A Nonparametric Approach to the Estimation of Diffusion Processes, with an Application to a Short-Term Interest Rate Model

George J. Jiang; John L. Knight

Econometric Theory, Vol. 13, No. 5. (Oct., 1997), pp. 615-645.

Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199710%2913%3A5%3C615%3AANATTE%3E2.0.CO%3B2-X

Econometric Theory is currently published by Cambridge University Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/cup.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org
Sun Nov 25 21:07:40 2007

Electronic copy available at: http://ssrn.com/abstract=2149914

Econometric Theory, **13**, 1997, 615–645. Printed in the United States of America.

A NONPARAMETRIC APPROACH TO THE ESTIMATION OF DIFFUSION PROCESSES, WITH AN APPLICATION TO A SHORT-TERM INTEREST RATE MODEL

GEORGE J. JIANG
University of Groningen

JOHN L. KNIGHT
University of Western Ontario

In this paper, we propose a nonparametric identification and estimation procedure for an Itô diffusion process based on discrete sampling observations. The nonparametric kernel estimator for the diffusion function developed in this paper deals with general Itô diffusion processes and avoids any functional form specification for either the drift function or the diffusion function. It is shown that under certain regularity conditions the nonparametric diffusion function estimator is pointwise consistent and asymptotically follows a normal mixture distribution. Under stronger conditions, a consistent nonparametric estimator of the drift function is also derived based on the diffusion function estimator and the marginal density of the process. An application of the nonparametric technique to a short-term interest rate model involving Canadian daily 3-month treasury bill rates is also undertaken. The estimation results provide evidence for rejecting the common parametric or semiparametric specifications for both the drift and diffusion functions.

1. INTRODUCTION

The purpose of this paper is twofold. First, we propose a nonparametric identification and estimation procedure for both the drift and diffusion functions of an Itô diffusion process based on a set of discrete sampling observations. Second, we illustrate the technique with an application to a short-term interest rate model involving Canadian daily 3-month treasury bill rates.

We thank Peter Phillips and two anonymous referees for their helpful comments and suggestions. We acknowledge also the comments of Victor Aguirregabiria, Phelim Boyle, Robin Carter, Joel Fried, Reg Kulperger, and workshop participants in both the Department of Economics and the Department of Statistical and Actuarial Science at the University of Western Ontario. We are particularly grateful to Yanquin Fan for her detailed comments, which significantly improved the exposition of Section 3. Bruce Montador of the Bank of Canada kindly supplied the data on the Canadian 3-month treasury bill rate. The usual disclaimer applies. The second author acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC). Address correspondence to: John L. Knight, Department of Economics, Social Science Centre, University of Western Ontario, London, Ontario N6A 5C2, Canada; e-mail: knight@sscl.uwo.ca.

© 1997 Cambridge University Press 0266-4666/97 $9.00 + .10

An important and attractive property of the continuous-time Itô diffusion process is its closure under quite general nonlinear transformations; that is, under mild regularity conditions nonlinear functions of Itô diffusion processes are also Itô diffusion processes. Because derivative security prices can be expressed as functions of the prices of underlying assets (e.g., bond prices and bond option prices as functions of the spot interest rates), then via Itô's stochastic calculus a partial differential equation (PDE) for derivative security prices can be easily derived as a result of no-arbitrage arguments or a general equilibrium approach. Based on this PDE, the prices of derivative securities can be routinely solved with certain boundary and initial (or final) conditions. From this point of view, the asset pricing and derivative security evaluation problems in finance are very well suited to the application of diffusion processes, and the modeling of asset prices, spot interest rates, exchange rates, and so forth is generally much more tractable and convenient in a continuous-time framework than through binomial or other discrete approximations.

However, identification, estimation, and hence the investigation of the asymptotic sampling properties of the continuous-time diffusion process estimators have proved to be quite difficult. These problems arise mainly due to the unavailability of the continuous sampling observations. The estimation of stochastic differential equations has been considered in the literature for many years, with most of the papers only concerned with continuous sampling observations. For instance, the nonparametric drift and/or diffusion function estimators have been proposed by Banon (1978), Geman (1979), Pham Dinh (1981), and Banon and Nguyen (1981), whereas the parametric drift and/or diffusion function estimators have been proposed by Brown and Hewitt (1975), Lanska (1979), and Kutoyants (1984). Estimation from a set of discrete observations can of course be undertaken by maximum likelihood (ML) estimation if the exact transitional density function or marginal density function is known. The first paper to deal with parametric estimation of the diffusion process from a set of discretely sampled observations is that by Dacunha-Castelle and Florens-Zmirou (1986). Dohnal (1987) also considered the parametric estimation of the diffusion term and proved the local asymptotic mixed normality property of its likelihood function. Lo (1988) proposed an ML estimation method for a jump-diffusion process. Pedersen (1995) suggested an approximate maximum likelihood (AML) parameter estimator for multidimensional diffusion processes, but his framework was purely theoretical. Chan, Karolyi, Longstaff, and Sanders (1992) adopted Hansen's (1982) generalized method of moments (GMM) for the parametric estimation. Duffie and Singleton (1993) proposed an estimation method based on simulated moments of the processes. Hansen and Scheinkman (1995), in a recent paper, formally derived moment conditions for the continuous-time diffusion process based on the infinitesimal generator. Nonparametric estimation with discrete sampling observations was first proposed for the diffusion term by Florens-Zmirou (1993), where the drift term was left unidentified and was restriction-free. Aït-Sahalia (1996) proposed a semiparametric estimation procedure for the diffusion term based on a parametric drift function. Stanton (1996) constructed approximations to the true drift and diffusion.

In this paper, we develop nonparametric estimators for both the diffusion and drift functions. The nonparametric estimator for the diffusion function follows that of Florens-Zmirou (1993), where the kernel method is used in place of the naive method (i.e., the indicator function estimator) to achieve better properties of the estimator on the one hand and to construct a nonparametric drift function estimator on the other. The proposed estimator of the drift function is new in that it combines our aforementioned estimator of the diffusion function along with the proposal from Banon (1978) and Banon and Nguyen (1981), who developed a nonparametric drift function estimator based on known diffusion function and continuous sampling observations. The nonparametric diffusion function estimator deals with general Itô diffusion processes and avoids any functional form specification for either the drift function or the diffusion function. The estimator is based on the local-time properties of the diffusion process, derived from the expansion of the transitional density for small changes in time. It allows the model to capture the true volatility of the process.[1] It is shown that under certain regularity conditions the nonparametric diffusion function estimator is pointwise consistent and asymptotically follows a normal mixture distribution. The consistent nonparametric drift function estimator is derived based on the nonparametric diffusion function estimator and the marginal density of the process under stronger conditions.

Section 2 details the model along with the basic assumptions and the relationships connecting the drift term, the diffusion term, and the transitional or marginal densities. The proposed nonparametric estimators for the diffusion and drift functions are developed in Section 3. Section 4 gives the estimation results from an application of the nonparametric technique to a short-term interest rate model after first discussing various alternative specifications used in the literature. A brief conclusion is contained in Section 5.

2. THE MODEL: ITÔ DIFFUSION PROCESS

The model specified as the underlying process of the state of asset prices, exchange rates, or spot interest rates $\{X_t, t \geq t_0\}$ in most continuous-time finance literature is a time-homogeneous Itô diffusion process represented by the following stochastic differential equation (SDE):

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

with initial condition

$$X_{t_0} = X, \tag{2}$$

where $\{W_t, t \geq t_0\}$ is a standard Brownian motion process or a Wiener process. The functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ are, respectively, the drift (or instantaneous mean) and the diffusion (or instantaneous variance) functions of the process. The stochastic process $\{X_t, t \geq t_0\}$ defined in (1) is a one-dimensional, regular, strong Markov process with continuous sample paths and time-stationary transition probabilities. The precise interpretation of the drift and diffusion terms is

$$\mu(X_t) = \lim_{h\to 0} E\left\{ \frac{X_{t+h} - X_t}{h} \middle| X_t(\omega) = X_t \right\}, \tag{3}$$

$$\sigma^2(X_t) = \lim_{h\to 0} E\left\{ \frac{[X_{t+h} - X_t]^2}{h} \middle| X_t(\omega) = X_t \right\}, \tag{4}$$

which are means and variances of the basic random variable X_t for infinitesimal changes in time. The drift term describes the movement of the process due to time change, whereas the diffusion term measures the magnitude of the random fluctuations around the drift. In terms of a local approximation, the random property of the underlying stochastic process of (1) is characterized by a Brownian motion process path. Despite the continuity, the Brownian paths are almost surely (a.s.) nowhere differentiable. Technically, it is to say that a Brownian motion process path is a.s. of "infinite variation" on any finite interval.

To ensure that the stochastic process defined in (1) can be applied with Itô stochastic integration, to ensure the existence and uniqueness of the solution, and also to ensure its underlying process is a regular Markov process, we need to impose some regularity conditions. Let $T > 0$ and $\mu(X_t): [0,T] \times \mathcal{R} \to \mathcal{R}$, $\sigma(X_t): [0,T] \times \mathcal{R} \to \mathcal{R}^+$ defined as the limits in (3) and (4) and $P(X_{t+h}|X_t)$ be the transitional probability governing the Markov process X_t; we assume the following.

A1. Let $\{F_t : t \in [0,T]\}$ denote a sequence of right-continuous filtrations of sub-σ-algebras of the σ-field F such that (i) $F_t \subset F_s$ for $t \le s$ and (ii) $F_t = \cap_{\tau > t} F_\tau$, and let W_t be F_t-measurable for all $t \in [0,T]$.

A2. As time-homogeneous functions, $\mu(\cdot)$ and $\sigma(\cdot)$ are continuously differentiable and $\mathcal{B}$-measurable functions for $-\infty < X < +\infty$, where $\mathcal{B}$ is the σ-field of the Borel sets on $\mathcal{R}$, and

$$\int_0^T |\mu(X_t)|\, dt < \infty, \qquad \int_0^T |\sigma^2(X_t)|\, dt < \infty$$

a.s.

A3. The initial random variable X is defined on the same space as X_t in (1) and is of second order: $EX^2 < \infty$. $\mu(\cdot)$ and $\sigma(\cdot)$ satisfy the linear growth condition; that is, there exists a positive constant C for which

$$|\mu(x)| \le C(1 + x^2)^{1/2}, \qquad 0 \le \sigma(x) \le C(1 + x^2)^{1/2}$$

a.s. and also satisfy the uniform Lipschitz condition in x, that is, for $x, y \in \mathcal{R}$, $t \in [0,T]$

$$|\mu(x) - \mu(y)| \le D|x - y|, \qquad |\sigma(x) - \sigma(y)| \le D|x - y|$$

for some positive constant D.

A4. The transitional probability $P(X_{t+h}|X_t)$, which governs the underlying Markov process X_t, is a differentiable, $\mathcal{B}$-measurable function of X_t for fixed X_{t+h} and $\mathcal{B}$-measurable function of X_{t+h} for fixed X_t and satisfies the Chapman–Kolmogorov equation

$$P(X_{t+h}|X_t) = \int_{X_\tau} P(X_{t+h}|X_\tau)P(dX_\tau|X_t), \qquad t<\tau<t+h$$

and $\lim_{h\to 0}(1/h)P(|X_{t+h}-X_t| \geq \epsilon|X_t(\omega)=X_t) = 0$ a.s. for every $\epsilon > 0$, where ω denotes a sample path of X_t.

Conditions A1 and A2 ensure that the SDE defined in (1) can be interpreted by Itô stochastic calculus. Conditions A1–A3 ensure that the SDE has a unique solution $X_t(\omega)$ that satisfies

$$X_t = X_0 + \int_0^t \mu(X_t)\,dt + \int_0^t \sigma(X_t)\,dW_t \tag{5}$$

a.s., each component of which is t-continuous over $[0,T]$, with $\int_0^T E[X_t^2(\omega)]\,dt < \infty$. Conditions A1–A4 ensure that the underlying Markov process of (1) is regular such that both the Kolmogorov backward and forward equations hold true a.s. Under the preceding regularity conditions for the drift and diffusion functions $\mu(\cdot)$ and $\sigma(\cdot)$ and the underlying Markov process, it is shown (e.g., see Karlin and Taylor, 1981) that, under certain terminal conditions, the transitional probability density function $p(X_t = x|X_{t_0} = x_0)$ is the unique solution of both the Kolmogorov backward equation,

$$\frac{1}{2}\sigma^2(x_0)\frac{d^2p(X_t=x|X_{t_0}=x_0)}{dx_0^2} + \mu(x_0)\frac{dp(X_t=x|X_{t_0}=x_0)}{dx_0} = -\frac{dp(X_t=x|X_{t_0}=x_0)}{dt_0},$$

and the Kolmogorov forward (or Fokker–Planck) equation,

$$\frac{1}{2}\frac{d^2(\sigma^2(x)p(X_t=x|X_{t_0}=x_0))}{dx^2} - \frac{d(\mu(x)p(X_t=x|X_{t_0}=x_0))}{dx} = \frac{dp(X_t=x|X_{t_0}=x_0)}{dt},$$

where $p(X_t = x|X_{t_0} = x_0)$ is the transitional probability density function for $X_t = x$ at time t conditional on $X_{t_0} = x_0$, and $p(X_{t_0} = x|X_{t_0} = x_0) = \delta(x - x_0)$ (the Dirac delta function). This implies that the transitional probability function of the underlying Markov process of (1) is fully characterized by its coefficients under the regularity conditions.

Stronger relations can be derived under the condition that the stochastic process is stationary in the strict sense or, equivalently, that there exists a stationary initial probability density $p(X_{t_0})$ such that $p(X_t = x) = \int p(X_t = x | X_{t_0} = u) p(X_{t_0} = u)\, du = p(X_{t_0} = x)$ for any x in the state space. Under the preceding condition, let $p(X_{t+h}|X_t)$ be the time-stationary transitional density, which is invariant of time t, of the Markov process X_t; the left-hand side of the Kolmogorov forward equation becomes zero. By multiplying both terms in the right-hand side with the marginal density $p(X_{t_0})$ and then integrating with respect to X_{t_0}, for $0 < X_t < +\infty$ (e.g., interest rates are distributed over $(0,+\infty)$), we can solve for the solution of $p(X_t)$ as

$$p(X_t) = \frac{A}{\sigma^2(X_t)} \exp\left\{ 2 \int_{X^0}^{X_t} \frac{\mu(U)}{\sigma^2(U)}\, dU \right\} \tag{6}$$

with the boundary conditions $p(+\infty) = p'(+\infty) = 0$ or $\sigma^2(+\infty) = \sigma^{2\prime}(+\infty) = 0$, where A is the normalizing constant and X^0 is an arbitrary interior point of the state space, that is, $0 < X^0 < +\infty$. This implies further that the marginal density of the underlying Markov process of (1) is fully characterized by its coefficients under the regularity conditions. Recall that for a normal random variable its distribution is entirely characterized by its first two moments, mean and variance. Even though the underlying process generated by (1) is in general not normally distributed, because the Brownian or Wiener increments are, it turns out that an analogous property will hold for a stationary Itô diffusion process: the distribution of the process (marginal and transitional densities) is entirely characterized by the first two moments of the process, here the drift and diffusion functions. Also, with the boundary condition $p(+\infty) = p'(+\infty) = 0$, we have

$$\mu(X_t) = \frac{1}{2p(X_t)} \frac{d}{dX_t} [\sigma^2(X_t) p(X_t)] \tag{7}$$

or

$$\sigma^2(X_t) = \frac{2}{p(X_t)} \int_0^{X_t} \mu(U) p(U)\, dU. \tag{8}$$

That is, with any functional form specification for either the drift or the diffusion term, the other term will be specified given the marginal density function of the diffusion process.

3. IDENTIFICATION AND ESTIMATION FROM DISCRETELY SAMPLED OBSERVATIONS

Before embarking on the discussion of estimation with discretely sampled data, we will briefly review the procedure followed when the sampling observations are continuous. Suppose that X_t can be observed continuously throughout the time interval $0 \leq t \leq T$. Observations of this kind enable the true diffusion function $\sigma^2(X_t)$ to be determined (at least for those states x visited by X_t during $[0,T]$) through

$$\lim_{n\to\infty}\sum_{j=1}^{2^n}(X_{jt2^{-n}}-X_{(j-1)t2^{-n}})^2=\int_0^t\sigma^2(X_s)\,ds, \tag{9}$$

which holds a.s. for all $t\in[0,T]$, where $\{t_j=jt2^{-n}, j=0,1,2,\ldots,2^n\}$ is a sequence of divisions of the interval $[0,t]$ such that $\max\{(t_j-t_{j-1}), 1\le j\le 2^n\}\to 0$ as $n\to\infty$ (see, e.g., Brown and Hewitt, 1975). Based on the fact that $\mu(X_t)$ is the limit of a conditional expectation, Geman (1979) considered the commonsense nonparametric estimator of the drift function of the following form:

$$\mu_n(x)=n^{-1}\sum_{i=1}^{n}\tau_i^{-1}(X_{t_i+\tau_i}-x), \tag{10}$$

where $0<\tau_i\to 0$ as $i\to+\infty$ and $\{t_i\}$ is a sequence of random times defined by $t_1=\inf\{t\ge 0:X_t=x\}$ and $t_{i+1}=\{t\ge t_i+\tau_i:X_t=x\}$. Geman (1979) proved the consistency and asymptotic normality of this estimator.

In most practical situations, however, continuous sampling of the stochastic process is impossible because the characteristic dynamics of the system can be much faster than the sampling rate. The need to estimate the parameters of a stochastic differential equation from discrete-time observations is the situation arising frequently in practical applications. The first paper to deal with parametric estimation of the coefficients of a stationary diffusion process from discrete sampling observations is that by Dacunha-Castelle and Florens-Zmirou (1986), in which they also derived the measure of the amount of information lost due to discretization. Dohnal (1987) showed that with discrete observations of the process parameter estimation based on a formula like (9) or (10) does not give us the best results. He proved the local asymptotic mixed normality property of the likelihood function of the diffusion function and showed that we can obtain better results using this property. The method used in both papers is the expansion of the transitional density of the underlying Markov process for small changes in time. Under the condition derived in Dacunha-Castelle and Florens-Zmirou (1986), that is, $n\Delta_n\to 0$ as $n\to\infty$, where Δ_n is the sampling interval, ML estimators of the parameters of the diffusion process can be defined based on the expansion of the transitional density functions. For diffusion processes with jumps, Lo (1988) derived the ML estimation method of the parameters based on the Markovian properties of the process. Lo's method, however, requires that a PDE be solved numerically per observation in order to get the transitional density of the process unless its functional form is explicitly known. Pedersen's (1995) AML estimator is applicable to multidimensional Itô parametric diffusion processes; however, his framework is purely theoretical. Stanton (1996) developed approximations to the true drift and diffusion functions and estimated these approximations nonparametrically. Also developed is the alternative "indirect inference" approach to estimating nonlinear SDE's; see Monfort (1996) for a recent review with focus on misspecified models.[2] A particular example of the "indirect inference" approach is the simulated GMM proposed by Duffie and Singleton (1993) based on simulated sampling paths from given parameter values. The parameter estimates are derived from minimizing the difference between the simulated moments and the sample moments. Their method requires that new sample paths be simulated every time when the parameter estimates are adjusted.

The nonparametric estimation of the diffusion term suggested by Aït-Sahalia (1996) is virtually a semiparametric approach to the identification of the strictly stationary diffusion processes. It specifies the drift as a linear mean-reverting function, that is, $\mu(X_t) = \beta(\alpha - X_t)$. Then, based on (8), $\sigma^2(\cdot)$ can be characterized through the marginal density $p(\cdot)$ and the estimates of α and β. The transitional distribution can be used to obtain consistent estimates of α and β through

$$E[X_{t+\rho}|X_t] = \alpha + e^{-\rho\beta}(X_t - \alpha),$$

which is derived from the Kolmogorov backward equation. Aït-Sahalia (1996) showed that under certain regularity conditions the estimator $\hat{\sigma}^2(X_t)$ is pointwise consistent and asymptotically normal. However, this estimation procedure has to rely on the parametric specification of the drift term and works only for the strictly stationary diffusion processes.

Another commonly used method of estimating the diffusion process relies on "discretizing" the model in order to estimate the parameters using, for example, Hansen's (1982) GMM technique (e.g., Chan et al., 1992). As noted in the study of linear SDE's, any attempts to estimate the exact discrete model (EDM) after a reparameterization cannot be recommended, because it may cause severe misspecification due to the "aliasing problem" (Phillips, 1973; Hansen and Sargent, 1983).[3] Lo (1988) and others have provided the examples of inconsistent estimators based on "discretization" of Itô diffusion processes. Discretization can cause problems for GMM apart from the fact that it requires stationarity of the process for its optimal asymptotic property. Only recently, Hansen and Scheinkman (1995) derived moment conditions based on an infinitesimal generator. However, they did not discuss the issues of statistical efficiency, optimal moment conditions, or inference using moment conditions.

Similar to the "aliasing problem" for linear SDE's, identifying both the drift and diffusion functions of the Itô diffusion process without any restrictions from discretely sampled data in a finite sampling period is impossible in general. Phillips (1973) pointed out the fact that, unless there are sufficient a priori restrictions on the parameters of the linear SDE's, we cannot distinguish between structures generating cycles whose frequencies differ by integer multiples of the reciprocal of the observation period. In particular, on a fixed time interval, the drift of a nonlinear SDE (univariate or multivariate) cannot be directly identified because the Cameron–Martin–Girsanov transformation can always be applied to give an otherwise unnoticeable change in the drift. Furthermore, following Girsanov's Theorem (see Øksendal, 1992), direct identification of the drift is impossible on a fixed sampling period, no matter how small the sampling intervals are. One hope to identify the drift term, therefore, is from continuous sampling observations.

An alternative way to identify the process that is proposed in this paper is to identify the diffusion term first. The drift function can then be indirectly identified, under certain conditions based on discrete sampling observations, through the diffusion function and the information contained in the marginal density function of the process. In terms of local-time properties, the drift and diffusion functions can also be expressed as

$$\mu(X_t) = \lim_{h\to 0} \frac{1}{h} \int_{[|X_{h+t}-X_t|<\epsilon]} (X_{t+h} - X_t)\, dP(X_{h+t}|X_t(\omega) = X_t), \tag{11}$$

$$\sigma^2(X_t) = \lim_{h\to 0} \frac{1}{h} \int_{[|X_{h+t}-X_t|<\epsilon]} (X_{t+h} - X_t)^2\, dP(X_{t+h}|X_t(\omega) = X_t), \tag{12}$$

for all $\epsilon > 0$ and all X_t in the state space. It is noted that in (1) the drift term is of order dt and the diffusion term is of order $\sqrt{dt}$, as $(dW_t)^2 = dt + O((dt)^2)$. That is, the diffusion term has lower order than the drift term for infinitesimal changes in time, and the local-time dynamics of the sampling path reflects more of the properties of the diffusion term than those of the drift term. Therefore, when sampling intervals are small, identification of the diffusion term is easier than that of the drift term, which suggests the possibility of identifying the diffusion term first from high-frequency discrete sampling observations.

3.1. The Nonparametric Estimator of $\sigma^2(X_t)$

We propose here a nonparametric estimator for the diffusion function $\sigma^2(X_t)$ of a general diffusion process based on observing X_t at $\{t = t_1, t_2, \ldots, t_n\}$ in the time interval $[0,T]$, with $T \geq T_0 > 0$, where T_0 is a positive constant. Although all our results could be derived for nonequispaced data, the only complication being notational, for ease of presentation we will discuss only the equispaced data case. Consequently, we let $\{X_t = X_{\Delta_n}, X_{2\Delta_n}, \ldots, X_{n\Delta_n}\}$ be n equispaced observations at $\{t_1 = \Delta_n, t_2 = 2\Delta_n, \ldots, t_n = n\Delta_n\}$, where $\Delta_n = T/n$. The nonparametric diffusion function estimator is developed without imposing any restrictions on the drift term. Discrete approximations of the local time of the diffusion process and the expansion of the transitional density for infinitesimal changes in time lead to the estimator. The methodology is along the lines of Dacunha-Castelle and Florens-Zmirou (1986), Dohnal (1987), and Florens-Zmirou (1993). For the purpose of estimation, we impose more conditions on the SDE in (1).

A5. $\mu(\cdot)$ is a bounded function, twice continuously differentiable, with bounded derivatives; $\sigma(\cdot)$ is a bounded function, three times continuously differentiable, with bounded derivatives such that there exist two constants k and K with $0 < k \leq \sigma(x) \leq K$, that is, we exclude the case that at some set the diffusion process is singular with $\sigma(X_t) = 0$;

A6. Let $K(\cdot) \in L^2(\mathcal{R})$ be a positive kernel, that is, $\int_{-\infty}^{\infty} K(x)\, dx = 1$, and continuously differentiable, with $\lim_{x\to+\infty} K(x) = \lim_{x\to-\infty} K(x) = 0$.

Note that under conditions A1–A5, the stochastic process in (1) is nonsingular and has a time-stationary transitional probability function. Condition A6 imposes regularity conditions on $K(\cdot)$.

To define the local time process associated with X_t, let $I_A(\xi)$ be the indicator function of set A. The occupation time of X_t in the set A up to time t is defined by $L_A(t) = \int_0^t I_A(X_\tau)\, d\tau$. Let $A_\Delta = (x - \Delta, x + \Delta)$, and consider the limit

$$L_t(x) = \lim_{\Delta \to 0} \frac{1}{2\Delta} \int_0^1 I_{A_\Delta}(X_\tau)\, d\tau,$$

which defines a family of random variables for $t > 0$ and any interior state point x. $L_t(x)$ is called the local time process, which is well defined as an ordinary integral along each continuous sampling realization path of the diffusion process. For any positive sequence $\{h_n > 0, n = 1,2,\ldots\}$, such that $nh_n \to \infty$ and $h_n \to 0$ as $\Delta_n \to 0$, we can show that

$$L_t^n(x) = \frac{1}{nh_n} \sum_{i=1}^{[t/\Delta_n]} K\left(\frac{X_{i\Delta_n} - x}{h_n}\right)$$

is a discrete approximation for $L_t(x)$ in the sense that, for each t, $L_t^n(x) \to L_t(x)$ a.s. as $nh_n \to \infty$ and $h_n \to 0$. Furthermore, we construct

$$V_t^n(x) = \frac{1}{Th_n} \sum_{i=1}^{[t/\Delta_n]-1} K\left(\frac{X_{i\Delta_n} - x}{h_n}\right)[X_{(i+1)\Delta_n} - X_{i\Delta_n}]^2,$$

which converges in the L^2 sense to $\sigma^2(x)L_t(x)$ if the sequence $\{h_n > 0, n = 1,2,\ldots\}$ is such that nh_n^4 tends to zero when $\Delta_n \to 0$. When the sampling path of the diffusion process visits x, we estimate $\sigma^2(x)$ by

$$S_n(x) = \frac{\displaystyle\sum_{i=1}^{n-1} nK\left(\frac{X_{i\Delta_n} - x}{h_n}\right)[X_{(i+1)\Delta_n} - X_{i\Delta_n}]^2}{\displaystyle\sum_{i=1}^{n} TK\left(\frac{X_{i\Delta_n} - x}{h_n}\right)}. \tag{13}$$

The asymptotic distribution of $S_n(x)$ is given in Theorem 1, the proof of which is in the Appendix.

THEOREM 1. *Under conditions* A1–A6, *as* $h_n \to 0$, $\Delta_n \to 0$, $nh_n \to \infty$, *and* $nh_n^3 \to 0$, $\hat\sigma^2(x) = S_n(x)$ *is a pointwise consistent estimator of* $\sigma^2(x)$ *and asymptotically normally distributed. That is, conditional on the event that x is visited,* $\sqrt{nh_n}(S_n(x)/\sigma^2(x) - 1)$ *converges in distribution to a mixture of normal law* $[L_T(x)]^{-1/2}Z$, *where Z is a standard normal random variable independent of the local time* $L_T(x)$.[4] *The variance of* $\hat\sigma^2(x)$ *can be consistently estimated by* $\hat V[\hat\sigma^2(x)] = \hat\sigma^4(x)/\sum_{i=1}^n K((X_{i\Delta_n} - x)/h_n)$.

Remark 1.1. It is noted that, in this estimation procedure of the diffusion function, the drift term $\mu(\cdot)$ is a nuisance coefficient function with no restrictions imposed on its functional form. All the regularity conditions are mild, which ensures that the nonparametric diffusion function estimator developed here works for very general diffusion processes.

Remark 1.2. The estimator developed here is clearly consistent with that proposed by Florens-Zmirou (1993), in which he defined the discrete approximation of the local time process as $L_t^n(x) = (1/2nh_n)\sum_{i=1}^{[t/\Delta_n]} I_A(X_{i\Delta_n})$, where $A = (x - h_n,\ x + h_n)$, and the corresponding nonparametric estimator is constructed from the indicator function rather than a general kernel function, that is, the naive method instead of the kernel method. Even though there appears to be little discrimination among different nonparametric functional estimators on the basis of integrated mean squared error (IMSE), Kumar and Markman (1975) concluded, based on their Monte Carlo studies, that the kernel estimator with a standard normal kernel performed as well as the estimator with the optimal kernel of Epanechnikov but better than the naive estimator of Rosenblatt and the orthogonal-series estimator of Kronmal–Tartr. A well-known serious drawback of the naive method is that it is by definition not a continuous function but has jumps at the endpoints of the window and zero derivatives everywhere else. The discontinuity of the naive method could cause extreme difficulty when constructing the nonparametric drift function estimator of the diffusion process, as we shall see later in this section. In contrast, the normal kernel functional estimator has derivatives of all orders and causes no problem in the derivation of the nonparametric drift function estimator. In addition, compared to the kernel method, the naive method imposes very strong restrictions on the choice of window width, because the denominator of $S_n(x)$ has to be bounded away from zero.[5]

3.2. The Nonparametric Estimator of $\mu(X_t)$

To estimate $\mu(\cdot)$, one could try to use the fact that $\mu(\cdot)$ can be defined as the limit in (11) or use the nonparametric estimator defined in (10) with continuous sampling observations. With discrete observations, however, these approaches are extremely sensitive to the sampling intervals and the length of the total sampling period. Our approach here is based on equation (7) or, equivalently,

$$\mu(x) = \frac{1}{2}\left[\frac{d\sigma^2(x)}{dx} + \sigma^2(x)\frac{p'(x)}{p(x)}\right]. \tag{14}$$

That is, we aim to identify $\mu(x)$ by using the additional information contained in the marginal density function $p(x)$, along with the estimated $\hat{\sigma}^2(x)$. To ensure that equation (14) holds, we need to impose the following stronger conditions on the diffusion process.

A7. The function $\mu(\cdot)$ and $\sigma(\cdot)$ are such the solutions of the equation

$$\tfrac{1}{2}d(\sigma^2(x)p(x))/dx = \mu(x)p(x)$$

are bounded and integrable on $\mathcal{R}$ and $\{X_t, t \geq t_0\}$ is an asymptotically uncorrelated process.[6] The derivatives $\mu'(\cdot), \sigma'(\cdot), \sigma''(\cdot)$ satisfy the uniform Lipschitz and linear growth conditions.

A8. $p(\cdot)$ is positive on the interior of its support, twice continuously differentiable with bounded derivatives.

Conditions A1–A5 and A7 ensure that the Kolmogorov forward equation has a unique fundamental solution for the marginal density and that the transitional density function converges, as the initial time $t_0 \to -\infty$, to a bounded and continuous limiting density, say, $\lim_{h\to+\infty} p(X_t|X_{t-h}) = p(X_t)$, which is the asymptotic marginal density of the process defined in (1) in the steady state. The condition that there exists a stationary limiting probability density is weaker than the condition that the process is stationary in the strict sense. For asymptotic results, however, these two conditions can be considered as equivalent with the necessary condition that the initial stationary probability distribution is set to be identical to the final limiting distribution. Thus, all results obtained in Section 2 under stationarity can be used here in the asymptotic sense. Furthermore, from condition A7, a necessary and sufficient condition for a finite invariant measure, that is, a stationary limiting probability distribution $p(X_t)$, to exist is that the waiting time for a return to any interior point x of the stochastic process be finite or, in other words, the process is strongly recurrent (positive ergodic). An analytical condition for the existence of a stationary distribution consists in the $\mathcal{L}^1$-integrability of the function $G(x) = 2/[\sigma^2(x)R(x)]$ (see Prohorov and Rozanov, 1969), that is, $\int_{x_1}^{x_2} G(x)\,dx < +\infty$, where $R(x) = \exp\{-\int_{x_0}^{x}[(2\mu(u))/(\sigma^2(u))]\,du\}$. Condition A8 imposes regularity conditions on $p(\cdot)$.

With equation (14), because the nonparametric kernel estimator $\hat{\sigma}^2(x)$ is differentiable, estimating $\mu(\cdot)$ amounts to estimating $p(x)$ and $p'(x)$, or estimating $Q(x) = p'(x)/p(x)$. Banon and Nguyen (1981) proposed a strongly consistent estimator of $Q(x)$ with continuous sampling observations. Given equispaced discrete observations of $\{X_t = X_{\Delta_n}, X_{2\Delta_n}, \ldots, X_{n\Delta_n}\}$ at $\{t_1 = \Delta_n, t_2 = 2\Delta_n, \ldots, t_n = n\Delta_n\}$ in the time interval $[0,T]$ with $T \geq T_0 > 0$, where T_0 is a positive constant, we develop here the standard kernel estimators for $p(x)$ and $p'(x)$ based on discrete sampling observations, that is, $p_n(x) = \sum_{i=1}^{n}(1/nh_n)K((X_{i\Delta_n} - x)/h_n)$ and $p_n'(x) = \sum_{i=1}^{n}(1/nh_n^2)K'((X_{i\Delta_n} - x)/h_n)$, which are consistent estimators of $p(x)$ and $p'(x)$, respectively. With this approximation, we can construct a consistent estimator of $Q(x)$ from discrete sampling observations, such that $Q(x)$ can be nonparametrically estimated by

$$q_n(x) = \frac{\sum_{i=1}^{n} \frac{1}{h_n} K'\left(\frac{X_{i\Delta_n} - x}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{X_{i\Delta_n} - x}{h_n}\right)} \tag{15}$$

given that x is visited by the stochastic process and $p(x) \neq 0$. The pointwise consistency of $q_n(x)$ and $\hat{\mu}(\cdot)$ are detailed in Theorem 2 and its corollary, the proofs of which are in the Appendix.

THEOREM 2. *Under conditions* A1–A8, *as* $h_n \to 0$, $\Delta_n \to 0$, $nh_n \to +\infty$, *and* $n\Delta_n \to +\infty$, $\hat{p}(x) = p_n(x)$ *is a pointwise consistent estimator of* $p(x)$ *and asymptotically normally distributed. That is, conditional on the event that x is visited,* $\sqrt{nh_n}(p_n(x) - p(x))$ *converges in distribution to* $N(0, V[\hat{p}(x)]$. *The asymptotic* *variance of $\hat{p}(x)$ is $V[\hat{p}(x)] = (nh_n)^{-1}p(x)\int_{-\infty}^{\infty}K^2(u)\,du$, which can be consistently estimated by $\hat{V}[\hat{p}(x)] = (nh_n)^{-1}\hat{p}(x)\int_{-\infty}^{\infty}K^2(u)\,du$. Furthermore, $\hat{p}'(x) = p_n'(x)$ is also a consistent estimator of $p'(x)$, and so is $\hat{Q}(x) = q_n(x)$ a consistent estimator of $Q(x)$ given that x is visited by X_t and $p(x) \neq 0$.*

With $q_n(x)$ as the consistent estimator of $Q(x) = p'(x)/p(x)$ and $S_n(x)$ as the consistent estimator of $\sigma^2(x)$, $\mu(x)$ can be consistently estimated from (14) as

$$\hat{\mu}(x) = \frac{1}{2}\left[\frac{dS_n(x)}{dx} + S_n(x)\frac{\sum_{i=1}^{n}\frac{1}{h_n}K'\left(\frac{X_{i\Delta_n} - x}{h_n}\right)}{\sum_{i=1}^{n}K\left(\frac{X_{i\Delta_n} - x}{h_n}\right)}\right]. \tag{16}$$

Note, with conditions A1–A8, $q_n(\cdot)$ is well defined and $S_n(\cdot)$ is pointwise differentiable. We have the following.

COROLLARY. *Under the conditions of Theorem* 1 *and Theorem* 2, *$\hat{\mu}(x)$ estimated from $\hat{\sigma}^2(x) = S_n(x)$ and $\hat{Q}(x) = q_n(x)$ based on* (16) *is a pointwise consistent estimator of $\mu(x)$.*

Remark 2.1. The preceding identification and estimation procedure requires minimum conditions imposed on the diffusion processes. In particular, the estimation of the diffusion function places no restrictions on the functional form of the drift term, which allows the model to capture the true volatility of the process.

Remark 2.2. Compared to the semiparametric diffusion function estimator suggested by Aït-Sahalia (1996), in which the asymptotic distribution of the estimator is derived by letting the sample size increase through prolonging the observation period, that is, $T \to +\infty$, the nonparametric diffusion function estimator proposed in this paper requires the sampling frequency to increase in order to obtain asymptotic results; that is, instead of $T \to +\infty$, it requires the sampling interval $\Delta_n = T/n \to 0$. The reason for different requirements here is that in the semiparametric model, with the drift function specified, the perturbation caused by the diffusion term is actually implicitly identified. In our case, where no *prior* restrictions are imposed on the drift term, the identification of the diffusion term has to rely on the local-time properties of the process, that is, the evolution of the stochastic process for small changes in time.

Remark 2.3. The asymptotic distribution of $\hat{\mu}(x)$ is complicated due to the functional form of its estimator. The variance of $\hat{\mu}(x)$, however, can be obtained using the δ method conditional on either $\hat{p}(x)$ or $\hat{\sigma}^2(x)$, or otherwise unconditionally if the covariance of $\hat{p}(x)$ and $\hat{\sigma}^2(x)$ is known. In practice, a bootstrapping technique could be used to derive the standard error of $\hat{\mu}(x)$.

Remark 2.4. With respect to the linear mean-reverting specification for the drift function, in addition to the restriction imposed through its functional form, our simulation results in Jiang and Knight (1996) indicated that the parameter estimators of the function are not robust in that they are very sensitive to the sampling path and/or the discretely sampled observations along the sampling path of the diffusion process.

Remark 2.5. In Jiang and Knight (1996), a Monte Carlo simulation study is performed comparing the performance of the nonparametric diffusion function and drift function estimators along with other commonly used parametric estimators in models with explicit transitional density functions. It is noted that both the nonparametric diffusion function estimator and the drift function estimator performed extremely well in these comparisons.

4. APPLICATION OF THE NONPARAMETRIC ESTIMATION TO A SHORT-TERM INTEREST RATE MODEL

Many models have specified the interest rate diffusion process parametrically or semiparametrically in order to implement common estimators—for example, the GMM, nonlinear least-squares, or ML method. Table 1 lists alternative specifications of the short-term interest rate models in the literature. A few of these models are worth a little discussion here. The Vasicek (1977) model is an Ornstein–Uhlenbeck process and has Gaussian transitional densities. The Merton (1973) model is a Brownian motion with drift process and nested within the Vasicek model by the parameter restriction $b = 0$. Both models specify the instantaneous volatility of the spot interest rate as flat, that is, $\sigma(X_t) = \sigma$. The Vasicek and Merton models are often criticized for allowing negative interest rates. The Cox, Ingersoll, and Ross (1985) squared-root diffusion term model assumes that the instantaneous variance is a linear function of the level of interest rates, so the standard deviation is a square root, that is, $\sigma(X_t) = \sigma X_t^{1/2}$. This stochastic differential equation yields a noncentral chi-square transitional distribution whereas

TABLE 1. Alternative specifications of interest rate models

Model and authors	Drift term	Diffusion term
Merton (1973)	a	σ
Vasicek (1977)	$a + bX_t$	σ
CIR SR[a] (1985)	$a + bX_t$	$\sigma X_t^{1/2}$
Dothan (1978)	0	σX_t
GBM[b] Black–Scholes (1973)	bX_t	σX_t
Brennan–Schwartz (1977,1979,1980)	$a + bX_t$	σX_t
CIR VR[c] (1980)	0	$\sigma X_t^{3/2}$
CEV[d] Cox (1975) and Cox and Ross (1976)	bX_t	σX_t^{γ}
Aït-Sahalia (1996)	$a + bX_t$	semiparametric

[a]The Cox–Ingersoll–Ross (1985) squared-root diffusion function model.

[b]The geometric Brownian motion model introduced by Black and Scholes (1973) and also by Marsh and Rosenfeld (1983).

[c]The Cox–Ingersoll–Ross (1980) variable-rate model.

[d]The constant elasticity of variance model introduced by Cox (1975) and Cox and Ross (1976).

the marginal density is a gamma distribution in the steady state. The Black and Scholes (1973) model, also considered by Marsh and Rosenfeld (1983), is a geometric Brownian motion (GBM) process and assumes that the short-term interest rate has log-normal transitional densities. A more general parametric model for the diffusion term is the constant elasticity of variance model introduced by Cox (1975) and Cox and Ross (1976). All these parametric models are nested within the following model with various parameter restrictions:

$$dX_t = (a + bX_t)\,dt + \sigma X_t^{\gamma}\,dW_t. \tag{17}$$

The functional forms of the transitional densities corresponding to this model, however, are not known explicitly. For a more detailed comparison of the parametric models, refer to Chan et al. (1992). The Aït-Sahalia (1996) model specifies the drift term as a mean-reverting linear function, as in most of the literature, and the diffusion term as a semiparametric function of X_t. It is noted that some of the preceding processes are nonstationary (e.g., the Merton model, the Dothan model, the Black–Scholes model, the Cox–Ingersoll–Ross variable-rate model, the constant elasticity of variance model), whereas some of them are stationary in the steady state when $t_0 \to -\infty$ (e.g., the Vasicek model with $b < 0$, the Cox–Ingersoll–Ross squared-root model with $b < 0$, $X_t \geq 0$, $2a > \sigma^2$, the Brennan–Schwartz model with $a > 0$, $X_t \geq 0$). The Aït-Sahalia model is stationary by construction.

The implications of different specifications for the underlying process can be derived from our discussion in Section 2. For the diffusion process defined in (1), under regularity and certain terminal conditions, there is a unique representation of the transitional probability function $P(X_{t+h}|X_t)$ given the drift and diffusion functions $\mu(\cdot)$ and $\sigma(\cdot)$. For instance, the Brownian motion process, the Brownian motion with drift process, and the Ornstein–Uhlenbeck process all have Gaussian transitional densities determined by the parameters of the drift and diffusion functions; the GBM process has log-normal transitional densities; and the Cox–Ingersoll–Ross squared-root process has noncentral chi-square transitional probability density with X_t taking nonnegative values.

Furthermore, if the process is stationary in the strict sense, with any specification of either the drift or the diffusion term, the other will also be specified through equation (7) or (8). For instance, if we specify the drift term as a constant, that is, $\mu(X_t) = \mu$, it implies that $\sigma^2(X_t) = 2\mu[(P(X_t))/(p(X_t))]$, where $P(X_t)$ and $p(X_t)$ are, respectively, the marginal cumulative distribution and probability density functions. If we impose the linear specification of the drift term, that is, $\mu(X_t) = a + bX_t$, it implies that $\sigma^2(X_t) = [(2P(X_t))/(p(X_t))][a + bE_{Y_t}[y_t]]$, where Y_t is a random variable with probability density function $f_{Y_t}(y_t) = p(y_t)/P(X_t)$, $0 \leq y_t \leq X_t$. Similarly, any specification of the diffusion term would lead to the corresponding specification for the drift term. For instance, if we specify the diffusion term as a constant, that is, $\sigma(X_t) = \sigma$, it implies that $\mu(X_t) = (\sigma^2/2)[(p'(X_t))/(p(X_t))]$. If we specify the diffusion term as the squared root of the level of the stochastic process, that is, $\sigma(X_t) = \sigma X_t^{1/2}$, it implies that $\mu(X_t) = (\sigma^2/2)(1 + X_t[(p'(X_t))/(p(X_t))]$.

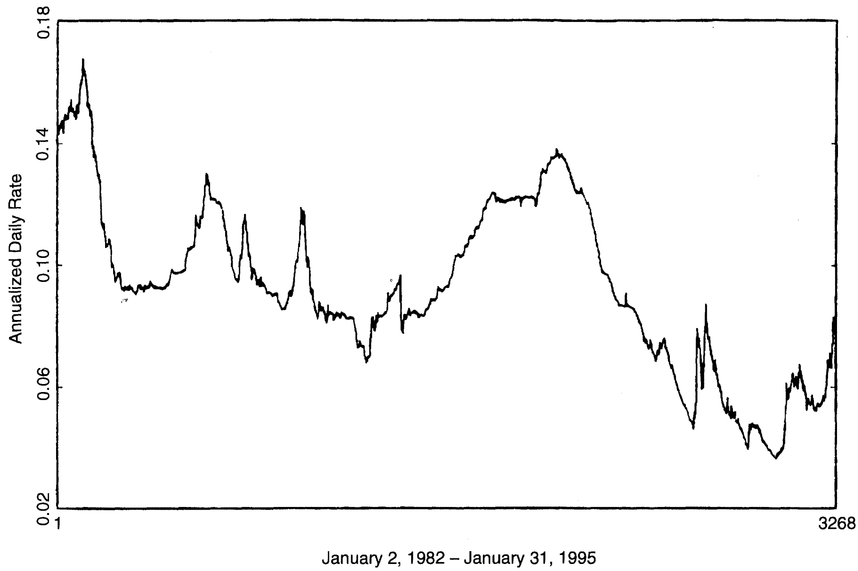

FIGURE 1. Canadian daily 3-month T-bill rate.

Most of the models specify the drift term as a linear mean-reverting function $\beta(\alpha - X_t)$ with $\beta > 0$, which was first proposed by Merton (1973). The rationale behind this specification is that the interest rate is elastically attracted to its equilibrium value; in other words, the instantaneous drift $\beta(\alpha - X_t)$ represents a force that keeps pulling the process toward its long-term mean α, with magnitude proportional to the deviation of the process from the mean. It is the diffusion element that causes the process to fluctuate around the level α in an erratic but continuous fashion. However, it is never claimed that this identification for the stochastic process represents the best description of the spot rate behavior. Actually, for any stationary diffusion process X_t, because $E[dX_t|X_t] = \mu(X_t)\,dt$ and $E[dX_t] = 0$, we can conclude that any linear specification of the drift term, $a + bX_t$ with $b < 0$, will always lead to mean reversion. The drawback of this specification is that with the diffusion term specified to describe the local behavior of the "fluctuations" of the stochastic process, the drift term thus specified complies more with the long-term feature of the stochastic process. It makes these two specifications inconsistent and thus fails to provide a desirable representation of the local-time behavior of the process. The main reasons that most models impose strong parametric restrictions on the diffusion process are the following. First, to derive the exact pricing formula for the derivative security, the explicit functional form of the Markov transitional probability density is required. Second, from an econometric point of view, little progress has been made with the identification and estimation of a general continuous-time diffusion process. Thus, most researchers have constrained their models to be simple in order to use available estimation methods.

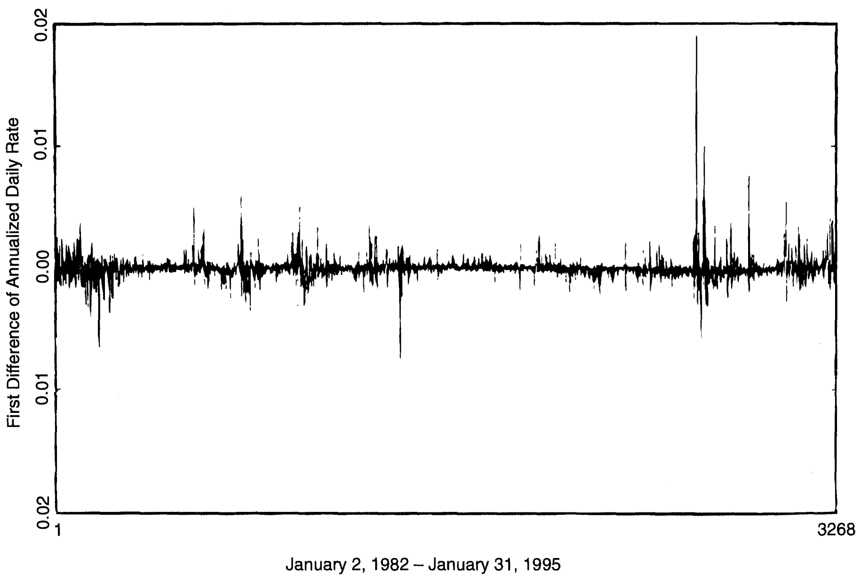

FIGURE 2. First difference of Canadian daily 3-month T-bill rate.

In Section 3, we developed nonparametric estimators for both the diffusion function and the drift function. Thus, the spot interest rate model can also be identified nonparametrically based on discretely observed data. In this section, we apply the nonparametric technique to a short-term interest rate model. The short-term interest rate used for the model estimation is the Canadian 3-month treasury bill rate. The data are daily and cover the period from January 2, 1982, to January 31, 1995, providing 3,268 observations in total.[7] All the data are quoted as the midpoints between bid and asked prices at the close of the business day to capture the volatility of the short-term interest rates from day to day. All spot rates are expressed in annualized form, representing the continuously compounded rates computed daily by the Bank of Canada.[8] Monday is taken as the first day after Friday. Although weekend effects have been documented extensively for stock prices, there does not seem to be a conclusive weekend effect in spot interest rates. The time series daily data and its first difference are plotted in Figures 1 and 2.

Table 2 shows the means, standard deviations, and part of the first 11 autocorrelations of the daily rates and the daily changes in the spot rates. The unconditional average level of the daily rate is 0.0926, with a standard deviation of 0.0274. Although the autocorrelations in the interest rate level decays very slowly, those of the day-to-day changes are generally small and are not consistently positive or negative. The results of a formal augmented Dickey–Fuller nonstationarity test are also reported in Table 2. The null hypothesis of nonstationarity is rejected at the 10% significance level. Because the test is known to have low power, which

TABLE 2. Summary statistics of the data and stationarity test[a]

	Summary statistics								
	N	Mean	SD	ρ_1	ρ_3	ρ_5	ρ_7	ρ_9	ρ_{11}
X_t	3,268	0.0926	0.0274	0.999	0.996	0.993	0.990	0.986	0.983
$X_{t+1}-X_t$	3,267	$-0.19E-6$	0.92E-5	0.145	0.075	0.118	-0.028	0.083	-0.047

Augmented Dickey–Fuller stationarity test

H_0:	Test statistic	Critical value (10%)
Nonstationarity	-2.72	-2.57

[a] X_t denotes the daily 3-month treasury bill rate. $X_{t+1} - X_t$ denotes day-to-day change of the daily 3-month treasury bill rate. ρ_j denotes the autocorrelation coefficient of order j. The augmented Dickey–Fuller test statistic is computed as $\hat{\tau} = \hat{\alpha}_1/ase(\hat{\alpha}_1)$ in the model $\Delta X_t = \alpha_0 + \alpha_1 X_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta X_{t-j} + \epsilon_t$ (see, e.g., Harvey, 1993). The value of p is set as the highest significant lag order (using an approximate 95% confidence interval) from either the autocorrelation function or the partial autocorrelation function of the first differenced series (up to a maximum lag order of $\sqrt{n}$). The justification for using the Dickey–Fuller table when the residuals are heteroskedastic and possibly serially dependent is provided by Phillips (1987).

is the probability of rejecting the null hypothesis when it is not true, a rejection suggests that stationarity of the series is very likely.

The nonparametric kernel estimator of the marginal density of the daily spot interest rate, with 99% pointwise confidence band, is reported in Figure 3. The

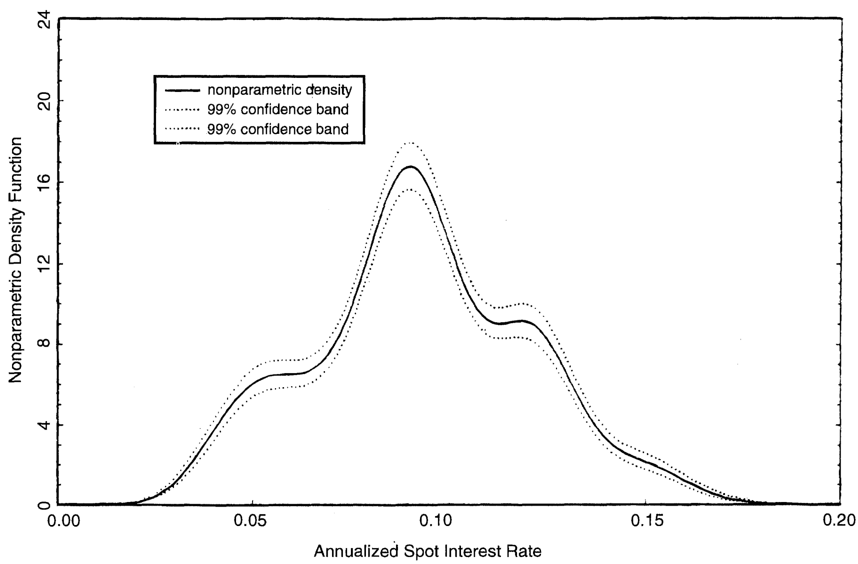

FIGURE 3. Nonparametric marginal density of spot interest rate.

nonparametric density is estimated with Gaussian kernel and the optimal window width, which minimizes the IMSE. The large sample size makes it possible to estimate the density very precisely. Noticeable features of the nonparametric density function include its nonnormality and the flat tails on both sides. It could be interpreted as a mixture of four unimodal densities with modes around the values of 0.06, 0.09, 0.12, and 0.15. Two unimodal densities on the tails look flatter than the other two in the center. Most of the low observations in the data set were recorded after 1992, and most of the high observations in the data set were recorded before 1983 and around 1989. Whether these subperiods should be included in the sample depends on one's view regarding the likelihood that these low or high interest rates would occur again in the future. Figure 4 also reports the Gaussian and gamma marginal densities of the spot rate corresponding to the Ornstein–Uhlenbeck model and Cox–Ingersoll–Ross squared-root model; the respective parameter values are obtained from the GMM estimation of the models. Compared to the Gaussian and gamma densities, the nonparametric density exhibits more variations and two longer tails for low and high values of interest rates.

Figure 5 plots the nonparametric estimator, with 90% pointwise confidence band, of the diffusion function of the short-term interest rate model. It has three important features. First, the 90% pointwise confidence band, which reflects the quality of the estimation, is narrower in the middle but tends to get wider dramatically toward the two ends for the lack of enough observations around the high and low levels of interest rates. Second, the diffusion function exhibits noticeable variations from low to high values, which provides the evidence to reject

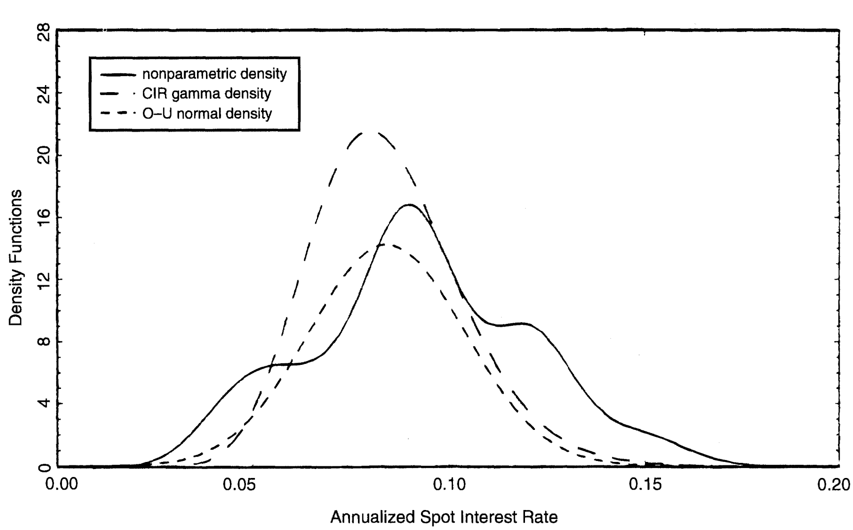

FIGURE 4. Marginal density functions of spot interest rate: nonparametric, Ornstein–Uhlenbeck normal, and Cox–Ingersoll–Ross gamma densities.

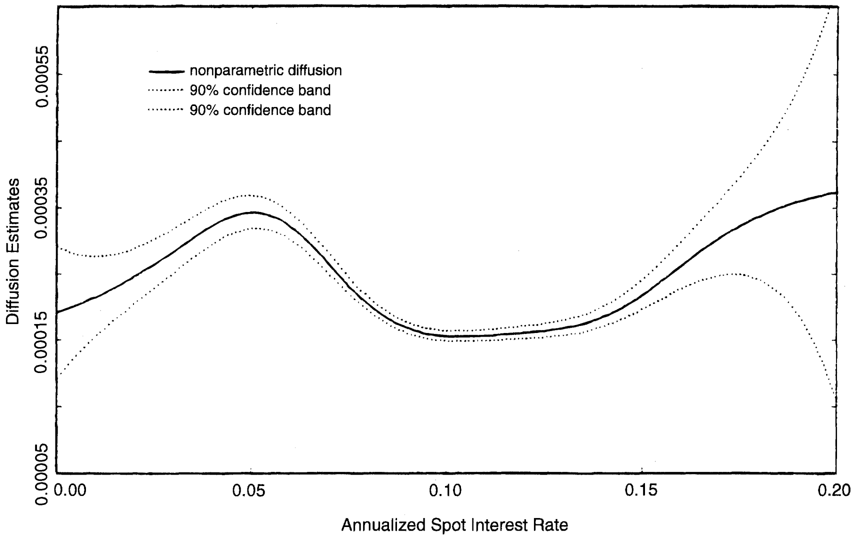

FIGURE 5. Nonparametric diffusion function.

the constant or flat diffusion function specification as in the Vasicek (1977) model. Third, not only does the diffusion function not look flat, it does not look linear either. Rather, it shows the shape of a "saddle" or "smile," with both low and high interest rates showing higher volatilities and the medium interest rates showing relatively lower volatilities. This provides the evidence to reject the Cox–Ingersoll–Ross (1985) squared-root model specification, which expects that high interest rates should vary more than low interest rates.

Given the estimated nonparametric marginal density function of the spot interest rate and the nonparametric diffusion function, the estimated nonparametric drift function is given in Figure 7. The pointwise confidence band of the drift function estimator is estimated here using the boot-strapping method. Compared to the linear mean reverting specification in the Ornstein–Uhlenbeck process, the Cox–Ingersoll–Ross squared-root process, and the Aït-Sahalia (1996) semiparametric process, the nonparametric drift function shows more variations and exhibits no consistent overall mean reverting property. Mean reversion occurs only when interest rate levels are very high or very low. Table 3 and Figures 6 and 8 give the comparisons of the estimates of the diffusion function and drift function for different models with different estimation methods, that is, the GMM estimation results for the Ornstein–Uhlenbeck process, the GMM estimation results for the Cox–Ingersoll–Ross squared-root model, the nonlinear least-squares and semiparametric estimation results for the Aït-Sahalia (1996) semiparametric model, with the nonparametric estimation results of both the diffusion function and drift function for the nonparametric model.

TABLE 3. Estimates of alternative short-term interest rate models[a]

Model	Estimation method	Drift function	Diffusion function
Ornstein–Uhlenbeck process	GMM	$\alpha = 0.0628$ (3.387) $\beta = 0.3429$ (1.906)	$\sigma^2 = 2.106\ 10^{-4}$ (9.521)
Cox–Ingersoll–Ross process	GMM	$\alpha = 0.0524$ (1.095) $\beta = 0.2407$ (1.356)	$\sigma^2 = 1.961\ 10^{-3}$ (9.362)
Aït-Sahalia (1996) semiparametric process	Nonlinear least squares and semiparametric	$\alpha = 0.0762$ (7.228) $\beta = 0.4636$ (2.813)	Semiparametric diffusion (Figure 6)
Nonparametric process	Nonparametric	Nonparametric drift (Figure 7)	Nonparametric diffusion (Figure 5)

[a]The numbers in parentheses are t-ratios of the preceding estimates.

5. CONCLUSION

In this paper, we have proposed a consistent nonparametric diffusion function kernel estimator with asymptotic normal mixture distribution for a general Itô diffusion process under mild regularity conditions. The estimator is developed without imposing any restrictions on the functional form of the drift term, so that it captures the true volatilities of the process. Under the condition that the process

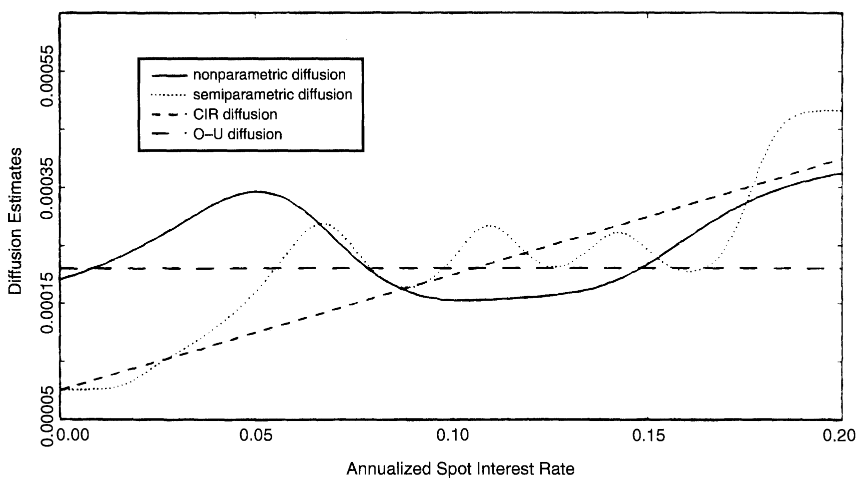

FIGURE 6. Alternative diffusion functions: nonparametric, semiparametric, linear, and constant diffusion functions.

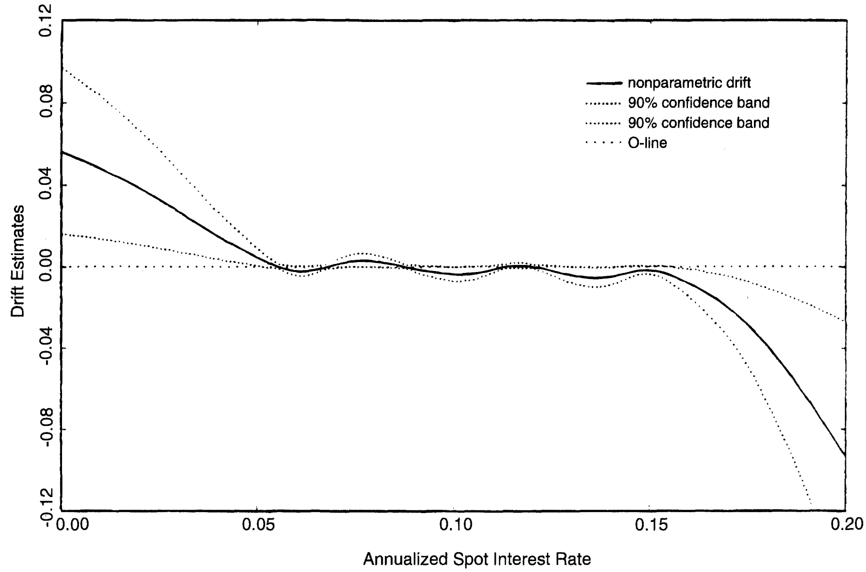

FIGURE 7. Nonparametric drift function.

is strictly stationary or approaches a limiting stationary distribution, a consistent nonparametric drift function estimator is also derived based on the nonparametric diffusion function estimator and the nonparametric kernel estimator of the marginal density function.

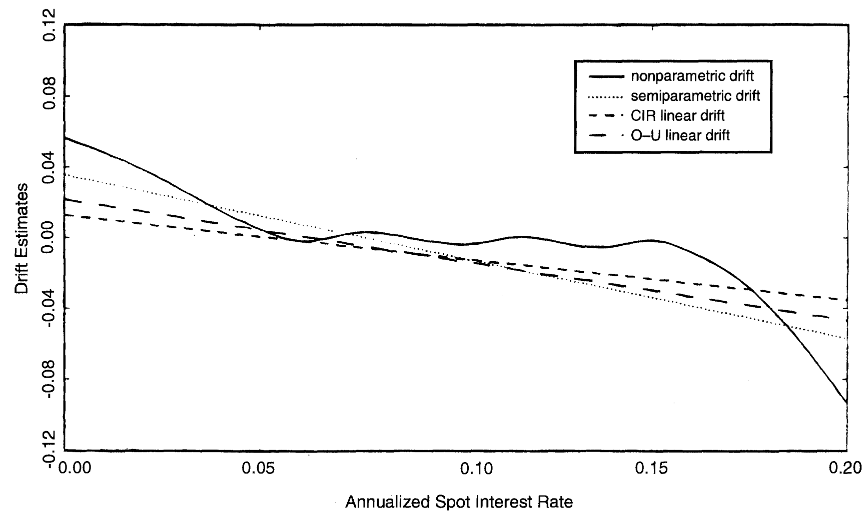

FIGURE 8. Alternative drift functions: nonparametric, semiparametric, Cox–Ingersoll–Ross, and O-U drift functions.

An application of the nonparametric identification and estimation technique for the diffusion process is undertaken to a short-term interest rate model involving the Canadian daily 3-month treasury bill rate. The nonparametric estimators of the drift function and diffusion function provide evidence for rejecting the common parametric specifications of both functions. This suggests that it will be worthwhile to apply the nonparametric technique to identify and estimate the underlying stochastic process of asset prices or interest rates in order to more precisely price derivative securities, evaluate contingent claims, design optimal hedging strategies, and so on. Further research can also be undertaken in extending the nonparametric identification and estimation technique to the multidimensional Itô diffusion process.

NOTES

1. This property is of primary importance in the modeling of the spot interest rate process because, as Chan et al. (1992) also pointed out, the volatility of the riskless interest rate is one of the key factors determining the value of contingent claims. In addition, optimal hedging strategies for risk-averse investors depend critically on term structure volatility.
2. We thank Peter Phillips and one of the referees for bringing to our attention the papers by Monfort (1996) and Stanton (1996).
3. Much earlier, Bartlett (1946) considered the problem of estimating the parameters of the general single first- and second-order SDE's from samples of discrete observations. He showed that estimates obtained through discretization are seriously biased and the bias does not tend to zero as the observation interval does.
4. The rate of convergence to the normal law is a random variable linked to the local time of the diffusion or to its suitable discrete approximation. This can also be interpreted as a convergence to a mixture of normal law.
5. As one of the referees points out, according to variance, the optimal kernel is the indicator function by Gaussian-Markov–type arguments, but according to MSE, a general kernel function is usually superior to the indicator function. Further study could be undertaken to characterize the asymptotic bias and variance of the diffusion function estimator in order to obtain the optimal bandwidth to balance bias and standard deviation.
6. Banon and Nguyen (1981) showed that a stationary Markov process satisfying a certain mixing condition, namely, the G_2 condition of Rosenblatt (1971), can be interpreted as an asymptotically uncorrelated process. The condition G_2 is used to obtain asymptotic properties of the marginal density estimator based on continuous observations.
7. It is believed that there is no regime switch for Canadian monetary policy over this period.
8. It is noted that the measurement error of the daily rates due to different maturity lengths within 1 week is minimal and can be ignored.

REFERENCES

Aït-Sahalia, Y. (1996) Nonparametric pricing of interest rate derivative securities. *Econometrica* 64 (3), 527–560.

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

Banon, G. & H.T. Nguyen (1981) Recursive estimation in diffusion models. *SIAM Journal of Control and Optimization* 19, 676–685.

Bartlett, M.S. (1946) On the theoretical specification and sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society Supplement* 8, 27–41.

Black, F. & M. Scholes (1973) The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637–654.

Brennan, M.J. & E.S. Schwartz (1977) Saving bonds, retractable bonds, and callable bonds. *Journal of Financial Economics* 5, 67–88.

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

Brennan, M.J. & E.S. Schwartz (1980) Analyzing convertible bonds. *Journal of Financial and Quantitative Analysis* 15, 907–929.

Brown, B.M. & J.I. Hewitt (1975) Asymptotic likelihood theory for diffusion process. *Journal of Applied Probability* 12, 228–238.

Chan, K.C., G.A. Karolyi, F.A. Longstaff, & A.B. Sanders (1992) An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* XLVII (3), 1209–1227.

Conley, T.G., L.P. Hansen, E.G.J. Luttmer, & J.A. Scheinkman (1995) Short-Term Interest Rate as Subordinated Diffusions. Mimeo, University of Chicago.

Cox, J.C. (1975) Notes on Option Pricing I: Constant Elasticity of Variance Diffusion. Working Paper, Stanford University.

Cox, J.C., J.E. Ingersoll, & S.A. Ross (1980) An analysis of variable rate loan contracts. *Journal of Finance* XXXV (2), 389–403.

Cox, J.C., J.E. Ingersoll, & S.A. Ross (1985) A theory of the term structure of interest rates. *Econometrica* 53 (2), 385–407.

Cox, J.C. & S.A. Ross (1976) The valuation of options for alternative stochastic processes. *Journal of Financial Economics* 3, 145–166.

Dacunha-Castelle, D. & D. Florens-Zmirou (1986) Estimation of the coefficient of a diffusion from discrete observations. *Stochastics* 19, 263–284.

Delgado, M.A. & P.M. Robinson (1992) Nonparametric and semiparametric methods for economic research. *Journal of Economic Surveys* 6(3), 201–249.

Dohnal, G. (1987) On estimating the diffusion coefficient. *Journal of Applied Probability* 24, 105–114.

Doob, J.L. (1953) *Stochastic Processes.* New York: John Wiley & Sons.

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

Duffie, D. & K.J. Singleton (1993) Simulated moments estimation of Markov models for asset prices. *Econometrica* 61, 929–952.

Florens-Zmirou, D. (1993) On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability* 30, 790–804.

Geman, S.A. (1979) On a Common Sense Estimator for the Drift of a Diffusion. Pattern Analysis 79, Division of Applied Mathematics, Brown University, Providence, Rhode Island.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

Hansen, L.P. & T.J. Sargent (1983) The dimensionality of the aliasing problem. *Econometrica* 51, 377–388.

Hansen, L.P. & J.A. Scheinkman (1995) Back to the future: Generating moment implications for continuous-time Markov processes. *Econometrica* 63 (4), 767–804.

Harvey, A.C. (1993) *Time Series Models*, 2nd ed. Cambridge: MIT Press.

Ingersoll, J.E., Jr. (1987) *Theory of Financial Decision Making.* London: Rowman & Littlefield.

Jiang, G.J. & J.L. Knight (1996) Parametric Versus Nonparametric Estimation of Diffusion Processes—A Monte Carlo Comparison. Mimeo, University of Western Ontario.

Karlin, S. & H.M. Taylor (1981) *A Second Course in Stochastic Processes.* New York: Academic Press.

Knight, F.B. (1976) *A Reduction of Continuous Square Integrable Martingales to Brownian Motion.* Lecture Notes in Mathematics 190. Berlin: Springer-Verlag.

Kumar, T.K. & J.M. Markman (1975) Estimation of Probability Density Function: A Monte Carlo Comparison of Parametric and Nonparametric Methods. Mimeo.

Kutoyants, Y. (1984) *Parameter Estimation for Stochastic Processes.* Heldermann.
Lanska, V. (1979) Minimum contrast estimation in diffusion processes. *Journal of Applied Probability* 16, 65–75.
Lo, A.W. (1988) Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory* 4, 231–247.
Marsh, T.A. & E.R. Rosenfeld (1983) Stochastic processes for interest rates and equilibrium bond prices. *Journal of Finance* 38, 635–646.
Merton, R.C. (1973) Theory of rational option pricing. *Bell Journal of Economics and Management Science* 4, 141–183.
Monfort, A. (1996) A reappraisal of misspecified econometric models. *Econometric Theory* 12. 597–619.
Øksendal, B. (1992) *Stochastic Differential Equations.* New York: Springer-Verlag.
Pedersen, A.R. (1995) Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* 1 (3), 257–279.
Pham Dinh, T.D. (1981) Non-parametric estimator of the drift coefficient in the diffusion equation. *Math. Operations Statist. Ser Statist.* 12, 61–73.
Phillips, P.C.B. (1973) The problem of identification in finite parameter continuous-time models. *Journal of Econometrics* 1, 351–362.
Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55 (2), 277–310.
Prakasa Rao, B.L.S. (1983) *Nonparametric Functional Estimation.* London: Academic Press.
Prakasa Rao, B.L.S. (1990) Nonparametric density estimation for stochastic processes from sampled data. *Public Inst. Statist. Univ. Paris* XXXV (3), 51–83.
Prohorov, Yu.V. & Yu.A. Rozanov (1969) *Probability Theory.* Translated by K. Krickeberg & H. Urmitzer, New York: Springer-Verlag.
Rosenblatt, M. (1971) *Markov Processes, Structure and Asymptotic Behavior.* New York: Springer-Verlag.
Silverman, B.W. (1986) *Density Estimation for Statistical and Data Analysis.* London: Chapman-Hall.
Stanton, R. (1996) A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. Working paper, Haas School of Business, University of California, Berkeley.
Vasicek, O. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.

APPENDIX: PROOFS OF THEOREMS AND COROLLARY

A.1. Proof of Theorem 1. To prove Theorem 1, we need to use the expansion of the transitional density of the underlying Markov process and to first prove two propositions. For the diffusion process described by SDE (1) with $t \in [0,1]$ (i.e., without loss of generality, we let $T = 1$; therefore, $\Delta_n = 1/n$) and transitional density $p_t(x;y)$, let $s(x) = \int_0^x (dy/\sigma(y))$ and $g(x) = s^{-1}(x)$ be the inverse function of $s(x)$. The process $Y_t = s(X_t)$ satisfies the equation $dY_t = b(Y_t)\,dt + dW_t$, with $b = (\mu/\sigma) \circ g - \frac{1}{2}\sigma' \circ g$, which is a bounded function of class C^2 under condition A5 about $\mu(\cdot)$ and $\sigma(\cdot)$. From this and an expansion of the transition density $p_t(x;y)$ in y—for example, used in Dacunha-Castelle and Florens-Zmirou (1986)—we obtain

$$p_t(x;y) = \frac{1}{\sqrt{2\pi t}}\frac{1}{\sigma(y)}\exp\left(-\frac{(s(y)-s(x))^2}{2t}\right)U_t(s(x),s(y)) \qquad \textbf{(A.1.1)}$$

with $U_t(x,y) = H_t(x,y)\exp[G(y) - G(x)]$ and

$$H_t(x,y) = E\left[\exp\left(-t\int_0^1 h(x + z(y-x) + \sqrt{t}B_z)\,dz\right)\right],$$

where B_z is a standard Brownian bridge, $G(x) = \int_0^x b(u)\,du$, and $h = \frac{1}{2}(b^2 + b')$. We now introduce the following notation,

$$V_t^n(x) = \frac{1}{h_n}\sum_{i=1}^{[nt]-1} K\left(\frac{X_{i/n} - x}{h_n}\right)[X_{(i+1)/n} - X_{i/n}]^2,$$

$$V^n(x) = V_1^n(x), \qquad V^n = V^n(0), \qquad L^n = L_1^n(0),$$

and prove the following two propositions; the structure of proof follows that of Florens-Zmirou (1993).

PROPOSITION 1. *If the sequence h_n is such that nh_n^4 tends to zero as n tends to infinity, then $V_t^n(x)$ converges, in the $\mathcal{L}^2$ sense, to $\sigma^2(x)L_t(x)$. That is, $S_n(x)$ is a consistent estimator of $\sigma^2(x)$.*

Proof. Under assumption A8 on $K(\cdot)$, considering the fact that the local-time random variable $L_t(x)$ can be constructed as the density of the occupation time $L_A(t)$ in the sense that $L_A(t) = \int_A L_t(u)\,du$, it is easy to see that $L_t^n(x)$ converges a.s. to $L_t(x)$ when n tends to infinity and h_n tends to zero (sufficiently h_n tends to zero and nh_n tends to infinity); therefore, we shall only need to prove that

$$\lim_{n\to\infty} E_{x_0}[V_t^n(x) - \sigma^2(x)L_t^n(x)]^2 = 0,$$

that is, $V_t^n(x)$ converges in mean squared error to $\sigma^2(x)L_t^n(x)$ and without loss of generality, and take $x = 0$ and prove that

$$\lim_{n\to\infty} E_{x_0}[V^n - \sigma^2(0)L^n]^2 = 0.$$

We set

$$E_{x_0}[V^n - \sigma^2(0)L^n]^2 = \frac{1}{n^2h_n^2}I_n,$$

where

$$I_n = E_{x_0}\left[\sum_i K\left(\frac{X_{i/n}}{h_n}\right)\{n(X_{(i+1)/n} - X_{i/n})^2 - \sigma^2(0)\}\right]^2.$$

Expanding I_n, we can write $I_n = \sum_{i,j} I_n(i,j)$. For $j > i+1$,

$$I_n(i,j) = \int_{-\infty}^{\infty} p_{i/n}(x_0;x)K\left(\frac{x}{h_n}\right)dx\left[\int_{-\infty}^{\infty} p_{1/n}(x,y)[n(y-x)^2 - \sigma^2(0)]\,dy\right]$$
$$\times \int_{-\infty}^{\infty} p_{(j-i-1)/n}(y;z)K\left(\frac{z}{h_n}\right)dz\left[\int_{-\infty}^{\infty} p_{1/n}(z,v)[n(v-z)^2 - \sigma^2(0)]\,dv\right].$$

From the definition of $s(\cdot)$ and the assumptions on $\sigma(\cdot)$, we have $s(\infty) = \infty$, $s(-\infty) = -\infty$. Using the expansion of transitional density in (A.1.1) and through integration transformation, we have

$$
\begin{aligned}
I_n(i,j) = &\int_{-\infty}^{\infty} p_{i/n}(x_0;x)g'(x)K\left(\frac{x}{h_n}\right)dx \\
&\times \int_{-\infty}^{\infty} \exp(-u^2/2)\,U_{1/n}(x,x+u/\sqrt{n})[n(g(x+u/\sqrt{n})-g(x))^2-\sigma^2(0)]\,du \\
&\times \int_{-\infty}^{\infty} p_{(j-i-1)/n}(g(x+u/\sqrt{n});g(z))K\left(\frac{z}{h_n}\right)dz \\
&\times \int_{-\infty}^{\infty} \exp(-v^2/2)\,U_{1/n}(z,z+v/\sqrt{n})[n(g(z+v/\sqrt{n})-g(z))^2-\sigma^2(0)]\,dv.
\end{aligned}
$$

Because $\lim_{x\to+\infty}K(x/h_n) = \lim_{x\to-\infty}K(x/h_n) = 0$, and the second and fourth integrals of the preceding expression are finite, we can modify the preceding expression by replacing the bounds of the first and third integrals with $-Ch_n$ and Ch_n ($C > 0$) and define $I_n(C,i,j)$ as

$$
\begin{aligned}
I_n(C,i,j) = &\int_{-Ch_n}^{Ch_n} p_{i/n}(x_0;x)g'(x)K\left(\frac{x}{h_n}\right)dx \\
&\times \int_{-\infty}^{\infty} \exp(-u^2/2)\,U_{1/n}(x,x+u/\sqrt{n})[n(g(x+u/\sqrt{n})-g(x))^2-\sigma^2(0)]\,du \\
&\times \int_{-Ch_n}^{Ch_n} p_{(j-i-1)/n}(g(x+u/\sqrt{n});g(z))K\left(\frac{z}{h_n}\right)dz \\
&\times \int_{-\infty}^{\infty} \exp(-v^2/2)\,U_{1/n}(z,z+v/\sqrt{n})[n(g(z+v/\sqrt{n})-g(z))^2-\sigma^2(0)]\,dv
\end{aligned}
$$

such that for every $\epsilon_n > 0$ and $h_n > 0$ there exists a constant $C > 0$ to satisfy

$$
|I_n(i,j) - I_n(C,i,j)| < \epsilon_n \qquad \textbf{(A.1.2)}
$$

with $\sum_n \epsilon_n < +\infty$. Furthermore, let $\alpha < \frac{1}{2}$. Denote by $I_n(C,\alpha,i,j)$ the preceding expression with the bounds of the second and the fourth integrals replaced by the bounds $-n^\alpha$ and n^α. That is,

$$
\begin{aligned}
I_n(C,\alpha,i,j) = &\int_{-Ch_n}^{Ch_n} p_{i/n}(x_0;x)g'(x)K\left(\frac{x}{h_n}\right)dx \\
&\times \int_{-n^\alpha}^{n^\alpha} \exp(-u^2/2)\,U_{1/n}(x,x+u/\sqrt{n})[n(g(x+u/\sqrt{n})-g(x))^2-\sigma^2(0)]\,du \\
&\times \int_{-Ch_n}^{Ch_n} p_{(j-i-1)/n}(g(x+u/\sqrt{n});g(z))K\left(\frac{z}{h_n}\right)dz \\
&\times \int_{-n^\alpha}^{n^\alpha} \exp(-v^2/2)\,U_{1/n}(z,z+v/\sqrt{n})[n(g(z+v/\sqrt{n})-g(z))^2-\sigma^2(0)]\,dv.
\end{aligned}
$$

As a result of assumption A5, that is, $\mu(\cdot)$ and $\sigma(\cdot)$ and their derivatives are bounded, it is easy to show that

$$|I_n(C,i,j) - I_n(C,\alpha,i,j)| \leq O(\exp(-C_1 n^{2\alpha})), \tag{A.1.3}$$

where C_1 is a positive constant independent of i and j. Because $K(\cdot)$ is bounded over $(-\infty,\infty)$, let $K = \sup_{-\infty<x<+\infty} K(x)$; we have

$$\begin{aligned} I_n(C,\alpha,i,j) \leq K^2 & \int_{-Ch_n}^{Ch_n} p_{i/n}(x_0;x) g'(x)\, dx \\ & \times \int_{-n^\alpha}^{n^\alpha} \exp(-u^2/2) U_{1/n}(x, x + u/\sqrt{n}) \\ & \qquad \times [n(g(x + u/\sqrt{n}) - g(x))^2 - \sigma^2(0)]\, du \\ & \times \int_{-Ch_n}^{Ch_n} p_{(j-i-1)/n}(g(x + u/\sqrt{n}); g(z))\, dz \\ & \times \int_{-n^\alpha}^{n^\alpha} \exp(-v^2/2) U_{1/n}(z, z + v/\sqrt{n})[n(g(z + v/\sqrt{n}) - g(z))^2 - \sigma^2(0)]\, dv. \end{aligned}$$

Up to this point, we can use the results in Lemma 1 and Proposition 3 of Florens-Zmirou (1993) and have

$$\frac{1}{n^2 h_n^2} \sum_{j>i+1} I_n(C,\alpha,i,j) = O(nh_n^4 + h_n^2). \tag{A.1.4}$$

Similarly, for $j < i - 1$, we have

$$\frac{1}{n^2 h_n^2} \sum_{j<i-1} I_n(C,\alpha,i,j) = O(nh_n^4 + h_n^2). \tag{A.1.5}$$

In the case $j = i$,

$$I_n(i,i) = \int_{-\infty}^{\infty} p_{i/n}(x_0,x) K\left(\frac{x}{h_n}\right) dx \int_{-\infty}^{\infty} p_{i/n}(x,y)[n(y-x)^2 - \sigma^2(0)]^2\, dy;$$

we can show in the same way that

$$I_n(C,\alpha,i,i) = O(1) h_n n^{1/2} i^{-1/2}$$

and consequently

$$\frac{1}{n^2 h_n^2} \sum_i I_n(C,\alpha,i,i) = O\left(\frac{1}{nh_n}\right). \tag{A.1.6}$$

Similarly, an application of Hölder's inequality proves that

$$\frac{1}{n^2 h_n^2} \sum_i |I_n(C,\alpha,i,i+1)| = O\left(\frac{1}{nh_n}\right). \tag{A.1.7}$$

Combining (A.1.2)–(A.1.3) and (A.1.4)–(A.1.7), we complete the proof of Proposition 1.

Furthermore, we introduce the following notations; for any $0 < t \leq 1$,

$$N_x^n = \sum_{i=1}^{[nt]} K\left(\frac{X_{i/n} - x}{h_n}\right),$$

$$m_{i+1} = \sqrt{\frac{n}{h_n}}\, K\left(\frac{X_{i/n} - x}{h_n}\right)[(X_{(i+1)/n} - X_{i/n})^2 - \sigma^2(x)/n]$$

$$M_t^n = \sum_{i=0}^{[nt]-1} m_{i+1} \tag{A.1.8}$$

$$W_t^n = \sum_{i=0}^{[nt]-1} (W_{(i+1)/n} - W_{i/n}) \tag{A.1.9}$$

where $W_t(\cdot)$ is a standard Brownian motion or Wiener process. We shall prove the following.

PROPOSITION 2. *If the sequence h_n is such that nh_n^3 tends to zero as n tends to infinity, the sequence of processes $(M_t^n, W_t^n)_{0<t\leq 1}$ defined by* (A.1.8) *and* (A.1.9) *converges in distribution to the process $(B_{\sigma^4(x)L_t(x)}, W_t)_{0<t\leq 1}$, where B_t and W_t are two independent, standard Brownian motions.*

Proof. Denoting by $E_{i,n}$ the conditional expectation with respect to the filtration $\mathcal{F}_{i/n}^X = \sigma(X_u; u \leq i/n)$, which is generated by X_t up to i/n, we need to use the following results: If h_n is a sequence such that nh_n^3 tends to zero as n tends to infinity, then

(a) $\sum_{i=0}^{[nt]-1} E_{i,n}(m_{i+1})$ converges in probability to zero,
(b) $\sum_{i=0}^{[nt]-1} E_{i,n}(m_{i+1}^2)$ converges in probability to $\sigma^4(x)L_t(x)$,
(c) $\sum_{i=0}^{[nt]-1} E_{i,n}|m_{i+1}|^3$ converges in probability to zero, and
(d) $\sum_{i=0}^{[nt]-1} E_{i,n}(w_{i+1}m_{i+1})$ converges in probability to zero, where $w_{i+1} = W_{(i+1)/n} - W_{i/n}$.

These results directly follow Lemma 2 of Florens-Zmirou (1993), simply replacing the indicator function by $K(\cdot)$ in the proof. The proof makes use of the Burkholder–Davis–Gundy inequality for the semimartingale $(X_{i/n+s} - X_{i/n})_{0\leq s\leq 1/n}$. From (a)–(c), we can verify that M_t^n converges to the martingale M_t with increasing process $\langle M_t \rangle = \sigma^4(x)L_t(x)$. Then, we can write $M_t = B_{\sigma^4(x)L_t(x)}$, where B_t is a Brownian motion process. If $\tau_t = \inf\{u/\sigma^4(x)L_u(x) > t\}$, then $B_t = M_{\tau_t}$. From Knight's (1971) theorem and (d), we have that B_t and W_t are independent Brownian motions. This completes the proof of Proposition 2.

We now have shown that under conditions A1–A5 and A8, sufficiently as $h_n \to 0$, $nh_n \to +\infty$, and $nh_n^3 \to 0$, $S_n(x)$ is a pointwise consistent estimator of $\sigma^2(x)$. To complete the proof of Theorem 1, we further show the convergence in distribution of $S_n(x)$ as n tends to infinity. We denote by $\mathcal{F}_t^W$ the filtration generated by W; assumptions A1 and A2 imply that $\mathcal{F}_t^X \in \mathcal{F}_t^W$. Therefore, following Proposition 2, we deduce that B_t and $L_t(x)$ are mutually independent. From the preceding, we also have that M_1^n converges in distribution to $\sqrt{L(x)}\sigma^2(x)Z$, where Z is a normal random variable independent of $L(x)$, as we can write

$$\sqrt{N_x^n}\left(\frac{S_n(x)}{\sigma^2(x)} - 1\right) = \frac{M_1^n}{\sigma^2(x)\sqrt{L^n(x)}}. \tag{A.1.10}$$

Because $L^n(x)$ converges in probability to $L(x)$, from (A.1.10) we can conclude that $M_1^n/(\sigma^2(x)\sqrt{L^n(x)})$ and therefore $\sqrt{N_x^n}([(S_n(x))/(\sigma^2(x))] - 1)$ converges in distribution to

Z. Alternatively, we can state that under the condition nh_n^3 tends to zero as n tends to infinity, then $\sqrt{nh_n}([(S_n(x))/(\sigma^2(x))] - 1)$ converges in distribution to $L(x)^{-1/2}Z$, where Z is a standard normal variable independent of $L(x)$. Because $V[S_n(x)] = (\sigma^4(x))/(nh_nL(x))$, its consistent estimator follows directly. ■

A.2. Proof of Theorem 2. For the diffusion process defined by (1) with $t \in (0,\infty)$ and marginal density $p_X(\cdot) = p(\cdot)$, follow Banon and Nguyen (1981); under conditions A1–A8, the continuous kernel estimators $[\int_0^t h(s)H(h(s))\,ds]^{-1}\int_0^t H(h(s))K((X_s - x)/(h(s)))\,ds$ and $[\int_0^t h(s)H(h(s))\,ds]^{-1}\int_0^t [(H(h(s)))/(h(s))]K'((X_s - x)/(h(s)))\,ds$ are the strong consistent estimators of $p(x)$ and $p'(x)$, respectively. That is,

$$\left[\int_0^t h(s)H(h(s))\,ds\right]^{-1}\int_0^t H(h(s))K\left(\frac{X_s - x}{h(s)}\right)ds \xrightarrow{\text{a.s.}} p(x), \tag{A.2.1}$$

$$\left[\int_0^t h(s)H(h(s))\,ds\right]^{-1}\int_0^t \frac{H(h(s))}{h(s)}K'\left(\frac{X_s - x}{h(s)}\right)ds \xrightarrow{\text{a.s.}} p'(x), \tag{A.2.2}$$

as $t \to +\infty$, where $h(\cdot): \mathcal{R}^+ \to \mathcal{R}^+ - \{0\}$ and $H(\cdot): \mathcal{R}^+ - \{0\} \to \mathcal{R}^+$ such that

(a) $h(s) \to 0$ as $s \to +\infty$,
(b) $h(\cdot)H(h(\cdot))$ is locally integrable on $\mathcal{R}^+$ and

$$\int_0^t h(s)H(h(s))\,ds \to +\infty$$

as $t \to +\infty$, and
(c) $H(h(s)) \le C_1 h(s)^k$, for $h(s) > 0$, with $k \in [-1,\frac{1}{2}]$,

$$\int_0^t h(s)H(h(s))\,ds = C_2 t^p, h(t)^{2k-1} = O(t^q)$$

as $t \to +\infty$, with $p \in ((q/2) + \frac{3}{4},1]$ and $q \in [0,\frac{1}{2}]$, where C_1, C_2 are constants.

Given equispaced discrete observations of X_t at $\{t_1 = \Delta_n, t_2 = 2\Delta_n, \ldots, t_n = n\Delta_n\}$ in the time interval $[0,T]$, with $\Delta_n = T/n$ and $T \ge T_0 > 0$, where T_0 is a positive constant, we define $H(h(s)) = 1/h(s)$ as in Banon (1978) and construct the following discrete kernel estimators $\sum_{i=1}^n (1/nh_n)K((X_{i\Delta_n} - x)/h_n)$ and $\sum_{i=1}^n (1/nh_n^2)K'((X_{i\Delta_n} - x)/h_n)$ as approximations of the continuous kernel estimators. With assumption A8 on $K(\cdot)$, it is easy to show that

$$\sum_{i=1}^n \frac{1}{nh_n}K\left(\frac{X_{i\Delta_n} - x}{h_n}\right) \xrightarrow{\text{a.s.}} \left[\int_0^t h(s)H(h(s))\,ds\right]^{-1}\int_0^t H(h(s))K\left(\frac{X_s - x}{h(s)}\right)ds, \tag{A.2.3}$$

$$\sum_{i=1}^n \frac{1}{nh_n^2}K'\left(\frac{X_{i\Delta_n} - x}{h_n}\right) \xrightarrow{\text{a.s.}} \left[\int_0^t h(s)H(h(s))\,ds\right]^{-1}\int_0^t \frac{H(h(s))}{h(s)}K'\left(\frac{X_s - x}{h(s)}\right)ds, \tag{A.2.4}$$

as Δ_n and h_n tend to zero (sufficiently $h_n \to 0$, $nh_n \to +\infty$ for finite T).

By combining (A.2.1)–(A.2.2) and (A.2.3)–(A.2.4), it can be shown by contradiction that $\sum_{i=1}^n (1/nh_n)K((X_{i\Delta_n} - x)/h_n)$ and $\sum_{i=1}^n (1/nh_n^2)K'((X_{i\Delta_n} - x)/h_n)$ are consistent estimators of $p(x)$ and $p'(x)$ as $\Delta_n \to 0$, $h_n \to 0$, and $n\Delta_n \to +\infty$, or sufficiently $\Delta_n \to 0$, $h_n \to 0$, $nh_n \to +\infty$, and $n\Delta_n \to +\infty$. So is $q_n(x)$ a consistent estimator of $Q(x)$ conditioning on

x is visited by X_t and $p(x) \neq 0$. The convergence in distribution and the asymptotic variance of $\hat{p}(\cdot)$ as well as its consistent estimator simply follow the standard results of nonparametric density estimator (see, e.g., Delgado and Robinson, 1992). ■

A.3. Proof of Corollary. Because $S_n(x) \xrightarrow{p} \sigma^2(x)$ for all x, and $[\sum_{i=1}^{n}(1/h_n)K'((X_{i\Delta_n} - x)/h_n)]/[\sum_{i=1}^{n}K((X_{i\Delta_n} - x)/h_n)] \xrightarrow{p} (p'(x)/p(x))$ for all x with $p(x) \neq 0$, by Slutsky's theorem, from (16) we have

$$\hat{\mu}(x) \xrightarrow{p} \frac{1}{2}\left[\frac{d\sigma^2(x)}{dx} + \sigma^2(x)\frac{p'(x)}{p(x)}\right] = \mu(x). \tag{A.3.1}$$

That is, $\hat{\mu}(x)$ is a pointwise consistent estimator of $\mu(x)$. ■

http://www.jstor.org

LINKED CITATIONS

You have printed the following article:

A Nonparametric Approach to the Estimation of Diffusion Processes, with an Application to a Short-Term Interest Rate Model

George J. Jiang; John L. Knight

Econometric Theory, Vol. 13, No. 5. (Oct., 1997), pp. 615-645.

Stable URL:

http://links.jstor.org/sici?sici=0266-4666%28199710%2913%3A5%3C615%3AANATTE%3E2.0.CO%3B2-X

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

Notes

[1] **An Empirical Comparison of Alternative Models of the Short-Term Interest Rate**

K. C. Chan; G. Andrew Karolyi; Francis A. Longstaff; Anthony B. Sanders

The Journal of Finance, Vol. 47, No. 3, Papers and Proceedings of the Fifty-Second Annual Meeting of the American Finance Association, New Orleans, Louisiana January 3-5, 1992. (Jul., 1992), pp. 1209-1227.

Stable URL:

http://links.jstor.org/sici?sici=0022-1082%28199207%2947%3A3%3C1209%3AAECOAM%3E2.0.CO%3B2-T

[2] **A Reappraisal of Misspecified Econometric Models**

Alain Monfort

Econometric Theory, Vol. 12, No. 4. (Oct., 1996), pp. 597-619.

Stable URL:

http://links.jstor.org/sici?sici=0266-4666%28199610%2912%3A4%3C597%3AAROMEM%3E2.0.CO%3B2-%23

References

The Pricing of Options and Corporate Liabilities

Fischer Black; Myron Scholes

The Journal of Political Economy, Vol. 81, No. 3. (May - Jun., 1973), pp. 637-654.

Stable URL:

http://links.jstor.org/sici?sici=0022-3808%28197305%2F06%2981%3A3%3C637%3ATPOOAC%3E2.0.CO%3B2-P

NOTE: *The reference numbering from the original has been maintained in this citation list.*

http://www.jstor.org

LINKED CITATIONS

Analyzing Convertible Bonds

Michael J. Brennan; Eduardo S. Schwartz

The Journal of Financial and Quantitative Analysis, Vol. 15, No. 4, Proceedings of 15th Annual Conference of the Western Finance Association, June 19-21, 1980, San Diego, California. (Nov., 1980), pp. 907-929.

Stable URL:

http://links.jstor.org/sici?sici=0022-1090%28198011%2915%3A4%3C907%3AACB%3E2.0.CO%3B2-1

Asymptotic Likelihood Theory for Diffusion Processes

B. M. Brown; J. I. Hewitt

Journal of Applied Probability, Vol. 12, No. 2. (Jun., 1975), pp. 228-238.

Stable URL:

http://links.jstor.org/sici?sici=0021-9002%28197506%2912%3A2%3C228%3AALTFDP%3E2.0.CO%3B2-2

An Empirical Comparison of Alternative Models of the Short-Term Interest Rate

K. C. Chan; G. Andrew Karolyi; Francis A. Longstaff; Anthony B. Sanders

The Journal of Finance, Vol. 47, No. 3, Papers and Proceedings of the Fifty-Second Annual Meeting of the American Finance Association, New Orleans, Louisiana January 3-5, 1992. (Jul., 1992), pp. 1209-1227.

Stable URL:

http://links.jstor.org/sici?sici=0022-1082%28199207%2947%3A3%3C1209%3AAECOAM%3E2.0.CO%3B2-T

An Analysis of Variable Rate Loan Contracts

John C. Cox; Jonathan E. Ingersoll, Jr.; Stephen A. Ross

The Journal of Finance, Vol. 35, No. 2, Papers and Proceedings Thirty-Eighth Annual Meeting American Finance Association, Atlanta, Georgia, December 28-30, 1979. (May, 1980), pp. 389-403.

Stable URL:

http://links.jstor.org/sici?sici=0022-1082%28198005%2935%3A2%3C389%3AAAOVRL%3E2.0.CO%3B2-C

A Theory of the Term Structure of Interest Rates

John C. Cox; Jonathan E. Ingersoll, Jr.; Stephen A. Ross

Econometrica, Vol. 53, No. 2. (Mar., 1985), pp. 385-407.

Stable URL:

http://links.jstor.org/sici?sici=0012-9682%28198503%2953%3A2%3C385%3AATOTTS%3E2.0.CO%3B2-B

NOTE: *The reference numbering from the original has been maintained in this citation list.*

http://www.jstor.org

LINKED CITATIONS

On Estimating the Diffusion Coefficient
Gejza Dohnal
Journal of Applied Probability, Vol. 24, No. 1. (Mar., 1987), pp. 105-114.
Stable URL:
http://links.jstor.org/sici?sici=0021-9002%28198703%2924%3A1%3C105%3AOETDC%3E2.0.CO%3B2-X

Simulated Moments Estimation of Markov Models of Asset Prices
Darrell Duffie; Kenneth J. Singleton
Econometrica, Vol. 61, No. 4. (Jul., 1993), pp. 929-952.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199307%2961%3A4%3C929%3ASMEOMM%3E2.0.CO%3B2-P

On Estimating the Diffusion Coefficient from Discrete Observations
Danielle Florens-Zmirou
Journal of Applied Probability, Vol. 30, No. 4. (Dec., 1993), pp. 790-804.
Stable URL:
http://links.jstor.org/sici?sici=0021-9002%28199312%2930%3A4%3C790%3AOETDCF%3E2.0.CO%3B2-0

Large Sample Properties of Generalized Method of Moments Estimators
Lars Peter Hansen
Econometrica, Vol. 50, No. 4. (Jul., 1982), pp. 1029-1054.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198207%2950%3A4%3C1029%3ALSPOGM%3E2.0.CO%3B2-O

The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities
Lars Peter Hansen; Thomas J. Sargent
Econometrica, Vol. 51, No. 2. (Mar., 1983), pp. 377-387.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198303%2951%3A2%3C377%3ATDOTAP%3E2.0.CO%3B2-3

Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes
Lars Peter Hansen; José Alexandre Scheinkman
Econometrica, Vol. 63, No. 4. (Jul., 1995), pp. 767-804.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199507%2963%3A4%3C767%3ABTTFGM%3E2.0.CO%3B2-N

NOTE: *The reference numbering from the original has been maintained in this citation list.*

http://www.jstor.org

LINKED CITATIONS

Minimum Contrast Estimation in Diffusion Processes
V#ra Lánska
Journal of Applied Probability, Vol. 16, No. 1. (Mar., 1979), pp. 65-75.
Stable URL:
http://links.jstor.org/sici?sici=0021-9002%28197903%2916%3A1%3C65%3AMCEIDP%3E2.0.CO%3B2-Y

Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data
Andrew W. Lo
Econometric Theory, Vol. 4, No. 2. (Aug., 1988), pp. 231-247.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28198808%294%3A2%3C231%3AMLEOGI%3E2.0.CO%3B2-I

Stochastic Processes for Interest Rates and Equilibrium Bond Prices
Terry A. Marsh; Eric R. Rosenfeld
The Journal of Finance, Vol. 38, No. 2, Papers and Proceedings Forty-First Annual Meeting American Finance Association New York, N.Y. December 28-30, 1982. (May, 1983), pp. 635-646.
Stable URL:
http://links.jstor.org/sici?sici=0022-1082%28198305%2938%3A2%3C635%3ASPFIRA%3E2.0.CO%3B2-Y

Theory of Rational Option Pricing
Robert C. Merton
The Bell Journal of Economics and Management Science, Vol. 4, No. 1. (Spring, 1973), pp. 141-183.
Stable URL:
http://links.jstor.org/sici?sici=0005-8556%28197321%294%3A1%3C141%3ATOROP%3E2.0.CO%3B2-0

A Reappraisal of Misspecified Econometric Models
Alain Monfort
Econometric Theory, Vol. 12, No. 4. (Oct., 1996), pp. 597-619.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199610%2912%3A4%3C597%3AAROMEM%3E2.0.CO%3B2-%23

Consistency and Asymptotic Normality of an Approximate Maximum Likelihood Estimator for Discretely Observed Diffusion Processes
Asger Roer Pedersen
Bernoulli, Vol. 1, No. 3. (Sep., 1995), pp. 257-279.
Stable URL:
http://links.jstor.org/sici?sici=1350-7265%28199509%291%3A3%3C257%3ACAANOA%3E2.0.CO%3B2-G

NOTE: *The reference numbering from the original has been maintained in this citation list.*

http://www.jstor.org

Time Series Regression with a Unit Root
P. C. B. Phillips
Econometrica, Vol. 55, No. 2. (Mar., 1987), pp. 277-301.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198703%2955%3A2%3C277%3ATSRWAU%3E2.0.CO%3B2-R

NOTE: *The reference numbering from the original has been maintained in this citation list.*