A selective overview of nonparametric methods in financial econometrics

Jianqing Fan
Department of Statistics
Chinese University of Hong Kong
Shatin, Hong Kong *

March 30, 2003

Abstract

This paper gives a brief overview on the nonparametric techniques that are useful for financial econometric problems. The problems include estimation and inferences of instantaneous returns and volatility functions, and estimation of transition densities and state price densities. We first briefly describe the problems and then outline main techniques and main results. Some useful probabilistic aspects of diffusion processes are also briefly summarized to facilitate our presentation and applications.

1 Introduction

Technological invention and trade globalization have brought into a new era of financial markets. Over the last three decades or so, enormous number of new financial products have been introduced to meet customers' demands. An important milestone is that in the year 1973, the world's first options exchange opened in Chicago. At the very same year, Black and Scholes (1973) published their famous paper on option pricing and Merton (1973) launched general equilibrium model for security prices, two landmarks for modern asset pricing. Since then, the derivative markets have experienced extraordinary growth. Professionals in finance now routinely use sophisticated statistical techniques and modern computation power in portfolio management, securities regulation, proprietary trading, financial consulting and risk management.

Financial econometrics is an active field of integration of finance, economics, probability, statistics, and applied mathematics. This is exemplified in the very book by Campbell *et al.*(1997).

*Fan was partially supported by NSF grant DMS-0204329 and a direct allocation RGC grant of the Chinese University of Hong Kong. He acknowledges gratefully various discussions with Professors Yacine Aït-Sahalia and Jia-an Yan.

Complex financial products pose new challenges on their valuation. Sophisticated stochastic models have been introduced to capture the salient features of underlying economic variables. Statistical tools are used to identify parameters of stochastic models and to simulate complex financial systems. Many pricing formulas require solving a system of differential equations, where techniques in applied mathematics play a pivotal role. This paper gives an overview on nonparametric techniques for financial econometrics.

Modern asset pricing theory allows one to value and hedge contingent claims via risk-neutral valuation, once a model for the dynamics of an underlying state variable is given. There are many books on mathematical finance. See, for example, Bingham and Kiesel (1998), Steele (2000), and Duffie (2001). Many of such models have been developed. Most of these stochastic models are simple and convenient parametric ones to facilitate pricing contingent claims and statistical estimation. They are not derived from any economics theory and hence can not be expected to fit all financial data. Thus, while the pricing theory gives spectacularly beautiful formulas when the underlying dynamics is correctly modeled, it offers little guidance in choosing a correct model nor validating a specific model. Hence there is always a danger that misspecification of a model leads to erroneous valuation and hedging. Various extensions to relaxing restrictive assumptions have been made. Beyond a specific form on a stochastic model, there are infinitely many possibilities. Nonparametric methods offer a unified and elegant treatment in stochastic modeling.

Nonparametric approaches have recently been introduced to estimate instantaneous return and volatility function and to estimate transition densities and state price densities. They can also be applied to examine the extent to which the coefficient functions vary over time. They have immediate applications to the valuation of bond price and stock options and the management of investment risks. They can also be employed to answer the questions such as if the geometric Brownian motion fits certain stock indices, whether the Cox-Ingsoll-Ross model fits financial data, and if interest rates dynamics evolve with time. Furthermore, based on empirical data, one can also fit directly the observed option prices with their associate characteristics such as strike price, the time to maturity, risk-less interest rate, dividend yield and see if the option prices are consistent with the theoretical ones. Needless to say, nonparametric techniques will play an increasingly important role in financial econometrics, thanks to the availability of modern computing power.

The paper is organized as follows. We first introduce in section 2 some useful stochastic models for modeling stock prices and bond yields and then briefly outline some probabilistic aspect of the model. In Section 3, we review nonparametric techniques used for estimating the drift and diffusion functions, based on either discretely or continuously observed data. In Section 4, we outline techniques for estimating state price densities and transition densities. Their applications in asset pricing and testing for parametric diffusion models are also introduced.

2 Diffusion models

Stochastic diffusion models have been widely used for describing the dynamics of underlying economic variables and asset prices. They form the basis of many spectacularly beautiful formulas for pricing contingent claims. See, for example, the books by Hull (1997), Bingham and Kiesel (1998), Steele (2000) and Duffie (2001).

2.1 One-factor diffusion models

Let X_t be an observed economic variable at time t. This can be the prices of a stock or a stock index, or the yields of a bond. A simple and frequently used stochastic model to describe the dynamics of the economic variable is to assume that it satisfies the following stochastic differential equation:

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

where $\{W_t\}$ is a standard one-dimensional Brownian motion. The function $\mu(\cdot)$ is frequently called a drift or instantaneous return function and $\sigma(\cdot)$ is referred to as the diffusion or volatility function, since

$$\lim_{\Delta\to 0} \Delta^{-1} E(X_{t+\Delta} - X_t | X_t) = \mu(X_t) \tag{2}$$

$$\lim_{\Delta\to 0} \Delta^{-1} \mathrm{var}(X_{t+\Delta} | X_t) = \sigma^2(X_t) \tag{3}$$

The time-homogeneous model (1) contains many famous one-factor models in financial econometrics. The celebrated Black-Scholes option price formula is derived based on the Osborne's assumption that the stock price follows the geometric Brownian:

$$dX_t = \mu X_t dt + \sigma X_t dW_t. \tag{4}$$

The model (1) also includes many celebrated models for short-term interest rate. Examples are

$$\text{VAS}: \quad dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t, \tag{5}$$

$$\text{CIR}: \quad dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^{1/2} dW_t, \tag{6}$$

$$\text{CKLS}: \quad dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^{\gamma} dW_t, \tag{7}$$

$$\text{CIR2}: \quad dX_t = \sigma X_t^{3/2}\, dW_t, \tag{8}$$

$$\text{AIT}: \quad dX_t = (\alpha_0 X_t^{-1} + \alpha_1 + \alpha_2 X_t + \alpha_2 X_t^2)dt + \sigma X_t^{\gamma} dW_t \tag{9}$$

introduced respectively by Vasicek (VAS) (1977), Cox *et al.*(CIR) (1985), Chan *et al.*(CKLS) (1992), Cox *et al.*(1980) and Aït-Sahalia (AIT) (1996b). The first three models contain a linear drift and different forms of the volatility functions. As the interest rate goes above the mean level α, there is negative drift that pulls the rate down, while when the interest rate goes below α, there is a positive force that drives the rate up. Such a property is called "mean reversion". The AIT model has nonlinear mean reversion: while interest rates remain in the middle part of their domain, there is little mean reversion and at the end of the domain, the strong nonlinear mean reversion emerges. Note that the CIR2 model is closely related to the model used by Ahn and Gao (1999) who model the interest by $Y_t = X_t^{-1}$, in which the X_t follows the CIR model. An application of the Itô's formula yields:

$$dY_t = Y_t\{\kappa - (\sigma^2 - \kappa\alpha)Y_t\}dt + \sigma Y_t^{3/2} dW_t.$$

Economic conditions vary over time. Thus, it is reasonable to expect that the instantaneous return and volatility depend on both time and price level for a given state variable, such as stock prices and bond yields. This leads to a further generalization of model (1) to allow the coefficients to depend on time t:

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t. \tag{10}$$

Since only a trajectory of the process is observed, there is no sufficient information to estimate the bivariate functions in (10) without further restrictions. A subclass of models where coefficient functions are estimable is

$$dX_t = \{\alpha_0(t) + \alpha_1(t)g(X_t)\}\,dt + \beta_0(t)h(X_t)^{\beta_1(t)}\,dW_t, \tag{11}$$

for some functions α_0, α_1, g, β_0, β_1 and h. A useful specification of model (10) is

$$dX_t = \{\alpha_0(t) + \alpha_1(t)X_t\}\,dt + \beta_0(t)X_t^{\beta_1(t)}\,dW_t. \tag{12}$$

This is an extension of the CKLS model by allowing the coefficients to depend on time and was introduced and studied by Fan *et al.*(2003). Models (11) and (12) include many commonly-used time-varying models for the yields of bonds, introduced by Ho and Lee (HL) (1986), Hull and White (HW) (1990), Black, Derman and Toy (BDT) (1990), and Black and Karasinski (BK) (1991). They assume respectively the following forms:

$$\begin{aligned}
\text{HL:}\quad & dX_t = \mu(t)\,dt + \sigma(t)\,dW_t,\\
\text{HW:}\quad & dX_t = \{\alpha_0(t) + \alpha_1(t)X_t\}\,dt + \sigma(t){X_t}^i\,dW_t, \quad i = 0 \text{ or } 0.5,\\
\text{BDT:}\quad & dX_t = \{\alpha_1(t)X_t + \alpha_2(t)X_t\log(X_t)\}\,dt + \beta_0(t)X_t\,dW_t,\\
\text{BK:}\quad & dX_t = \{\alpha_1(t)X_t + \alpha_2(t)X_t\log(X_t)\}\,dt + \beta_0(t)X_t\,dW_t, \text{ with } \alpha_2(t) = \frac{d\log\{\beta_0(t)\}}{dt}.
\end{aligned}$$

The experience in Fan *et al.*(2003) and other studies of the varying coefficient models (see e.g., Cleveland *et al.*1991, Chen and Tsay 1993, Hastie and Tibshirani, 1993, Hoover *et al.*1998, Cai *et al.*2000) shows that coefficient functions in (12) can not be estimated reliably due to the collinearity effect in local estimation. This leads Fan *et al.*(2003) to introduce the semiparametric model:

$$dX_t = \{\alpha_0(t) + \alpha_1(t)X_t\}\,dt + \beta_0(t)X_t^{\beta}\,dW_t, \tag{13}$$

a generalization of the HW model.

2.2 Existence of Solution

A question arises naturally when there exists a solution to the stochastic differential equation (SDE) (10). Such a program was first carried out by Itô(1942, 1946). For SDE (10), there are two different meanings of solutions: strong solution and weak solution. See Sections 5.2(page 285) and 5.3 (page 300) of Karatzas and Shreve (1991). Basically, for a given initial condition ξ, a strong solution requires that X_t is determined completely by the information up to time t, namely, the information contained in the driving Brownian motion $\{W_s, 0 \le s \le t\}$ and the initial condition ξ. The weak solutions relax the above requirement.

Under the Lipchitz and linear growth conditions:

$$\|\mu(t,x) - \mu(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \le K\|x-y\|$$

$$\|\mu(t,x)\|^2 + \|\sigma(t,x)\|^2 \le K(1 + \|x\|^2)$$

for every t, x and y, for every ξ that is independent of $\{W_s\}$, there exists a strong solution of equation (10). Such a solution is unique. See Theorem 2.9 of Karatzas and Shreve (1991). The above result holds for the process in d-dimensional Euclidean space.

For one-dimensional time-homogeneous diffusion process (1), weaker conditions can be obtained. Let us recall the Itô's formula: For process X_t in (10), for a sufficiently regular function f [see e.g. page 153 of Karatzas and Shreve (1991)],

$$df(X_t,t) = \left\{ \frac{\partial f(X_t,t)}{\partial t} + \frac{1}{2}\frac{\partial^2 f(X_t,t)}{\partial x^2}\sigma^2(X_t,t) \right\} dt + \frac{\partial f(X_t,t)}{\partial x} dX_t.$$

By an application of Itô's formula, for the process X_t in (1), the process $Y_t = p(X_t)$ satisfies

$$dY_t = (\mu p' + \frac{1}{2}\sigma^2 p'')dt + \sigma p' dB_t.$$

Hence, the drift term can be removed if one takes p to satisfy

$$\mu p' + \frac{1}{2}\sigma^2 p'' = 0$$

or

$$p'(x) = c \exp\left\{ -2\int_0^x \mu(u)/\sigma^2(u) du \right\},$$

for an arbitrary constant c. Note that if σ is strictly positive and differentiable, the process (1) can also be transformed to have unit diffusion via $Y_t = \int_\cdot^{X_t} \sigma(u)^{-1} du \equiv G(Y_t)$ for some nonnegative lower limit, which satisfies

$$dY_t = \mu_Y(Y_t)dt + dW_t, \tag{14}$$

where $\mu_Y(y) = (\mu/\sigma - \frac{1}{2}\sigma')(G^{-1}(y))$. Thus, we can consider without loss of generality that the drift in (1) is zero. Let

$$Z(\sigma) = \{x : \sigma(x) = 0\} \text{ and } I(\sigma) = \left\{ x : \int_{x-}^{x+} \sigma(y)^{-2} dy = \infty \right\}.$$

Note that if $\sigma(x)$ is continuous, then $I(\sigma) \subset N(\sigma)$. Engelbert and Schmidt (1984) prove that equation (1) with $\mu \equiv 0$ has a weak solution for every initial distribution if and only if $I(\sigma) \subset N(\sigma)$. In that case, the uniqueness in law holds for every initial distribution if and only if $I(\sigma) = N(\sigma)$. See also Theorem 5.5.4 (page 333) of Karatzas and Shreve (1991) and Theorem 23.1 of Kallenberg (2001).

2.3 Markovian property and transition density

According to Theorem 5.4.20 (page 322) of Karatzas and Shreve (1991), the solution X_t to equation (1) is Markovian, provided that the coefficient functions b and σ are bounded on compact subsets. Let $p_t(y|x)$ be the transition density: the conditional density of $X_{s+t} = y$ given $X_s = x$. The transition density must satisfy the forward Kolmogorov equation

$$\frac{\partial}{\partial t} p_t(y|x) = \frac{1}{2}\frac{\partial^2}{\partial y^2}[\sigma^2(y)p_t(y|x)] - \frac{\partial}{\partial y}[\mu(y)p_t(y|x)]$$

and backward Kolmogorov equation

$$\frac{\partial}{\partial t} p_t(y|x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x)p_t(y|x)] + \frac{\partial}{\partial x}[\mu(x)p_t(y|x)].$$

See page 282 of Karatzas and Shreve (1991).

In many financial applications, the process X_t is nonnegative. For simplicity, we assume that the range of X_t is $(0, \infty)$. Aït-Sahalia (1999, 2002) derives the Hermit expansion of the transition density around a normal density function. As discussed in §2.2, we assume without loss of generality (14). For small t, the transition density $p_t(y|y_0)$ of the Markov process $\{Y_t\}$ can be approximated, up to order K, by

$$p_t^{(K)}(y|y_0) = t^{-1/2} \exp\left(\int_{y_0}^{y} \mu_Y(w)dw\right)\phi((y - y_0)/t^{1/2}) \sum_{k=0}^{K} c_k(y|y_0)t^k/k!, \tag{15}$$

where $\phi(z)$ is the density function of the standard normal distribution, $c_0(y|y_0) = 1$ and

$$c_j(y|y_0) = j(y - y_0)^{-j} \int_{y_0}^{y} (w - y_0)^{j-1} \times \{\lambda_Y(w)c_{j-1}(w|y_0) + \frac{1}{2}\partial^2 c_{j-1}(w|y_0)/\partial w^2\}dw,$$

where $\lambda_Y(y) = -(\mu_Y^2(y) + \mu_Y'(y))/2$. Note that

$$c_1(y|y_0) = -\frac{1}{2(y - y_0)}\left\{\int_{y_0}^{y} \mu_Y(w)^2 dw + \mu_Y(y) - \mu_Y(y_0)\right\}.$$

By the change-of-variable formula, the transition density of the process (1) can be approximated by

$$\tilde{p}_t^{(K)}(x|x_0) = p_t^{(K)}(y|y_0)/\sigma(y), \tag{16}$$

where $y = -\int_{\cdot}^{x} \sigma(u)^{-1}du$. Aït-Sahalia (1999) gives the explicit approximations for commonly used models (4), (5), (6) and (7) and evaluates their accuracies. A closed-form approximation of the transition density for the time-inhomogeneous model (10) is given by Egorov *et al.*(2003).

Under the linear growth and Lipchitz's conditions, and additional conditions on the boundary behavior of functions μ and σ, the solution to equation (1) is positive and ergodic. One sufficient condition is at the left boundary, $\int_{0+} s(x)dx = +\infty$ and at the right boundary, $\int^{+\infty} s(x)dx = +\infty$, and

$$\int_0^\infty \sigma^{-2}(x)/s(x)dx < \infty,$$

where $s(x) = \exp(-2\int_\cdot^x \mu(y)\sigma^{-2}(y)dy)$ is the derivative of the scale function of the diffusion process (1). See §5.7 of Rogers and Williams (1987). Let $f(x)$ be the invariant density. Multiplying both sides of the forward Kolmogorov equation by a factor of $f(x)$ and integrating it out with respect to x and noticing $\int p_t(y|x)f(x)dx = f(y)$, we have

$$\frac{1}{2}\frac{\partial^2}{\partial y^2}[\sigma^2(y)f(y)] - \frac{\partial}{\partial y}[\mu(y)f(y)] = 0.$$

Solving this differential equation yields

$$f(x) = 2C_0\sigma^{-2}(x)\exp(2\int_\cdot^x \mu(y)\sigma^{-2}(y)dy), \tag{17}$$

where C_0 is a normalizing constant and the lower limit of the integral does not matter. If the initial distribution is taken from the invariant density, then the process $\{X_t\}$ is stationary. Let H_t be the operator defined by

$$(H_t f)(x) = E(f(X_t)|X_0 = x), \quad x \in R, \tag{18}$$

where f is a Borel measurable bounded function on R. The operator satisfies the Feller semigroup property: for $s, t \geq 0$, $H_{s+t} = H_s H_t = H_t H_s$.

A stationary process X_t is said to satisfy the condition $G_2(s, \alpha)$ of Rosenblatt (1970) if there exists an s such that

$$\|H_s\|_2^2 = \sup_{\{f: Ef(X)=0\}} \frac{E(H_s f)^2(X)}{Ef^2(X)} \leq \alpha^2 < 1,$$

namely the operator is a contraction. As a consequence of the semigroup and contraction properties, the condition G_2 implies (Banon, 1977) that for any $t \in [0, \infty)$, $\|H_t\| \leq \alpha^{t/s-1}$. The latter implies, by the Cauchy-Schwartz inequality, that

$$\rho(t) = \sup_{g_1, g_2} \text{corr}(g_1(X_0), g_2(X_t)) \leq \alpha^{t/s-1}, \tag{19}$$

That is, the ρ-mixing coefficient decays exponentially fast. Banon and Nguyen (1981) show further that for stationary Markov process, $\rho(t) \to 0$ is equivalent to (19).

2.4 Valuation of contingent claims

An important application of SDE is the pricing of financial derivatives, such as options and bonds. It forms beautiful modern asset pricing theory and provides useful guidances in practice. Hull (1997) and Duffie (2001) offer very nice introduction to the field.

The simplest financial derivative is probably the European call option. A call option is the right to buy an asset for a certain price K (strike price) before or at expiration time T. A put option gives the right to sell an asset for a certain price K (strike price) before or at expiration. European options allow option holders to exercise only at maturity, while American options can be exercised at any time before expiration. Most stock options are American, while options on stock indices are European. The pay-off for a European call option is $(X_T - K)_+$, where X_T is the price of the asset at expiration T. A European put option, on the other hand, has pay-off $(K - X_T)_+$. There are many other exotic options such as Asian options, look-back options and barrier options, which have different pay-off functions. See Chapter 18 of Hull (1997).

Suppose that the asset price satisfies the SDE (10) and there is a riskless investment alternative such as bond which earns compounding rate of interest r_t. Suppose that the underlying asset pays no dividend. Let β_t be the value of such an asset. Then, with initial investment β_0,

$$\beta_t = \beta_0 \exp(\int_0^t r_s ds).$$

Suppose that the probability measure Q that is equivalent to the original probability measure P, namely $P(A) = 0$ if and only if $Q(A) = 0$. The measure Q is called an equivalent martingale measure for the deflated price processes of given securities if these processes are martingales with respect to Q. An equivalent martingale measure is also referred to as a "risk-neutral" measure if the deflater is the bond price process. See Chapter 6 of Duffie (2001).

When the markets are dynamically complete, the price of the derivative security with payoff $\Psi(X_T)$ with initial price $X_0 = x_0$ is

$$P_0 = \exp(-\int_0^T r_s ds) E^Q(\Psi(X_T)|X_0 = x_0), \tag{20}$$

where Q is the equivalent martingale measure for the deflated price process X_t/β_t. Namely, it is the discounted value of the expected pay-off in the risk neutral world.

Suppose that the price process follows GBM (4) and the risk-free rate r is constant. Then, it is easy to show that under the measure Q, the price process follows

$$dX_t = rX_t dt + \sigma X_t dW_t$$

and by an application of the Itô formula,

$$\log X_t - \log X_0 = (r - \sigma^2/2)t + \sigma^2 W_t. \tag{21}$$

Note that given the initial price X_0, the price follows a log-normal distribution. Evaluation of the expectation of (20) for the European call option $\Psi(X_T) = (X_T - K)_+$, one obtains the Black-Scholes (1973) option pricing formula:

$$P_0 = x_0 \Phi(d_1) - K \exp(-rT)\Phi(d_2),$$

where $d_1 = \{\log(x_0/K) + (r + \sigma^2/2)T\}\{\sigma\sqrt{T}\}^{-1}$ and $d_2 = d_1 - \sigma\sqrt{T}$.

2.5 Simulation of stochastic models

Simulation methods provide useful tools for valuation of financial derivatives and other financial instruments, when analytic formulas are hard to obtain. They also provide useful tools for assessing performance of statistical methods.

The simplest method is perhaps the Euler scheme. The SDE (10) is approximated as

$$X_{t+\Delta} = X_t + \mu(t, X_t)\Delta + \sigma(t, X_t)\Delta^{1/2}\varepsilon_t, \tag{22}$$

where $\{\varepsilon_t\}$ is a sequence of independent random variables with the standard normal distribution. The time unit is usually a year. Thus, the monthly, weekly and daily data correspond, respectively, to $\Delta = 1/12, 1/52$, and $1/252$ (there are approximately 252 trading days per year). Given an initial value, one can recursively apply (22) to obtain a sequence of simulated data $\{X_{j\Delta}, j = 0, 1, 2\cdots\}$. The approximation error can be reduced if one uses a smaller step size Δ/M for a given integer M to obtain first a more detailed sequence $\{X_{j\Delta/M}, j = 0, 1, \cdots\}$ and then takes the subsequence $\{X_{j\Delta}, j = 0, 1, 2\cdots\}$. Since the step size Δ/M is smaller, the approximation (22) is more accurate. However, the computational cost is about a factor of M higher.

The Euler scheme has convergence rate $\Delta^{-1/2}$, which is called strong order 0.5 approximation by Kloeden *et al.*(1996). The higher order approximations can be obtained by the Itô-Taylor expansion (see Schurz, 2000, page 242). In particular, a strong order-one approximation is given by

$$X_{t+\Delta} = X_t + \mu(t, X_t)\Delta + \sigma(t, X_t)\Delta^{1/2}\varepsilon_t + \frac{1}{2}\sigma(t, X_t)\sigma'_x(t, X_t)\Delta\{\varepsilon_t^2 - 1\}, \tag{23}$$

where $\sigma'_x(t, x)$ is the partial derivative function with respect to x. For the time-homogeneous model (1), an alternative form, without evaluating the derivative function, is

$$X_{t+\Delta} = X_t + \mu(X_t)\Delta + \frac{1}{2}\{\sigma(X_t) + \sigma(X_t + \mu(X_t)\Delta + \sigma(X_t)\Delta^{1/2}\varepsilon_t)\}\Delta^{1/2}\varepsilon_t.$$

See (3.14) of Kloeden *et al.*(1996). More numerical methods can be found in Kloeden and Platen (1995).

The exact simulation method is available if one can simulate the data from the transition density. Given the current value $X_t = x_0$, one draws $X_{t+\Delta}$ from the transition density $p_\Delta(\cdot|x_0)$. The initial condition can either be fixed at a given value or be generated from the invariant density (17). In the latter case, the generated sequence is stationary.

For GBM, one can generate the sequence from the explicit solution (21), where the Brownian motion can be simulated from independent Gaussian increments. The conditional density of Vascicek's model (5) is Gaussian with mean $\alpha+(x_0-\alpha)\exp(-\kappa\Delta)$ and variance $\sigma_\Delta^2 = \sigma^2(1-\exp(-2\kappa\Delta))/(2\kappa)$. Indeed, the solution to the SDE (5) is

$$X_t = \alpha + (X_0 - \alpha)\exp(-\kappa t) + \sigma \exp(-2\kappa t)\int_0^t \exp(st)dW_s.$$

By (17), one can easily see that its invariant density is Gaussian with mean α and variance σ_Δ^2. Generate X_0 from the invariant density. With X_0, generate X_Δ from the normal distribution

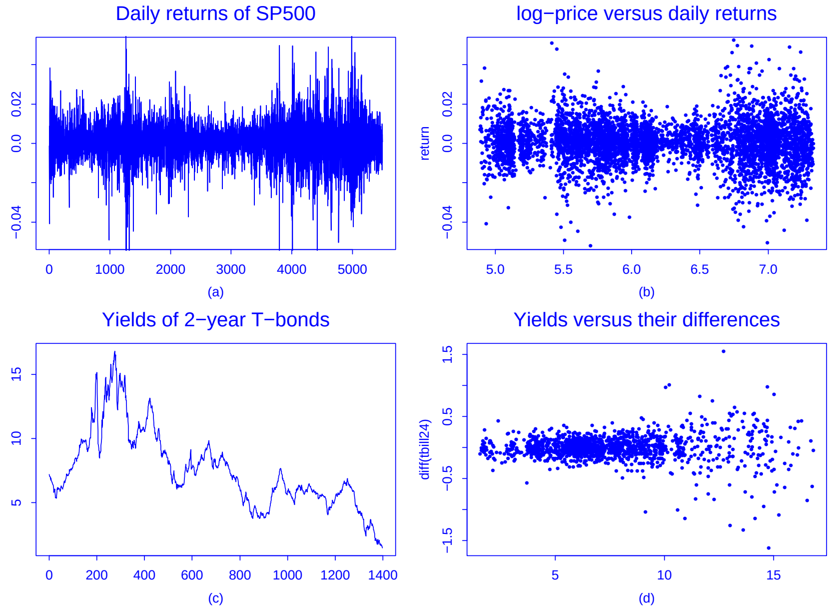

Figure 1: Simulated trajectories (multiplied by 100) using the Euler approximation and the strong order-one approximation for a CIR model. Top panel: solid-curve corresponds to the Euler approximation and the dashed curve is based on the order-one approximation. Botton panel: The difference between the order-one scheme and the Euler scheme.

with mean $\alpha + (X_0 - \alpha)\exp(-\kappa\Delta)$ and variance σ_Δ^2. With X_Δ, we generate $X_{2\Delta}$ from $\alpha + (X_\Delta - \alpha)\exp(-\kappa\Delta)$ and variance σ_Δ^2. Repeat this process until we obtain the desired length of the process.

For the CIR model (6), provided that $q = 2\kappa\alpha/\sigma^2 - 1 \geq 0$ (a sufficient condition for $X_t \geq 0$), the transition density is given by $p_\Delta(y|x) = c\exp(-u-v)(v/u)^{q/2}$ with $c = 2\kappa/\{\sigma^2(1-\exp(-\kappa\Delta))\}$, $u = cx_0\exp(k\Delta)$, $v = cx$ and

$$I_q(x) = \left(\frac{x}{2}\right)\sum_{i=0}^{\infty}\frac{1}{i!\Gamma(q+i+1)}\left(\frac{x}{2}\right)^{2i}$$

is the modified Bessel function of the first kind of order q. This distribution is often referred to as the noncentral χ^2 distribution. That is, given $X_t = x_0$, $2cX_{t+\Delta}$ has a noncentral χ^2 distribution with degrees of freedom $2q+2$ and noncentrality parameter $2u$. The invariant density is the Gamma distribution with shape parameter $q+1$ and the scale parameter $\sigma^2/(2\kappa)$.

As an illustration, we consider the CIR model (6) with parameters $\kappa = 0.21459$, $\alpha = 0.08571$, $\sigma = 0.07830$ and $\Delta = 1/12$. The model parameters are taken from Chapman and Pearson (2000). We simulated 1000 monthly data using both the Euler scheme (22) and strong order-one approximation (23) with the same random shocks. Figure 1 depicts one of their trajectories. The difference is negligible. This is in line with the observations made by Stanton (1997) that as long as data are sampled monthly or more frequently, the errors introduced by using the Euler approximation is very small for stochastic dynamics that are similar to the CIR model.

3 Estimation of return and volatility functions

There are a large literature on the estimation of the return and volatility functions. Early references include Pham (1981) and Prakasa Rao (1985). Some studies are based on continuously observed data, while others are based on discretely observed data. For the latter, some regard Δ tending to zero, while others regard Δ fixed. We briefly introduce some of the ideas.

3.1 Discretely observed data

Suppose that we have a sample $\{X_{i\Delta}, i = 0, \cdots, n\}$ from model (1). Then, the likelihood function, under the stationary condition, is

$$\log f(X_0) + \sum_{i=1}^{n} \log p_\Delta(X_{i\Delta}|X_{(i-1)\Delta}). \tag{24}$$

If the functions μ and σ are parameterized and the explicit form of the transition density is available, one can apply the maximum likelihood estimator. See, for example, Aït-Sahalia (2002) and references therein. For nonparametric estimation, the explicit form of the transition density is not available. Thus, a commonly-used technique is to rely on the Euler approximation scheme (22).

Let $Y_{i\Delta} = \Delta^{-1}(X_{(i+1)\Delta} - X_{i\Delta})$ and $Z_{i\Delta} = \Delta^{-1}(X_{(i+1)\Delta} - X_{i\Delta})^2$. Then,

$$E(Y_{i\Delta}|X_{i\Delta}) = \mu(X_{i\Delta}) + O(\Delta), \text{ and } E(Z_{i\Delta}|X_{i\Delta}) = \sigma^2(X_{i\Delta}) + O(\Delta).$$

Thus, $\mu(\cdot)$ and $\sigma^2(\cdot)$ can be approximately regarded as the regression functions of $Y_{i\Delta}$ and $Z_{i\Delta}$ on $X_{i\Delta}$, respectively. Stanton (1997) applies kernel regression (Wand and Jones, 1995; Simonoff, 1996) to estimate the return and volatility functions. Let $K(\cdot)$ be a kernel function and h be a bandwidth. Stanton's estimators are given by

$$\hat{\mu}(x) = \frac{\sum_{i=0}^{n-1} Y_{i\Delta} K_h(X_{i\Delta} - x)}{\sum_{i=0}^{n-1} K_h(X_{i\Delta} - x)}, \quad \text{and} \quad \hat{\sigma}^2(x) = \frac{\sum_{i=0}^{n-1} Z_{i\Delta} K_h(X_{i\Delta} - x)}{\sum_{i=0}^{n-1} K_h(X_{i\Delta} - x)},$$

where $K_h(u) = h^{-1}K(u/h)$ is a rescaled kernel. Independently, Fan and Yao (1998) apply the local linear technique (see §3.2 Fan and Gijbels, 1996 and §6.3 Fan and Yao 2003) to estimate the return and volatility functions, under a slightly different setup. The local linear estimator (Fan, 1992) is given by

$$\hat{\mu}(x) = \frac{\sum_{i=0}^{n-1} K_h(X_{i\Delta} - x) Y_{i\Delta} S_{n,2}(x) - \sum_{i=0}^{n-1} Y_{i\Delta} K_h(X_{i\Delta} - x)(X_{i\Delta} - x) S_{n,1}(x)}{S_{n,2}(x) S_{n,0}(x) - S_{n,1}(x)^2}, \tag{25}$$

where $S_{n,j}(x) = \sum_{i=0}^{n-1} K_h(X_{i\Delta} - x)(X_{i\Delta} - x)^j$, for $j = 0, 1$ and 2. Further, Fan and Yao (1998) use the squared residuals

$$\Delta^{-1}(X_{(i+1)\Delta} - X_{i\Delta} - \hat{\mu}(X_{i\Delta})\Delta)^2$$

rather than $Z_{i\Delta}$ to estimate the volatility function. They show further that the conditional variance function can be estimated as well as if the conditional mean function is known in advance.

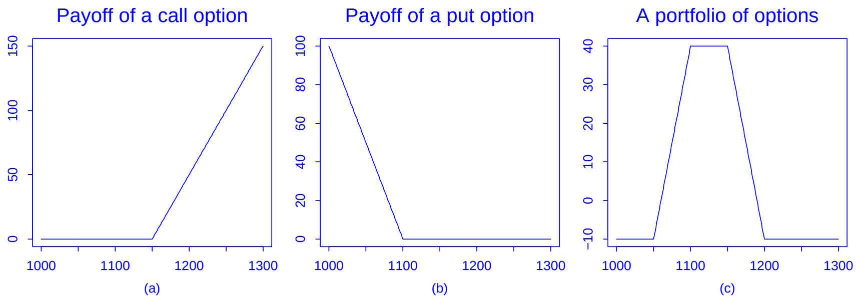

Figure 2: Nonparametric estimates of volatility based on orders 1 and 2 differences. The bars represent two standard deviations above and below the estimated volatility. Top panel: order 1 fit. Botton panel: order 2 fit.

Stanton (1997) derives higher order approximation scheme up to order 3. He suggested that higher order approximations must outperform lower order approximations. To verify such a claim, Fan and Zhang (2003) derive the following order k approximation scheme:

$$E(Y^*_{i\Delta}|X_{i\Delta}) = \mu(X_{i\Delta}) + O(\Delta^k), \text{ and } E(Z^*_{i\Delta}|X_{i\Delta}) = \sigma^2(X_{i\Delta}) + O(\Delta^k), \tag{26}$$

where with $a_{k,j} = (-1)^{j+1}\binom{k}{j}/j$,

$$Y^*_{i\Delta} = \Delta^{-1}\sum_{j=1}^{k} a_{k,j}\{X_{(i+j)\Delta} - X_{i\Delta}\} \text{ and } Z^*_{i\Delta} = \Delta^{-1}\sum_{j=1}^{k} a_{k,j}\{X_{(i+j)\Delta} - X_{i\Delta}\}^2.$$

While higher order approximations give better approximation errors, it is surprising that we have to pay a huge premium for variance inflation. Fan and Zhang (2003) show that

$$\begin{aligned}
\text{var}(Y^*_{i\Delta}|X_{i\Delta} = x_0) &= \sigma^2(x_0)V_1(k)\Delta^{-1}\{1 + O(\Delta)\},\\
\text{var}(Z^*_{i\Delta}|X_{i\Delta} = x_0) &= 2\sigma^4(x_0)V_2(k)\{1 + O(\Delta)\}.
\end{aligned}$$

The variance inflation factors $V_1(k)$ and $V_2(k)$ are explicitly given by Fan and Zhang (2003) and demonstrated to escalate exponentially fast (as k gets large). For example,

$$V_1(1) = 1, V_1(2) = 2.5, V_1(3) = 4.83, V_1(4) = 9.25, V_1(5) = 18.95,$$

while

$$V_2(1) = 1, V_2(2) = 3, V_2(3) = 8, V_2(4) = 21.66, V_2(5) = 61.50.$$

These theoretical results are also verified via empirical simulations in Fan and Zhang (2003). Therefore, the higher order difference methods should not be used unless the sampling interval is very wide (e.g. bi-annual data). It remains open whether it is possible to estimate the return and the volatility functions, without serious inflating the variance, with other higher approximation schemes.

As an illustration, we take the yields of the two-year Treasury notes from June 4, 1976 to March 7, 2003 with weekly frequency. Figure 2 presents nonparametric estimated volatility function based orders $k = 1$ and $k = 2$ approximations. The local linear fit is employed with the Epanechnikov kernel and bandwidth $h = 0.35$. It is evidenced that the order 2 approximation has higher variance than the order 1 approximation. In fact, the magnitude of variance inflation is in line with the theoretical result: the increase of the standard deviation from order 1 approximation to order 2 approximation is $\sqrt{3}$.

Stanton (1997) applies his kernel estimator to a Treasury's bill data set and observes nonlinear return function in his nonparametric estimates, particularly in region where the interest rate is high (over 14%, say). This led him to postulate the hypothesis that the return functions of short-term rates are nonlinear. Chapman and Pearson (2000) studied the finite sample properties of Stanton's estimator. By applying his procedure to the CIR model, they found that the Stanton's procedure produces spurious nonlinearity, due to the boundary effect and the mean reversion. Fan and Zhang (2003) develop a formal procedure, based on the generalized likelihood ratio (GLR) test of Fan *et al.*(2001), for checking whether the return and volatility functions possess certain parametric forms.

An alternative approach for discretely observed data is via a spectral analysis of the operator $H_\Delta = \exp(\Delta L)$, where the infinitesimal operator L is defined as

$$Lf(x) = \frac{\sigma^2(x)}{2} f''(x) + \mu(x) f'(x).$$

By Itô's formula, the operator L has the property:

$$Lf(x) = \lim_{\delta \to 0} \delta^{-1}[E\{f(X_{t+\delta})|X_t = x\} - f(x)].$$

The operator H_Δ is the transition operator in that [see also (18)]

$$H_\Delta f(x) = E\{f(X_\Delta)|X_0 = x\}.$$

For each given f, the above is a nonparametric regression problem and can be consistently estimated for fixed Δ (low-frequency data). The works of Hansen and Scheinkman (1995), Hansen *et al.*(1998) and Kessler and Sørensen (1999) consist of the following idea. The first step is to estimate the transition operator from the data. From the transition operator, one can identify the infinitesimal operator L and hence the functions $\mu(\cdot)$ and $\sigma(\cdot)$. Gobel *et al.*(2002) derive the optimal rate of convergence for such a scheme, using a wavelet basis. In particular, they show that for fixed Δ, the optimal rates of convergence for μ and σ are of orders $O(n^{-s/(2s+5)})$ and $O(n^{-s/(2s+3)})$, respectively, where s is the degree of smoothness of μ and σ.

Various discretization schemes and estimation methods have been proposed for the case with high frequency data over a long time horizon. More precisely, the studies are under the assumptions that $\Delta_n \to 0$ and $n\Delta_n \to \infty$. See for example, Dacunha-Castelle and Florens (1986), Yoshida (1992), Kessler (1997), Arfi (1998), Gobet (2003) and references therein.

There is a large statistical literature on estimating the mean and variance functions. See the books by Wand and Jones (1995), Fan and Gijbels (1996), Simonoff (1996) and Fan and Yao (2003). In particular, §8.7 of Fan and Yao (2003) give various methods for estimating the conditional variance function. For the estimation of constant variance with nonparametric regression function, Rice (1984), Gasser *et al.*(1986) and Hall *et al.*(1990) proposed various root-n consistent estimators. For the case where the conditional variance function is not constant, see Müller and Stadtmüller (1987, 1993), Hall and Carroll (1989), Ruppert *et al.*(1997) and Härdle and Tsybakov (1997).

3.2 Time-dependent model

The time dependent model (12) was introduced to accommodate the economy change over time. The coefficient functions in (12) are assumed to be slow time-varying and smooth. Nonparametric techniques can be applied to estimate these coefficient functions. See Fan *et al.*(2003).

We first estimate the coefficient functions $\alpha_0(t)$ and $\alpha_1(t)$. For each given time t_0, approximate the coefficient functions locally by constants: $\alpha(t) \approx a$ and $\beta(t) = b$ for t in a neighborhood of t_0. Using the Euler approximation (22), we run a local regression: Minimize

$$\sum_{i=0}^{n-1} (Y_{i\Delta} - a - bX_{i\Delta})^2 K_h(i\Delta - t_0) \tag{27}$$

with respect to a and b. This results in an estimate $\hat{\alpha}_0(t_0) = \hat{a}$ and $\hat{\alpha}_1(t_0) = \hat{b}$, where $\hat{a}$ and $\hat{b}$ are minimizers of the local regression (27). Fan *et al.*(2003) suggest using a one-sided kernel such as $K(u) = (1-u^2)I(-1 < u < 0)$ so that only the historical data in the time interval (t_0-h, t_0) are used in the above local regression. This facilitates forecasting and bandwidth selection. Our experience shows that there is not significant difference between nonparametric fitting with one-sided and two-sided kernels. We opt for local constant approximations instead of local linear approximations for estimating time varying functions, since the local linear fit can create artificial albeit insignificant linear trends when the underlying functions $\alpha_0(t)$ and $\alpha_1(t)$ are indeed time-independent.

We now turn to estimate the coefficient functions in the volatility via the pseudo-likelihood method. Let

$$\widehat{E}_t = \Delta^{-1/2}\{X_{t+\Delta} - X_t - (\widehat{\alpha}_0(t) + \widehat{\alpha}_1(t)X_t)\Delta\}$$

be the normalized residuals. Then, by (12) and (22), we have

$$\widehat{E}_t \approx \beta_0(t) X_t^{\beta_1(t)} \varepsilon_t. \tag{28}$$

The conditional log-likelihood of $\widehat{E}_t$ given X_t is, up to a location and a scale constant, approximately expressed as

$$-\log\{\beta_0^2(t) X_t^{2\beta_1(t)}\} - \frac{\hat{E}_t^2}{\beta_0^2(t) X_t^{2\beta_1(t)}}.$$

Using local constant approximations and incorporating the kernel weight, we obtain the local pseudo-likelihood at a time point t_0:

$$\ell(\beta_0, \beta_1; t_0) = -\sum_{i=0}^{n-1} K_h(i\Delta - t_0) \left(\log(\beta_0^2 X_{i\Delta}^{2\beta_1}) + \frac{\hat{E}_{i\Delta}^2}{\beta_0^2 X_{i\Delta}^{2\beta_1}} \right). \tag{29}$$

Maximizing (29) with respect to the local parameters β_0 and β_1, we obtain the estimates $\hat{\beta}_0(t_0) = \hat{\beta}_0$ and $\hat{\beta}_1(t_0) = \hat{\beta}_1$. This type of the local pseudo-likelihood method is related to the generalized method of moments of Hansen (1982), and the ideas of Florens-Zmirou (1993) and Genon-Catalot and Jacod (1993).

Since coefficient functions in both the return and volatility functions are estimated using only historical data, their bandwidths can be selected based on a form of the average prediction error and the maximum likelihood method. See Fan *et al.*(2003) for details. The local least-squares regression can also be applied to estimate the coefficient functions $\beta_0(t)$ and $\beta_1(t)$ via the transformed model [see (28)]

$$\log(\widehat{E}_t^2) \approx 2\log\beta_0(t) + \beta_1(t)\log(X_t^2) + \log(\varepsilon_t^2),$$

but we do not pursue along this direction, since the local least-squares estimate is known to be inefficient in the likelihood context and the exponentiation of an estimated coefficient function of $\log\beta_0(t)$ is unstable.

A question arises naturally if the coefficients in model (12) are really time-varying. This amounts for example to testing $H_0 : \beta_0(t) = \beta_0$ and $\beta_1(t) = \beta_1$. Based on the GLR technique, Fan *et al.*(2003) proposed a formal test for this kind of problems.

The coefficient functions in the semiparametric model (13) can also be estimated by using the profile pseudo-likelihood method. For each given β_1, by maximizing (29) with respect to β_0, we obtain

$$\hat{\beta}_0^2(t_0;\beta_1) = \sum_{i=0}^{n-1} K_h(i\Delta - t_0)\hat{E}_{i\Delta}^2 |X_{i\Delta}|^{-2\beta_1} \Big/ \sum_{i=0}^{n-1} K_h(i\Delta - t_0). \tag{30}$$

Now the coefficient β_1 can be obtained by maximizing the pseudo-likelihood [compare with (29)]

$$\ell(\beta_1) = -\sum_{i=0}^{n-1}\left(\log\{\hat{\beta}_0^2(i\Delta;\beta_1)X_{i\Delta}^{2\beta_1}\} + \frac{\hat{E}_{i\Delta}^2}{\hat{\beta}_0^2(i\Delta;\beta_1)X_{i\Delta}^{2\beta_1}}\right).$$

Note that $\hat{\beta}_1$ is estimated by using all the data points, while $\hat{\beta}_0(t) = \hat{\beta}_0(t;\hat{\beta}_1)$ is estimated by using only the local data points. See Fan *et al.*(2003) for details.

For other nonparametric methods of estimating of volatility in time inhomogeneous models, see Härdle *et al.*(2003) and Mercurio and Spokoiny (2003). Their methods are based on model (13) with $\beta_1 = 0$ and $\alpha_1(t) = 0$.

3.3 Continuously observed data

The case with continuously observed data assumes that the process $\{X_t\}$ is fully observable up to time T and asymptotic studies assume that $T \to \infty$, an infinite time horizon. Let us assume again that the observed process $\{X_t\}$ follows SDE (1). In this case, $\sigma^2(X_t)$ is the derivative of the quadratic variation process of X_t and hence is known up to time T. By (17), estimating the drift function $\mu(x)$ is equivalent to estimating the invariant density f. In fact,

$$\mu(x) = [\sigma^2(x)f(x)]'/[2f(x)]. \tag{31}$$

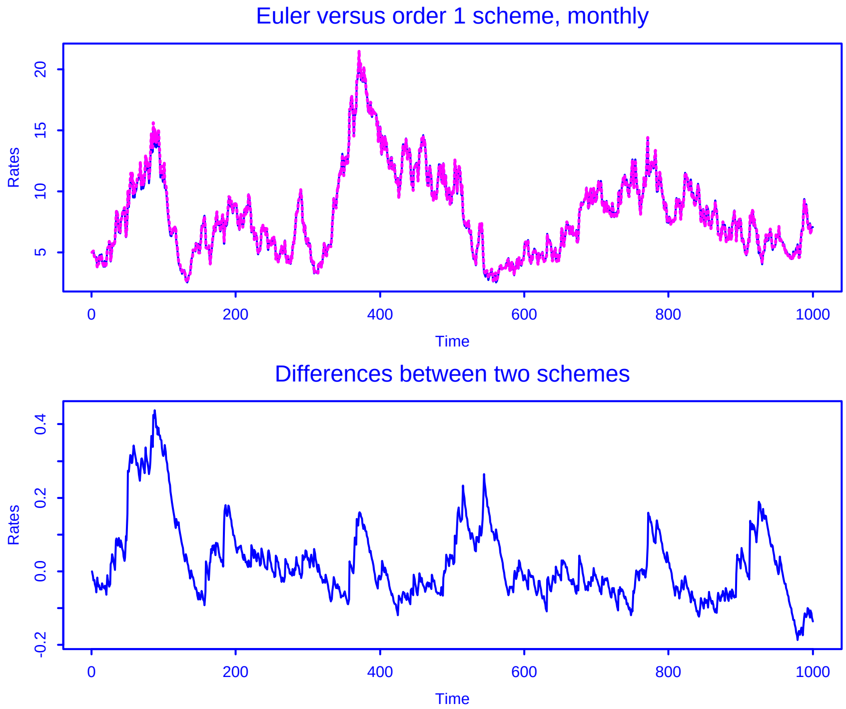

Figure 3: (a) Lag 1 scatterplot of the two-year Treasury note data. (b) Lag 1 scatterplot of those data falling in the neighborhood $8\% \pm 0.2\%$. The number in the scatterplot shows the indices of the data falling in the neighborhood. (c) Kernel density estimate of the invariant density.

The kernel density estimate for the invariant density is

$$\hat{f}^*(x) = n^{-1} \sum_{i=1}^{n} K_h(X_{i\Delta} - x), \qquad (32)$$

based on the discrete data $\{X_{i\Delta}, i = 1, \cdots, n\}$ with $n = T/\Delta$. In pricing the derivatives of interest rates, Aït-Sahalia (1996a) assumed $\mu(x) = k(\alpha - x)$. Using $\hat{f}$ and estimated κ and α from a least-squares method, he applied (17) to estimate $\sigma(\cdot)$, using discretely observed data: $\hat{\sigma}^2(x) = 2\int_0^x \hat{\mu}(u)\hat{f}^*(u)du/\hat{f}^*(x)$. He further established the asymptotic normality of such an estimate.

Before proceeding to the continuously observed data, we would like to note that nonparametric estimates use essentially only local data. The dependence of the local data is significantly weakened and hence many results on nonparametric estimates for independent data continue to hold for dependent data, as long as their mixing coefficients decay sufficiently fast. This can be understood graphically in Figure 3. Figure 3(a) shows that there is very strong serial correlation of the yields of the two-year treasury notes. However, this correlation is significantly weakened for the local data in the neighborhood of $8\% \pm 0.2\%$. In fact, as detailed in Figure 3(b), the indices of the data that fall in the local window are quite far apart. This in terms imply the week dependence for the data in the local window, i.e. "whitening by windowing". See §5.4 of Fan and Yao (2003) and Hart (1996) for further details. The effect of dependence structure on the kernel density estimation was thoroughly studied by Claeskens and Hall (2002).

When $\Delta \to 0$, the summation in (32) converges to

$$\hat{f}(x) = T^{-1} \int_0^T K_h(X_t - x)dt. \tag{33}$$

This forms a kernel density estimate of the invariant density based on the continuously observed data. Thus, an estimator for $\mu(x)$ can be obtained by substituting $\hat{f}(x)$ into (31). Such an approach has been employed by Kutoyants (1998) and Dalalyan and Kutoyants (2000, 2003). They established sharp asymptotic minimax risk for estimating the invariant density f and its derivative, as well as the drift function μ. In particular, the functions f, f' and μ can be estimated with rate $T^{-1/2}$, $T^{-2s/(2s+1)}$ and $T^{-2s/(2s+1)}$, respectively, where s is the degree of smoothness of μ. These are the optimal rates of convergence.

Spokoiny (2000) applies the local linear estimator to estimate the drift function. By taking the limit as $\Delta \to 0$, the estimator in (25) converges to

$$\hat{\mu}(x) = \frac{S_2(x) \int_0^T K_h(X_t - x)dX_t - S_1(x) \int_0^T K_h(X_t - x)(X_t - x)dX_t}{S_2(x)S_0(x) - S_1(x)^2}, \tag{34}$$

where $S_j(x) = \int_0^T K_h(X_t - x)(X_t - x)^j dt$. Spokoiny (2000) shows that this estimator attains the optimal rate of convergence. He established further a data-driven bandwidth such that the estimator (34) attains adaptive minimax rates.

The above local linear estimator can be derived directly from the local modeling technique (Fan and Gijbels, 1996). Note that

$$E\left\{\int_0^T f(X_t)^2 dt - 2\int_0^T f(X_t)dX_t\right\} = E\int_0^T (f(X_t) - \mu(X_t))^2 dt - \int_0^T \mu(X_t)^2 dt$$

is minimized at $f = \mu$. Thus, approximating $f(X_t)$ locally by $f(x) + f'(x)(X_t - x)$ for X_t around x, one would minimize

$$\int_0^T \{a + b(X_t - x)\}^2 K_h(X_t - x)dt - 2\int_0^T \{a + b(X_t - x)\}K_h(X_t - x)dX_t \tag{35}$$

with respect to a and b. Minimizing (35) results in the same estimator as (34).

4 Estimation of state price densities and transition densities

State-price density (SPD) is the probability density of the value of an asset under the risk-neutral world (20) [see Cox and Ross (1976)] or equivalent martingale measure (Harrison and Kreps, 1979). It is directly related to the pricing of financial derivatives. It is the transition density of X_T given $X_0 = x_0$ under the equivalent martingale Q. The SPD does not depend on the pay-off function and hence it can be used to evaluate other illiquid derivatives, once it is estimated from more liquid derivatives. The transition density characterizes the probability law of a Markovian process and hence is useful for validating Markovian properties and parametric models.

4.1 Estimation of state price density

For some specific models, the state price density can be formed explicitly. For example, for the GBM (4) with a constant risk-free rate r, according to (21), the SPD is log-normal, with mean $\log x_0 + (r - \sigma^2)/(2T)$ and variance σ^2.

Assume that the SPD f^* exists. Then, the European call option can be expressed as

$$C = \exp(-\int_0^T r_s ds) \int_K^\infty (x - K) f^*(x) dx.$$

See (20) (we have changed the notation from P to C to emphasize the price of the European call option). Hence,

$$f^*(K) = \exp(\int_0^T r_s ds) \frac{\partial^2 C}{\partial K^2}. \tag{36}$$

This was observed by Breeden and Litzenberger (1978). Thus, the state price density can be estimated from the European call options with different strike prices. With the estimated state price density, one can price new or less-liquid securities such as over the counter derivatives or nontraded options.

In general, the price of an European call option depends on the current stock price S, the strike price K, the time to maturity T, the risk-free interest rate r and yield rate δ. It can be written as $C(S, K, T, r, \delta)$. The exact form of C, in general, is hard to determine. Based on historical data $\{(C_i, S_i, K_i, T_i, r_i, \delta_i), i = 1, \cdots, n\}$, where C_i is the i^{th} traded-option price with associated characteristics $(S_i, K_i, T_i, r_i, \delta_i)$, Aït-Sahalia and Lo (1998) fit the following nonparametric regression

$$C_i = C(S_i, K_i, T_i, r_i, \delta_i) + \varepsilon_i$$

to obtain an estimate of the function C and hence the SPD f^*.

Due to the curse of dimensionality (see §7.1 of Fan and Gijbels, 1996), the five dimensional nonparametric function can not be estimated well with practical range of sample sizes. Aït-Sahalia and Lo (1998) realized that and proposed a few dimensionality reduction methods. First, by assuming that the option price depends only on the futures price $F = S\exp((r - \delta)T)$, namely,

$$C(S, K, T, r, \delta) = C(F, K, T, r)$$

(The Black-Scholes formula satisfies such an assumption), they reduced the dimensionality from 5 to 4. By assuming further that the option-pricing function is homogeneous of degree one in F and K, namely,

$$C(S, K, T, r, \delta) = C(F, K, T, r) = KC(F/K, T, r),$$

as in the Black-Scholes formula, they reduced the dimensionality to 3. Aït-Sahalia and Lo (1998) impose a semiparametric form on the pricing formula:

$$C(S, K, T, r, \delta) = C_{\mathrm{BS}}(F, K, T, r, \sigma(F, K, T)),$$

where $C_{\rm BS}(F,K,T,r,\sigma)$ is the Black-Scholes pricing formula given in §2 and $\sigma(F,K,T)$ is the implied volatility, computed by inverting the Black-Scholes formula. Thus, the problem becomes nonparametrically estimating the implied volatility function $\sigma(F,K,T)$. This can be estimated by using a nonparametric regression technique from historical data, namely

$$\sigma_i = \sigma(F_i, K_i, T_i) + \varepsilon_i,$$

where σ_i is the implied volatility of C_i by inverting the Black-Scholes formula. By assuming further that $\sigma(F,K,T) = \sigma(F/K,T)$, the dimensionality is reduced to 2. This is one of the options in Aït-Sahalia and Lo (1998).

The state price density f^* is non-negative and hence the function C should be convex in the strike K. Aït-Sahalia and Duarte (2003) proposed to estimate the option price under the convex constraint, using a local linear estimator. See also Härdle and Yatchew (2002) for a related approach.

4.2 Estimation of transition densities

The transition density of a Markov process characterizes the law of the process, except the initial distribution. It provides useful tools for checking whether or not such a process follows a certain SDE. It is the state price density of the price process under the risk neutral world. If such a process were observable, the state price density can be estimated using the methods to be introduced.

Assume that we have a sample $\{X_{i\Delta}, i = 0, \cdots, n\}$ from model (1). The "double-kernel" method of Fan *et al.*(1996) is to observe that

$$E\{W_{h_2}(X_{i\Delta} - y)|X_{(i-1)\Delta} = x\} \approx p_\Delta(y|x), \quad \text{as } h_2 \to 0, \tag{37}$$

for a kernel function W. Thus, the transition density $p_\Delta(y|x)$ can be regarded approximately as the nonparametric regression function of the response variable $K_{h_2}(X_{i\Delta} - y)$ on $X_{(i-1)\Delta}$. An application of the local linear estimator (Fan, 1992) yields [see (26)]

$$\hat{p}_\Delta(y|x) = \sum_{i=1}^n K_n(X_{(i-1)\Delta} - x) W_{h_2}(X_{i\Delta} - y), \tag{38}$$

where

$$K_n(u) = K_h(u) \frac{S_{n,2}(x) - uS_{n,1}(x)}{S_{n,2}(x)S_{n,0}(x) - S_{n,1}(x)^2}$$

is the equivalent kernel induced by the local linear fit (see §3.2 of Fan and Gijbels, 1996), where $S_{n,j}$ is defined in (25). For the homogeneity of notation, we will write h as h_1. Fan *et al.*(1996) establish the asymptotic normality of such an estimator under stationarity and ρ-mixing conditions [necessarily decaying at geometric rate for SDE (1)], which gives explicitly the asymptotic bias and variance of the estimator. See also §6.5 of Fan and Yao (2003).

Fan *et al.*(1996) propose a simple and ad-hoc method for selecting bandwidths. The resulting bandwidth tends to over smooth the underlying density function. An alternative method is to apply the bootstrap method of Hall *et al.*(1999). This method is also ad-hoc and can not be consistent. Here we introduce an alternative method which is related to the cross-validation idea of Rudemo (1982) and Bowman (1984).

Let $[a, b]$ be a given interval over which we wish to estimate the conditional density. Define the Integrated Square Error (ISE) as

$$\begin{aligned}
&\int_{-\infty}^{+\infty} \{\hat{p}_\Delta(y|x) - p_\Delta(y|x)\}^2 f(x) I(x \in [a,b]) dxdy \\
=& \int_{-\infty}^{+\infty} \hat{p}_\Delta(y|x)^2 f(x) I(x \in [a,b]) dxdy - 2\int_{-\infty}^{+\infty} \hat{p}_\Delta(y|x) p_\Delta(y|x) f(x) I(x \in [a,b]) dxdy \\
&+ \int_{-\infty}^{+\infty} p_\Delta(y|x)^2 f(x) I(x \in [a,b]) dxdy,
\end{aligned}$$

where f is the stationary density of $\{X_{i\Delta}\}$. Notice that the last term does not depend on bandwidths. The minimization of ISE with respect to h is equivalent to the minimization of

$$\int_{-\infty}^{+\infty} \hat{p}_\Delta(y|x)^2 f(x) I(x \in [a,b]) dxdy - 2\int \hat{p}_\Delta(y|x) p_\Delta(y|x) f(x) I(x \in [a,b]) dxdy. \tag{39}$$

A reasonable estimate of (39) is

$$\begin{aligned}
CV(h_1, h_2) &= \frac{1}{n}\sum_{i=1}^{n} I(X_{(i-1)\Delta} \in [a,b]) \int_{-\infty}^{+\infty} \hat{p}_{\Delta,-i}(y|X_{(i-1)\Delta})^2 dy \\
&\quad -\frac{2}{n}\sum_{i=1}^{n} \hat{p}_{\Delta,-i}(X_{i\Delta}|X_{(i-1)\Delta}) I(X_{(i-1)\Delta} \in [a,b])
\end{aligned} \tag{40}$$

where $\hat{p}_{\Delta,-i}(y|x)$ is the estimator (38) based on the sample $\{(X_{(j-1)\Delta}, X_{j\Delta}), j \neq i\}$ or more precisely

$$\left\{\left(X_{(j-1)\Delta}, K_{h_2}(X_{j\Delta} - y)\right), j \neq i\right\}.$$

The term "cross-validation" refers to setting a part of the data aside for validation of a model and using the remaining data to build the model. Minimizing $CV(h_1, h_2)$ over a region gives a data-driven bandwidth.

Direct use of $CV(h)$ requires very intensive computation, since n local linear estimators are needed. A simplified version is a 5-folded cross-validation as follow:

$$\begin{aligned}
CV_5(h_1, h_2) &= \frac{1}{5}\sum_{j=1}^{5} \frac{1}{|G(j)|} \sum_{i \in G(j)} I(X_{(i-1)\Delta} \in [a,b]) \int \hat{p}_{\Delta,-G(j)}(y|X_{(i-1)\Delta})^2 dy \\
&\quad -\frac{2}{5}\sum_{j=1}^{5} \frac{1}{|G(j)|} \sum_{i \in G(j)} \hat{p}_{\Delta,-G(j)}(X_{i\Delta}|X_{(i-1)\Delta}) I(X_{(i-1)\Delta} \in [a,b])
\end{aligned} \tag{41}$$

where $G(j) = \{5k + j : 5k + j \leq n\}$, $|G(j)|$ is the number of element in $G(j)$ and $\hat{p}_{\Delta,-G(j)}$ is the estimator (38) without using the pair of data $\{(X_{(i-1)\Delta}, X_{i\Delta}), i \in G(j)\}$. The procedure requires only to evaluate 5 local linear estimators.

The transition distribution can be estimated by integrating the estimator (38) over y. Alternative estimators can be obtained by an application of the local logistic regression and adjusted Nadaraya-Watson method of Hall *et al.*(1999).

Early references on the estimation of the transition distributions and densities include Roussas (1967, 1969) and Rosenblatt (1970).

4.3 Testing parametric diffusion models and Markovian properties

With estimated transition density, one can now verify whether models such as (4)–(7) are valid. Let $p_{\Delta,\theta}(y|x)$ be the transition density under a parametric diffusion model. For example, for the CIR model (6), the parameter $\theta = (\kappa, \alpha, \sigma)$. As in (24), ignoring the initial value X_0, the parameter θ can be estimated by maximizing

$$\ell(p_{\Delta,\theta}) = \sum_{i=1}^{n} \log p_{\Delta,\theta}(X_{i\Delta}|X_{(i-1)\Delta}).$$

Let $\hat{\theta}$ be the maximum likelihood estimator. By the spirit of the GLR of Fan *et al.*(2001), the GLR test for the null hypothesis

$$H_0 : p_\Delta(y|x) = p_{\Delta,\theta}(y|x)$$

is

$$\mathrm{GLR} = \ell(\hat{p}_\Delta) - \ell(p_{\Delta,\hat{\theta}}),$$

where $\hat{p}$ is a nonparametric estimate of the transition density. Since the transition density can not be estimated well over the region where data are sparse (usually at boundaries of the process), we need to truncate the nonparametric (and simultaneously parametric) evaluation of the likelihood at appropriate intervals.

In addition to employing the GLR test, one can also compare directly the difference between the parametric and nonparametric fits, resulting in test statistics such as $\|\hat{p}_\Delta - p_{\Delta,\hat{\theta}}\|^2$ for an appropriate norm $\|\cdot\|$. An alternative method is to apply the GLR of Fan *et al.*(2001) to separately test the forms of the drift and diffusion, as in Fan and Zhang (2003). The transition density approach appears more elegant, but more computationally intensive.

One can also use the transition density to test whether an observed series is Markovian (from personal communication with Yacine Aït-Sahalia). For example, if a process $\{X_{i\Delta}\}$ is Markovian, then

$$p_{2\Delta}(y|x) = \int_{-\infty}^{+\infty} p_\Delta(y|z)p_\Delta(z|x)dz.$$

Thus, one can compute the distance between $\hat{p}_{2\Delta}(y|x)$ and $\int_{-\infty}^{+\infty} \hat{p}_\Delta(y|z)\hat{p}_\Delta(z|x)dz$.

References

[1] Ahn, D.H. and Gao, B. (1999). A parametric nonlinear model of term structure dynamics. *Review of Financial Studies*, **12**, 721-762.

[2] Aït-Sahalia, Y. (1996a) Nonparametric pricing of interest rate derivative securities. *Econometrica*, **64**, 527-560.

[3] Aït-Sahalia, Y. (1996b). Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, **9**, 385-426.

[4] Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance*, **LIV**, 1361-1395.

[5] Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, **70**, 223-262.

[6] Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, to appear.

[7] Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, **53**, 499-547

[8] Arfi, M. (1998). Non-parametric Variance Estimation from Ergodic Samples. *Scandinavian Journal of Statistics*, **25**, 225–234.

[9] Banon, G. (1977). Estimation non paramétrique de densité de probabilité pour les processus de Markov. Thése, Université Paul Sabatier de Toulouse, France.

[10] Banon, G. (1978). Nonparametric Identification for Diffusion Processes. *SIAM J. Control and Optimization*, 16, 380-395.

[11] Banon, G., and Nguyen, H. T. (1981). Recursive Estimation in Diffusion Models. *SIAM J. Control and Optimization*, **19**, 676-685.

[12] Bingham, N.H. and Rüdiger, K. (1998). *Risk-Neutral Valuation: Pricing and hedging of financial derivatives*. Springer-verlag, New York.

[13] Black, F., Derman, E. and Toy, W. (1990). A one-factor model of interest rates and its application to treasury bond options. *Financial Analysts' Journal*, **46**, 33-39.

[14] Black, F. and Karasinski, P. (1991). Bond and option pricing when short rates are lognormal. *Financial Analysts' Journal*, **47**, 52-59.

[15] Black, F. and Scholes, M. (1973). The pricing of Options and Corporate Liabilities. *Journal of Political Economy*, **81**, 637-659.

[16] Breeden, D. and Litzenberger, R.H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, **51**, 621-651.

[17] Bowman, A.W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**, 353–360.

[18] Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of American Statistical Association*, **95**, 941-956.

[19] Campbell, J.Y., Lo, A. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton, N.J.

[20] Chan, K.C., Karolyi, A.G., Longstaff, F.A. and Sanders, A.B. (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, **47**, 1209-1227.

[21] Chapman, D.A. and Pearson, N.D. (2000). Is the short rate drift actually nonlinear ? *Journal of Finance*, **LV**, 355-388.

[22] Chen, R. and Tsay, R.J. (1993). Functional-coefficient autoregressive models. *J. Amer. Statist. Assoc.*, **88**, 298–308.

[23] Claeskens, G. and Hall, P. (2002). Effect of dependence on stochastic measures of accuracy of density estimators. *The Annals of Statistics*, **30**, 431 -454.

[24] Cleveland, W.S., Grosse, E. and Shyu, W.M. (1991). Local regression models. In *Statistical Models in S* (Chambers, J.M. and Hastie, T.J., eds), 309–376. Wadsworth & Brooks, Pacific Grove.

[25] Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1980). An analysis of variable rate loan contracts. *Journal of Finance*, **35**, 389-403.

[26] Cox, J.C., Ingersoll, J.E. and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrika*, **53**, 385-467.

[27] Cox, J.C. and Ross, S. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, **3**, 145-166.

[28] Dacunha-Castelle, D. and Florens, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, **19**, 263-284.

[29] Dalalyan, A.S. and Kutoyants, Y.A. (2000). Asymptotically efficient estimation of the derivative of the invariant density. Prepublication 00-4, Université du Maine, Le Mans.

[30] Dalalyan, A.S. and Kutoyants, Y.A. (2003). Asymptotically efficient trend coefficient estimation for ergodic diffusion. *Math. Methods of Statist.*, to appear.

[31] Duffie, D. (2001). *Dynamic Asset Pricing Theory* (Third Edition). Princeton University Press, Princeton, New Jersey.

[32] Egorov, A.V., Li, H. and Xu, Y. (2003). Maximum likelihood estimation of time-inhomogeneous diffusions. *Journal of Econometrics*, **114**, 107-139.

[33] Engelbert, H.J. and Schmidt, W. (1984). On one-dimensional stochastic differential equations with generalized drift. In *Lect. Notes in Control and Inform. Sci.*, **69**, 143-155. Springer, Berlin.

[34] Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, 998–1004.

[35] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

[36] Fan, J., Jiang, J., Zhang, C. and Zhou, Z. (2003). Time-dependent diffusion models for term structure dynamics and the stock price volatility. *Statistica Sinica*, to appear.

[37] Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, **85**, 645-660.

[38] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-verlag, New York.

[39] Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, **83**, 189–206.

[40] Fan, J. and Zhang, C.(2003). A re-examination of Stanton's diffusion estimations with applications to financial model validation. *Journal of American Statistical Association*, **98**, 118-134.

[41] Fan, J., Zhang, C.M. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics*, **29**, 153-193.

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

[43] Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73**, 625-33.

[44] Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincar Probab. Statist.*, **29**, 119-151.

[45] Gobet, E. (2003). LAN property for ergodic diffusions with discrete observations. *Annales de l'IHP–Probabilités et Statistique*, to appear.

[46] Gobet, E., Hoffmann, M. and Reiss, M. (2002) Nonparametric estimation of scalar diffusions based on low frequency data is ill-posed. *Manuscript*.

[47] Härdle, W. , Herwatz, H. Spokoiny, V. (2003). Time inhomogeneous multiple volatility modelling. *Financial Econometrics*, to appear.

[48] Härdle, W. and Yatchew, A. (2002). Dynamic nonparametric state price density estimation using constrained least-squares and the bootstrap. Discussion paper 16, Quantification and Simulation of Economics Processes, Humboldt-Universität zu Berlin.

[49] Härdle, W. and Tsybakov, A.B. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics*, **81**, 223–242.

[50] Hall, P. and Carroll, R.J. (1989). Variance function estimation in regression: the effect of estimation of the mean. *Journal of the Royal Statistical Society, Series B*, **51**, 3-14.

[51] Hall, P., Kay, J.W. and Titterington, D.M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, **77**, 521-8.

[52] Hall, P., Wolff, R. C. L., Yao, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.*, **94**, 154–163.

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

[54] Hansen, L.P., and J.A. Scheinkman. (1995), Back to the future: Generating moment implications for continuous-Time Markov processes, *Econometrica* 63, 767-804.

[55] Hansen, L.P., Scheinkman, J.A. and Touzi, N. (1998). Spectral methods for identifying scalar diffusions. *Journal of Econometrics*, **86**, 1-32.

[56] Harrison, J. M. and Kreps, D. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, **20**, 281-408.

[57] Hart, J. D. (1996). Some automated methods of smoothing time-dependent data, *Nonparametric Statistics*, 6, 115-142.

[58] Hastie, T.J. and Tibshirani, R. J. (1993). Varying-coefficient models (with discussion). *Jour. Roy. Statist. Soc. B.*, **55**, 757-796.

[59] Ho, T. S. Y. and Lee, S. B. (1986). Term structure movements and pricing interest rate contingent claims. *Journal of Finance*, **41**, 1011-1029.

[60] Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85**, 809-822.

[61] Hull, J. (1997). *Options, Futures, and Other Derivatives* (Third edition). Prentice Hall, Upper Saddle Review, New Jersey.

[62] Hull, J. and White, A. (1990). Pricing interest-rate derivative securities. *Review of Financial Studies*, **3**, 573-592.

[63] Itô, K. (1942). Differential equations determining Markov processes (in Japanese). *Zenkoku Shijo Sugaku Danwakai*, **244**, 1352-1400.

[64] Itô, K. (1946). On a stochastic integral equation. *Proc. Imp. Acad. Tokyo*, **22**, 32-35.

[65] Kallenberg, O. (2001). *Foundations of Modern Probability.* (2nd ed.). Springer, New York.

[66] Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus* (2nd ed.). Springer, New York.

[67] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scan. Jour. Statist.*, **24**, 211-229.

[68] Kessler, M. and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, **5**, 299-314.

[69] Kloeden, P.E. and Platen, E. (1995). *Numerical methods for stochastic differential equations* (2nd ed.), Springer, Berlin.

[70] Kloeden, P. E., Platen, E., Schurz, H., and Sørensen, M. (1996). On effects of discretization on estimators of drift parameters for diffusion processes. *Journal of Applied Probability*, **33**, 1061-1076.

[71] Kutoyants, Y.A. (1998). Efficient density estimation for ergodic diffusion. *Statistical Inference for Stochastic Processes*, **1**, 131-155.

[72] Mercurio, D. and Spokoiny, V. (2003). Statistical inference for time-inhomogeneous volatility models. *The Annals of Statistics*, to appear.

[73] Merton, R. (1973). Rational theory of option pricing. *Bell Journal of Economics and Management Science*, **4**, 141-183.

[74] Müller, H.G. and Stadtmüller U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.*, **15**, 610-25.

[75] Müller, H.G. and Stadtmüller U. (1993). On variance function estimation with quadratic forms. *J. Statist. Plann. Inf.*, **35**, 213-231.

[76] Øksendal, B. *Stochastic Differential Equations: An Introduction with Applications.* Springer-verlag, Italy.

[77] Osborne, M. F. M. (1959). Brownian Motion in the Stock Market. *Operations Research*, **7**, 145-173.

[78] Pham, D.T. (1981). Nonparametric estimation of the drift coefficient in the diffusion equation. *Math. Operationsforsch. Statist. Ser. Statist.*, **12**, 61–73.

[79] Prakasa Rao, B. L. S. (1985). Estimation of the drift for diffusion process. *Statistics*, **16**, 263–275.

[80] Rice, J.A. (1984). Boundary modification for nonparametric regression. *Commun. Statist. Theory Meth.*, **13**, 893–900.

[81] Rogers, L.C.G. and Williams, D. (1987). *Diffusions, Markov Processes and Martingales*, Vol. 2, Itô Calculus. Wiley, New York.

[82] Rosenblatt, M. (1970). Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference*, (M.L. Puri, ed.), 199–213, Cambridge University Press, London

[83] Roussas, G. (1967). Nonparametric estimation in Markov Processes. *Ann. Inst. Stat. Math.*, **21**, 73 – 87.

[84] Roussas, G.G. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.*, **40**, 1386–1400.

[85] Ruppert, D., Wand, M.P., Holst, U. and Hössjer, O. (1997). Local polynomial variance function estimation. *Technometrics*, **39**, 262-273.

[86] Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.*, **9**, 65–78.

[87] Schurz, H. (2000). *Numerical analysis of stochastic differential equations without tears.* In *Handbook of Stochastic Analysis and Applications* (Kannan, D. and Lakshmikantham, V., eds), 237 – 359, Marcel Dekker, New York.

[88] Simonoff, J. S. (1996). *Smoothing Methods in Statistics.* Springer, New York.

[89] Spokoiny, V. (2000). Drift estimation for nonparametric diffusion. *Then Annals of Statistics*, **28**, 815-836.

[90] Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *J. Finance*, **52**, 1973-2002.

[91] Steele, M.J. (2000). *Stochastic Calculus and Financial Applications.* Springer-verlag, New York.

[92] Vasicek, O.A. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**, 177-188.

[93] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* Chapman and Hall, London.

[94] Yoshida, N. (1992). Estimation for diffusion processes from discrete observations. *Jour. Multivar. Anal.*, **41**, 220-242.