

**NBER WORKING PAPER SERIES**

**TESTING CONTINUOUS-TIME MODELS  
OF THE SPOT INTEREST RATE**

**Yacine Aït-Sahalia**

**Working Paper 5346**

**NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
November 1995**

The comments and suggestions of Kerry Back and an anonymous referee were very helpful. I am also grateful to George Constantinides, Lars Hansen, Mahesh Maheswaran, José Scheinkman, Rob Vishny and seminar participants at Boston University, Harvard Business School, MIT, New York University, UCLA, and the Universities of Chicago, Minnesota, and Montréal. Financial support from the Center for Research in Security Prices and the Chicago Board of Trade Research Foundation is gratefully acknowledged. All errors are mine. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

© 1995 by Yacine Aït-Sahalia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

TESTING CONTINUOUS-TIME MODELS  
OF THE SPOT INTEREST RATE

ABSTRACT

Different continuous-time models for interest rates coexist in the literature. We test parametric models by comparing their implied parametric density to the same density estimated nonparametrically. We do not replace the continuous-time model by discrete approximations, even though the data are recorded at discrete intervals. The principal source of rejection of existing models is the strong nonlinearity of the drift. Around its mean, where the drift is essentially zero, the spot rate behaves like a random walk. The drift then mean-reverts strongly when far away from the mean. The volatility is higher when away from the mean.

Yacine Aït-Sahalia  
Graduate School of Business  
University of Chicago  
1101 East 58th Street  
Chicago, IL 60637  
and NBER

The continuous-time financial theory has developed extensive tools to price derivative securities when the underlying traded asset(s) or non-traded factor(s) follow stochastic differential equations. However, as a practical matter, how to specify an appropriate stochastic differential equation is for the most part an unanswered question. For example, many different continuous-time processes have been proposed to capture the dynamics of the instantaneous spot interest rate. These models are generally mutually exclusive and, when used to price derivative securities, generate significantly different prices. This paper develops a simple methodology for testing their specification, and evaluates various models for the short term spot interest rate.

A continuous-time model in finance typically rests on one or more stationary diffusion processes  $\{X_t, t \geq 0\}$ , with dynamics represented by Itô stochastic differential equations:

$$dX_t = \mu(X_t) dt + \sigma(X_t) dW_t \quad (1)$$

where  $\{W_t, t \geq 0\}$  is a standard Brownian motion. The functions  $\mu(\cdot)$  and  $\sigma^2(\cdot)$  are respectively the drift and the diffusion functions of the process. Most often they are parametrized:

$$\mu(x) = \mu(x, \theta) \text{ and } \sigma^2(x) = \sigma^2(x, \theta), \text{ where } \theta \in \Theta \subset \mathbb{R}^K. \quad (2)$$

The choice of the parametric drift and diffusion families is often arbitrary. For example, most interest rate models specify a linear mean-reverting drift,  $\mu(x, \theta) = \beta(\alpha - x)$ , where the spot rate moves in the direction of its mean  $\alpha$  at speed  $\beta$ . The specification of the diffusion function, however, tends to vary widely. The Vasicek (1977) model has  $\sigma^2(x, \theta) = \sigma^2$ , while the Cox-Ingersoll-Ross (1985) model has  $\sigma^2(x, \theta) = \sigma^2 x$  and in the Courtadon (1982) model the diffusion function is  $\sigma^2(x, \theta) = \sigma^2 x^2$ . Chan et al. (1992) utilized a discrete approximation to estimate the constant-elasticity-of-variance model where  $\sigma^2(x, \theta) = \sigma^2 x^\gamma$ . In the Constantinides (1992) model, the resulting specification is of the form  $\mu(x, \theta) = \alpha_0 + \alpha_1 x + \alpha_2 \sqrt{x - \alpha_3}$  and  $\sigma^2(x, \theta) = \sigma_0^2 (x - \alpha_3)^2$ , while Duffie and Kan (1993) specify  $\mu(x, \theta) = \beta(\alpha - x)$  and  $\sigma(x, \theta) = \sqrt{\sigma_0 + \sigma_1 x}$ . Aït-Sahalia (1992b) developed an estimator designed to take into account the discrete character of the data while maintaining the attractiveness of the continuous-time model, and estimate nonparametrically the diffusion function.

This far-from-exhaustive list (see also Table 2) shows that there is no shortage of proposed models for the continuous-time dynamics of the spot rate. We therefore focus on the following question: is it possible to distinguish between the various specifications? Or equivalently: when is a given parametric model (2) a “good” model for short term interest rate movements? For a given parametric model, the null hypothesis will therefore state that there exist parameter values for which the parametric model considered is an acceptable representation of the data. The alternative hypothesis, on the contrary, affirms that no parameter values are capable of reproducing the true  $\mu$  and  $\sigma^2$  functions.

One approach might be to attempt to choose among possible spot rate processes based on how well they price specific liquid derivatives, for example bonds, swaps or caps. Unfortunately this is often a difficult task in practice, and when at all feasible the answer tends to vary with the specific derivatives selected. This approach also implicitly assumes that the market prices these instruments correctly in the first place, so the resulting process can hardly be used to try to detect and exploit market biases in relatively illiquid markets. Furthermore, any such test is a joint test of the specification of the factor process, here the spot rate, and the pricing model --in particular the specification of the market price of spot rate risk. Any rejection is consequently hard to interpret.

It is therefore desirable to be able to test the specification of the actual spot rate process without making use of observations on a collection of derivatives. The test statistic in this paper is based on the following idea. An essential property of a diffusion process is that it is entirely characterized by its first two continuous-time moments, the  $\mu$  and  $\sigma^2$  functions<sup>2</sup>. So each parametric model for the  $\mu$  and  $\sigma^2$  functions, such as the Cox-Ingersoll-Ross model, will imply a certain family of density functions for the interest rate observations. The test statistic is based on a comparison of the density implied by the parametric model and a nonparametric estimator valid even if the parametric model is misspecified.

We empirically test several one-factor diffusion models for the seven-day Eurodollar deposit rate. We then identify the sources of the rejection of existing models. The linearity of the drift imposed in the literature appears to be the main source of misspecification. We find instead that the drift of the spot rate process is essentially zero as long as the rate is between 4 and 17%, but pulls it strongly towards this middle region whenever it escapes. This nonlinear mean-reversion effectively makes the process stationary, even though it is locally non-stationary on most of its support. We also find that the associated diffusion function is lower in the middle region, and higher at both extremes

--as opposed to the uniformly increasing pattern specified by the popular constant-elasticity-of-variance (CEV) model.

The paper is organized as follows. Section 1 presents the test statistic based on the unconditional information and gives its properties. Section 2 incorporates the conditional information contained in the interest rate transitions. Section 3 describes how the basic framework can be extended to regime switching models, multivariate diffusions and diffusions with jumps. Section 4 tests spot rate models proposed in the literature and determines a simple specification that cannot be rejected by the data. Section 5 summarizes the results and concludes. Technical assumptions and proofs are in the Appendix.

## 1. Testing the Parametric Specification of Diffusion Processes

### 1.1 Null Hypothesis and Test Statistics

The main focus of the test is on models for the spot interest rate which belong to the class of (P1) Markov processes of (P2) the univariate diffusion type, which are (P3) strictly stationary. The Markov property P1 is fundamental for what follows. When a model does not yield Markov dynamics, it is sometimes possible to redefine the state space so that the Markov property holds [see e.g., Cox and Miller (1965) pages 18-19]. Property P3 in particular precludes the  $\mu$  and  $\sigma^2$  functions from depending directly on time in addition to being functions of the state  $X^3$ . Strict stationarity means that the joint distributions of the process satisfy for every  $m$ , every  $0 < t_1 < \dots < t_m < \infty$  and every  $x_1, \dots, x_m$  in the state space:  $P(X_{t_1} \leq x_1, \dots, X_{t_m} \leq x_m) = P(X_{t_1+\Delta} \leq x_1, \dots, X_{t_m+\Delta} \leq x_m)$ . In particular, applying this to  $m=1$  and  $2$ , the transition density of the process (the conditional density of  $X_{t+\Delta}$  given  $X_t$ ) can depend on  $\Delta$  but not on  $t$ . And the process admits a stationary marginal density<sup>4</sup>. Section 4 shows how to relax property P2 to cover multivariate diffusions and diffusions with jumps, and P3 to cover regime switching models.

We further restrict attention (property P4) to models for the spot interest rate such that zero and infinity are unattainable in finite expected time, starting from any point in the interior of the state space  $(0, \infty)$ . We finally require (property P5) that classical asymptotic theory be applicable to the discrete data generated by this continuous-time process, i.e., that the discrete data be mixing at a sufficiently fast rate. We give sufficient conditions on the drift and diffusion in the Appendix to insure that the model considered is the unique strong solution of (1) and satisfies P1-P5.

Suppose for now under P1-P5 that the true process of the spot interest rate is the unique strong solution of (1) for given “true” drift and diffusion functions  $(\mu_0(\cdot), \sigma_0^2(\cdot))$ . Consider then a specific parametrization, that is, a joint parametric family:

$$P \equiv \{(\mu(\cdot, \theta), \sigma^2(\cdot, \theta)) / \theta \in \Theta\} \quad (3)$$

where  $\Theta$  is a compact subset of  $\mathbb{R}^k$ . An example is the Cox-Ingersoll-Ross (1985) model where  $P \equiv \{(\mu(x, \theta), \sigma^2(x, \theta)) = (\beta(\alpha - x), \sigma^2 x) / \theta \equiv (\alpha, \beta, \sigma^2) \in \Theta\}$ . The question to be addressed is the following: *are there values of the parameters in  $\Theta$  for which this parametric model is an acceptable representation of the true process*, i.e., do the functions  $(\mu_0(\cdot), \sigma_0^2(\cdot))$  belong to the parametric space  $P$ ? Formally, our null and alternative hypotheses are:

$$\begin{cases} H_0: \exists \theta_0 \in \Theta / \mu(\cdot, \theta_0) = \mu_0(\cdot) \text{ and } \sigma^2(\cdot, \theta_0) = \sigma_0^2(\cdot) \\ H_1: (\mu_0(\cdot), \sigma_0^2(\cdot)) \notin P \end{cases} \quad (4)$$

Except in special cases, it is quite difficult to estimate directly the drift and diffusion functions. While the most common approach in the literature<sup>5</sup>, discretization-based methods encounter many problems: the discrete scheme is only an approximation to the continuous-time model<sup>6</sup>, and in general the first and second conditional moments of the data generated by the model (1)-(2) over discrete time intervals are not given by  $\mu$  and  $\sigma^2$  respectively<sup>7</sup>. The approximation is valid only as the time interval shrinks to zero. It may be difficult to justify for monthly, weekly or even daily observations of financial time series: their intra-day volatility is often substantial. So while the functions  $\mu$  and  $\sigma^2$  may be correctly specified by (2) the discrete estimates obtained under the approximation that daily observations represent a continuous record of observations would fail to reflect this volatility. In light of this problem, one possible option would be to collect data sampled very frequently. However, moving to transaction data, or other forms of high frequency observations, introduces microstructure issues which then have to be addressed: discreteness of price and therefore interest rate changes, price jumping from the bid to the ask, asynchronous trading, etc. The diffusion model may not any longer be an adequate model for these data or, at least, these effects would need to be modeled explicitly.

For these reasons it is difficult to envision a direct way of testing (4). We will be looking for a characterization of the process (1) equivalent to (3), but which relies on quantities that can actually be estimated under the constraints imposed by the discrete data. Under the assumptions made, the  $\mu$  and  $\sigma^2$  functions of the process will characterize uniquely its marginal (or unconditional) and transitional (or conditional) densities over

discrete time intervals. The joint parametrizations for  $\mu$  and  $\sigma^2$  adopted in the literature indeed imply specific forms for the densities of the process. For example, the Ornstein-Uhlenbeck process  $dX_t = \beta(\alpha - X_t)dt + \gamma dW_t$  specified by Vasicek (1977) generates Gaussian marginal and transitional densities. The square-root process  $dX_t = \beta(\alpha - X_t)dt + \gamma X_t^{1/2} dW_t$  used by Cox, Ingersoll and Ross (1985) yields a Gamma marginal and non-central chi-squared transitional densities [Feller (1951)].

More generally, any parametrization  $P$  of  $\mu$  and  $\sigma^2$  corresponds to a parametrization of the marginal and transitional densities:

$$\Pi \equiv \left\{ (\pi(\cdot, \theta), p(\cdot, \cdot | \cdot, \cdot, \theta)) / (\mu(\cdot, \theta), \sigma^2(\cdot, \theta)) \in P, \theta \in \Theta \right\} \quad (5)$$

where  $\pi(x, \theta)$  is the marginal density at  $x$  and  $p(s, y | t, x, \theta)$  the transition probability density from  $x$  at time  $t$  to  $y$  at time  $s$ . This is the conditional density that  $X_s = y$  given  $X_t = x$ . While the direct estimation of  $\mu$  and  $\sigma^2$  with discrete data is problematic, the estimation of the densities explicitly take into account the discreteness of the data. The basic idea of the paper is to use the mapping between the drift and diffusion on the one hand, and the marginal and transitional densities on the other, to test the model's specification.

We start by concentrating on the marginal density. The marginal density corresponding to the pair  $(\mu, \sigma^2)$  is<sup>8</sup>:

$$\pi(x, \theta) = \frac{\xi(\theta)}{\sigma^2(x, \theta)} \exp \left\{ \int_{x_0}^x \frac{2\mu(u, \theta)}{\sigma^2(u, \theta)} du \right\} \quad (6)$$

where the process is distributed on  $D = (\underline{x}, \bar{x})$  with  $-\infty \leq \underline{x} < \bar{x} \leq +\infty$  (for example our interest rate process has  $D = (0, +\infty)$ ). The choice of the lower bound of integration in the interior of  $D$  is irrelevant. It is absorbed in the normalization constant  $\xi(\theta)$  determined to insure that the density integrates to one.

Define  $\Pi_M$  to be the space of density functions of the form (6) corresponding to pairs  $(\mu, \sigma^2)$  in  $P$ ,  $\Pi_M \equiv \left\{ \pi(\cdot, \theta) / (\mu(\cdot, \theta), \sigma^2(\cdot, \theta)) \in P, \theta \in \Theta \right\}$ , and let the true marginal density of the process be  $\pi_0(x) \equiv (\xi_0 / \sigma_0^2(x)) \exp \left\{ \int_{x_0}^x (2\mu_0(u) / \sigma_0^2(u)) du \right\}$ . Formally, we test in this section the null versus alternative hypothesis that:

$$\begin{cases} H_{M0}: \exists \theta_0 \in \Theta / \pi(\cdot, \theta_0) = \pi_0(\cdot) \\ H_{M1}: \pi_0(\cdot) \notin \Pi_M \end{cases} \quad (7)$$

It is necessary that  $H_{M0}$  be true for  $H_0$  to be true. If the true density  $\pi_0(\cdot)$  were known, we could simply check to see if it belonged to the proposed parametric class. Since it is unknown, we must estimate it, and do so with an estimator that does not already assume that the null hypothesis is correct (otherwise there is obviously no way of testing it). We use for that purpose a nonparametric estimator --that is, free of all parametric assumptions regarding  $\mu$  and  $\sigma^2$ -- which will converge to the true density whether or not the parametric model (3) is correctly specified. Now consider a parametric estimator of the implied density  $\pi(\cdot, \theta_0)$ . It will converge to the true density only if the model is correctly specified. Therefore the parametric and nonparametric density estimators should be close together if the parametric model is correct, and far from each other otherwise. A measure of distance  $M$  between the two density estimates provides a natural statistic to test the null hypothesis of correct parametric specification<sup>9</sup>. We suggest to test  $H_0$  vs.  $H_1$  using the distance measure between the densities:

$$M \equiv \min_{\theta \in \Theta} \int_{\bar{x}} (\pi(u, \theta) - \pi_0(u))^2 \pi_0(u) du = \min_{\theta \in \Theta} E[(\pi(X, \theta) - \pi_0(X))^2] \quad (8)$$

The distance  $M$  weighs the difference between the parametric and the nonparametric estimators according to their relevance, putting more weight on values of the process more likely to appear ( $\pi_0(u)$  high). Similarly when the density is very low, that is for interest rate regions not likely to be visited, differences in the true density and the implied parametric density are not penalized heavily (other distance measures could be used as well). From the discussion above, under the null hypothesis  $M$  is small, while it is large under the alternative:

	Under the Null (Parametric Model Correctly Specified)	Under the Alternative (Parametric Model Misspecified)
Parametric Density Estimator	Consistent	Inconsistent
Nonparametric Density Estimator	Consistent	Consistent
Distance (Parametric , Nonparametric)	Approximately zero	Large

The proposed test statistic is the (properly normalized) sample analog<sup>10</sup> of (8):

$$\hat{M} \equiv n h_n \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n (\pi(r_i, \theta) - \hat{\pi}_0(r_i))^2 \quad (9)$$

where  $\hat{\pi}_0(\cdot)$  is a nonparametric estimator of the density function with bandwidth  $h_n$  and  $n h_n$  is a normalizing constant. The null hypothesis is therefore rejected when the test



statistic  $\hat{M}$  is large enough. The test statistic has a minimum-distance flavor. It considers the scenario that is the most favorable to the parametric model, to maximize power, evaluating the distance between the densities at the “best possible” parametric estimator:

$$\hat{\theta}_M \equiv \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n (\pi(r_i, \theta) - \hat{\pi}_0(r_i))^2. \quad (10)$$

The identification of the parameter vector is discussed in the Appendix [see assumption A0]. Before investigating the distribution of the test statistic and its other properties, we next describe the nonparametric estimator  $\hat{\pi}_0(\cdot)$  of the density function.

## 1.2 Nonparametric Estimation of the Density

The data used to form these density estimators consists of discrete observations of the short interest rate  $\{r_1, \dots, r_n\}$  sampled at interval  $\Delta$  (for example  $\Delta=1$  day). These discrete data are used to estimate the continuous density of the process without replacing the continuous-time model with a discrete approximation. The asymptotic properties of the estimators are derived for an expanding sampling period, i.e., the interest rate is observed over an increasingly long period of time. Our method does not require that the sampling interval  $\Delta$  be small: for example weekly or monthly data could be used as well.

The nonparametric kernel estimator of the marginal density is given by:

$$\hat{\pi}_0(u) \equiv \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} K\left(\frac{u - r_i}{h_n}\right) \quad (11)$$

where  $K(\cdot)$  is the kernel function and  $h_n$  the bandwidth. Silverman (1986) provides an introduction to kernel estimation, while Scott (1992) contains more details. One can think of (11) as being a “smooth histogram.” Regularity conditions and restrictions on the choice of the kernel and bandwidth are given in the Appendix.

## 1.3 Distribution of the Test Statistics

It is shown in the Appendix that under  $H_0$ , the estimated parameter  $\hat{\theta}_M$  has the distribution:

$$n^{1/2} \{\hat{\theta}_M - \theta_0\} \xrightarrow{d} N(0, \Omega_M) \quad (12)$$

with  $\Omega_M$  and a consistent estimator given in the Appendix.

The test statistic  $\hat{M}$  is distributed as:

$$h_n^{-1/2} \{\hat{M} - E_M\} \xrightarrow{d} N(0, V_M) \quad (13)$$

with  $E_M$ ,  $V_M$  given in the Appendix.

To test the null hypothesis at the level  $\alpha$ , we would therefore use the critical region:

$$\text{Reject } H_0 \text{ when } \hat{M} \geq \hat{c}(\alpha) \equiv \hat{E}_M + h_n^{1/2} z_{1-\alpha} / \left\{ \hat{V}_M \right\}^{1/2} \quad (14)$$

where consistent estimators of  $E_M$  and  $V_M$  are given by:

$$\begin{cases} \hat{E}_M \equiv \left( \int_{-\infty}^{+\infty} K^2(x) dx \right) \left( \frac{1}{n} \sum_{i=1}^n \hat{\pi}_0(r_i) \right) \\ \hat{V}_M \equiv 2 \left( \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} K(u)K(u+x) du \right\}^2 dx \right) \left( \frac{1}{n} \sum_{i=1}^n \hat{\pi}_0^3(r_i) \right) \end{cases} \quad (15)$$

This test is locally strictly unbiased. Any other consistent estimates of  $E_M$  and  $V_M$  could be used. Set  $z_{1-\alpha} = 1.96$  to obtain a test at the 95% level. For the Gaussian kernel  $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$ , the two kernel constants are  $\left( \int_{-\infty}^{+\infty} K^2(x) dx \right) = 1/(2\sqrt{\pi})$  and  $\left( \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} K(u)K(u+x) du \right\}^2 dx \right) = 1/(2\sqrt{2\pi})$ . The optimal bandwidth can be set in the form  $h_n = c_n n^{-1/5}$ , where  $c_n = c \left[ \sum_{i=1}^n (r_i - \bar{r})^2 / n \right]^{1/2}$ . The constant  $c$  is chosen by cross-validation to minimize the integrated square error of the estimator. This completes the description of the marginal-based test.

## 2. Exploiting the Information Contained in the Transitions of the Process

The stationary marginal density of the process does not summarize all the information available in the data. For our specification test, it is clear that a rejection based on the marginal density ( $\hat{M}$  large) already invalidates the candidate parametrization of the drift and the diffusion, and we need not go any further. While the empirical results of this paper suggest that it is unlikely in practice, we could possibly be unable to reject an inadequate parametrization based on the marginal density alone. For that reason, it is worthwhile to examine the transitions of the process over finite intervals of time, corresponding to the sampled data (say, daily). Hansen and Scheinkman (1995) proposed to estimate diffusion processes based on moment conditions which must be satisfied over time intervals by arbitrary functions. Their method, along with an extension of the method of Aït-Sahalia (1992b), is implemented empirically by Conley et al. (1995). We treat the Markovian character of the process as a maintained hypothesis, so there is no need to look at transitions over time intervals longer than the sampled frequency of the data (say, two-day when daily data are available)<sup>11</sup>.

We propose to exploit stationarity by combining the forward and backward equations characterizing the transitional density  $p_0(s, y | t, x)$  in a way that eliminates the time derivatives of  $p_0$  --which are unobservable with discrete data. Consider the forward Kolmogorov equation, with the natural initial condition ( $x$  and  $t$  are fixed)<sup>12</sup>:

$$\frac{\partial p_0(s, y | t, x)}{\partial s} = -\frac{\partial}{\partial y}(\mu(y, \theta)p_0(s, y | t, x)) + \frac{1}{2} \frac{\partial^2}{\partial y^2}(\sigma^2(y, \theta)p_0(s, y | t, x)) \quad (16)$$

for all  $y$  in the interior of the domain of the diffusion  $D=(\underline{x}, \bar{x})$ , and  $s$  such that  $s > t$ . The backward equation ( $y$  and  $s$  are fixed) is<sup>13</sup>:

$$-\frac{\partial p_0(s, y | t, x)}{\partial t} = \mu(x, \theta) \frac{\partial}{\partial x}(p_0(s, y | t, x)) + \frac{1}{2} \sigma^2(x, \theta) \frac{\partial^2}{\partial x^2}(p_0(s, y | t, x)). \quad (17)$$

for all  $x$  in  $D$  and  $t$  such that  $0 \leq t < s$ .

Unfortunately, these two equations cannot be used as such to estimate the parameters because their left-hand-side contains the derivative of the transition density with respect to time. There is clearly no way to estimate this derivative without observations on interest rate changes over small intervals of time. The idea here is to work around this problem by getting rid of the time derivatives with the following trick. By stationarity:  $p_0(s, y | t, x) = p_0(s-t, y | 0, x) \equiv p_0(s-t, y | x)$ , and therefore:  $\partial p_0 / \partial s = -\partial p_0 / \partial t$ . Combining the two equations then yields (with  $\Delta \equiv s-t$ ) what we might call the "transition discrepancy":

$$\begin{aligned} m(\Delta, y, x, \theta) \equiv & \left\{ \frac{1}{2} \frac{\partial^2}{\partial y^2}(\sigma^2(y, \theta)p_0(\Delta, y | x)) - \frac{\partial}{\partial y}(\mu(y, \theta)p_0(\Delta, y | x)) \right\} \\ & - \left\{ \mu(x, \theta) \frac{\partial}{\partial x}(p_0(\Delta, y | x)) + \frac{1}{2} \sigma^2(x, \theta) \frac{\partial^2}{\partial x^2}(p_0(\Delta, y | x)) \right\} \end{aligned} \quad (18)$$

which, for every  $(x, y)$  in  $D^2$  and  $\Delta > 0$ , must be zero at  $\theta = \theta_0$  under the null of correct parametric specification<sup>14</sup>.

The essential fact is that  $m(\Delta, y, x, \theta_0) = 0$  must hold for every time interval  $\Delta$ , not just small ones. Let  $\Delta$  be the time interval between successive observations in the data  $(r_i, r_{i+1})$ ,  $i=1, \dots, n$  (for example  $\Delta=1$  day). A simple consistent method to estimate  $\theta_0$  based on the transition data for intervals of length  $\Delta$  is to solve:

$$T \equiv \min_{\theta \in \Theta} \int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} \{m(\Delta, y, x, \theta)\}^2 p_0(\Delta, y, x) dx dy. \quad (19)$$

where  $p_0(\Delta, y, x)$  is the joint density of observations separated by a time interval  $\Delta$  (while  $p_0(\Delta, y | x)$  is the conditional density).

The normalized sample analog is our transition-based test statistic:

$$\hat{T} \equiv nh_n^6 \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \left\{ \hat{m}(\Delta, r_{i+1}, r_i, \theta) \right\}^2 \quad (20)$$

where  $\hat{m}(\Delta, y, x, \theta)$  is analogous to (18) except that the unknown joint and transition densities have been replaced by their kernel estimators:

$$\begin{cases} \hat{p}_0(\Delta, y, x) = \left\{ \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n^2} \mathbf{K}\left(\frac{x - r_i}{h_n}\right) \mathbf{K}\left(\frac{y - r_{i+1}}{h_n}\right) \right\} \\ \hat{p}_0(\Delta, y | x) = \hat{p}_0(\Delta, y, x) / \hat{\pi}_0(x) \end{cases} \quad (21)$$

The asymptotic distribution of the test statistic based on the transition densities is:

$$h_n^{-1} \left\{ \hat{T} - E_T \right\} \xrightarrow{d} N(0, V_T) \quad (22)$$

where  $E_T$ ,  $V_T$  and their consistent estimators are given in the Appendix.

The transition-based test can be derived as in (14). Additionally, it can be noted that  $\hat{\theta}_T \equiv \arg \min_{\theta \in \Theta} \hat{T}$  provides a consistent (under the null of correct specification) and asymptotically normal estimator of the drift and diffusion parameters.

### 3. Extensions

Recent papers in the term structure literature have modeled discrete regime shifts in the spot interest rate process [e.g., Hamilton (1988), Naik-Lee (1993), Gray (1994) and Das (1994)]. These models typically posit a spot interest rate process which can shift randomly between two or more regimes: for example, a “normal” low-mean regime and an exceptional high-mean regime [e.g., 1980-82 in the US]. They typically keep the same  $\mu$  and  $\sigma^2$  functions, but with different parameter values for each regime. This makes the process time-inhomogenous, and therefore violates property P3. The rigorous estimation of such processes is difficult.

A simpler modeling alternative can be implemented to capture some of the same economic effects, without introducing different drift and diffusion functions. A process switching randomly between say two possible regimes will have a bimodal marginal density, with: (i) the relative height of each mode reflecting the amount of time spent by the process in each regime, and (ii) the relative peakness of each mode reflecting the volatility of the corresponding regime.

Consider the typical configuration in Figures 1-abc for a basic diffusion of the type (1). The density in Figure 1a is bimodal. Instead of two separate pairs (drift,diffusion)

with linear drift for each regime, the regime-switching effect can be modeled by a diffusion process with a “two-regime potential” drift function (Figure 1b). The type of density in Figure 1a would be the product of a kernel estimator with a bandwidth set lower than is optimal to artificially emphasize the multimodality (i.e., the multiplicity of regimes). The specification of the diffusion function in Figure 1c is a simple CEV function, for illustrative purposes. This specification of this model is as readily testable as any other diffusion of the type (1).

In other words, instead of two separate linear drift functions which results in a discontinuous global drift, we could specify a single continuous but nonlinear drift function which preserves time-homogeneity. The mean-time spent in each regime is easily computed, as well as the probability of switching from one regime to the other [see Gardiner (1990), 9.1]. These can then be used to calibrate the model to the empirical observations on regime switches.

We finally note that the test can be extended to cover multivariate diffusions and jump-diffusion models. Indeed both classes of models are characterized by forward and backward equations that are straightforward extensions of (16)-(17). The extension to jump-diffusion presents no fundamental difficulty. However, the extension to multivariate diffusions is likely to be more difficult to implement in practice due to the “curse of dimensionality.” This refers to the fact that nonparametric kernel estimators converge at a speed which decreases as the dimensionality of the problem increases, in contrast to parametric estimators which always converge at speed root-n. Estimating nonparametrically higher-dimensional densities therefore requires larger sample sizes.

## **4. Empirical Tests of Spot Rate Models**

### **4.1 The Data**

The short term rate used is the seven-day Eurodollar deposit spot rate, bid-ask midpoint. The  $n=5505$  observations are daily from 6/1/1973 to 2/25/1995. Weekends and holidays have not been treated specifically (Monday is taken as the next day after Friday). Whereas weekend effects have been documented for stock prices [e.g., French and Roll (1986)], there does not seem to be a conclusive weekend effect in money market instruments<sup>15</sup>. Table 1 reports the descriptive statistics. Dickey-Fuller nonstationarity tests have been conducted, and the presence of a unit root is rejected, both for the full period and the subperiod 1983-94. Since the test is known to have low power, even a slight rejection means that existence of a unit root is unlikely.

A time-series plot of the data is provided in Figure 2. Due to the shift in monetary policy, the 1980-82 years are characterized by substantially higher interest rate levels than the rest of the sample period. There exists evidence<sup>16</sup> that the parameters of a linear drift-CEV diffusion are different during the 1980-82 than the rest of the sample [e.g., Bliss and Smith (1994)]. More precisely, the evidence means that this particular parametrization is not time-homogenous. For example, the mean  $\alpha$  of the process with drift  $\mu(r, \theta) = \beta(\alpha - r)$  estimated over 1980-82 is significantly higher than the mean estimated on the rest of the sample. As we shall see, a better functional dependence of the drift on the spot rate --with no time-varying parameters-- can capture the dynamics of the process since the shift is associated with higher interest rate levels. Misspecified models, like linear drift with CEV diffusion, will mask nonlinearities as changes in parameters.

Two additional series have also been used to check the robustness of the results. The other proxies for the spot rate are the overnight yield for Federal Funds, computed daily by the Federal Reserve Bank of New York [source: Telerate Systems Inc.], and the one-month continuously-compounded bond-equivalent yield on Treasury bills [source: Bank of America]. All three series produce qualitatively similar results, with the provision that one-month rates tend to exhibit substantially less short-term volatility than one week rates, which in turn are more stable than overnight rates<sup>17</sup>.

Choosing a seven-day rate --such as the seven-day Eurodollar-- as the underlying factor for pricing derivatives is a necessary compromise between: (i) literally taking an “instantaneous” rate that does not already embed the result of a pricing operator and (ii) avoiding some of the spurious microstructure effects associated with overnight rates. For example, the second Wednesday settlement effect in the Federal Funds market creates a spike in the raw Federal Funds data that has to be smoothed [see Hamilton (1994)]. More generally short-term supply and demand effects in overnight markets can create excess volatility at the short end of the yield curve that is often irrelevant for the rest of the curve.

## 4.2 Parametric Specifications of the Short Rate Process

The drift-diffusion parametrization considered for the empirical test is:

$$\begin{cases} \mu(r, \theta) = \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3 / r \\ \sigma^2(r, \theta) = \beta_0 + \beta_1 r + \beta_2 r^{\beta_3} \end{cases} \quad (23)$$

The parametrization (23) nests the single-factor diffusion models considered in Table 2. This general parametric model allows for departures from linearity in the drift, and at the same time sufficient variability in the shapes of the diffusion function. The parameter vector is:  $\theta \equiv (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2, \beta_3)$ . The implied marginal density is given by (6).

Natural restrictions have to be imposed on the parameter values to satisfy requirements P1-P5 from Section 2, and are derived in the Appendix. They are:

$$\beta_0 \geq 0 \text{ (and } \beta_2 > 0 \text{ if } \beta_0 = 0 \text{ and } 0 < \beta_3 < 1, \text{ or } \beta_1 > 0 \text{ if } \beta_0 = 0 \text{ and } \beta_3 > 1) \quad (24a)$$

$$\beta_2 > 0 \text{ if either } \beta_3 > 1 \text{ or } \beta_1 = 0, \text{ and } \beta_1 > 0 \text{ if either } 0 < \beta_3 < 1 \text{ or } \beta_2 = 0 \quad (24b)$$

$$\alpha_2 \leq 0 \text{ and } \alpha_1 < 0 \text{ if } \alpha_2 = 0 \quad (24c)$$

$$\alpha_3 > 0 \text{ and } 2\alpha_3 \geq \beta_0 \geq 0, \text{ or } \alpha_3 = 0, \alpha_0 > 0, \beta_0 = 0, \beta_3 > 1 \text{ and } 2\alpha_0 \geq \beta_1 > 0 \quad (24d)$$

Restriction (24a) (resp. (24b)) is necessary for  $\sigma^2$  to be positive in the neighborhood of the zero (resp. infinity) boundary. (24c) insures that the drift is mean-reverting at high interest rate values. It is sufficient here to make infinity unattainable. (24d) guarantees that zero is unattainable.

### 4.3 Results of the Specification Test

The results of the test are in Table 2. First, models that are strictly more parsimonious than (23) are rejected even by the marginal-density-based test. Second, the estimated  $\mu$  and  $\sigma^2$  functions of the spot rate have features that had not been previously documented. The ability to reject non-trivial parametric models such as linear drift with CEV diffusion against a nonparametric alternative suggests that the power of the test is significant at the sample sizes considered.

Figures 3-abc correspond to the model obtained under the restriction that the drift be linear mean-reverting ( $\alpha_2 = \alpha_3 = 0$ ) with a CEV diffusion ( $\beta_0 = \beta_1 = 0$ ) under (24). This model is rejected at the 95% level. The rejection is not surprising given the nonparametric shape of the diffusion found in Aït-Sahalia (1992b) when the drift is linear. A CEV diffusion function constrains the volatility to be uniformly increasing. In a less constrained model the diffusion function is found to be non-monotonic, whether the drift is linear [Aït-Sahalia (1992b)] or not (Figure 4c).

We find strong evidence in the data that CEV diffusions are not a perfect match for linear drifts. There are two reasons for this. First the behavior of the diffusion for large values of the spot rate constrains the parameter  $\beta_3$ . The shape of the marginal density requires a relatively low rate of growth of the diffusion as the interest rate grows ( $\beta_3$  small), because of the slow mean-reversion effect imposed by a linear drift. A rapidly increasing diffusion function would generate a probability of large values, given the linearity of the drift, that would be higher than observed in the data. The second constraint is due to the unattainability of zero. This on the contrary imposes a fast rate of decay of the diffusion as interest rates become close to zero ( $\beta_3$  large). The CEV specification obviously has  $\beta_3$  constant, and these two effects are incompatible in the data.

This incompatibility is exemplified by the comparison of the Vasicek, Cox-Ingersoll-Ross and Courtadon models. In the Vasicek model ( $\beta_3=0$ ), zero is attainable, and the model is easily rejected but by a smaller margin than the Courtadon model ( $\beta_3=2$ ). The best fit is provided by the Cox-Ingersoll-Ross model ( $\beta_3=1$ ), which is a compromise between these two competing effects.

#### 4.4 A Nonlinear Spot Rate Model

The empirical evidence so far suggests that misspecification of the models in the literature is caused jointly by the linearity of the drift and the CEV specification of the diffusion. We now relax these two assumptions and examine the features of the resulting model. Figures 4-abc plot respectively the density ( $\pi$ ), drift ( $\mu$ ) and diffusion ( $\sigma^2$ ) functions corresponding to the most parsimonious model that cannot be rejected at the 95% level. The drift and diffusion are as (23), unconstrained other than (24). The first interesting feature of Figure 4b is that in the middle region of the rate spectrum (4 to 17%) the drift is essentially zero. In this region, the spot rate is locally nonstationary as it behaves like a random walk, with no drift and a non-zero diffusion component. This explains why it is difficult, on relatively short time periods, to reject the existence of a unit root in time series of spot rates. This finding is consistent with the insights of the literature on vector autoregressions in macroeconomics where it is common to model high frequency interest rate data as nonstationary [e.g., Stock and Watson (1993) page 800].

However the second finding is that the nonlinear mean-reverting drift pulls the rate back strongly into this middle region whenever it wanders outside (for example below 4% in 1993-early 1994 or above 17% in 1981-82). This makes the process globally stationary. Overall, the process estimated from the data is very similar to the target band model used in the exchange rate literature: a Brownian motion with two reflecting barriers. We would obviously expect a different result in a hyperinflationary economy.

The diffusion function in Figure 4c is larger when far away from the central region 4-17%. It has long been thought that the volatility of the spot rate should increase with the level of the rate: beyond avoiding negative nominal rates, this was one of the rationales for the Cox-Ingersoll-Ross specification compared to Vasicek's. The finding here is that the spot rate is more volatile --albeit less so-- also when below its long-term mean. This documents yet another aspect of the process. Loosely speaking, markets may become more nervous outside the central region, at both ends of the interest rate spectrum. Market participants may expect the Federal Reserve to credibly return the short-term interest rate to its middle range at some point, but are uncertain about the precise timing of the



intervention. This rather speculative interpretation would also be consistent with the strong pull-back drift of Figure 4b at either end of the spectrum.

We also report results for the 1983-95 subsample (Figures 5-abc). The shapes of the drift and diffusion functions are remarkably similar to the full 1973-95 sample, although shifted to the left by the absence of high interest rates observed during this period. The parameter estimates corresponding to all these configurations are in Table 3.

#### 4.5. Transition Densities

Figure 6a reports the nonparametric estimate of the joint density of the spot rate at one day intervals --the basic input for the method proposed in Section 2. The joint density brings some interesting information regarding the interest rate process. Most of the mass is around the 45-degree line, since most day-to-day changes are small. Intuitively, the strength of the drift can then be read from the steepness of the off-diagonal slopes. The unconstrained model (23) is not rejected by the transition-based test. To see how close the answer from the two tests can be, Figure 6b reports the transition discrepancy  $(r_t, r_{t+\Delta}) \mapsto \hat{m}(\Delta, r_t, r_{t+\Delta}, \hat{\theta}_M)$  defined in (18)-(20) and evaluated under the unconstrained parametrization (23)-(24) at the parameter values (10) estimated in the course of running the marginal density test. Figure 6c plots the transition discrepancy for the linear drift - CEV diffusion model, also evaluated at the parameters estimated from the marginal test. From (20) and (22), large discrepancies lead to rejection of a given model. We find that the magnitude of the discrepancy in Figure 6c is significantly larger than that of Figure 6b. Like the marginal-based test, the transition-based test does reject the linear drift-CEV diffusion model, but not the general parametric model.

#### 4.6. A Discrete Look at Nonlinearities in the Drift

Figures 7-ab show estimates of the  $\mu$  and  $\sigma^2$  functions obtained from discretizing the continuous-time moments in (23). Feasible generalized least squares is a two-stage estimator, where residuals are computed from a first-stage ordinary least squares regression of the spot rate change. The specification of the regression is the crude discretization of the drift in (23):  $E[r_{t+1} - r_t | r_t] = \alpha_0 + (\alpha_1 - 1)r_t + \alpha_2 r_t^2 + \alpha_3 / r_t$ . The squared residuals  $\varepsilon_{t+1}^2$  from this first-stage regression are then regressed by nonlinear least squares, with a discretization of the diffusion in (23):  $E[\varepsilon_{t+1}^2 | r_t] = \beta_0 + \beta_1 r_t + \beta_2 r_t^3$ .

The second stage regression for the drift uses the fitted values from the diffusion regression to form the weighting matrix for the generalized least squares estimation of the discretized drift. The resulting FGLS drift parameters are in Figure 7a, along with a scatter

plot of the actual spot rate changes. To avoid cluttering the graphs, each point in Figures 7-abc represents 50 daily observations. Figure 7b reports the fitted values from the diffusion equation next to a scatter plot of the first-stage squared residuals. Figure 7c reports the first-stage residuals scaled by the fitted standard errors ( $\epsilon_{t+1} / \sqrt{\hat{\beta}_0 + \hat{\beta}_1 r_t + \hat{\beta}_2 r_t^3}$ ). This plot shows that the heteroskedasticity apparent in Figure 7b is removed by the FGLS procedure: the scaled residuals in Figure 7c are essentially white noise. A Chow test to detect parameter shifts between the 1980-82 period and the rest of the sample does not reject the null hypothesis of parameter equality. The same test for the linear drift - CEV diffusion model rejects the null, emphasizing again the misspecification of this parametrization.

This FGLS estimation would be asymptotically efficient if the discretized moments were exact. But this is not the case with a crude discretization, and the difference introduces a discretization bias --likely to be less severe with daily data than monthly data. In general, there are no closed-form expressions available for the moments of  $r_{t+1}$  given  $r_t$  so no better discretization can be proposed<sup>18</sup>. For this reason, this section is provided only for comparison purposes with the other results in the literature [e.g., Chan et al. (1992)]. The test of the continuous time model with discrete data test in this paper, and the nonparametric estimation of the diffusion in Aït-Sahalia (1992b), both take explicitly into account the effect of estimating a continuous-time model sampled at a discrete frequency and do not suffer from discretization bias.

## 5. Conclusions

Stochastic differential equations are a powerful tool in the hands of finance theorists. But it is often hard to justify on empirical grounds the particular specifications being adopted in any given model. This paper provides a first step in that direction. Many other processes of interest in finance, for example stock returns or exchange rates, could be subjected to the same type of inquiry.

We test the specification of parametric interest rate models by examining how closely the parametric model can reproduce features of its nonparametric counterpart. We examine the sources of the rejection of the models tested. We propose a simple model that captures well the dynamics of the spot rate. This model has the appealing feature of being a random walk for middle values of the spot rate, while still being globally stationary. The paper provides evidence of nonlinear effects in the dynamics of the spot interest rate.

The results in this paper can be generalized. For example, certain situations may dictate the choice of the distance metric between the parametric and nonparametric densities. We may be particularly interested in how the model performs in the tails of the distribution. Then the class of distance metrics  $\min_{\theta \in \Theta} \int_{\underline{x}}^{\bar{x}} (\pi(u, \theta) - \pi_0(u))^2 \omega_0(u) du$  can be used where the weighting function  $\omega_0$  is chosen to emphasize the tails (as opposed to  $\omega_0 = \pi_0$  which has the opposite effect). As another example, if the current spot rate level is  $r_t$  and we are interested in the performance of the model over the next few weeks, it may be natural to minimize the distance  $\min_{\theta \in \Theta} \int_{\underline{x}}^{\bar{x}} \{m(\Delta, y, r_t, \theta)\}^2 p_0(\Delta, y, r_t) dy$ , i.e., conditioning on the current value.

## Appendix

### Assumptions

*Assumption A0:* The parameter space  $\Theta \subset \mathbb{R}^K$  is compact. In a neighborhood of the true parameter  $\theta_0$ ,  $\pi(x, \theta)$  is twice continuously differentiable in  $\theta$ ,  $E[(\partial\pi(x, \theta)/\partial\theta)(\partial\pi(x, \theta)/\partial\theta')]$  has full rank and  $\partial^2\pi(x, \theta)/\partial\theta_i\partial\theta_j$  is bounded in absolute value for all  $\theta$  in  $\Theta$ ,  $x$  in  $D$ ,  $i$  and  $j$ .

This assumption insures the identifiability of  $\theta_0$  in the parametric class of densities considered. The relative scale of  $\mu$  and  $\sigma^2$  is arbitrary and  $\theta_0$  is defined to incorporate only the identified parameters. For identification purposes,  $\beta_3$  can be neither 0 (unless  $\beta_0=0$ ) nor 1 (unless  $\beta_1=0$ ).

We first impose constraints on the signs of the parameters to insure that:  $\sigma^2(\cdot, \theta) > 0$  on  $D$  and  $\lim_{x \rightarrow 0} \mu(x, \theta) > 0$ ,  $\lim_{x \rightarrow +\infty} \mu(x, \theta) < 0$ . In particular for model (23) it is necessary that: (i)  $\alpha_2 \leq 0$  (and  $\alpha_1 < 0$  if  $\alpha_2 = 0$ ); (ii)  $\alpha_3 \geq 0$  (and  $\alpha_0 > 0$  if  $\alpha_3 = 0$ ); (iii)  $\beta_0 \geq 0$  (and  $\beta_2 > 0$  if  $\beta_0 = 0$  and  $0 < \beta_3 < 1$ , or  $\beta_1 > 0$  if  $\beta_0 = 0$  and  $\beta_3 > 1$ ); (iv)  $\beta_2 > 0$  if  $\beta_3 > 1$  or  $\beta_1 = 0$ ; (v)  $\beta_1 > 0$  if  $0 < \beta_3 < 1$  or  $\beta_2 = 0$ .

*Assumption A1:* For every  $\theta$  in  $\Theta$ :

(i) The drift and the diffusion functions are twice continuously differentiable in  $x$  on  $D=(0, \infty)$ , and  $\sigma^2 > 0$  on  $(0, \infty)$ .

(ii) The integral of  $m(v, \theta) \equiv (1/\sigma^2(v, \theta)) \exp\left\{-\int_v^{\bar{e}} [2\mu(u, \theta)/\sigma^2(u, \theta)] du\right\}$ , the speed measure  $M$ , converges at both boundaries of  $D$ .

(iii) The integral of  $s(v, \theta) \equiv \exp\left\{\int_v^{\bar{e}} [2\mu(u, \theta)/\sigma^2(u, \theta)] du\right\}$ , the scale measure  $S$ , diverges at both boundaries of  $D$ .

In A1(ii)-(iii),  $\bar{e}$  is fixed in  $D$  and its particular choice is irrelevant. A1(i)-(ii) guarantees that stochastic differential equation (1) admits a unique strong solution, which is a Markov process with time-homogenous transition densities and a stationary marginal density (note that (6) is proportional to the speed measure). A1(i)-(ii) is considerably less restrictive than the *global* Lipschitz and growth conditions usually imposed on  $\mu$  and  $\sigma^2$  to obtain the existence and unicity of a solution. A1(i) implies *local* Lipschitz and growth properties (on every compact in  $D$ ) by the mean-value theorem. The local Lipschitz condition is sufficient to insure strong uniqueness of the solution. For example in (23) any

of the nonlinear  $\mu$  models ( $\alpha_2 < 0$  and/or  $\alpha_3 > 0$ ) or nonlinear  $\sigma^2$  (with  $\beta_3 > 1$ ) fail the global growth condition.

A1(iii) guarantees that starting from any point in the interior of the state space  $D=(0,\infty)$ , the boundaries  $\underline{x} = 0$  and  $\bar{x} = +\infty$  cannot be attained in finite expected time. If in addition we were to assume (iv): starting from either of the boundaries  $\underline{x} = 0$  and  $\bar{x} = +\infty$ , an interior point in  $D=(0,\infty)$  is reached in finite expected time, then the boundaries are entrance boundaries. Otherwise, the process cannot restart from the boundaries and the boundaries are natural<sup>19</sup> (see Feller's classification of boundaries in Karlin and Taylor (1981), Lemma 15.6.3 and Table 15.6.2). How the interest rate process would restart from the boundaries is not particularly relevant from the economic standpoint, and we therefore allow both type of behaviors.

We now derive the parameter restrictions imposed by A1 on model (23). We first examine the boundary behavior of the speed and scale measures near zero. When  $\alpha_3 > 0$  and  $\beta_0 > 0$ ,  $\mu(u,\theta)/\sigma^2(u,\theta) \propto (1/\beta_0)(\alpha_3/u)$  so  $m(v,\theta) \propto (1/\beta_0)v^{2\alpha_3/\beta_0}$  and  $s(v,\theta) \propto (1/v)^{2\alpha_3/\beta_0}$ . Therefore in this case we require  $2\alpha_3 \geq \beta_0 > 0$ . When  $\alpha_3 > 0$  and  $\beta_0 = 0$ , A1 is satisfied. The only other admissible case is  $\alpha_3 = 0$ ,  $\alpha_0 > 0$ ,  $\beta_0 = 0$ ,  $\beta_3 > 1$  and  $2\alpha_0 \geq \beta_1 > 0$  (then  $\mu(u,\theta)/\sigma^2(u,\theta) \propto \alpha_0/(\beta_1 u)$ ). Note in particular that the Cox-Ingersoll-Ross model satisfies A1 when  $2\alpha_0 \geq \beta_1$ , while the Vasicek model does not: A1(iii) is violated and zero --as well as negative values-- are attainable.

Near infinity, if  $\alpha_2 < 0$  and  $\beta_3 > 1$  then  $\mu(u,\theta)/\sigma^2(u,\theta) \propto (\alpha_2 u^2)/(\beta_2 u^{\beta_3})$  so  $s(v) \propto \exp\left(-\left(2\alpha_2/\beta_2(3-\beta_3)\right)v^{3-\beta_3}\right)$  and  $m(v,\theta) \propto (1/\beta_2 v^{\beta_3})(1/s(v))$  and A1 is satisfied<sup>20,21</sup>. If  $\alpha_2 < 0$  and  $0 < \beta_3 < 1$ , A1 is satisfied. If  $\alpha_2 = 0$ ,  $\alpha_1 < 0$ , and  $\beta_3 > 1$  then A1 is similarly satisfied since then  $s(v,\theta) \propto \exp\left(-\left(2\alpha_1/\beta_2(2-\beta_3)\right)v^{2-\beta_3}\right)$ . If  $\alpha_2 = 0$ ,  $\alpha_1 < 0$  and  $0 < \beta_3 < 1$ , A1 is satisfied since  $s(v,\theta) \propto \exp\left(-\left(2\alpha_1/\beta_1\right)v\right)$ .

The next assumption restricts the amount of dependence allowed in the discrete interest rate observations, so classical asymptotic theory can be applied (P5).

*Assumption A2:* The observed data sequence  $\{r_t\}$  is a strictly stationary  $\beta$ -mixing sequence satisfying:  $k^\delta \beta(k) \xrightarrow[k \rightarrow +\infty]{} 0$  for some fixed  $\delta > 1$ .

A condition on the drift and diffusion sufficient to insure that the discrete observations will satisfy A2 is the following<sup>22</sup>:

*Assumption A2':* Assume that:

$$(i) \quad \lim_{x \rightarrow 0 \text{ or } x \rightarrow \infty} \sigma(x,\theta)\pi(x,\theta) = 0$$

$$(ii) \lim_{x \rightarrow 0 \text{ or } x \rightarrow \infty} \left| \sigma(x, \theta) / \{2\mu(x, \theta) - \sigma(x, \theta)[\partial\sigma(x, \theta)/\partial x]\} \right| < \infty.$$

For the model (23), this condition is always satisfied under the parameter restrictions already made.

*Assumption A3:*

(i) The true density  $\pi_0$  is positive on the interior of its support  $D$  and both  $\pi_0$  and  $\pi_0^2$  are in  $L^2(D)$ .

(ii) The initial random variable  $r_0$  is distributed as  $\pi_0$ .

(iii) The true drift and diffusion functions satisfy A1 and A2.

Note that under A1(i),  $\pi_0$  will be twice continuously differentiable on  $R$ . A3(i) is required to derive that the asymptotic distributions of the test statistics. A3(ii) insures that the process is initialized with the appropriate distribution to be stationary.

*Assumption A4:* The kernel function  $K(\cdot)$  integrates to one and is of order<sup>23</sup>  $r=2$ .  $K$  is bounded, twice continuously differentiable on  $R$ .  $K$  and its derivatives are in  $L^2(R)$ .

The next assumption determines how fast the bandwidth  $h_n$  should go to zero as the sample size  $n$  increases.

*Assumption A5:* To estimate the marginal density, the bandwidth is chosen such that  $\lim_{n \rightarrow \infty} h_n/n^{-1/5}$  is a finite strictly positive constant.

The empirical results are obtained with a Gaussian kernel  $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$  of order  $r=2$ . The quality of density estimates is now widely recognized to be determined primarily by the choice of the bandwidth rather than that of the kernel. The optimal bandwidth here is  $h_n = c_n n^{-1/5}$ , where  $c_n = c$  times the standard deviation of the spot rate time series  $\{r_t\}$ . The constant  $c$  is chosen by cross-validation to minimize the integrated square error of the estimator. In the empirical application, we find  $c=2.1$  for the full 1973-95 sample and  $c=3.5$  for the subsample 1983-95. The results are qualitatively robust to the choice of the kernel as well as small changes in the bandwidth parameter around the optimal value. Monte Carlo experiments suggest that the use of the asymptotic distribution in this setup is warranted (coverage probabilities are all higher than .85).

*Assumption A5':* To estimate the joint density, the bandwidth is chosen such that  $\lim_{n \rightarrow \infty} h_n/n^{-1/5}$  is a finite strictly positive constant. To estimate its first (resp. second) derivatives, set instead  $\lim_{n \rightarrow \infty} h_n/n^{-1/6}$  (resp.  $\lim_{n \rightarrow \infty} h_n/n^{-1/10}$ ) to be finite strictly positive constants.

The empirical results for the transition density are obtained with a product Gaussian kernel,  $k(u,v)=K(u)K(v)$ . The joint density is estimated with optimal bandwidth  $h_n = c_n n^{-1/6}$  where  $c_n = c$  times the standard deviation of the spot rate time series  $\{r_i\}$ . The constant  $c$  is again chosen by cross-validation to minimize the integrated square error of the estimator and the results are similarly robust.

## Proofs

**Proof of (12):** In this proof and the next,  $F_0(\cdot)$  will denote the cumulative density function associated with the density  $\pi_0(\cdot)$ . To obtain the limiting distribution of  $\hat{\theta}_M$ , we define the functional:

$$\Theta_M[F] \equiv \arg \min_{\theta \in \Theta} \int_{\underline{x}}^{\bar{x}} (\pi(x, \theta) - \pi_0(x))^2 \pi_0(x) dx$$

We need to compute the functional derivative of the Kx1 parameter vector  $\Theta_M(\cdot)$ . The functional and its perturbation by H (with density  $h=H'$ ) around  $F_0$  are the implicit solutions of the first order conditions:

$$\begin{cases} \int_{\underline{x}}^{\bar{x}} \frac{\partial \pi(x, \Theta_M[F_0])}{\partial \theta} (\pi(x, \Theta_M[F_0]) - \pi_0(x)) \pi_0(x) dx \equiv 0 \\ \int_{\underline{x}}^{\bar{x}} \frac{\partial \pi(x, \Theta_M[F_0 + H])}{\partial \theta} (\pi(x, \Theta_M[F_0 + H]) - (\pi_0 + h)(x)) (\pi_0 + h)(x) dx \equiv 0 \end{cases}$$

where the gradients  $\frac{\partial \pi(x, \theta)}{\partial \theta}$  are Kx1. To compute the Kx1 functional derivative  $\varphi(F_0)$ , consider the expansion:  $\Theta_M[F_0 + H] = \Theta[F_0] + \int_{\underline{x}}^{\bar{x}} \varphi[F_0](u) h(u) du + O(\|h\|_{L^2}^2)$ . The  $L^2$  norm of  $h$  is  $\|h\|_{L^2} = \left\{ \int_{\underline{x}}^{\bar{x}} |h(x)|^2 dx \right\}^{1/2}$ . The derivative of  $\Theta_M$  is computed as:

$$\begin{aligned} \varphi[F_0](u) &= \left\{ \int_{\underline{x}}^{\bar{x}} \frac{\partial \pi(x, \Theta_M[F_0])}{\partial \theta} \frac{\partial \pi(x, \Theta_M[F_0])}{\partial \theta'} \pi_0(x) dx \right\}^{-1} \frac{\partial \pi(u, \Theta_M[F_0])}{\partial \theta} \pi_0(u) \\ &= E \left[ \frac{\partial \pi(x, \theta_0)}{\partial \theta} \frac{\partial \pi(x, \theta_0)}{\partial \theta'} \right]^{-1} \frac{\partial \pi(u, \theta_0)}{\partial \theta} \pi_0(u) \end{aligned}$$

and therefore  $\varphi[F_0](\cdot)$  is continuous in  $F_0$ . The asymptotic distribution of  $\hat{\theta}_M = \Theta_M[\hat{F}_0]$  follows immediately from Ait-Sahalia (1992a):  $n^{1/2} \{\hat{\theta}_M - \theta_0\} \xrightarrow{d} N(0, \Omega_M)$ , with asymptotic variance given by:

$$\Omega_M \equiv \text{VAR} (\varphi[F](x_t)) + \sum_{k=1}^{+\infty} \left\{ \text{COV} (\varphi[F](x_t), \varphi[F](x_{t+k})) + \text{COV} (\varphi[F](x_{t+k}), \varphi[F](x_t)) \right\}.$$

A consistent estimate of  $\Omega_M$  can be obtained by replacing in  $\varphi[F](\cdot)$  the unknown parameter  $\theta_0$  by  $\hat{\theta}_M$ , and by truncating the infinite sum à-la Newey-West.

**Proof of (13):** We now need to compute a Taylor expansion for the functional  $M(\cdot)$  defined as:

$$M[F_0] \equiv \min_{\theta \in \Theta} \int_{\underline{x}}^{\bar{x}} (\pi(x, \theta) - \pi_0(x))^2 \pi_0(x) dx = \int_{\underline{x}}^{\bar{x}} (\pi(x, \Theta_M[F_0]) - \pi_0(x))^2 \pi_0(x) dx$$

It can easily be shown that under the null hypothesis the functional Taylor expansion of  $M(\cdot)$  has the form:

$$\begin{aligned} M[F_0 + H] &= M[F_0] + \int_{\underline{x}}^{\bar{x}} \left\{ \frac{\partial \pi(x, \theta_0)}{\partial x} \int_{\underline{x}}^{\bar{x}} \varphi[F_0](u) h(u) du - h(x) \right\}^2 \pi_0(x) dx + O(\|h\|_{L^2}^3) \\ &\equiv M[F_0] + \int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} \ell[F_0](x, y) dH(x) dH(y) + O(\|h\|_{L^2}^3) \end{aligned}$$

The linear term in the expansion is zero because under  $H_0$ :  $\pi(x, \theta_0) - \pi_0(x) = 0$ . Therefore the term driving the asymptotic distribution of  $\Lambda$  in the delta method will be quadratic. Now the form of  $\ell[F](x, y)$  is  $\ell[F](x, y) = g[F](x, y) + \pi_0(x) \delta_{(x)}(y)$  where the functional  $g$  is continuous. When  $h$  is replaced by  $\hat{\pi}_0 - \pi_0$  by the continuous mapping theorem, it can be shown that the term  $\int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} g[F](x, y) dH(x) dH(y)$  will converge at speed  $n$  (to a sum of independent chi-squared distributions for iid data, a classical U-statistics result, see e.g., Serfling (1980), Theorem B page 226-227; the speed of convergence is still  $n$  with dependent data). However, when  $h$  is replaced by  $\hat{\pi}_0 - \pi_0$  the term  $\int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} \pi_0(x) \delta_{(x)}(y) dH(x) dH(y) = \int_{\underline{x}}^{\bar{x}} \pi_0(x) h(x)^2 dx$  will converge (i) at the slower speed  $n h_n$  and (ii) to a normal distribution (by extending to dependent data the result of Bickel and Rosenblatt (1973) page 1073; it can be shown with the tools of Aït-Sahalia (1992a) that the asymptotic distribution of this term is unchanged for dependent data).

Therefore, recalling that  $M[F_0] = 0$ ,  $\hat{M} = n h_n M[\hat{F}_0]$  converges at speed  $h_n^{-1/2}$ , and to the same asymptotic distribution as  $n h_n \int_{\underline{x}}^{\bar{x}} \{\hat{\pi}_0(x) - \pi_0(x)\}^2 \pi_0(x) dx$ . The asymptotic distribution is:  $h_n^{-1/2} \{\hat{M} - E_M\} \xrightarrow{d} N(0, V_M)$  where:



$$\begin{cases} E_M \equiv \left( \int_{-\infty}^{+\infty} K^2(x) dx \right) \left( \int_{\underline{x}}^{\bar{x}} \pi_0^2(x) dx \right) \\ V_M \equiv 2 \left( \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} K(u)K(u+x) du \right\}^2 dx \right) \left( \int_{\underline{x}}^{\bar{x}} \pi_0^4(x) dx \right) \end{cases}$$

**Proof of (22):** This proof is similar to that of (13), albeit more complicated. Compute an expansion of (19):  $T[F_0] \equiv \min_{\theta \in \Theta} \int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} \{m(\Delta, y, x, \theta)\}^2 p_0(\Delta, y, x) dx dy$  at  $F_0+H$ , with  $F$  now being the joint cdf of  $X_t$  and  $X_{t+\Delta}$ . First replace the conditional density appearing in  $m(\cdot)$  with the joint and marginal densities and then expand  $p_0(\Delta, y | x) = p_0(\Delta, y, x)/\pi_0(x)$ . Next note that:

$$\frac{\partial^2}{\partial y^2} (\sigma^2(y, \theta) h(\Delta, y, x)) = h(\Delta, y, x) \frac{\partial^2 \sigma^2(y, \theta)}{\partial y^2} + 2 \frac{\partial \sigma^2(y, \theta)}{\partial y} \frac{\partial h(\Delta, y, x)}{\partial y} + \sigma^2(y, \theta) \frac{\partial^2 h(\Delta, y, x)}{\partial y^2}$$

Higher order derivatives of  $h$  converge slower so only the last of the three terms on the right-hand-side above contributes to the asymptotic distribution. Thus the term in the expansion giving the asymptotic distribution of the  $\hat{T}$  is:

$$I \equiv \int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} \left\{ \frac{\sigma^2(y, \theta)}{2\pi_0(x)} \frac{\partial^2 h(\Delta, y, x)}{\partial y^2} - \frac{\sigma^2(x, \theta)}{2\pi_0(x)} \frac{\partial^2 h(\Delta, y, x)}{\partial x^2} \right\}^2 p_0(\Delta, y, x) dx dy$$

as the marginal density  $\pi_0(\cdot)$  also converges faster than the conditional density and its derivatives.

Evaluated at  $h(\Delta, y, x) = \hat{p}_0(\Delta, y, x) - p_0(\Delta, y, x)$ , the term  $I$  gives the asymptotic distribution of  $\hat{T} = nh_n^6 I$ :  $h_n^{-1} \{\hat{T} - E_T\} \xrightarrow{d} N(0, V_T)$ , where:

$$\begin{cases} E_T \equiv \left( \int_{-\infty}^{+\infty} K^2(x) dx \int_{-\infty}^{+\infty} [K^{(2)}(x)]^2 dx \right) \left( \int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} \{ \sigma^4(y, \theta) + \sigma^4(x, \theta) \} \frac{p_0^2(\Delta, y, x)}{\pi_0^2(x)} dy dx \right) \\ \quad - 2 \left( \int_{-\infty}^{+\infty} K(x) K^{(2)}(x) dx \right) \left( \int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} \{ \sigma^2(y, \theta) \sigma^2(x, \theta) \} \frac{p_0^2(\Delta, y, x)}{\pi_0^2(x)} dy dx \right) \\ V_T \equiv \int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} \{ \gamma(x, y, x_1, y_1, x_2, y_2) \}^2 \frac{p_0^4(\Delta, y_1, x_1)}{\pi_0^4(x_1)} dy_1 dx_1 dy_2 dx_2 \end{cases}$$

with:

$$\gamma(x, y, x_1, y_1, x_2, y_2) \equiv \int_{\underline{x}}^{\bar{x}} \int_{\underline{x}}^{\bar{x}} \{ \sigma^2(y_1, \theta) K^{(2)}(y) K(x) - \sigma^2(x_1, \theta) K^{(2)}(x) K(y) \}.$$

$$\cdot \{ \sigma^2(y + y_2, \theta) K^{(2)}(y + y_2) K(x + x_2) - \sigma^2(x + x_2, \theta) K^{(2)}(x + x_2) K(y + y_2) \} dy dx$$

To obtain consistent estimates of  $E_T$  and  $V_T$  replace  $\pi_0$  and  $p_0$  by their kernel estimates,  $\theta_0$  by its estimate, plug-in the kernel function  $K$  and evaluate the integrals.

<sup>1</sup> Other papers considering a similar class of models are: Brenner et al. (1994), Broze et al. (1995), Koejdik et al. (1995), Pagan et al. (1995), Pfann et al. (1995).

<sup>2</sup> The continuous-time process therefore has the same characterization as a discrete-time Gaussian random variable, even though its distribution is not necessarily Gaussian.

<sup>3</sup> "Arbitrage-free" models of the term structure [e.g., Ho-Lee (1986), Hull-White (1990)] make some or all the parameters functions of time. They are not meant to describe the time series properties of the short term interest rate --which is the focus here-- but rather force the derivative pricing model to fit at a given instant the underlying term structures of bond prices, volatilities, caps, etc. The functions of time change and must be re-fitted continually to maintain the term structures' fit.

<sup>4</sup> A process can have time-homogenous transition densities and nevertheless not be stationary (for example a Brownian motion).

<sup>5</sup> See Melino (1994) for a review of the empirical literature.

<sup>6</sup> The convergence of discrete approximations to the continuous-time model has been studied in the finance literature by Nelson (1990), He (1990) and Nelson and Ramaswamy (1990). The approximation can only be made exact in a few cases (typically when the likelihood function is known in closed form, see Wong (1964) for a list).

<sup>7</sup> For example, unless (3) specifies a linear drift function, the conditional mean of the process at date  $s$  given where it was at date  $t$  depends on *both* the  $\mu$  and  $\sigma$  functions.

<sup>8</sup> This equation for the density follows easily from the forward Kolmogorov equation [e.g., Karlin and Taylor (1981), (15.6.22)].

<sup>9</sup> Comparing different types of estimators of a same quantity goes back to Hausman (1978) in the context of parameters of simultaneous equations. Here we must use a nonparametric estimator because no consistent parametric estimator is available if the model is misspecified. See also Bickel and Rosenblatt (1973).

<sup>10</sup> Any other measure of the integral in (2.6) could be used: for example a Riemann sum  $\frac{1}{J} \sum_{j=1}^J (\pi(x_j, \theta) - \hat{\pi}_0(x_j))^2 \hat{\pi}_0(x_j) \{x_{j+1} - x_j\}$  where the  $x_j$ 's form a grid on  $(\underline{x}, \bar{x})$ .

<sup>11</sup> As an aside, note that a simple test of the Markovian property of the data consists in examining whether the two-day transitions can be derived from the one-day transitions (say  $\Delta =$  one day) according to the Chapman-Kolmogorov Equation. This equation can be found for example in Gardiner (1990), Section 3.2.1. In particular, note that knowledge of the one-day transition density is enough to characterize the marginal density since we have for all  $n$ :  $p(n\Delta, y | x) = \int_{\underline{x}}^{\bar{x}} p(\Delta, y | z) p((n-1)\Delta, z | x) dz$ , which can be iterated recursively from  $p(\Delta, y | z)$ . Then by ergodicity  $\lim_{n \rightarrow \infty} p(n\Delta, y | z) = \pi(y)$ .

<sup>12</sup> See e.g., Gardiner (1990), Chapter 5. The forward and backward equations cannot be solved explicitly for the transition density, except in a small number of special cases [see Wong (1964) for a list]. The initial condition for the forward equation at  $s=t$  is  $p(s, y | s, x) = \delta(y-x)$ . The boundary condition for an absorbing barrier is:  $p(s, y | t, x) = 0$  when  $y$  is at the barrier  $\underline{x}$  or  $\bar{x}$ . For a reflecting barrier:  $\mu(y, \theta)p(s, y | t, x) - (1/2)(\partial/\partial y)(\sigma^2(y, \theta)p(s, y | t, x)) = 0$  when  $y$  is at the barrier. The marginal density  $\pi$  is the only solution of (3.1) which does not depend on time (and integrates to one), hence (6).

<sup>13</sup> The final condition for the backward equation at  $t=s$  is  $p(t, y | t, x) = \delta(y-x)$ . The boundary condition for an absorbing barrier is:  $p(s, y | t, x) = 0$  when  $x$  is at the barrier  $\underline{x}$  or  $\bar{x}$ . For a reflecting barrier:  $\partial p(s, y | t, x) / \partial x = 0$  when  $x$  is at the barrier, unless  $\sigma(\cdot)$  vanishes there.

<sup>14</sup> There is a link between (3.3) and the second moment condition in Hansen and Scheinkman (1995). As they show, their condition is equivalent to the fact that the conditional expectation operator and the infinitesimal generator commute [for a proof of this fact, see e.g., Revuz and Yor (1988)]. Our condition is also equivalent to this commutativity property. Indeed we have just shown that  $m(\Delta, y, x, \theta_0) = 0$  for all  $x$  and  $y$  in  $D$ , so if we take an arbitrary twice-continuously differentiable function  $\varphi(y, x)$  in  $L^2(D^2)$ :

$$\begin{aligned}
0 &= \iint m(\Delta, y, x, \theta_0) \varphi(y, x) dy dx \\
&= \iint \left\{ \frac{1}{2} \frac{\partial^2}{\partial y^2} (\sigma^2(y, \theta_0) p_0(\Delta, y | x)) - \frac{\partial}{\partial y} (\mu(y, \theta_0) p_0(\Delta, y | x)) \right\} \varphi(y, x) dy dx \\
&\quad - \iint \left\{ \mu(x, \theta_0) \frac{\partial}{\partial x} (p_0(\Delta, y | x)) + \frac{1}{2} \sigma^2(x, \theta_0) \frac{\partial^2}{\partial x^2} (p_0(\Delta, y | x)) \right\} \varphi(y, x) dx dy \\
&= \iint \left\{ \frac{1}{2} \sigma^2(y, \theta_0) \frac{\partial^2 \varphi(y, x)}{\partial y^2} + \mu(y, \theta_0) \frac{\partial \varphi(y, x)}{\partial y} \right\} p_0(\Delta, y | x) dy dx \\
&\quad - \iint \left\{ \frac{1}{2} \frac{\partial^2}{\partial x^2} (\sigma^2(x, \theta_0) \varphi(y, x)) - \frac{\partial}{\partial x} (\mu(x, \theta_0) \varphi(y, x)) \right\} p_0(\Delta, y | x) dx dy
\end{aligned}$$

so by integrating twice by parts we recover the commutativity property.

<sup>15</sup> The ratio of weekend to weekday mean and standard deviation of S&P500 returns are respectively -0.652 and 1.311 between 6/2/1973 and 12/30/1994. The corresponding numbers are 1.004 and 1.003 for the seven-day Eurodollar rate.

<sup>16</sup> Typically by Chow tests on a discrete approximation with monthly data.

<sup>17</sup> Duffee (1994) documents the idiosyncratic variation of several short maturity yields and recommends the adoption of Eurodollar yields instead of Treasury bill yields for the purpose of calibrating term structure models.

<sup>18</sup> Cases where closed-form expressions are available include the Vasicek [(1977), (25)-(26) p.185] and Cox-Ingersoll-Ross [(1985), (19) p.392] models. Every model with a linear drift has the same conditional mean, regardless of its diffusion [Ait-Sahalia (1992b), (2.4)].

<sup>19</sup> It can be established that a process with entrance boundaries must have a stationary marginal distribution (that is, A1(ii) is automatically satisfied) --and is also ergodic: that is,  $\lim_{\Delta \rightarrow \infty} p(\Delta, x | y, \theta) = \pi(x, \theta)$ . It is not necessary however to have entrance boundaries for the stationary distribution to exist. Natural boundaries cannot be reached from the interior of the state space but the process cannot be started from natural boundaries. A stationary distribution may or may not exist if both boundaries are natural (depending upon whether A1(ii) is satisfied).

<sup>20</sup> It is counterintuitive at first that A1 is still satisfied when  $\beta_3 > 3$ . The intuition provided by Conley et al. (1995) is the following ("diffusion-induced-mean-reversion"): when  $\sigma^2$  is large at high spot rate values, moves to higher interest rate levels are more likely, but so are large downward moves following a large upward move.

<sup>21</sup> When  $\beta_3 > 3$  no restrictions on the drift coefficients are needed to satisfy A1. When  $\beta_3 = 3$ , only  $2\alpha_2 < \beta_2$  is necessary. It is only when  $\beta_3 < 3$  that the leading term of the drift must be negative. We will however always impose that either  $\alpha_2 < 0$  or  $\alpha_1 < 0$  if  $\alpha_2 = 0$ , that is, the process does mean-revert at high values of the spot rate.

<sup>22</sup> Under this assumption, the conditional expectation operator is a contraction [Hansen and Scheinkman (1995)] and then the desired mixing property of the discrete observations follows [see Ait-Sahalia (1992b)].

---

23 The order of a kernel  $K(\cdot)$  is defined as the integer  $r$  such that:  $\int_{-\infty}^{+\infty} x^i K(x) dx = 0$   $i = 1, \dots, r-1$ , and:  
 $\int_{-\infty}^{+\infty} x^r K(x) dx \neq 0$ ,  $\int_{-\infty}^{+\infty} |x|^r |K(x)| dx < \infty$ .

## References

- Aït-Sahalia, Y., 1992a, "The Delta and Bootstrap Methods for Nonlinear Kernel Functionals," working paper, University of Chicago.
- Aït-Sahalia, Y., 1992b, "Nonparametric Pricing of Interest Rate Derivative Securities," working paper, University of Chicago; forthcoming in *Econometrica*.
- Aït-Sahalia, Y., Bickel, P.J. and T.M. Stoker, 1994, "Goodness-of-Fit Tests for Regression Using Kernel Methods," working paper, University of Chicago.
- Bickel, P.J. and M. Rosenblatt, 1973, "On Some Global Measures of the Deviations of Density Function Estimates," *The Annals of Statistics*, 1, 6, 1071-1095.
- Bliss, R.R. and D.C. Smith, 1994, "Stability of Interest Rate Processes," working paper, Indiana University.
- Brennan, M.J. and E.S. Schwartz, 1979, "A Continuous-Time Approach to the Pricing of Bonds," *Journal of Banking and Finance*, 3, 133-155.
- Brenner, R.J., R.H. Harjes and K.F. Kroner, 1994, "Another Look at Alternative Models of the Short-Term Interest Rate," working paper, University of Arizona.
- Broze, L., O. Scaillet and J.M. Zakoïan, 1995, "Testing for Continuous Time Models of the Short Term Interest Rate," working paper, Center for Operations Research and Econometrics, Université Catholique de Louvain.
- Brown, S.J. and P.H. Dybvig, 1986, "The Empirical Implications of the Cox, Ingersoll and Ross Theory of the Term Structure of Interest Rates," *Journal of Finance*, 41, 3, 617-632.
- Chan, K.C., G.A. Karolyi, F.A. Longstaff and A.B. Sanders, 1992, "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 47, 3, 1209-1227.
- Conley, T., L.P. Hansen, E. Luttmer and J. Scheinkman, 1995, "Estimating Subordinated Diffusions from Discrete Time Data," working paper, University of Chicago.
- Constantinides, G.M., 1992, "A Theory of the Nominal Term Structure of Interest Rates," *Review of Financial Studies*, 5, 4, 531-552.
- Courtadon, G., 1982, "The Pricing of Options on Default-Free Bonds," *Journal of Financial and Quantitative Analysis*, 17, 1, 75-100.
- Cox, D.R. and H.D. Miller, 1965, *The Theory of Stochastic Processes*, Chapman & Hall, London.
- Cox, J.C., J.E. Ingersoll and S.A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 2, 385-407.

- Das, S.R., 1994, "Mean Rate Shifts and Alternative Models of the Interest Rate: Theory and Evidence," working paper, New York University.
- Dothan, U.L., 1978, "On the Term Structure of Interest Rates," *Journal of Financial Economics*, 7, 59-69.
- Duffee, G.R., 1994, "Idiosyncratic Variation of Treasury Bill Yields," Finance and Economics Discussion Paper 94-28, Federal Reserve Board.
- Duffie, D. and R. Kan, 1993, "A Yield Factor Model of Interest Rates," working paper, Stanford University.
- Feller, W., 1951, "Two Singular Diffusion Problems," *Annals of Mathematics*, 54, 2, 173-182.
- French, K.R. and R. Roll, 1986, "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics*, 17, 5-26.
- Gardiner, C.W., 1990, *Handbook of Stochastic Methods*, Springer-Verlag, New-York.
- Gibbons, M.R. and K. Ramaswamy, 1993, "A Test of the Cox, Ingersoll and Ross Model of the Term Structure," *The Review of Financial Studies*, 6, 3, 619-658.
- Gray, S.F., 1994, "Regime Switching Models: A New Approach," working paper, Stanford University.
- Hamilton, J.D., 1988, "Rational Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates," *Journal of Economic Dynamics and Control*, 12, 385-423.
- Hamilton, J.D., 1994, "The Daily Market for Fed Funds," working paper, University of California, San Diego.
- Hansen, L.P. and J.A. Scheinkman, 1995, "Back to the Future: Generating Moment Implications for Continuous Time Markov Processes," *Econometrica*, 63, 4, 767-804.
- Hausman, J., 1978, "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1272.
- He, H., 1990, "Convergence from Discrete to Continuous-Time Contingent Claims Prices," *Review of Financial Studies*, 3, 4, 523-546.
- Heston, S., 1993, "A Model of Discontinuous Interest Rate Behavior," working paper, Washington University in St Louis.
- Ho, T.S.Y. and S.-B. Lee, 1986, "Term Structure Movements and Pricing Interest Rate Contingent Claims," *Journal of Finance*, 41, 1011-1029.
- Hull, J.C. and A. White, 1990, "Pricing Interest-Rate Derivative Securities," *Review of Financial Studies*, 3, 4, 573-592.

- Karlin, S. and H.M. Taylor, 1981, *A Second Course in Stochastic Processes*, Academic Press, New-York.
- Koejdik, K.G., P.C. Schotman and C.C.P. Wolff, 1995, "The Dynamics of Short-Term Interest Rate Volatility Reconsidered," working paper, University of Limburg.
- Marsh, T.A. and E.R. Rosenfeld, 1983, "Stochastic Processes for Interest Rates and Equilibrium Bond Prices," *Journal of Finance*, 38, 2, 635-646.
- Melino, A., 1994, "Estimation of Continuous-Time Models in Finance," in C. Sims (ed.), *Advances in Econometrics, Sixth World Congress, Volume II*, Cambridge University Press, Cambridge, England..
- Merton, R.C., 1990, *Continuous-Time Finance*, Blackwell, Cambridge.
- Merton, R.C., 1973, "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4, 141-183.
- Naik, V. and M.H. Lee, 1993, "The Yield Curve and Bond Option Prices with Discrete Shifts in Economic Regimes, working paper, University of British Columbia.
- Nelson, D.B., 1990, "ARCH Models as Diffusion Approximations," *Journal of Econometrics*, 45, 1/2, 7-38.
- Nelson, D.B. and K. Ramaswamy, 1990, "Simple Binomial Processes as Diffusion Approximations in Financial Models," *Review of Financial Studies*, 3, 3, 393-430.
- Pagan, A.R., A.D. Hall and V. Martin, 1995, "Modeling the Term Structure," working paper, Australian National University.
- Pearson, N.D. and T.-S. Sun, 1994, "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model," *Journal of Finance*, 49, 4, 1279-1304.
- Pfann, G.A., P.C. Schotman and R. Tschernig, 1995, "Nonlinear Interest Rate Dynamics and Implications for the Term Structure," working paper, University of Limburg.
- Phillips, P.C.B., 1987, "Time Series Regression with a Unit Root," *Econometrica*, 55, 2, 277-310.
- Revuz, D. and M. Yor, 1988, *Continuous Martingale Calculus*, Springer-Verlag, Berlin.
- Scott, D.W., 1992, *Multivariate Density Estimation: Theory, Practice and Visualization*, Wiley, New-York.
- Serfling, R.J., 1980, *Approximation Theorems of Mathematical Statistics*, Wiley, New-York.
- Silverman, B.W., 1986, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Stock, J.H. and M.W. Watson, 1993, "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 4, 783-820.

Vasicek, O., 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177-188.

Wong, E., 1964, "The Construction of a Class of Stationary Markov Processes," in *Stochastic Processes in Mathematical Physics and Engineering*, Proceedings of Symposia in Applied Mathematics, XVI, 264-276.



**Table 1: Descriptive Statistics**

<b>Source</b>	<b>Bank of America 7-Day Eurodollar (Deposit Rate Midpoint Bid-Ask)</b>	
<b>Frequency</b>	Daily	
<b>Sample Period</b>	6/1/1973 - 2/25/1995	
<b>Sample Size</b>	5505 observations	
<b>Type</b>	Continuously-Compounded Yield-to-Maturity (annualized rate)	
	<b>Spot Interest Rate</b>	<b>First Difference of Spot Interest Rate</b>
<b>Mean</b>	0.08362	-0.0000035
<b>Standard Deviation</b>	0.03591	0.004063
<b>Monthly <math>\rho_1</math></b>	0.9389	0.02136
<b><math>\rho_2</math></b>	0.8785	-0.00689
<b><math>\rho_3</math></b>	0.8300	-0.01658
<b><math>\rho_4</math></b>	0.8014	0.00242
<b><math>\rho_5</math></b>	0.7783	0.00858
<b><math>\rho_6</math></b>	0.7715	0.01573
<b><math>\rho_7</math></b>	0.7361	0.00056
<b>Augmented Daily Dickey-Fuller</b>	-2.60	
<b><math>H_0</math>: Nonstationary</b>	Reject at 90% (critical value = -2.57)	

In the table, the coefficient  $\rho_i$  designates the autocorrelation of the series at lag  $i$ . The augmented Dickey-Fuller test statistic is computed as  $\hat{\tau}_\mu = \hat{\phi} / \text{ase}(\hat{\phi})$  in the model:  $\Delta r_t = \mu + \phi r_{t-1} + \sum_{j=1}^p \phi_j \Delta r_{t-1} + u_t$ , with  $p=5$  lags (see e.g., Harvey (1993), section 5.4).

**Table 2: Tests of Parametric Specifications of the Continuous-Time Spot Rate Process**

<b>Drift Function</b> $\mu(x, \theta)$	<b>Diffusion</b> <b>Function</b> $\sigma^2(x, \theta)$	<b>Reference</b>	<b>Test</b> <b>Stat.</b> $\hat{M}$	<b>95%</b> <b>Crit.</b> <b>Value</b>	<b>95%</b> <b>Result</b>
$\alpha_0 + \alpha_1 x$	$\beta_0$	Vasicek (1977) (0 attainable)	15.09	2.28	Reject
$\alpha_0 + \alpha_1 x$	$\beta_1 x$	Cox-Ingersoll-Ross (1985) Gibbons-Ramaswamy (1993) Pearson-Sun (1994)	42.01	2.28	Reject
$\alpha_0 + \alpha_1 x$	$\beta_2 x^2$	Brennan-Schwartz (1979) Courtadon (1982)	129.49	2.28	Reject
$\alpha_0 + \alpha_1 x$	$\beta_2 x^{\beta_3}$	Chan et al. (1992) (0 unattainable: $\beta_3 \geq 1$ ) <b>Figures 3-abc</b>	38.56	2.28	Reject
$\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 / x$	$\beta_2 x^{\beta_3}$	General Drift, CEV Diffusion	6.79	2.28	Reject
$\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 / x$	$\beta_0 + \beta_1 x + \beta_2 x^{\beta_3}$	General Parametric Model <b>73-95: Figures 4-abc</b> <b>83-95: Figures 5-abc</b>	0.350 0.289	2.28	Cannot Reject

Table 2 reports the value  $\hat{M}$  of the M test statistic and the result of the test for a sample of the models considered.

**Table 3: Estimated Parameter Values for the Continuous-Time Spot Rate Process**

Estimated Parameter Values	Linear Drift, CEV Diffusion	Linear Drift, CEV Diffusion, 0 unattainable	General Parametric Model 1973-95	General Parametric Model 1983-95	
$\hat{\theta}_M$		Figures 3-abc Figure 6c	Figures 4-abc Figure 6b	Figures 5-abc	
$\mu(x, \theta)$	$\alpha_0$	2.935 $10^{-3}$	1.224 $10^{-3}$	-4.643 $10^{-3}$	-5.652 $10^{-3}$
	$\alpha_1$	-3.688 $10^{-2}$	-1.400 $10^{-2}$	4.333 $10^{-2}$	9.648 $10^{-2}$
	$\alpha_2$	constrained = 0	constrained = 0	-1.143 $10^{-1}$	-5.349 $10^{-1}$
	$\alpha_3$	constrained = 0	constrained = 0	1.304 $10^{-4}$	1.041 $10^{-4}$
$\sigma^2(x, \theta)$	$\beta_0$	constrained = 0	constrained = 0	1.108 $10^{-4}$	1.099 $10^{-4}$
	$\beta_1$	constrained = 0	constrained = 0	-1.883 $10^{-3}$	-2.007 $10^{-3}$
	$\beta_2$	1.796 $10^{-4}$	4.902 $10^{-4}$	9.681 $10^{-3}$	1.329 $10^{-2}$
	$\beta_3$	0.29 (unconstrained)	1.04 constrained $\geq 1$	2.073	2.051
<b>Test Statistic</b>		13.15	38.56	0.350	0.289
$\hat{M}$					

The estimated parameter vector reported in Table 3 is from (10). These parameter values correspond to a spot rate which is continuously compounded, annualized and expressed as a decimal number (0.07 for 7%). All the coefficients are significant at the 95% level. The table also indicates the parameter values corresponding to the relevant figures. The relative scale of the drift and diffusion,  $\mu/\sigma^2$ , is arbitrary. To report consistent results across models we fix the scale by matching the average level of interest rate volatility observed in the sample. The otherwise unconstrained estimates of the general parametric model (23) reported in the last two columns of the table satisfy the parameter restrictions (24).

Figure 1a  
Bimodal Density of a Regime-Switching Model

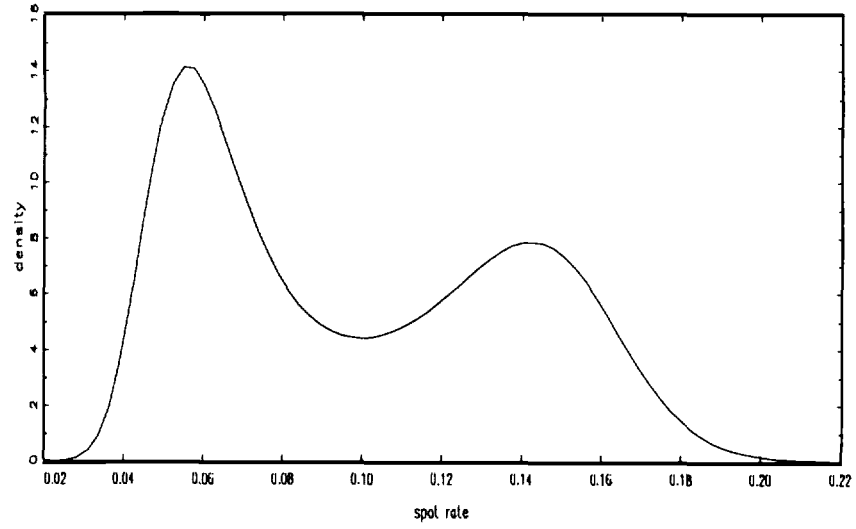


Figure 1b  
Double-Well Potential Drift

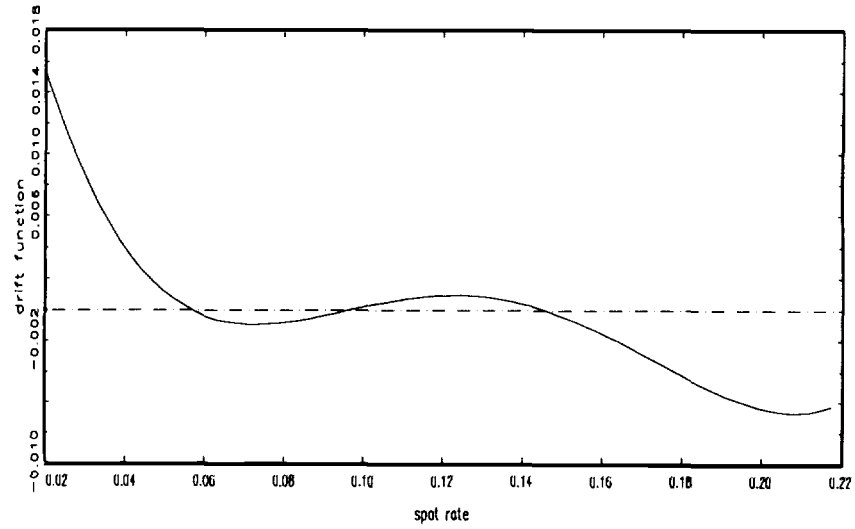


Figure 1c  
Diffusion Function

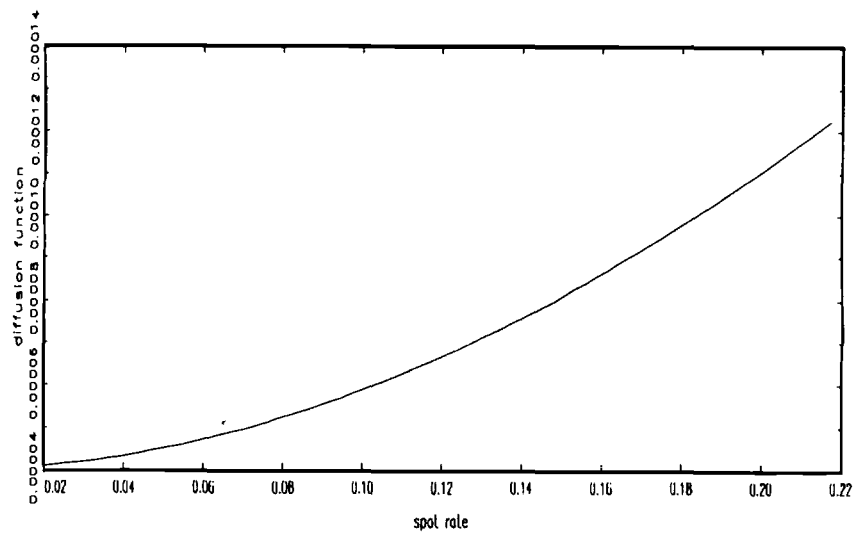


Figure 2  
Spot Interest Rate

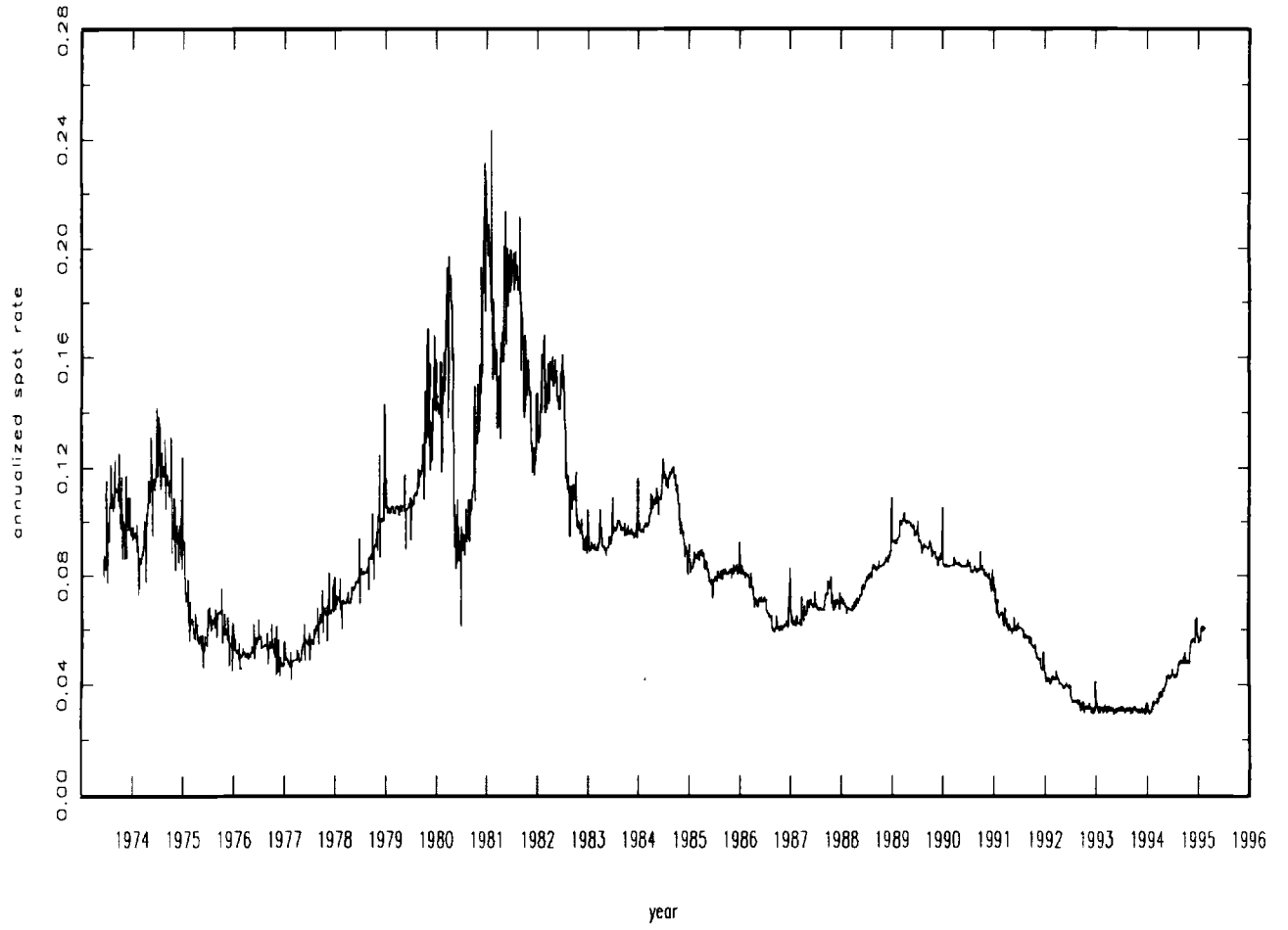


Figure 3a  
Density Comparison with Linear Drift, 1973-95

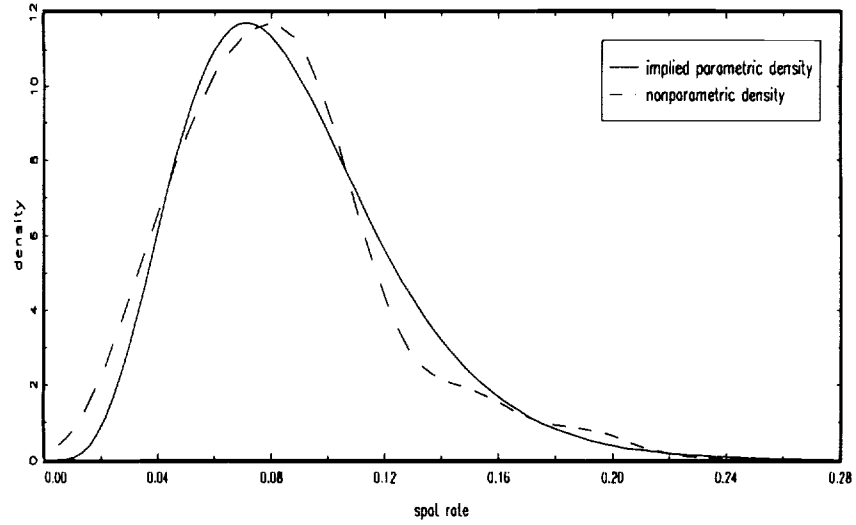


Figure 3b  
Estimated Parametric Linear Drift, 1973-95

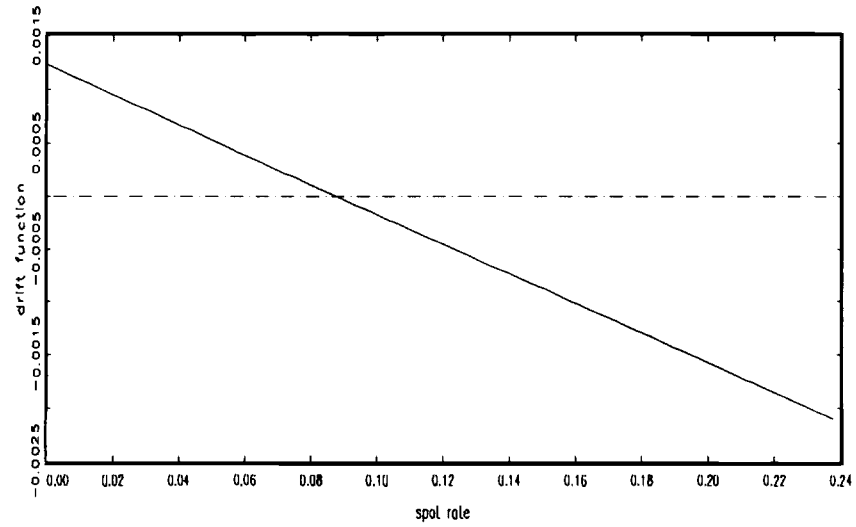


Figure 3c  
Estimated Parametric CEV Diffusion, 1973-95

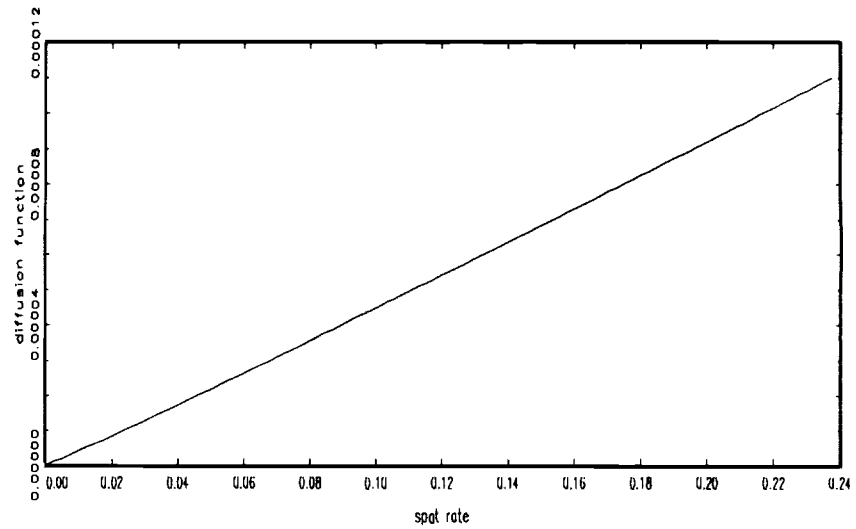


Figure 4a  
Density Comparison with Nonlinear Drift, 1973-95

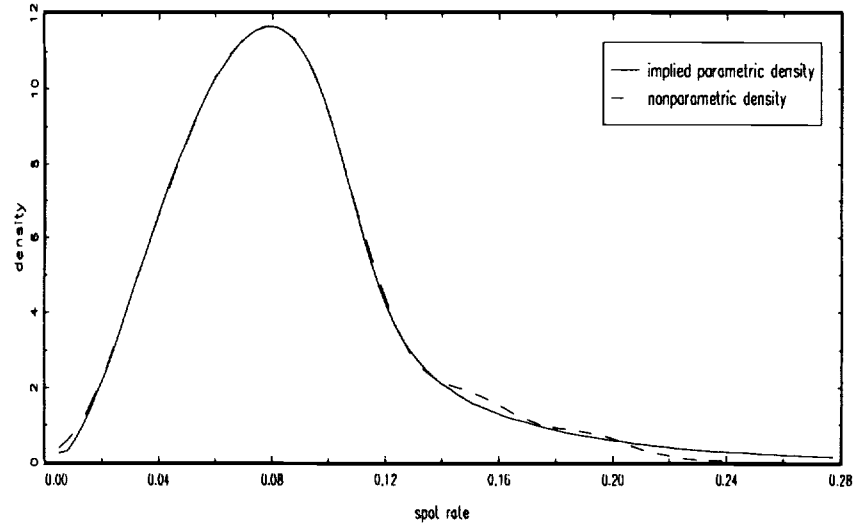


Figure 4b  
Estimated Parametric Nonlinear Drift, 1973-95

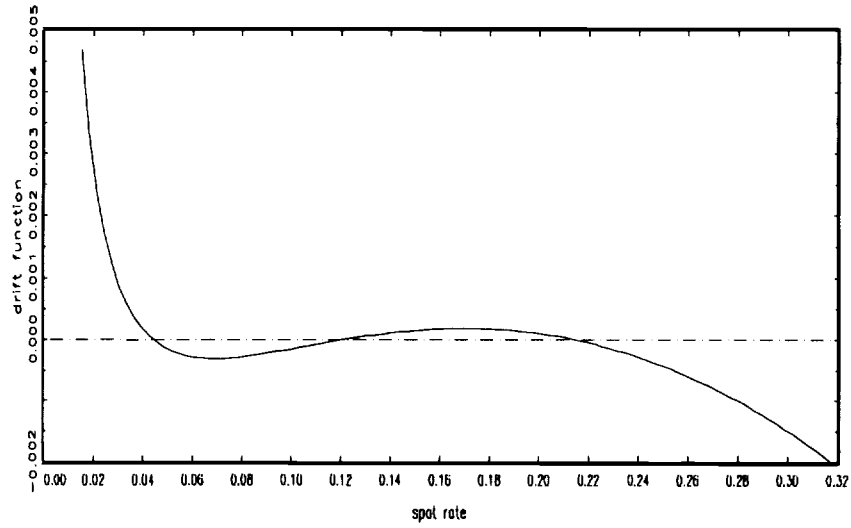


Figure 4c  
Estimated Parametric Nonlinear Diffusion, 1973-95

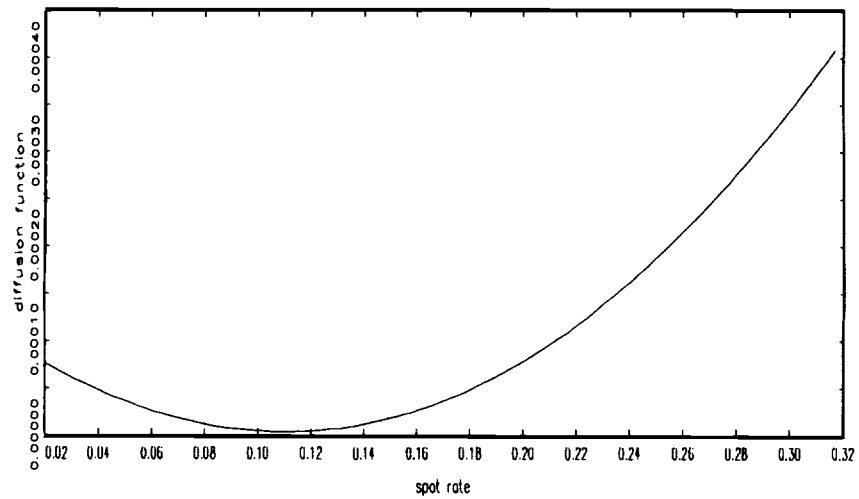


Figure 5a  
Density Comparison with Nonlinear Drift, 1983-95

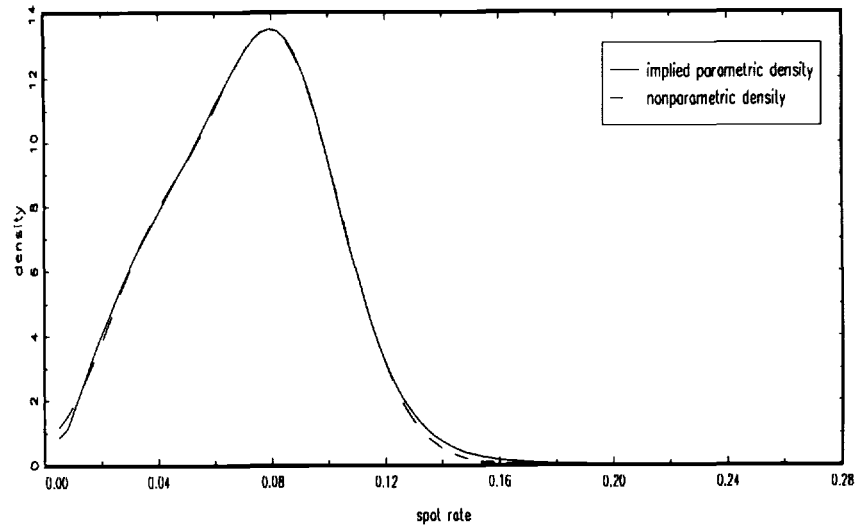


Figure 5b  
Estimated Parametric Nonlinear Drift, 1983-95

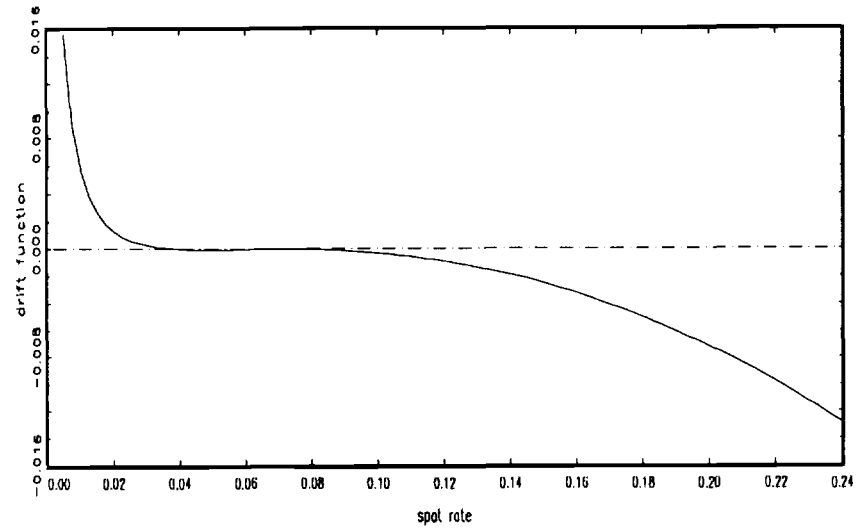


Figure 5c  
Estimated Parametric Nonlinear Diffusion, 1983-95

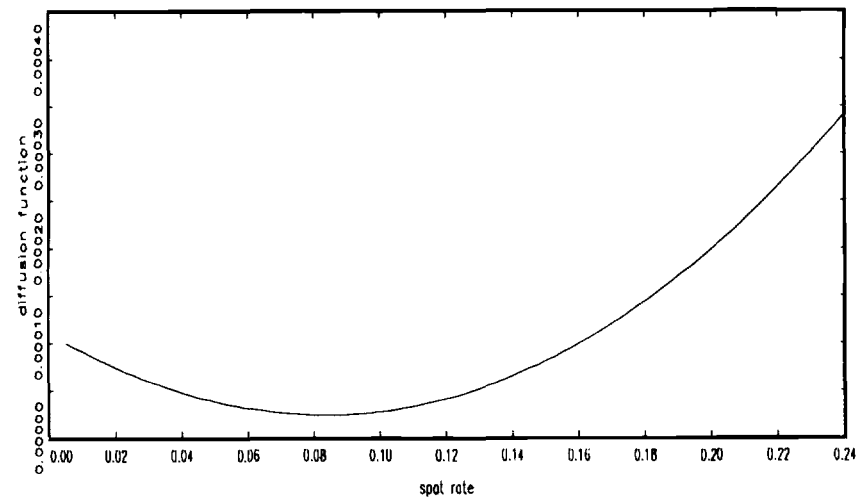




Figure 6a  
Nonparametric Transition Densities, 1973-95

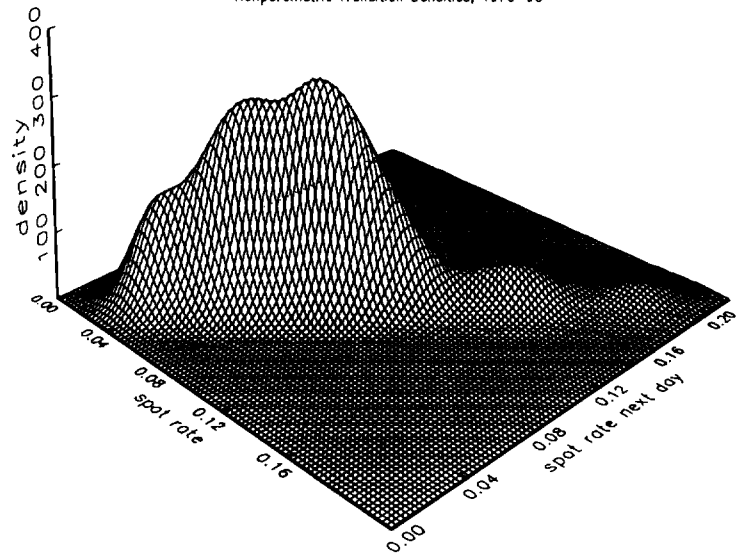


Figure 6b  
Transition Discrepancy  $m(\tau, x, y, \theta)$  for Nonlinear Drift Model

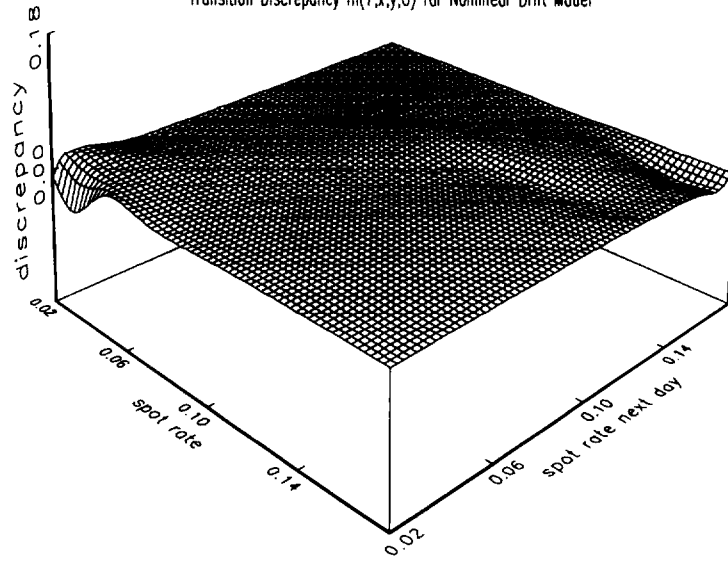


Figure 6c  
Transition Discrepancy  $m(\tau, x, y, \theta)$  for Linear Drift and CEV Diffusion Model

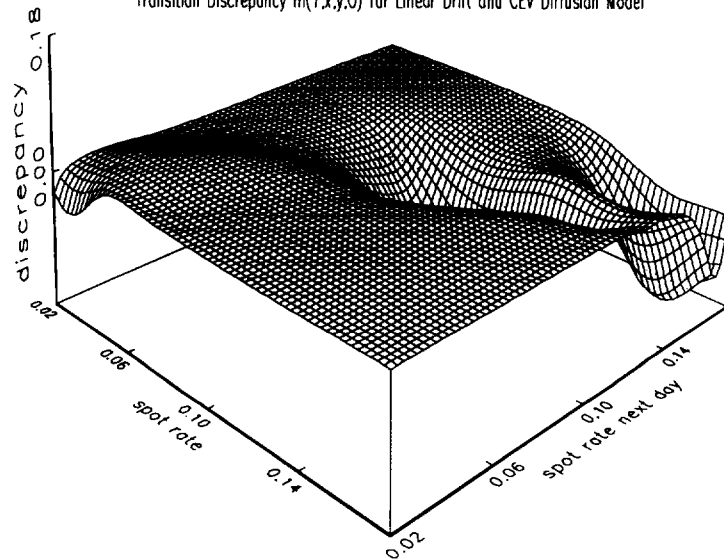


Figure 7a  
Drift FGLS Regression, 1973-95

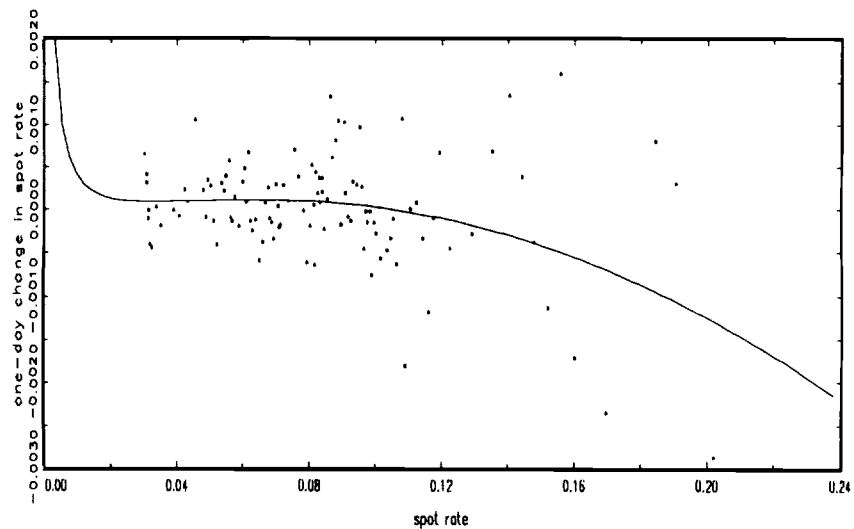


Figure 7a  
Diffusion NLLS Regression, 1973-95

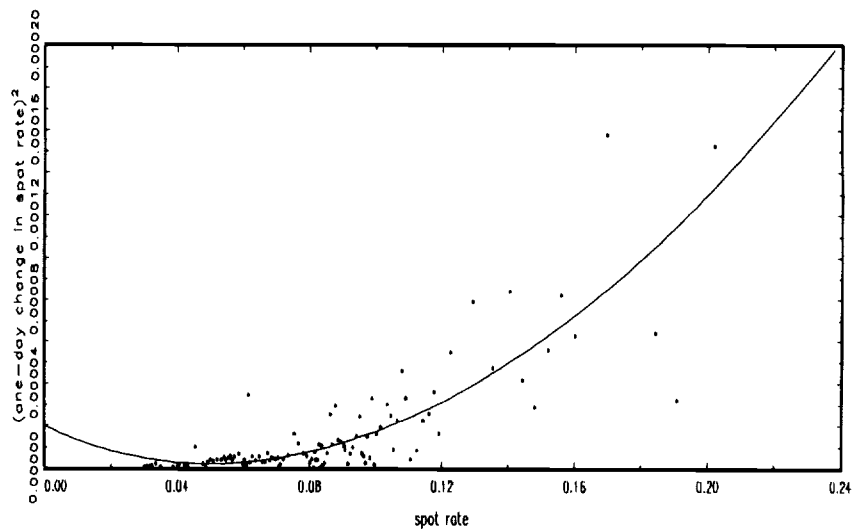


Figure 7c  
Scaled Residuals from Drift Regression, 1973-95

