Brownian motion, and μ and σ, the drift and diffusion of the process $\{X_t\}$, are functions only of the contemporaneous value of X_t. Examples include the interest rate models of Cox, Ingersoll and Ross (CIR) (1985), Vasicek (1977), Brennan and Schwartz (1979, 1982), and Chan, Karolyi, Longstaff and Sanders (CKLS) (1992). The main difference between these models (and many others) lies in their assumed functional forms for μ and σ in equation (1). For example,

$$\begin{aligned}
&\text{CIR:} && dr_t = (\alpha_0 + \alpha_1 r_t)\,dt + \sigma\sqrt{r_t}\,dZ_t,\\
&\text{Vasicek:} && dr_t = (\alpha_0 + \alpha_1 r_t)\,dt + \sigma\,dZ_t,\\
&\text{CKLS:} && dr_t = (\alpha_0 + \alpha_1 r_t)\,dt + \sigma r_t^{\gamma}\,dZ_t.
\end{aligned}$$

Some recent interest rate models (including Hull and White (1990), Ho and Lee (1986), Black, Derman and Toy (1990), and Black and Karasinski (1991)), while possessing continuous-time representations similar to equation (1), do *not* assume time-homogeneity. Instead, they allow arbitrary time dependence in certain parameters to match the current term structure exactly. Although such "arbitrage-free" models are in common use by practitioners, they often imply counterfactual behavior for the future behavior of interest rates. In addition, to retain their exact fit to the term structure they need to be re-estimated each period, undermining the assumptions under which they are valid (see Backus, Foresi and Zin (1995) and Canabarro (1995)). This article concentrates only on time-homogeneous representations.

A. *Parametric Models*

In estimating the functions μ and σ in equation (1), the usual approach is first to specify parametric forms for the functions μ and σ (as in the examples above), then to estimate the values of the parameters. Given functions μ and σ, the transition density from value x at time t to value y at some later time s, $p(s, y \mid t, x)$, must satisfy the Kolmogorov forward equation,

$$\frac{\partial p(s,y \mid t,x)}{\partial s} = -\frac{\partial}{\partial y}\left(\mu(y)p(s,y \mid t,x)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y)p(s,y \mid t,x)\right), \tag{2}$$

and the Kolmogorov backward equation (see Øksendal (1985)),

$$-\frac{\partial p(s,y \mid t,x)}{\partial t} = \mu(x)\frac{\partial}{\partial x}\left(p(s,y \mid t,x)\right) + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}\left(p(s,y \mid t,x)\right). \tag{3}$$

In principle, for a given parametrization of μ and σ, we can solve equation (2) for the conditional density p as a function of the parameters, then use maximum likelihood to estimate the model's parameters (see, for example, Lo (1988)). This approach was followed by Pearson and Sun (1994) in estimating the parameters of the CIR interest rate model, using the fact that, under this process, interest rates are conditionally distributed as a multiple of a non-central χ^2 random variable (see Feller (1951)).

Unfortunately, except in a few cases such as this, equation (2) can only be solved numerically, making implementation of maximum likelihood extremely inconvenient. Hansen's (1982) Generalized Method of Moments (GMM) can often be used instead of full maximum likelihood, either when the full likelihood function is too complicated or time-consuming to calculate, or where we wish to specify only certain properties of the distribution, rather than the full likelihood function. Duffie and Singleton (1993) use simulation to calculate arbitrary population moments as functions of the parameters of the process being estimated, which can be compared with sample moments estimated from the data. Gallant and Tauchen (1994) also use simulation, generating moment conditions from the score function of an auxiliary (quasi) maximum likelihood estimation. Hansen and Scheinkman (1995) show how to derive analytic moment restrictions from equation (1) using the infinitesimal generator (see Øksendal (1985)) of X_t, $\mathcal{L}$, defined by

$$\begin{aligned}
\mathcal{L}f(x,t) &= \lim_{\tau \downarrow t} \frac{E\left(f(X_\tau,\tau) \mid X_t = x\right) - f(x,t)}{\tau - t}, \\
&= \frac{\partial f(x,t)}{\partial t} + \frac{\partial f(x,t)}{\partial x}\mu(x) + \frac{1}{2}\frac{\partial^2 f(x,t)}{\partial x^2}\sigma^2(x).
\end{aligned} \tag{4}$$

For example, their first class of moment conditions (C1) can be obtained by noting that, if X_t is stationary, $E\left[\phi(X_t)\right]$ must be independent of calendar time for any function ϕ. This implies that its unconditional expected rate of change must be zero, i.e.

$$\begin{aligned}
\text{C1:} \quad E\left[\mathcal{L}\phi(X_t)\right] &= E\left[\phi'(X_t)\mu(X_t) + \tfrac{1}{2}\phi''(X_t)\sigma^2(X_t)\right], \\
&= 0.
\end{aligned}$$

While these moment conditions are less computationally intensive than those of Duffie and Singleton (1993) or Gallant and Tauchen (1994), they do not take advantage of all of the information contained in the discretely observed data. An alternative approach is to use GMM with *approximate* moment conditions. A well-known example is Chan, Karolyi, Longstaff and Sanders (1992). In estimating their continuous-time interest rate model,

$$dr_t = (\alpha + \beta r_t)\, dt + \sigma r^{\gamma}\, dZ_t, \tag{5}$$

they use approximate conditional moments of the form

$$\begin{aligned} E_t\left(\epsilon_{t+\Delta}\right) &= 0, && (6)\\ E_t\left(\epsilon_{t+\Delta}^2\right) &= \sigma^2 r_t^{2\gamma}\Delta, \qquad \text{where} && (7)\\ \epsilon_{t+\Delta} &= r_{t+\Delta} - r_t - (\alpha + \beta r_t)\,\Delta. && (8) \end{aligned}$$

While these are only approximately correct, this approach is the simplest of all to implement, and as we shall show later, the approximation errors introduced are likely to be small in practice, as long as reasonably frequent data are available.

B. Nonparametric Methods

One potentially serious problem with any parametric model, particularly when there is no economic reason why we should prefer one functional form over another, is misspecification. Even if a model fits interest rate movements well in-sample, this does not necessarily imply that it will price securities well. This is because the price today of an interest rate dependent security depends not on past interest rates, but on the entire distribution of possible *future* interest rates between today and the maturity of the security. Fitting historical data well is no guarantee of matching this entire distribution, leading to the possibility of large pricing and hedging errors. In fact, existing parametric interest rate models do not even fit *historical* data well. Aït-Sahalia (1996b) rejects "…every parametric model of the spot rate [previously] proposed in the literature" by comparing the marginal density implied by each model with that estimated from the data. The results of Backus, Foresi and Zin (1995) and Canabarro (1995) further show that misspecification of the underlying interest rate model can lead to serious pricing and hedging errors.

To avoid misspecification, recent research has used *nonparametric* estimation techniques in order to avoid having to specify (arbitrary) functional forms for μ and/or σ. By letting $t \to -\infty$ in the Kolmogorov forward equation (2), we obtain

$$\frac{d^2}{dx^2}\left(\sigma^2(x)\pi(x)\right) = 2\frac{d}{dx}\left(\mu(x)\pi(x)\right), \tag{9}$$

a relationship between the drift (μ), the diffusion (σ), and the stationary density (π). Given any two of these functions, equation (9) allows us to determine the third. For example, Banon (1978) integrates this equation to obtain

$$\mu(x) = \frac{1}{2\pi(x)} \frac{d}{dx} \left[\sigma^2(x)\pi(x) \right], \tag{10}$$

which allows us to estimate the drift nonparametrically, given a nonparametric estimate of the stationary density, π, but only if we know the diffusion, σ. Aït-Sahalia (1996a) assumes a linear drift,

$$\mu(x) = \kappa \left[\theta - x \right],$$

which can be estimated using OLS, then estimates σ nonparametrically, by integrating equation (10) one more time to obtain[4]

$$\sigma^2(x) = \frac{2}{\pi(x)} \int_0^x \mu(u)\pi(u)du. \tag{11}$$

However, just as we have no economic motivation for fully parametrizing both μ and σ, we have no particular basis for parametrizing either one. While misparametrizing the drift would not matter (at least for pricing purposes) if the underlying state variable were itself an asset price, this is not the case otherwise.[5] In particular, both the drift and the diffusion are important in pricing interest rate derivatives. Moreover, there is mounting evidence that the drift of the short rate is *not* linear (see, for example, Conley, Hansen, Luttmer and Scheinkman (1996), Aït-Sahalia (1996b), and Pfann, Schotman and Tschernig (1996)). To avoid imposing parametric restrictions on either μ *or* σ, the next section derives a family of approximations to these functions, which

1. Can easily be identified and estimated (parametrically or nonparametrically) given only discrete observations on X_t.

2. Converge pointwise to the true functions μ and σ at a rate Δ^k, where Δ is the time between successive observations, and k is an arbitrary positive integer.

---

[4] With this parametrization of the drift,

$$E_t \left(X_{t+\Delta} \right) = \theta \left[1 - e^{-\kappa\Delta} \right] + e^{-\kappa\Delta} X_t,$$

independent of σ.

[5] When the underlying state variable is an asset price, its drift is replaced by the riskless interest rate in the partial differential equation for derivative security prices. However, as Lo and Wang (1995) point out, misspecifying the drift may lead to misestimation of the diffusion.

II. Constructing Approximations to μ and σ

A. *Taylor Series*

Consider a diffusion process, X_t, which satisfies the stochastic differential equation

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dZ_t.$$

Under suitable restrictions on μ, σ, and an arbitrary function f (see, for example, Hille and Phillips (1957), Chapter 11), we can write the conditional expectation $E_t\left[f(X_{t+\Delta}, t)\right]$ in the form of a Taylor series expansion,

$$E_t\left[f(X_{t+\Delta}, t+\Delta)\right] = f(X_t,t) + \mathcal{L}f(X_t,t)\Delta + \frac{1}{2}\mathcal{L}^2 f(X_t,t)\Delta^2 + \ldots + \frac{1}{n!}\mathcal{L}^n f(X_t,t)\Delta^n + O(\Delta^{n+1}), \tag{12}$$

where $\mathcal{L}$ is the infinitesimal generator of the process $\{X_t\}$ (see equation (4)). Note that, if we set $\sigma \equiv 0$, this reduces to a standard non-stochastic Taylor series.

The most common use of equation (12) is in the construction of numerical approximations to the expectation on the left hand side, given known functions μ and σ (see, for example, Milshtein (1978), and Duffie and Singleton (1993)). Our aim here is, in a sense, the opposite. We do *not* know the functions μ and σ, but given a long enough interest rate series we can estimate the expectation on the left hand side of equation (12). Given suitable choices of the function f, equation (12) can then be used to construct approximations to μ and σ.

B. *Approximating μ and σ*

We can rewrite equation (12) in the following form,

$$\mathcal{L}f(X_t,t) = \frac{1}{\Delta}E_t\left[f(X_{t+\Delta}, t+\Delta) - f(X_t,t)\right] - \frac{1}{2}\mathcal{L}^2 f(X_t,t)\Delta - \frac{1}{6}\mathcal{L}^3 f(X_t,t)\Delta^2 - \ldots. \tag{13}$$

Ignoring all terms except the first on the right hand side gives us a first order approximation for $\mathcal{L}f$,

$$\mathcal{L}f(X_t,t) = \frac{1}{\Delta}E_t\left[f(X_{t+\Delta}, t+\Delta) - f(X_t,t)\right] + O(\Delta). \tag{14}$$

We can also construct higher order approximations, despite the fact that higher order terms in equation (13) involve derivatives of μ and σ, which are themselves unknown. To see this, consider equation (13) with a time step of 2Δ,

$$\mathcal{L}f(X_t,t) = \frac{1}{2\Delta}E_t\left[f(X_{t+2\Delta},t+2\Delta) - f(X_t,t)\right] - \frac{1}{2}\mathcal{L}^2 f(X_t,t)(2\Delta) - \frac{1}{6}\mathcal{L}^3 f(X_t,t)(2\Delta)^2 - \ldots. \tag{15}$$

Multiplying equation (13) by 2, and subtracting equation (15), yields the second order approximation

$$\begin{aligned}\mathcal{L}f(X_t,t) &= \frac{1}{2\Delta}\left\{4E_t\left[f(X_{t+\Delta},t+\Delta) - f(X_t,t)\right] - E_t\left[f(X_{t+2\Delta},t+2\Delta) - f(X_t,t)\right]\right\} \\ &\quad + O(\Delta^2),\end{aligned} \tag{16}$$

an approximation to $\mathcal{L}f$ in terms of expectations of functions of only observed values of $\{X_t\}$, which converges to the true function at a rate Δ^2 as $\Delta \to 0$. We can continue this process, generating approximations of successively higher order. For example, repeating the process once more, using a time step of 3Δ, yields the third order approximation

$$\begin{aligned}\mathcal{L}f(X_t,t) &= \frac{1}{6\Delta}\left\{18E_t\left[f(X_{t+\Delta},t+\Delta) - f(X_t,t)\right] - 9E_t\left[f(X_{t+2\Delta},t+2\Delta) - f(X_t,t)\right]\right. \\ &\quad \left. + 2E_t\left[f(X_{t+3\Delta},t+3\Delta) - f(X_t,t)\right]\right\} + O(\Delta^3).\end{aligned} \tag{17}$$

To approximate a particular function $g(x,t)$, we now need merely to find a specific function f satisfying

$$\mathcal{L}f(x,t) = g(x,t). \tag{18}$$

Drift To derive approximations to the drift, μ, consider the function

$$f_{(1)}(x,t) \equiv x. \tag{19}$$

From the definition of $\mathcal{L}$, we have

$$\mathcal{L}f_{(1)}(x,t) = \mu(x). \tag{20}$$

Substituting successively into equations (14), (16) and (17) leads to the following approximations for μ,

$$\mu(X_t) = \frac{1}{\Delta}E_t\left[X_{t+\Delta} - X_t\right] + O(\Delta), \tag{21}$$

$$\mu(X_t) = \frac{1}{2\Delta}\left\{4E_t\left[X_{t+\Delta} - X_t\right] - E_t\left[X_{t+2\Delta} - X_t\right]\right\} + O(\Delta^2), \tag{22}$$

$$\mu(X_t) = \frac{1}{6\Delta}\left\{18E_t\left[X_{t+\Delta} - X_t\right] - 9E_t\left[X_{t+2\Delta} - X_t\right] + 2E_t\left[X_{t+3\Delta} - X_t\right]\right\}$$

$$+ O(\Delta^3), \tag{23}$$

etc.

Diffusion To construct approximations to the diffusion, σ, consider the function

$$f_{(2)}(x,t) \equiv (x - X_t)^2 . \tag{24}$$

From the definition of $\mathcal{L}$, we have

$$\mathcal{L}f_{(2)}(x,t) = 2(x - X_t)\mu(x) + \sigma^2(x), \qquad \text{and so} \tag{25}$$

$$\mathcal{L}f_{(2)}(X_t,t) = \sigma^2(X_t). \tag{26}$$

Substituting into equations (14), (16) and (17) yields approximations for σ^2,

$$\sigma^2(X_t) = \frac{1}{\Delta}E_t\left[(X_{t+\Delta} - X_t)^2\right] + O(\Delta), \tag{27}$$

$$\sigma^2(X_t) = \frac{1}{2\Delta}\left\{4E_t\left[(X_{t+\Delta} - X_t)^2\right] - E_t\left[(X_{t+2\Delta} - X_t)^2\right]\right\} + O(\Delta^2), \tag{28}$$

$$\sigma^2(X_t) = \frac{1}{6\Delta}\left\{18E_t\left[(X_{t+\Delta} - X_t)^2\right] - 9E_t\left[(X_{t+2\Delta} - X_t)^2\right] + 2E_t\left[(X_{t+3\Delta} - X_t)^2\right]\right\} + O(\Delta^3). \tag{29}$$

Note that the first order approximations, equations (21) and (27), correspond to the "naive" discretization used by many previous authors, such as Chan, Karolyi, Longstaff and Sanders (1992).[6] Taking square roots leads to approximations of the same order for $\sigma(X_t)$,[7]

$$\sigma(X_t) = \sqrt{\frac{1}{\Delta}E_t\left[(X_{t+\Delta} - X_t)^2\right]} + O(\Delta), \tag{30}$$

$$\sigma(X_t) = \sqrt{\frac{1}{2\Delta}\left\{4E_t\left[(X_{t+\Delta} - X_t)^2\right] - E_t\left[(X_{t+2\Delta} - X_t)^2\right]\right\}} + O(\Delta^2), \tag{31}$$

[6]In related work, Siddique (1994) constructs an alternative first order approximation in his nonparametric estimation of interest rate dynamics. Chesney, Elliott, Madan and Yang (1993) use Milshtein (1978) approximations to generate point by point estimates of diffusion coefficients.

[7]Suppose $\sigma^2(X_t) = A + O(\Delta^k)$. Then, using a binomial approximation,

$$\begin{aligned} \sigma(X_t) &= \sqrt{A}\left[1 + O(\Delta^k)\right]^{\frac{1}{2}}, \\ &= \sqrt{A}\left[1 + \tfrac{1}{2}O(\Delta^k) - \tfrac{1}{8}O(\Delta^{2k}) + \ldots\right], \\ &= \sqrt{A} + O(\Delta^k), \end{aligned}$$

an approximation to σ with the same rate of convergence.

$$\sigma(X_t) = \sqrt{\frac{1}{6\Delta}\left\{18E_t\left[(X_{t+\Delta}-X_t)^2\right] - 9E_t\left[(X_{t+2\Delta}-X_t)^2\right] + 2E_t\left[(X_{t+3\Delta}-X_t)^2\right]\right\}} + O(\Delta^3). \tag{32}$$

As an alternative, we can replace the terms in $E_t\left[(X_{t+j\Delta}-X_t)^2\right]$ with the conditional variance, $\mathrm{var}_t(X_{t+j\Delta})$, leading to the following set of approximations for $\sigma(X_t)$,

$$\sigma(X_t) = \sqrt{\frac{1}{\Delta}\mathrm{var}_t\left(X_{t+\Delta}\right)} + O(\Delta), \tag{33}$$

$$\sigma(X_t) = \sqrt{\frac{1}{2\Delta}\left[4\,\mathrm{var}_t\left(X_{t+\Delta}\right) - \mathrm{var}_t\left(X_{t+2\Delta}\right)\right]} + O(\Delta^2), \tag{34}$$

$$\sigma(X_t) = \sqrt{\frac{1}{6\Delta}\left[18\,\mathrm{var}_t\left(X_{t+\Delta}\right) - 9\,\mathrm{var}_t\left(X_{t+2\Delta}\right) + 2\,\mathrm{var}_t\left(X_{t+3\Delta}\right)\right]} + O(\Delta^3), \tag{35}$$

etc. This does not affect the order of convergence, but may improve the approximation for a given time step, Δ. We shall use these latter approximations in the empirical analysis below.

C. Performance of the Approximations

The higher the order of the approximation, the faster it will converge to the true drift and diffusion of the process given in equation (1), as we observe the variable X_t at finer and finer time intervals. Eventually, if we can sample arbitrarily often, higher order approximations must outperform lower order approximations. However, we may only observe X_t relatively infrequently, say monthly. Even when we have more frequent observations, we may wish to avoid market microstructure issues (such as bid-ask bounce) by not sampling too frequently. It is therefore useful to know how the approximations perform over different sampling intervals. In general, the approximation error of Taylor series expansions involves higher order derivatives of the functions μ, σ and f, which we do not know. To get some idea for the performance of the approximations with interest rate models that we might wish to estimate, we consider the Cox, Ingersoll and Ross (1985) model, and a version of the Black, Derman and Toy (1990) model. These take the form:

$$\text{CIR: } dr_t = \kappa\left(\theta - r_t\right)dt + s\sqrt{r_t}\,dZ_t, \tag{36}$$

$$\text{BDT: } dy_t = \kappa\left(\theta - y_t\right)dt + s\,dZ_t, \qquad \text{where} \tag{37}$$

$$y_t = \ln r_t, \qquad \text{so} \tag{38}$$

$$dr_t = r\left[\kappa\left(\theta - \ln r_t\right) + \frac{1}{2}s^2\right]dt + s\,r\,dZ_t. \tag{39}$$

Under the CIR model, interest rates are conditionally distributed as a multiple of a non-central χ^2 random variable. In the BDT model, interest rates are conditionally lognormal. For both models, we can therefore calculate conditional means and variances in closed form, allowing us to compare the approximations derived above with the true values of μ and σ.

Table I shows first, second, and third order approximations to the drift of the CIR process with parameter values $\kappa = 0.5, \theta = 0.07$, and $s = 0.1$ (these roughly match estimates obtained by previous authors). Using daily data, it can be seen that first, second, and third order approximations are all essentially indistinguishable from the true drift for values of r_t between 1 and 30 percent. As the sampling frequency decreases, the performance of all of the approximations deteriorates, although this deterioration is worse for the lower order approximations. Even so, with weekly and even monthly data, the second and third order approximations are still indistinguishable from the true drift; even the first order approximation is extremely close for values of r_t below about 10 percent.

Table II shows first, second, and third order approximations to the drift of the BDT process with parameter values $\kappa = 0.5, \theta = -2.75$, and $s = 0.43$. Despite the nonlinearity of the drift, the results are very similar to those for the CIR model. Again the higher order approximations do better. For daily data, all results are indistinguishable from the true drift. For weekly data, the second and third order approximations still match the true drift very closely. For monthly data, the third order approximation is still almost identical to the true μ, and the second order approximation is very close.

Tables III and IV look at the corresponding approximations to the diffusion, $\sigma(r)$, for the two interest rate processes. The results are similar to those for the drift. For the CIR model (Table III), all of the approximations are virtually indistinguishable from the true diffusion function as long as the data are sampled at least weekly. The second and third order approximations are indistinguishable from the true σ even with monthly data. For the BDT model (Table IV), the results are similar. Again, the third order approximation is almost indistinguishable from the true diffusion for data observed at least monthly.

Overall, these results indicate that, at least for models similar to these two, as long as we sample the data monthly or better, the errors introduced by using approximations rather than the true drift and diffusion are extremely small, especially when compared with the likely magnitude of estimation error. They also show that, even with daily or weekly data, we can achieve gains by using higher order approximations compared with the traditional first order discretizations, equations (21) and (33).[8]

[8] The performance of the approximations to both the drift and diffusion of the nonparametric model estimated below, in Section V, is very similar to that obtained here for the CIR and BDT models.

III. Valuation and the Market Price of Risk

Knowing the process governing movements in a state variable such as the short rate is *not* enough by itself to allow us to price contingent claims whose payoffs depend on that variable. Focusing on interest rates, assume an asset's price can be written as $V(r,t)$, depending only on the short rate and time. Then, by Ito's Lemma,

$$\frac{dV(r,t)}{V(r,t)} = m(r,t)\,dt + s(r,t)\,dZ, \tag{40}$$

where

$$\begin{aligned} m(r,t)\,V &= V_t + \mu(r)V_r + \frac{1}{2}\sigma^2(r)V_{rr}, && (41)\\ s(r,t)\,V &= \sigma(r)V_r. && (42)\end{aligned}$$

This equation holds for any asset V. With a one factor interest rate model, the instantaneous returns on all interest rate contingent claims are perfectly correlated. As a result, the risk premium on any asset must be proportional to the standard deviation of its return, to prevent arbitrage.[9] If the asset pays out dividends at rate d, we can thus write

$$m = r - \frac{d}{V} + \lambda(r,t)\frac{V_r}{V}, \tag{43}$$

where $\lambda(r,t)$ is the price of interest rate risk. Substituting equation (43) into equation (41), assuming that the market price of interest rate risk is a function only of r_t, yields a partial differential equation that must be satisfied by any interest rate contingent claim,

$$\frac{1}{2}\sigma(r)^2V_{rr} + [\mu(r) - \lambda(r)]\,V_r + V_t - rV + d = 0, \tag{44}$$

subject to appropriate boundary conditions, and some technical regularity conditions (see, for example, Duffie (1988)). To price interest rate dependent assets, we therefore need to know not only the functions μ and σ, but also the price of risk, λ. While previous authors have assumed various functional forms for $\lambda(r)$, the fact that μ and λ only appear in equation (44) as the difference $(\mu - \lambda)$ means that misspecifying λ can lead to pricing and hedging errors of exactly the same magnitude as misspecifying μ.[10]

[9] Suppose this did not hold for two risky assets. We could then create a riskless portfolio of these two assets with a return strictly greater than r, leading to an arbitrage opportunity (see Ingersoll (1987)).

[10] For example, in Chan, Karolyi, Longstaff and Sanders (1992), $\lambda(r) = 0$; in Cox, Ingersoll and Ross (1985), $\lambda(r) = qr$; in Vasicek (1977), $\lambda(r) = q$; in Aït-Sahalia (1996a), $\lambda(r) = q\sigma(r)$.

A. *Approximating the Market Price of Risk*

The approximation methodology developed in Section II can be used to construct approximations to $\lambda(r)$ in a manner analogous to those for μ and σ. Remembering that $\lambda(r)$ is related to the excess return on interest rate dependent securities, let $V^{(1)}(r,t)$ and $V^{(2)}(r,t)$ be the prices of two non-dividend paying securities, and consider the function

$$f_{(3)}(r,s) \equiv \frac{V^{(1)}(r,s)}{V^{(1)}(r_t,t)} - \frac{V^{(2)}(r,s)}{V^{(2)}(r_t,t)}, \tag{45}$$

the excess return on asset 1 over asset 2. From the definition of $\mathcal{L}$, we have

$$\begin{aligned} \mathcal{L}f_{(3)}(r_t,t) &= \frac{1}{V^{(1)}(r_t,t)}\left[V_t^{(1)} + V_r^{(1)}\mu(r_t) + \frac{1}{2}V_{rr}^{(1)}\sigma^2(r_t)\right] \\ &\quad - \frac{1}{V^{(2)}(r_t,t)}\left[V_t^{(2)} + V_r^{(2)}\mu(r_t) + \frac{1}{2}V_{rr}^{(2)}\sigma^2(r_t)\right]. \end{aligned} \tag{46}$$

Using equation (44) to substitute for each of the bracketed expressions, we obtain

$$\mathcal{L}f_{(3)}(r_t,t) = \lambda(r_t)\left[\frac{V_r^{(1)}}{V^{(1)}} - \frac{V_r^{(2)}}{V^{(2)}}\right]. \tag{47}$$

Using equation (42), this can be rewritten as

$$\mathcal{L}f_{(3)}(r_t,t) = \frac{\lambda(r_t)}{\sigma(r_t)}\left[\sigma^{(1)}(r_t) - \sigma^{(2)}(r_t)\right], \tag{48}$$

where

$$\sigma^{(i)}(r_t) \equiv \frac{V_r^{(i)}\sigma(r_t)}{V^{(i)}}, \tag{49}$$

the instantaneous volatility of asset i. Substituting in equation (12), and rearranging terms, yields

$$\lambda(r_t) = \frac{\sigma(r_t)}{\sigma^{(1)}(r_t) - \sigma^{(2)}(r_t)}\left[\frac{1}{\Delta}E_t\left(R_{t,t+\Delta}^{(1)} - R_{t,t+\Delta}^{(2)}\right) - \frac{1}{2}\mathcal{L}^2 f_{(3)}(r_t,t)\Delta - \frac{1}{6}\mathcal{L}^3 f_{(3)}(r_t,t)\Delta^2 - \ldots\right], \tag{50}$$

where $R_{t,t+\Delta}^{(i)}$ is the holding period return on asset i between times t and $t+\Delta$. This leads immediately to a first order approximation for λ,

$$\lambda(r_t) = \frac{\sigma(r_t)}{\Delta\left(\sigma^{(1)}(r_t) - \sigma^{(2)}(r_t)\right)}E_t\left(R_{t,t+\Delta}^{(1)} - R_{t,t+\Delta}^{(2)}\right) + O(\Delta). \tag{51}$$

Higher order approximations can be obtained as above.[11]

B. Absence of Arbitrage

If the short term riskless rate, r_t, really does follow the one factor model described by equation (1), and if the market price of interest rate risk really is a function only of r_t, then given enough data our estimates will get arbitrarily close to the true functions μ, σ and λ, which must automatically preclude any possibility of arbitrage. However, given only a finite amount of data, we only have estimates of these functions, and it is well-known that an arbitrary specification of the function $\lambda(r)$ may lead to arbitrage opportunities (see, for example, Cox, Ingersoll and Ross (1985), p. 398, or Ingersoll (1987), pp. 400–401). Fortunately, we can quite simply avoid this possibility. To preclude arbitrage in a one factor world, the excess return on any asset must be proportional to its standard deviation. From equations (41) and (42), this is equivalent to the drift restriction given in equation (43), *as long as* $\sigma(r) \neq 0$. However, from equation (42), when $\sigma(r) = 0$, the value V_r/V is indeterminate, yet the standard deviation of any interest rate contingent claim is zero. Its excess return must therefore also be zero. To preclude arbitrage, we therefore need the *pair* of restrictions

$$\begin{aligned} m &= r - \frac{d}{V} + \lambda(r)\frac{V_r}{V}, && (52) \\ \lambda(r) &= 0 \qquad\qquad \text{if } \sigma(r) = 0. && (53) \end{aligned}$$

The arbitrage opportunity noted by Cox, Ingersoll and Ross (1985) arises through failure to impose this second condition. Fortunately, the estimate of $\lambda(r)$ in equation (51) automatically satisfies this second restriction, since it involves multiplying by the corresponding estimator of $\sigma(r)$. As a result, we can be sure that, in using the estimated functions μ, σ and λ to price interest rate derivatives, we are using a model which precludes the possibility of arbitrage.

[11] We do not actually know the functions $\sigma(r_t)$ and $\sigma^{(i)}(r_t)$ which appear on the right hand side of equation (51). However, as long as we have $O(\Delta)$ approximations of these functions, the error term in equation (51) is still $O(\Delta)$. In constructing a second order approximation to λ, we need $O(\Delta^2)$ approximations for $\sigma(r_t)$ and $\sigma^{(i)}(r_t)$, and so on.

IV. Implementation

A. *Kernel Density Estimation*

To implement the procedure, we need a means of nonparametrically estimating the conditional expectations (regression functions) in equations (21), (22), (23), (33)–(35), and (51). We use a kernel estimation procedure for doing this. Kernel estimation is a nonparametric method for estimating the joint density of a set of random variables. Given m-dimensional vectors $\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T$, from an unknown density $f(\mathbf{z})$, a kernel estimator of this density is

$$\hat{f}(\mathbf{z}) = \frac{1}{Th^m} \sum_{t=1}^{T} K\left(\frac{\mathbf{z} - \mathbf{z}_t}{h}\right), \tag{54}$$

where $K(\cdot)$ is a suitable kernel function and h is the window width or smoothing parameter. The density at any point is estimated as the average of densities centered at the actual data points. The further away a data point is from the estimation point, the less it contributes to the estimated density, and hence the estimated density is highest near high concentrations of data points, and lowest when observations are sparse. The kernel density estimator in equation (54) is similar to a (multidimensional) histogram. Just as with a histogram, for each of the T points in the sample, we add a "block" of total volume $1/T$. Unlike a histogram, however, the blocks are not (in general) rectangular, and they are centered at each data point, rather than at the center of a fixed number of bins.

The econometrician has at his or her discretion the choice of $K(\cdot)$ and h. Results in the kernel estimation literature suggest that any reasonable kernel gives almost optimal results (see Epanechnikov (1969)). Here we use a normal kernel function. The other parameter, the window width, is chosen based on the dispersion of the observations. Scott (1992) suggests the window width,

$$\hat{h}_i = \hat{\sigma}_i T^{\frac{-1}{m+4}}, \tag{55}$$

where $\hat{\sigma}_i$ is the standard deviation estimate of each variable z_i, T the number of observations, and m the dimension of the variables. This window width has the property that, for certain joint distributions of the variables, it minimizes the asymptotic mean integrated squared error of the estimated density function.

Given the estimated density in equation (54), we can calculate any moments we desire from the distribution. For example, the conditional expectation which appears in the first order approximation to μ (equation (21)) is estimated as

$$E\left[r_{\tau+\Delta}-r_t | r_\tau = r\right] \approx \frac{\sum_{t=1}^{T-1}\left(r_{t+\Delta}-r_t\right) K\left(\frac{r-r_t}{h}\right)}{\sum_{t=1}^{T-1} K\left(\frac{r-r_t}{h}\right)}, \tag{56}$$

where $K(z)=(2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}z^2}$. This is a weighted average of the observed interest rate changes, the weight depending on how far each observation is from the value r. Similar expressions yield estimates for the other conditional expectations required, as well as other statistics, such as variances, covariances and other moments. For further discussion of kernel estimators and their properties, see Scott (1992) or Silverman (1986).

B. *Data Description*

To apply the methodology to interest rate dynamics, daily values of the secondary market yields on three and six month Treasury Bills between January 1965 and July 1995 are obtained, and converted from discounts to annualized interest rates. The three month yield is used in estimating the dynamics of the short term riskless rate, and the market price of interest rate risk is calculated using the six month and three month Treasury Bills as assets 1 and 2 respectively in equation (51). No specific adjustments are made for weekends or holidays.

The annualized three month yield is plotted in Figure 1, and Figure 2 shows a scatter plot of daily changes in the short rate plotted against the previous day's rate.[12] It shows distinct evidence of heteroskedasticity, with the range of the changes increasing markedly as the level of interest rates increases. As an illustration of the kernel methodology, and to give some idea for the distribution of interest rates, Figure 3 shows the estimated marginal density of the short rate, calculated using equation (54). This figure also shows pointwise 95 percent confidence bands for the estimated density, calculated using 10,000 iterations of the Künsch (1989) block bootstrap algorithm.

V. Results

A. *Estimation of μ, σ and λ*

First, second, and third order approximations to the drift, $\mu(r_t)$, diffusion, $\sigma(r_t)$, and market price of interest rate risk, $\lambda(r_t)$, together with pointwise 95 percent confidence bands, are

[12]Note that estimating the conditional expectation $E[r_{t+\Delta}-r_t \mid r_t = r]$ in equation (56), where Δ corresponds to one day, is equivalent to determining a (nonlinear) regression line through Figure 2.

shown in Figures 4 and 5.[13] The first thing to notice about both figures is that the first, second, and third order approximations all yield very similar estimates, which is what we would expect if the approximation errors are as small as the experiments above suggest.

Looking first at Figure 4, we see that the estimated drift does not look linear. For low and medium values of the interest rate r_t, there is only very slight mean reversion. For example, the expected change in r_t over the next year, conditional on its being 0 percent this period, is only 0.44 percent. As r_t increases beyond about 14 percent, however, the estimated drift drops sharply. Figure 5 shows the estimated diffusion function, $\sigma(r_t)$. This looks neither linear (as in Vasicek (1977)), nor like a square root function (as in Cox, Ingersoll and Ross (1985)), but appears closer to the $r^{1.5}$ estimated by Chan, Karolyi, Longstaff and Sanders (1992). However, our drift estimate is very different. Their estimated diffusion, combined with the assumed linear drift, results in an interest rate process that is actually non-stationary. The dramatic decline in the drift that we estimate at high interest rates has the effect of preventing interest rates from exploding towards infinity, despite the increase in volatility.

Figure 6 shows the estimated market price of interest rate risk, $\lambda(r_t)$. This function takes on negative values (corresponding to a positive premium for bearing interest rate risk), is decreasing in r_t, and decreases more rapidly as r_t increases (thus behaving differently from the linear specification in Cox, Ingersoll and Ross (1985)).

Looking at the 95 percent pointwise confidence bands around the estimated functions, we see that the bands are fairly tight for all three functions at low interest rates (where there are many observations). However, as interest rates rise and the data become sparser, the bands become much wider, especially for the market price of risk, reflecting the lower confidence we have in our estimates here. This emphasizes the greater data requirements of nonparametric techniques compared with their parametric counterparts.

B. Economic Significance of Price of Risk

The estimated market price of interest rate risk is different from zero.[14] However, this does not necessarily translate into a significant difference in pricing interest rate dependent assets. To examine this, and to demonstrate how the estimated interest rate process can be used to price assets, the process estimated above, and shown in Figures 4 and 5, is used to value zero-coupon bonds of different maturities. Given this interest rate model, the price of a zero

[13] Confidence bands are calculated (for the first order approximation) using 10,000 iterations of the Künsch (1989) block bootstrap algorithm; results for second and third order approximations are similar.

[14] This is consistent with the results of Ronn and Wadhwa (1995), and with the extensive literature documenting relationships between bond maturity and expected return (see, for example, Fama (1984), Fama and Bliss (1987), and Fama and French (1993)).

coupon bond with a payoff of \$1 at time T, $P_t(T)$, can be written in the form

$$P_t(T) = E_t\left[e^{-\int_t^T \widehat{r}_u\,du}\right], \tag{57}$$

where

$$\begin{aligned} \widehat{r}_t &= r_t, & (58)\\ d\widehat{r}_t &= [\mu(\widehat{r}_t) - \lambda(\widehat{r}_t)]\,dt + \sigma(\widehat{r}_t)\,dZ_t. & (59)\end{aligned}$$

The expectation in equation (57) can be calculated numerically using Monte Carlo simulation. This entails repeatedly simulating paths for the risk-adjusted interest rate process $\widehat{r}_t$ using (a discrete version of) the dynamics described by equations (58) and (59), calculating the integral inside the expectation in equation (57) for each path, and averaging over the values obtained for each path to obtain an estimate of the value $P_t(T)$.[15] Table V compares the prices obtained using the price of risk estimated and plotted in Figure 6, with prices obtained using the same interest rate model, but assuming $\lambda(r_t) = 0$. It can be seen that incorporating the estimated market price of interest rate risk makes a significant difference to the results, the size of the difference increasing as the maturity of the bond increases. One year bond prices differ very little; however, there is greater dispersion in the two year bond prices, and a difference of 5 percent in the prices of three year bonds when the short rate is 5 percent. If zero-coupon bond prices differ this much, we can expect even greater differences for more complex interest rate derivatives.

C. *Constraints on Volatility*

By not specifying a particular parametric form, nonparametric techniques avoid the possibility of misspecification, but at the expense of greater estimation error than their parametric counterparts. Given enough data, our estimates will eventually satisfy any constraints satisfied by the true functions, μ, σ and λ. However, this is not necessarily true in any particular finite sample. This is a generic feature of nonparametric estimators, and is the price that we pay for their flexibility.

On the other hand, if we know that the functions ought to satisfy some constraints, and if we can impose these constraints on our estimation, this can only improve our results. As an example, there is currently some very interesting research underway (see Goldman and Ruud (1995)) into nonparametric estimation subject to monotonicity and convexity

[15]In performing the simulations here, 10,000 interest rate paths are simulated using 100 time periods per day, and the variability of the results is reduced using the antithetic variate approach (see Boyle (1977)).

constraints. We cannot be sure, in the present setting, that the functions μ, σ and λ necessarily satisfy either of these constraints, but it *is* natural to assume that $\sigma(0) = 0$, a condition which prevents interest rates from becoming negative.[16] This constraint is not imposed in the estimation above and, indeed, the estimated value of $\sigma(0)$ is greater than zero. However, we can easily modify the estimation procedure to impose this constraint. Instead of approximating σ^2 using the function $f_{(2)}$ defined in Section II, define the related function

$$f_{(2')}(x) \equiv \frac{(x - X_t)^2}{X_t}. \tag{60}$$

From the definition of $\mathcal{L}$, we have

$$\mathcal{L}f_{(2')}(x) = 2(x - X_t)/X_t \mu(x) + \sigma^2(x)/X_t, \qquad \text{and so} \tag{61}$$

$$\mathcal{L}f_{(2')}(X_t) = \sigma^2(X_t)/X_t. \tag{62}$$

Using this function in the approximation methodology of Section II, we obtain a series of approximations to $\sigma^2(r_t)/r_t$, which can be converted to approximations to $\sigma^2(r_t)$ by multiplying by r_t. This guarantees that the estimated value for $\sigma^2(0)$ must be zero. Figure 7 shows the result of using this procedure to estimate the diffusion, subject to the constraint that $\sigma(0) = 0$.

VI. Summary

It is convenient to represent financial series, such as interest rates, as continuous-time diffusion processes, satisfying stochastic differential equations of the form

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dZ_t. \tag{63}$$

However, the specific parametric forms selected for μ and σ often reflect considerations of analytic tractability or computational convenience, more than any real economic motivation. Even for the short term riskless interest rate, one of the most modeled series, empirical tests of existing models have yielded disappointing results. This article presents a technique for *nonparametrically* estimating continuous-time diffusion processes which are observed only at discrete time intervals, and illustrates the methodology with an application to the term structure of interest rates. The procedure involves estimating approximations to the true drift and diffusion. We derive a family of such approximations, which converge to the true functions at a rate Δ^k, where Δ is the time between successive observations, and k

[16]Negative nominal interest rates are ruled out by a simple arbitrage argument.

is an arbitrary positive integer. For some commonly used parametric interest rate models, the approximations and the true functions are almost indistinguishable for any observed interest rate value, as long as the sampling frequency is monthly or greater. This suggests that the approximation errors introduced may not be of practical relevance. Using these approximations to estimate the drift and diffusion of the short rate, we find that while the diffusion function looks similar to the (parametric) function estimated by Chan, Karolyi, Longstaff and Sanders (1992), the drift exhibits substantial nonlinearity. For low and medium interest rates, there is only very slight mean reversion. However, as interest rates continue to climb, the degree of mean reversion increases dramatically.

While knowing the process governing movements in a financial variable is important for forecasting and hedging, it is *not* enough by itself to price contingent claims whose payoffs depend on that variable (such as interest rate derivatives). We also need to know the market price of risk, the excess return required for an investor to bear each additional unit of risk. Previous research has either assumed a specific functional form for the price of risk, or assumed it to be identically zero. Our nonparametric procedure for estimating μ and σ can also be used to estimate the market price of risk associated with a variable. We illustrate this by nonparametrically estimating the functional relationship between the market price of interest rate risk and the level of interest rates, looking at excess returns on six month versus three month Treasury Bills over the period 1965 to 1995. We combine the estimated price of risk with the estimated short rate model to price interest rate dependent securities, and find that the results are very different from those obtained assuming the price of risk to be identically zero.

There are many economic series which are much less studied than the three month Treasury Bill rate, and for which we have little idea of the appropriate model, yet are still potentially important for asset pricing. For example, empirical evidence suggests that bond default likelihood and severity are related to macroeconomic factors, which should therefore be important in pricing defaultable bonds and other credit related assets. In most cases, however, we do not have a good model for the dynamics of these factors, let alone the associated price of risk. Besides shedding valuable new light on the dynamics of the short rate and the market price of interest rate risk, the estimation procedure described in this article (which can easily be extended to a multivariate context) offers a convenient way of estimating such models.

References

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

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

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

Banon, G., 1978, Nonparametric identification for diffusion processes, *SIAM Journal of Control and Optimization* 16, 380–395.

Black, Fischer, Emanuel Derman, and William Toy, 1990, A one factor model of interest rates and its application to Treasury bond options, *Financial Analysts' Journal* 46, 33–39.

Black, Fischer, and Piotr Karasinski, 1991, Bond and option pricing when short rates are lognormal, *Financial Analysts' Journal* 47, 52–59.

Boyle, Phelim P., 1977, Options: A Monte Carlo approach, *Journal of Financial Economics* 4, 323–338.

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

Brennan, Michael J., and Eduardo S. Schwartz, 1982, An equilibrium model of bond pricing and a test of market efficiency, *Journal of Financial and Quantitative Analysis* 17, 210–239.

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

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

Chesney, Marc, Robert J. Elliott, Dilip Madan, and Hailiang Yang, 1993, Diffusion coefficient estimation and asset pricing when risk premia and sensitivities are time varying, *Mathematical Finance* 3, 85–99.

Conley, Timothy G., Lars Peter Hansen, Erzo G. J. Luttmer, and José A. Scheinkman, 1996, Short-term interest rates as subordinated diffusions, Working paper, Northwestern University.

Cox, John C., Jonathan E. Ingersoll, Jr., and Stephen A. Ross, 1981, A re-examination of traditional hypotheses about the term structure of interest rates, *Journal of Finance* 36, 769–799.

Cox, John C., Jonathan E. Ingersoll, Jr., and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–467.

Duffie, Darrell, 1988, *Security Markets: Stochastic Models* (Academic Press, Boston, MA).

Duffie, Darrell, and Kenneth J. Singleton, 1993, Simulated moments estimation of Markov models of asset prices, *Econometrica* 61, 929–952.

Epanechnikov, V. A., 1969, Nonparametric estimates of multivariate probability density, *Theory of Probability and Applications* 14, 153–158.

Fama, Eugene F., 1984, Term premiums in bond returns, *Journal of Financial Economics* 13, 529–546.

Fama, Eugene F., and Robert R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680–692.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Feller, William, 1951, Two singular diffusion problems, *Annals of Mathematics* 54, 173–182.

Gallant, A. Ronald, and George E. Tauchen, 1994, Which moments to match?, *Econometric Theory,* forthcoming.

Goldman, Steven, and Paul A. Ruud, 1995, Nonparametric multivariate regression subject to constraint, Working paper, U.C. Berkeley.

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

Hansen, Lars Peter, and José A. Scheinkman, 1995, Back to the future: Generating moment implications for continuous-time Markov processes, *Econometrica* 63, 767–804.

Hille, Einar, and Ralph S. Phillips, 1957, *Functional Analysis and Semigroups* (American Mathematical Society, Providence, RI).

Ho, Thomas S. Y., and Sang Bin Lee, 1986, Term structure movements and pricing interest rate contingent claims, *Journal of Finance* 41, 1011–1029.

Hull, John, and Alan White, 1990, Pricing interest rate derivative securities, *Review of Financial Studies* 3, 573–592.

Ingersoll, Jonathan E., Jr., 1987, *Theory of Financial Decision Making* (Rowman and Littlefield, Totowa, NJ).

Künsch, H. R., 1989, The jackknife and the bootstrap for general stationary observations, *Annals of Statistics* 17, 1217–1241.

Lo, Andrew W., 1988, Maximum likelihood estimation of generalized Ito processes with discretely sampled data, *Econometric Theory* 4, 231–247.

Lo, Andrew W., and Jiang Wang, 1995, Implementing option pricing models when asset returns are predictable, *Journal of Finance* 50, 87–129.

Merton, Robert C., 1992, *Continuous-Time Finance* (Blackwell, Cambridge, MA).

Milshtein, G. N., 1978, A method of second-order accuracy integration of stochastic differential equations, *Theory of Probability and its Applications* 23, 396–401.

Øksendal, Bernt, 1985, *Stochastic Differential Equations: An Introduction with Applications*, third edition (Springer-Verlag, New York, NY).

Pearson, Neil D., and Tong-Sheng Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model, *Journal of Finance* 49, 1279–1304.

Pfann, Gerard A., Peter C. Schotman, and Rolf Tschernig, 1996, Nonlinear interest rate dynamics and implications for the term structure, *Journal of Econometrics* 74, 149–176.

Ronn, Ehud I., and P. Wadhwa, 1995, On the relationship between expected returns and implied volatility of interest rate-dependent securities, Working paper, U.T. Austin.

Scott, David W., 1992, *Multivariate Density Estimation: Theory, Practice and Visualization* (John Wiley, New York, NY).

Siddique, Akhtar R., 1994, Nonparametric estimation of mean and variance and pricing of securities, Working paper, Duke University.

Silverman, B. W., 1986, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London, U.K.).

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

Table I
CIR Drift Approximations

The table shows first, second, and third order Taylor series approximations to the drift of the Cox, Ingersoll and Ross interest rate process,

$$dr_t = \kappa\left(\theta - r_t\right)dt + s\sqrt{r_t}dZ_t,$$

with parameter values $\kappa = 0.5, \theta = 0.07$, and $s = 0.1$. Approximations used are:

$$\begin{aligned} 1: \quad \mu(r_t) &= \tfrac{1}{\Delta}E_t(r_{t+\Delta} - r_t) + O(\Delta), \\ 2: \quad \mu(r_t) &= \tfrac{1}{2\Delta}\left[4E_t\left(r_{t+\Delta} - r_t\right) - E_t\left(r_{t+2\Delta} - r_t\right)\right] + O(\Delta^2), \\ 3: \quad \mu(r_t) &= \tfrac{1}{6\Delta}\left[18E_t\left(r_{t+\Delta} - r_t\right) - 9E_t\left(r_{t+2\Delta} - r_t\right) + 2E_t\left(r_{t+3\Delta} - r_t\right)\right] + O(\Delta^3). \end{aligned}$$

		Approximation to value of drift, $\mu(r_t)$				
	Order of	$\Delta = 0.004$	$\Delta = 0.02$	$\Delta = 0.08$	$\Delta = 1.0$	$\Delta = 5.0$
r_t	Approximation	(daily)	(weekly)	(monthly)	(annual)	(5 year)
0.0100	1	0.0300	0.0299	0.0294	0.0236	0.0110
	2	0.0300	0.0300	0.0300	0.0283	0.0161
	3	0.0300	0.0300	0.0300	0.0295	0.0192
	Limiting Value	0.0300	0.0300	0.0300	0.0300	0.0300
0.0500	1	0.0100	0.0100	0.0098	0.0079	0.0037
	2	0.0100	0.0100	0.0100	0.0094	0.0054
	3	0.0100	0.0100	0.0100	0.0098	0.0064
	Limiting Value	0.0100	0.0100	0.0100	0.0100	0.0100
0.1000	1	−0.0150	−0.0149	−0.0147	−0.0118	−0.0055
	2	−0.0150	−0.0150	−0.0150	−0.0141	−0.0080
	3	−0.0150	−0.0150	−0.0150	−0.0147	−0.0096
	Limiting Value	−0.0150	−0.0150	−0.0150	−0.0150	−0.0150
0.1500	1	−0.0400	−0.0398	−0.0392	−0.0315	−0.0147
	2	−0.0400	−0.0400	−0.0400	−0.0377	−0.0214
	3	−0.0400	−0.0400	−0.0400	−0.0393	−0.0256
	Limiting Value	−0.0400	−0.0400	−0.0400	−0.0400	−0.0400
0.2000	1	−0.0649	−0.0647	−0.0637	−0.0512	−0.0239
	2	−0.0650	−0.0650	−0.0650	−0.0612	−0.0348
	3	−0.0650	−0.0650	−0.0650	−0.0639	−0.0415
	Limiting Value	−0.0650	−0.0650	−0.0650	−0.0650	−0.0650
0.3000	1	−0.1149	−0.1144	−0.1127	−0.0905	−0.0422
	2	−0.1150	−0.1150	−0.1149	−0.1083	−0.0616
	3	−0.1150	−0.1150	−0.1150	−0.1130	−0.0735
	Limiting Value	−0.1150	−0.1150	−0.1150	−0.1150	−0.1150

Table II
Black, Derman and Toy Drift Approximations

The table shows first, second, and third order Taylor series approximations to the drift of the Black, Derman and Toy interest rate process,

$$\begin{aligned} dy_t &= \kappa\,(\theta - y_t)\,dt + s dZ_t, \qquad \text{where} \\ y_t &= \ln r_t, \end{aligned}$$

with parameter values $\kappa = 0.5, \theta = -2.75$, and $s = 0.43$. Approximations used are:

$$\begin{aligned} 1: \quad \mu(r_t) &= \tfrac{1}{\Delta} E_t(r_{t+\Delta} - r_t) + O(\Delta), \\ 2: \quad \mu(r_t) &= \tfrac{1}{2\Delta}\left[4E_t\left(r_{t+\Delta} - r_t\right) - E_t\left(r_{t+2\Delta} - r_t\right)\right] + O(\Delta^2), \\ 3: \quad \mu(r_t) &= \tfrac{1}{6\Delta}\left[18E_t\left(r_{t+\Delta} - r_t\right) - 9E_t\left(r_{t+2\Delta} - r_t\right) + 2E_t\left(r_{t+3\Delta} - r_t\right)\right] + O(\Delta^3). \end{aligned}$$

		Approximation to value of drift, $\mu(r_t)$				
r_t	Order of Approximation	$\Delta = 0.004$ (daily)	$\Delta = 0.02$ (weekly)	$\Delta = 0.08$ (monthly)	$\Delta = 1.0$ (annual)	$\Delta = 5.0$ (5 year)
0.0100	1	0.0102	0.0102	0.0104	0.0120	0.0100
	2	0.0102	0.0102	0.0102	0.0115	0.0141
	3	0.0102	0.0102	0.0102	0.0105	0.0163
	Limiting Value	0.0102	0.0102	0.0102	0.0102	0.0102
0.0500	1	0.0108	0.0107	0.0106	0.0084	0.0037
	2	0.0108	0.0108	0.0108	0.0101	0.0055
	3	0.0108	0.0108	0.0108	0.0106	0.0065
	Limiting Value	0.0108	0.0108	0.0108	0.0108	0.0108
0.1000	1	−0.0131	−0.0131	−0.0130	−0.0111	−0.0055
	2	−0.0131	−0.0131	−0.0131	−0.0130	−0.0080
	3	−0.0131	−0.0131	−0.0131	−0.0134	−0.0095
	Limiting Value	−0.0131	−0.0131	−0.0131	−0.0131	−0.0131
0.1500	1	−0.0500	−0.0498	−0.0487	−0.0363	−0.0150
	2	−0.0501	−0.0501	−0.0500	−0.0450	−0.0220
	3	−0.0501	−0.0501	−0.0501	−0.0479	−0.0264
	Limiting Value	−0.0501	−0.0501	−0.0501	−0.0501	−0.0501
0.2000	1	−0.0954	−0.0947	−0.0923	−0.0646	−0.0246
	2	−0.0956	−0.0956	−0.0954	−0.0819	−0.0363
	3	−0.0956	−0.0956	−0.0956	−0.0886	−0.0437
	Limiting Value	−0.0956	−0.0956	−0.0956	−0.0956	−0.0956
0.3000	1	−0.2037	−0.2019	−0.1954	−0.1269	−0.0441
	2	−0.2042	−0.2041	−0.2036	−0.1649	−0.0653
	3	−0.2042	−0.2042	−0.2041	−0.1813	−0.0788
	Limiting Value	−0.2042	−0.2042	−0.2042	−0.2042	−0.2042

Table III
CIR Diffusion Approximations

The table shows first, second, and third order Taylor series approximations to the diffusion of the Cox, Ingersoll and Ross interest rate process,

$$dr_t = \kappa\left(\theta - r_t\right)dt + s\sqrt{r_t}dZ_t,$$

with parameter values $\kappa = 0.5, \theta = 0.07$, and $s = 0.1$. Approximations used are:

$$\begin{aligned}
1: \quad \sigma(r_t) &= \sqrt{\tfrac{1}{\Delta}\mathrm{var}_t(r_{t+\Delta})} + O(\Delta),\\
2: \quad \sigma(r_t) &= \sqrt{\tfrac{1}{2\Delta}\left[4\mathrm{var}_t\left(r_{t+\Delta}\right) - \mathrm{var}_t\left(r_{t+2\Delta}\right)\right]} + O(\Delta^2),\\
3: \quad \sigma(r_t) &= \sqrt{\tfrac{1}{6\Delta}\left[18\mathrm{var}_t\left(r_{t+\Delta}\right) - 9\mathrm{var}_t\left(r_{t+2\Delta}\right) + 2\mathrm{var}_t\left(r_{t+3\Delta}\right)\right]} + O(\Delta^3).
\end{aligned}$$

		Approximation to value of diffusion, $\sigma(r_t)$				
r_t	Order of Approximation	$\Delta = 0.004$ (daily)	$\Delta = 0.02$ (weekly)	$\Delta = 0.08$ (monthly)	$\Delta = 1.0$ (annual)	$\Delta = 5.0$ (5 year)
0.0100	1	0.0100	0.0101	0.0104	0.0125	0.0110
	2	0.0100	0.0100	0.0100	0.0122	0.0131
	3	0.0100	0.0100	0.0100	0.0115	0.0142
	Limiting Value	0.0100	0.0100	0.0100	0.0100	0.0100
0.0500	1	0.0223	0.0223	0.0220	0.0186	0.0115
	2	0.0224	0.0224	0.0223	0.0209	0.0140
	3	0.0224	0.0224	0.0224	0.0217	0.0154
	Limiting Value	0.0224	0.0224	0.0224	0.0224	0.0224
0.1000	1	0.0316	0.0314	0.0309	0.0242	0.0122
	2	0.0316	0.0316	0.0316	0.0283	0.0150
	3	0.0316	0.0316	0.0316	0.0299	0.0167
	Limiting Value	0.0316	0.0316	0.0316	0.0316	0.0316
0.1500	1	0.0387	0.0385	0.0378	0.0287	0.0128
	2	0.0387	0.0387	0.0387	0.0341	0.0160
	3	0.0387	0.0387	0.0387	0.0363	0.0180
	Limiting Value	0.0387	0.0387	0.0387	0.0387	0.0387
0.2000	1	0.0447	0.0444	0.0436	0.0326	0.0134
	2	0.0447	0.0447	0.0447	0.0390	0.0169
	3	0.0447	0.0447	0.0447	0.0418	0.0191
	Limiting Value	0.0447	0.0447	0.0447	0.0447	0.0447
0.3000	1	0.0547	0.0544	0.0533	0.0392	0.0144
	2	0.0548	0.0548	0.0547	0.0474	0.0185
	3	0.0548	0.0548	0.0548	0.0509	0.0213
	Limiting Value	0.0548	0.0548	0.0548	0.0548	0.0548

Table IV
Black, Derman and Toy Diffusion Approximations

The table shows first, second, and third order Taylor series approximations to the diffusion of the Black, Derman and Toy interest rate process,

$$\begin{aligned} dy_t &= \kappa\left(\theta - y_t\right) dt + s dZ_t, \qquad \text{where} \\ y_t &= \ln r_t, \end{aligned}$$

with parameter values $\kappa = 0.5, \theta = -2.75$, and $s = 0.43$. Approximations used are:

$$\begin{aligned} 1: \quad \sigma(r_t) &= \sqrt{\tfrac{1}{\Delta}\mathrm{var}_t(r_{t+\Delta})} + O(\Delta), \\ 2: \quad \sigma(r_t) &= \sqrt{\tfrac{1}{2\Delta}\left[4\mathrm{var}_t\left(r_{t+\Delta}\right) - \mathrm{var}_t\left(r_{t+2\Delta}\right)\right]} + O(\Delta^2), \\ 3: \quad \sigma(r_t) &= \sqrt{\tfrac{1}{6\Delta}\left[18\mathrm{var}_t\left(r_{t+\Delta}\right) - 9\mathrm{var}_t\left(r_{t+2\Delta}\right) + 2\mathrm{var}_t\left(r_{t+3\Delta}\right)\right]} + O(\Delta^3). \end{aligned}$$

		Approximation to value of diffusion, $\sigma(r_t)$				
r_t	Order of Approximation	$\Delta = 0.004$ (daily)	$\Delta = 0.02$ (weekly)	$\Delta = 0.08$ (monthly)	$\Delta = 1.0$ (annual)	$\Delta = 5.0$ (5 year)
0.0100	1	0.0043	0.0044	0.0046	0.0077	0.0121
	2	0.0043	0.0043	0.0043	0.0037	0.0140
	3	0.0043	0.0043	0.0043	0.0000	0.0146
	Limiting Value	0.0043	0.0043	0.0043	0.0043	0.0043
0.0500	1	0.0215	0.0215	0.0215	0.0206	0.0138
	2	0.0215	0.0215	0.0215	0.0223	0.0168
	3	0.0215	0.0215	0.0215	0.0225	0.0184
	Limiting Value	0.0215	0.0215	0.0215	0.0215	0.0215
0.1000	1	0.0429	0.0427	0.0419	0.0313	0.0146
	2	0.0430	0.0430	0.0429	0.0372	0.0180
	3	0.0430	0.0430	0.0430	0.0398	0.0201
	Limiting Value	0.0430	0.0430	0.0430	0.0430	0.0430
0.1500	1	0.0644	0.0638	0.0618	0.0400	0.0151
	2	0.0645	0.0645	0.0642	0.0493	0.0188
	3	0.0645	0.0645	0.0645	0.0541	0.0211
	Limiting Value	0.0645	0.0645	0.0645	0.0645	0.0645
0.2000	1	0.0858	0.0848	0.0815	0.0477	0.0155
	2	0.0860	0.0860	0.0855	0.0598	0.0194
	3	0.0860	0.0860	0.0859	0.0667	0.0219
	Limiting Value	0.0860	0.0860	0.0860	0.0860	0.0860
0.3000	1	0.1285	0.1267	0.1203	0.0610	0.0160
	2	0.1290	0.1289	0.1277	0.0783	0.0202
	3	0.1290	0.1290	0.1288	0.0888	0.0229
	Limiting Value	0.1290	0.1290	0.1290	0.1290	0.1290

Table V
Effect of Price of Risk on Bond Valuation

The table shows values of zero coupon bonds of different maturities, calculated using Monte Carlo simulation from the following formula for the value of a bond paying \$1 at time T:

$$P_t(T) = E_t\left[e^{-\int_t^T \widehat{r}_u\, du}\right],$$

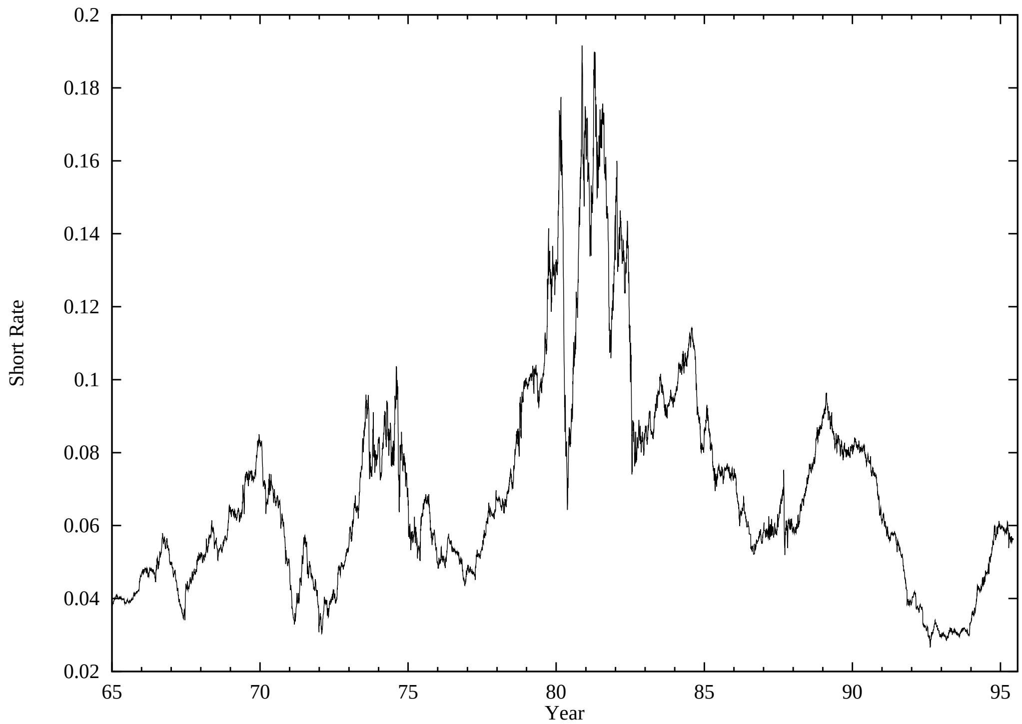
where

$$\begin{aligned} \widehat{r}_t &= r_t, \\ d\widehat{r}_t &= [\mu(\widehat{r}_t) - \lambda(\widehat{r}_t)]\ dt + \sigma(\widehat{r}_t)\, dZ_t. \end{aligned}$$

Values in the first column are calculated imposing $\lambda \equiv 0$. Values in the second column use the estimated function $\lambda(r)$, shown in Figure 6.

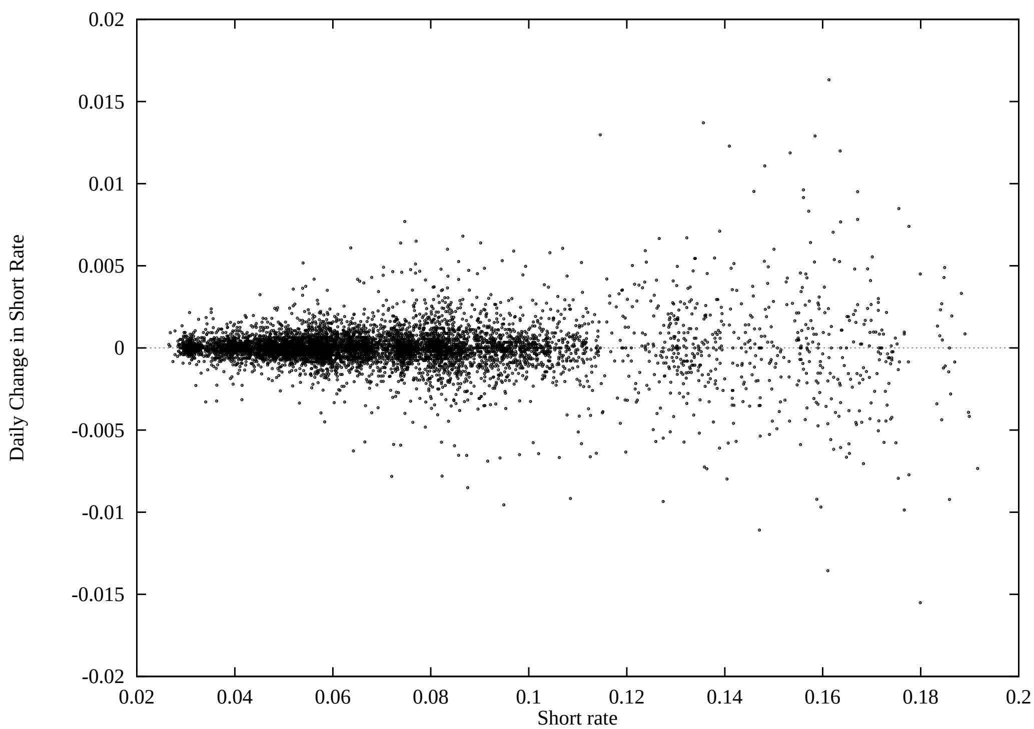
		Value of Bond	
Maturity	r	Zero Price of Risk	Estimated Price of Risk
1 year	1%	0.9885	0.9870
	5%	0.9500	0.9456
2 years	1%	0.9737	0.9668
	5%	0.9001	0.8817
3 years	1%	0.9558	0.9390
	5%	0.8509	0.8115

Figure 1
Three month Treasury Bill rate, January 1965 to July 1995



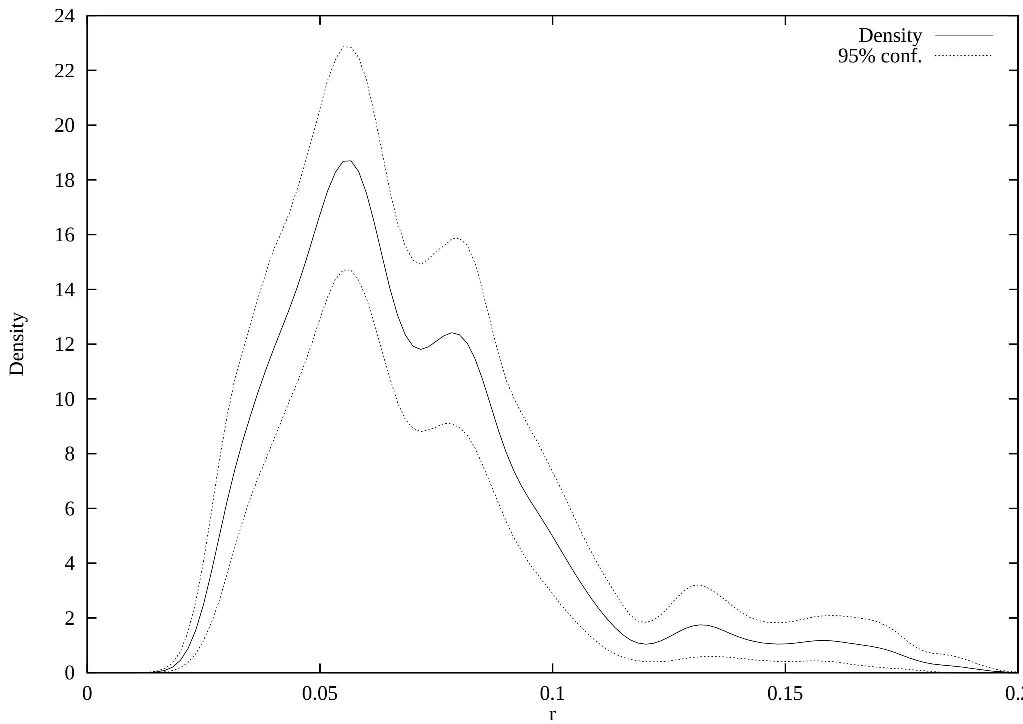
Annualized yield on three month Treasury bills, January 1965 to July 1995.

Figure 2
Daily interest rate changes, January 1965 to July 1995



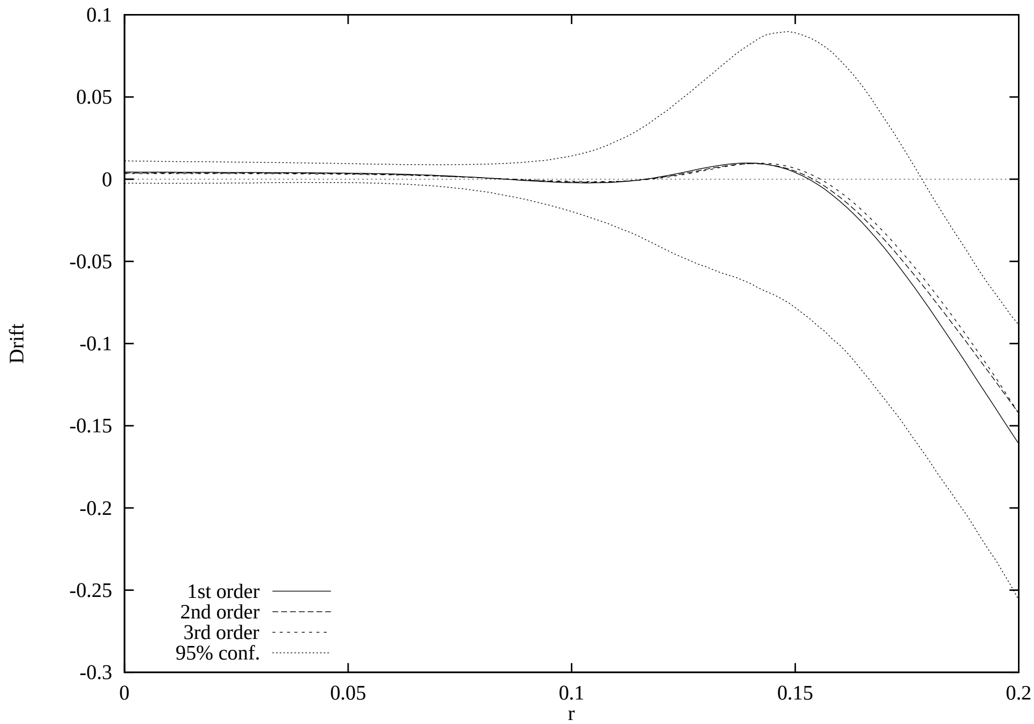
Daily changes in the three month Treasury Bill yield plotted against yield on preceding day.

Figure 3
Estimated marginal density of short rate



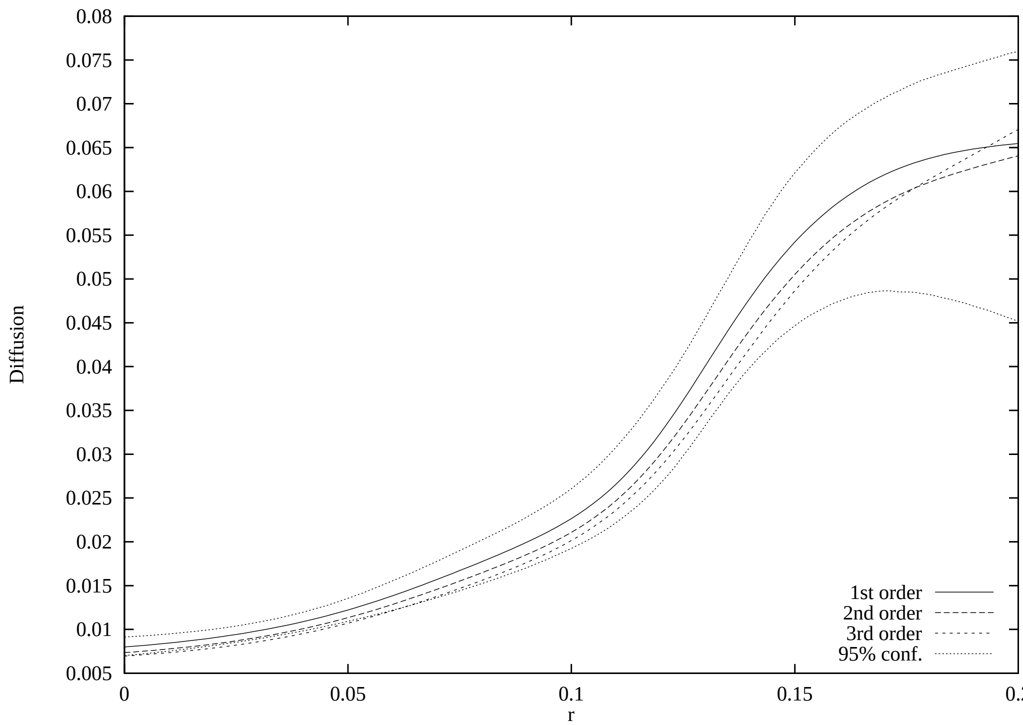
Estimated marginal density of the three month Treasury Bill yield, calculated using daily data, January 1965 to July 1995. Pointwise 95 percent confidence bands are calculated using 10,000 iterations of the Künsch (1989) block bootstrap algorithm.

Figure 4
Estimated drift of short rate process



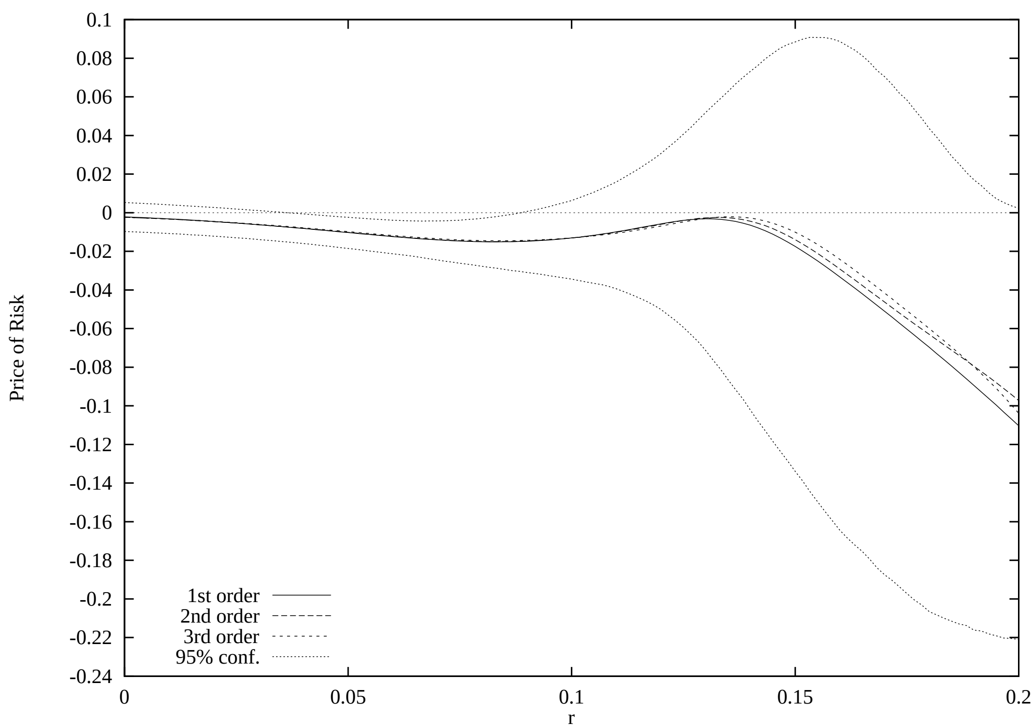
Estimates of first, second, and third order approximations to the drift, $\mu(r)$, are shown, calculated using daily data, January 1965 to July 1995. Pointwise 95 percent confidence bands (for the first order approximation) are calculated using 10,000 iterations of the Künsch (1989) block bootstrap algorithm.

Figure 5
Estimated diffusion of short rate process



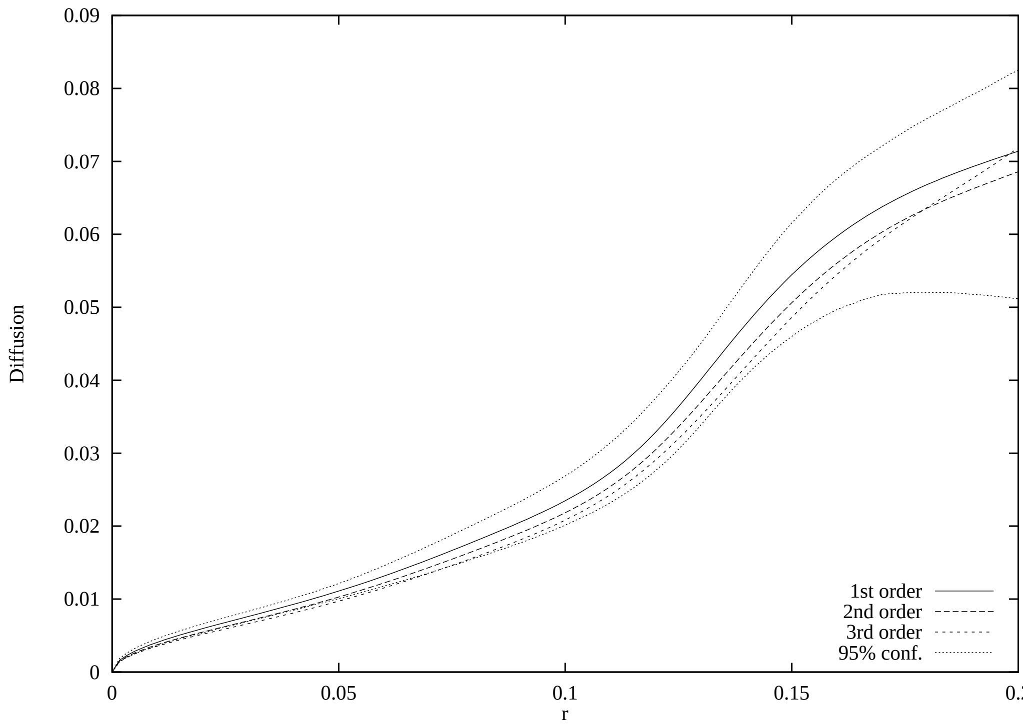
Estimates of first, second, and third order approximations to the diffusion, $\sigma(r)$, are shown, calculated using daily data, January 1965 to July 1995. Pointwise 95 percent confidence bands (for the first order approximation) are calculated using 10,000 iterations of the Künsch (1989) block bootstrap algorithm.

Figure 6
Estimated market price of interest rate risk



Estimates of first, second, and third order approximations to the market price of interest rate risk, $\lambda(r)$, are shown, calculated from daily excess returns on six month Treasury Bills, January 1965 to July 1995. Pointwise 95 percent confidence bands (for the first order approximation) are calculated using 10,000 iterations of the Künsch (1989) block bootstrap algorithm.

Figure 7
Constrained estimate of short rate diffusion



Estimates of first, second, and third order approximations to the diffusion, $\sigma(r)$, are shown, calculated using daily data, January 1965 to July 1995. All estimates are constrained to be zero at $r = 0$. Pointwise 95 percent confidence bands (for the first order approximation) are calculated using 10,000 iterations of the Künsch (1989) block bootstrap algorithm.