Statistical Science
2005, Vol. 20, No. 4, 317–337
DOI 10.1214/088342305000000412

A Selective Overview of Nonparametric Methods in Financial Econometrics

Jianqing Fan

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

Key words and phrases: Asset pricing, diffusion, drift, GLR tests, simulations, state price density, time-inhomogeneous model, transition density, volatility.

1. INTRODUCTION

Technological innovation and trade globalization have brought us into a new era of financial markets. Over the last three decades, a large number of new financial products have been introduced to meet customers' demands. An important milestone occurred in 1973 when the world's first options exchange opened in Chicago. That same year, Black and Scholes [23] published their famous paper on option pricing and Merton [90] launched the general equilibrium model for security pricing, two important 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 computational power in portfolio management, securities regulation, proprietary trading, financial consulting and risk management.

Financial econometrics is an active field that integrates finance, economics, probability, statistics and applied mathematics. This is exemplified by the books by Campbell, Lo and MacKinlay [28], Gouriéroux and Jasiak [60] and Cochrane [36]. Financial activities generate many new problems, economics provides useful theoretical foundation and guidance, and quantitative methods such as statistics, probability and applied mathematics are essential tools to solve the quantitative problems in finance. To name a few, complex financial products pose new challenges on their valuation and risk management. Sophisticated stochastic models have been introduced to capture the salient features of underlying economic variables and to price derivatives of securities. Statistical tools are used to identify parameters of stochastic models, to simulate complex financial systems and to test economic theories via empirical financial data.

An important area of financial econometrics is study of the expected returns and volatilities of the price dynamics of stocks and bonds. Returns and volatilities are directly related to asset pricing, proprietary trading, security regulation and portfolio management. To achieve these objectives, the stochastic dynamics of underlying state variables should be correctly specified. For example, option pricing theory allows one to value stock or index options and hedge against the risks of option writers once a model for the dynamics of underlying state variables is given. See, for example, the books on mathematical finance by Bingham and Kiesel [20], Steele [105] and Duffie [42]. Yet many of the stochastic models in use are simple and convenient ones to facilitate mathematical derivations and statistical inferences. They are not derived from any economics theory and hence cannot be expected to fit all financial data. Thus, while the pricing theory gives

Jianqing Fan is Professor, Benheim Center of Finance and Department of Operations Research and Financial Engineering, Princeton University, Princeton, New Jersey 08544, USA (e-mail: jqfan@princeton.edu).

spectacularly beautiful formulas when the underlying dynamics is correctly specified, it offers little guidance in choosing or validating a model. There is always the danger that misspecification of a model leads to erroneous valuation and hedging strategies. Hence, there are genuine needs for flexible stochastic modeling. Nonparametric methods offer a unified and elegant treatment for such a purpose.

Nonparametric approaches have recently been introduced to estimate return, volatility, transition densities and state price densities of stock prices and bond yields (interest rates). They are also useful for examining the extent to which the dynamics of stock prices and bond yields vary over time. They have immediate applications to the valuation of bond price and stock options and management of market risks. They can also be employed to test economic theory such as the capital asset pricing model and stochastic discount model [28] and answer questions such as if the geometric Brownian motion fits certain stock indices, whether the Cox–Ingersoll–Ross model fits yields of bonds, and if interest rate dynamics evolve with time. Furthermore, based on empirical data, one can also fit directly the observed option prices with their associated characteristics such as strike price, the time to maturity, risk-free 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 and the development of financial econometrics.

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 aspects of the models. 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. Section 5 makes some concluding remarks.

2. STOCHASTIC DIFFUSION MODELS

Much of financial econometrics is concerned with asset pricing, portfolio choice and risk management. 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. For an introduction to financial derivatives, see Hull [78].

2.1 One-Factor Diffusion Models

Let $S_{t\Delta}$ denote the stock price observed at time $t\Delta$. The time unit can be hourly, daily, weekly, among others. Presented in Figure 1(a) are the daily log-returns, defined as

$$\log(S_{t\Delta}) - \log\bigl(S_{(t-1)\Delta}\bigr) \approx \bigl(S_{t\Delta} - S_{(t-1)\Delta}\bigr)/S_{(t-1)\Delta},$$

of the Standard and Poor's 500 index, a value-weighted index based on the prices of the 500 stocks that account for approximately 70% of the total U.S. equity (stock) market capitalization. The styled features of the returns include that the volatility tends to cluster and that the (marginal) mean and variance of the returns tend to be constant. One simplified model to capture the second feature is that

$$\log(S_{t\Delta}) - \log\bigl(S_{(t-1)\Delta}\bigr) \approx \mu_0 + \sigma_0\varepsilon_t,$$

where $\{\varepsilon_t\}$ is a sequence of independent normal random variables. This is basically a random walk hypothesis, regarding the stock price movement as an independent random walk. When the sampling time unit Δ gets small, the above random walk can be regarded as a random sample from the continuous-time process:

$$d\log(S_t) = \mu_0 + \sigma_1\,dW_t, \tag{1}$$

where $\{W_t\}$ is a standard one-dimensional Brownian motion and $\sigma_1 = \sigma_0/\sqrt{\Delta}$. The process (1) is called geometric Brownian motion as S_t is an exponent of Brownian motion W_t. It was used by Osborne [92] to model the stock price dynamic and by Black and Scholes [23] to derive their celebrated option price formula.

Interest rates are fundamental to financial markets, consumer spending, corporate earnings, asset pricing, inflation and the economy. The bond market is even bigger than the equity market. Presented in Figure 1(c) are the interest rates $\{r_t\}$ of the two-year U.S. Treasury notes at a weekly frequency. As the interest rates get higher, so do the volatilities. To appreciate this, Figure 1(d) plots the pairs $\{(r_{t-1}, r_t - r_{t-1})\}$. Its dynamic is very different from that of the equity market. The interest rates should be nonnegative. They possess heteroscedasticity in addition to the mean-revision property: As the interest rates rise above the mean level α, there is a negative drift that pulls the rates down; while when the interest rates fall below α, there is a positive force that drives the rates up. To capture these two

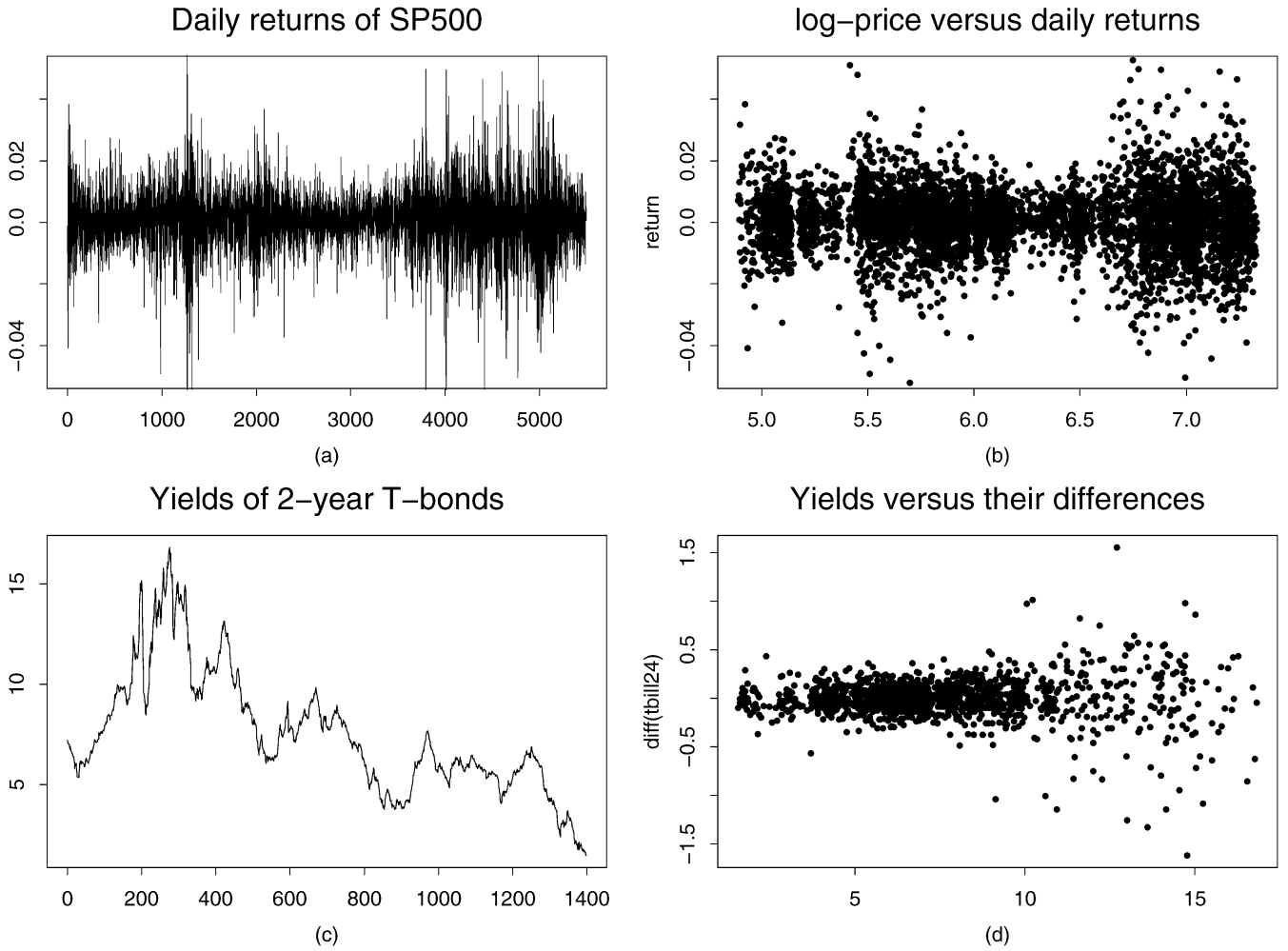

FIG. 1. (a) *Daily log-returns of the Standard and Poor's* 500 *index from October 21, 1980 to July 29, 2004.* (b) *Scatterplot of the returns against logarithm of the index (price level).* (c) *Interest rates of two-year U.S. Treasury notes from June 4, 1976 to March 7, 2003 sampled at weekly frequency.* (d) *Scatterplot of the difference of yields versus the yields.*

main features, Cox, Ingersoll and Ross [37] derived the following model for the interest rate dynamic:

$$dr_t = \kappa(\alpha - r_t)\,dt + \sigma r_t^{1/2}\,dW_t. \tag{2}$$

For simplicity, we will refer it to as the CIR model. It is an amelioration of the Vasicek model [106],

$$dr_t = \kappa(\alpha - r_t)\,dt + \sigma\,dW_t, \tag{3}$$

which ignores the heteroscedasticity and is also referred to as the Ornstein–Uhlenbeck process. While this is an unrealistic model for interest rates, the process is Gaussian with explicit transition density. It fact, the time series sampled from (3) follows the autoregressive model of order 1,

$$Y_t = (1-\rho)\alpha + \rho Y_{t-1} + \varepsilon_t, \tag{4}$$

where $Y_t = r_{t\Delta}$, $\varepsilon \sim N(0, \sigma^2(1-\rho^2)/(2\kappa))$ and $\rho = \exp(-\kappa\Delta)$. Hence, the process is well understood and usually serves as a test case for proposed statistical methods.

There are many stochastic models that have been introduced to model the dynamics of stocks and bonds. Let X_t be an observed economic variable at time t. This can be the price of a stock or a stock index, or the yield of a bond. A simple and frequently used stochastic model is

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t. \tag{5}$$

The function $\mu(\cdot)$ is often called a drift or instantaneous return function and $\sigma(\cdot)$ is referred to as a diffusion or volatility function, since

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

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

The time-homogeneous model (5) contains many famous one-factor models in financial econometrics. In an effort to improve the flexibility of modeling interest dynamics, Chan et al. [29] extends the CIR model (2) to the CKLS model,

$$dX_t = \kappa(\alpha - X_t)\,dt + \sigma X_t^{\gamma}\,dW_t. \tag{6}$$

Aït-Sahalia [3] introduces a 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, a strong nonlinear mean reversion emerges. He imposes the nonlinear drift of the form $(\alpha_0 X_t^{-1} + \alpha_1 + \alpha_2 X_t + \alpha_2 X_t^2)$. See also Ahn and Gao [1], which models the interest rates by $Y_t = X_t^{-1}$, in which the X_t follows the CIR model.

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 (5) to allow the coefficients to depend on time t:

$$dX_t = \mu(X_t, t)\,dt + \sigma(X_t, t)\,dW_t. \tag{7}$$

Since only a trajectory of the process is observed [see Figure 1(c)], there is not sufficient information to estimate the bivariate functions in (7) without further restrictions. [To consistently estimate the bivariate volatility function $\sigma(x,t)$, we need to have data that eventually fill up a neighborhood of the point (t,x).] A useful specification of model (7) is

$$dX_t = \{\alpha_0(t) + \alpha_1(t)X_t\}\,dt + \beta_0(t)X_t^{\beta_1(t)}\,dW_t. \tag{8}$$

This is an extension of the CKLS model (6) by allowing the coefficients to depend on time and was introduced and studied by Fan et al. [48]. Model (8) includes many commonly used time-varying models for the yields of bonds, introduced by Ho and Lee [75], Hull and White [79], Black, Derman and Toy [21] and Black and Karasinski [22], among others. The experience in [48] and other studies of the varying coefficient models [26, 31, 74, 76] shows that coefficient functions in (8) cannot be estimated reliably due to the collinearity effect in local estimation: localizing in the time domain, the process $\{X_t\}$ is nearly constant and hence $\alpha_0(t)$ and $\alpha_1(t)$ and $\beta_0(t)$ and $\beta_1(t)$ cannot easily be differentiated. This leads Fan et al. [48] to introduce the semiparametric model

$$dX_t = \{\alpha_0(t) + \alpha_1 X_t\}\,dt + \beta_0(t)X_t^{\beta}\,dW_t \tag{9}$$

to avoid the collinearity.

2.2 Some Probabilistic Aspects

The question when there exists a solution to the stochastic differential equation (SDE) (7) arises naturally. Such a program was first carried out by Itô [80, 81]. For SDE (7), there are two different meanings of solution: strong solution and weak solution. See Sections 5.2 and 5.3 of [84]. Basically, for a given initial condition ξ, a strong solution requires that X_t is determined completely by the information up to time t. Under Lipschitz and linear growth conditions on the drift and diffusion functions, for every ξ that is independent of $\{W_s\}$, there exists a strong solution of equation (7). Such a solution is unique. See Theorem 2.9 of [84].

For the one-dimensional time-homogeneous diffusion process (5), weaker conditions can be obtained for the so-called weak solution. By an application of the Itô formula to an appropriate transform of the process, one can make the transformed process have zero drift. Thus, we can consider without loss of generality that the drift in (5) is zero. For such a model, Engelbert and Schmidt [45] give a necessary and sufficient condition for the existence of the solution. The continuity of σ suffices for the existence of the weak solution. See Theorem 5.5.4 of [84], page 333, and Theorem 23.1 of [83].

We will use several times the Itô formula. For the process X_t in (7), for a sufficiently regular function f ([84], page 153),

$$\begin{aligned} df(X_t,t) = &\left\{\frac{\partial f(X_t,t)}{\partial t}\right. \\ &\left. + \frac{1}{2}\frac{\partial^2 f(X_t,t)}{\partial x^2}\sigma^2(X_t,t)\right\}dt \\ &+ \frac{\partial f(X_t,t)}{\partial x}\,dX_t. \end{aligned} \tag{10}$$

The formula can be understood as the second-order Taylor expansion of $f(X_{t+\Delta}, t+\Delta) - f(X_t,t)$ by noticing that $(X_{t+\Delta} - X_t)^2$ is approximately $\sigma^2(X_t, t)\Delta$.

The Markovian property plays an important role in statistical inference. According to Theorem 5.4.20 of [84], the solution X_t to equation (5) is Markovian, provided that the coefficient functions μ and σ are bounded on compact subsets. Let $p_\Delta(y|x)$ be the transition density, the conditional density of $X_{t+\Delta} = y$ given $X_t = x$. The transition density must satisfy the forward and backward Kolmogorov equations ([84], page 282).

Under the linear growth and Lipschitz conditions, and additional conditions on the boundary behavior of

the functions μ and σ, the solution to equation (1) is positive and ergodic. The invariant density is given by

$$f(x) = 2C_0\sigma^{-2}(x) \cdot \exp\left(-2\int_{\cdot}^{x} \mu(y)\sigma^{-2}(y)\,dy\right), \tag{11}$$

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 with the marginal density f and transition density p_Δ.

Stationarity plays an important role in time series analysis and forecasting [50]. The structural invariability allows us to forecast the future based on the historical data. For example, the structural relation (e.g., the conditional distribution, conditional moments) between X_t and $X_{t+\Delta}$ remains the same over time t. This makes it possible to use historical data to estimate the invariant quantities. Associated with stationarity is the concept of mixing, which says that the data that are far apart in time are nearly independent. We now describe the conditions under which the solution to the SDE (1) is geometrically mixing.

Let H_t be the operator defined by

$$(H_t g)(x) = E\big(g(X_t)|X_0 = x\big), \quad x \in R, \tag{12}$$

where f is a Borel measurable bounded function on R. A stationary process X_t is said to satisfy the condition $G_2(s,\alpha)$ of Rosenblatt [95] if there exists an s such that

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

namely, the operator is contractive. As a consequence of the semigroup ($H_{s+t} = H_s H_t$) and contraction properties, the condition G_2 implies [16, 17] that for any $t \in [0,\infty)$, $\|H_t\|_2 \le \alpha^{t/s-1}$. The latter implies, by the Cauchy–Schwarz inequality, that

$$\rho(t) = \sup_{g_1,g_2} \operatorname{corr}\big(g_1(X_0), g_2(X_t)\big) \le \alpha^{t/s-1}, \tag{13}$$

that is, the ρ-mixing coefficient decays exponentially fast. Banon and Nguyen [18] show further that for a stationary Markov process, $\rho(t) \to 0$ is equivalent to (13), namely, ρ-mixing and geometric ρ-mixing are equivalent.

2.3 Valuation of Contingent Claims

An important application of SDE is the pricing of financial derivatives such as options and bonds. It forms a beautiful modern asset pricing theory and provides useful guidance in practice. Steele [105], Duffie [42] and Hull [78] offer very nice introductions to the field.

The simplest financial derivative is the European call option. A call option is the right to buy an asset at a certain price K (strike price) before or at expiration time T. A put option gives the right to sell an asset at 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 payoff for a European call option is $(X_T - K)_+$, where X_T is the price of the stock at expiration T. When the stock rises above the strike price K, one can exercise the right and make a profit of $X_T - K$. However, when the stock falls below K, one renders one's right and makes no profit. Similarly, a European put option has payoff $(K - X_T)_+$. See Figure 2. By creating a portfolio with different maturities and different strike prices, one can obtain all kinds of payoff functions. As an example, suppose that a portfolio of options consists of contracts of the S&P 500 index maturing in six months: one call option with strike price \$1,200, one put option with strike price \$1,050 and \$40 cash, but with short position (borrowing or -1 contract) on a call option with strike price \$1,150 and on a put option with strike price \$1,100. Figure 2(c) shows the payoff function of such a portfolio of options at the expiration T. Clearly, such an investor bets the S&P 500 index will be around \$1,125 in six months and limits the risk exposure on the investment (losing at most \$10 if his/her bet is wrong). Thus, the European call and put options are fundamental options as far as the payoff function at time T is concerned. There are many other exotic options such as Asian options, look-back options and barrier options, which have different payoff functions, and the payoffs can be path dependent. See Chapter 18 of [78].

Suppose that the asset price follows the SDE (7) and there is a riskless investment alternative such as a bond which earns compounding rate of interest r_t. Suppose that the underlying asset pays no dividend. Let β_t be the value of the riskless bond at time t. Then, with an initial investment β_0,

$$\beta_t = \beta_0 \exp\left(\int_0^t r_s\,ds\right),$$

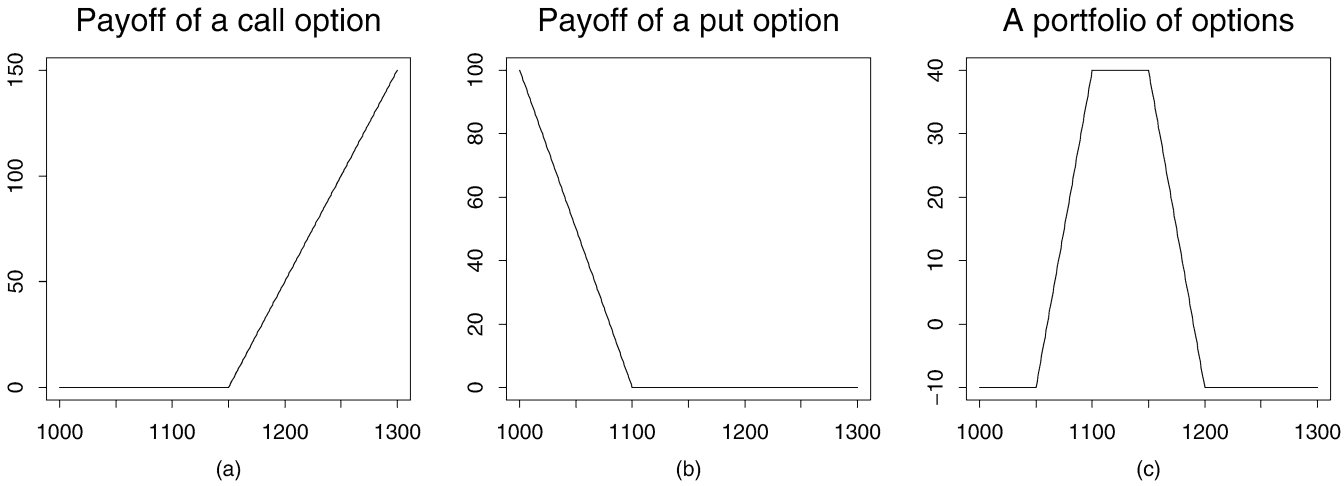

FIG. 2. (a) *Payoff of a call option*. (b) *Payoff of a put option*. (c) *Payoff of a portfolio of four options with different strike prices and different* (*long and short*) *positions*.

thanks to the compounding of interest. Suppose that a probability measure Q 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 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 [42].

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

$$P_0 = \exp\left(-\int_0^T r_s\,ds\right) E^Q\big(\Psi(X_T)|X_0 = x_0\big), \tag{14}$$

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 payoff in the risk neutral world. The formula is derived by using the so-called relative pricing approach, which values the price of the option from given prices of a portfolio consisting of a risk-free bond and a stock with the identical payoff as the option at the expiration.

As an illustrative example, suppose that the price of a stock follows the geometric Brownian motion $dX_t = \mu X_t\,dt + \sigma X_t\,dW_t$ and that the risk-free rate r is constant. Then the deflated price process $Y_t = \exp(-rt)X_t$ follows the SDE

$$dY_t = (\mu - r)Y_t\,dt + \sigma Y_t\,dW_t.$$

The deflated price process is not a martingale as the drift is not zero. The risk-neutral measure is the one that makes the drift zero. To achieve this, we appeal to the Girsanov theorem, which changes the drift of a diffusion process without altering the diffusion via a change of probability measure. Under the "risk-neutral" probability measure Q, the process Y_t satisfies $dY_t = \sigma Y_t\,dW_t$, a martingale. Hence, the price process $X_t = \exp(rt)Y_t$ under Q follows

$$dX_t = rX_t\,dt + \sigma X_t\,dW_t. \tag{15}$$

Using exactly the same derivation, one can easily generalize the result to the price process (5). Under the risk-neutral measure, the price process (5) follows

$$dX_t = rX_t\,dt + \sigma(X_t)\,dW_t. \tag{16}$$

The intuitive explanation of this is clear: all stocks under the "risk-neutral" world are expected to earn the same rate as the risk-free bond.

For the geometric Brownian motion, by an application of the Itô formula (10) to (15), we have under the "risk-neutral" measure

$$\log X_t - \log X_0 = (r - \sigma^2/2)t + \sigma^2 W_t. \tag{17}$$

Note that given the initial price X_0, the price follows a log-normal distribution. Evaluating the expectation of (14) for the European call option with payoff $\Psi(X_T) = (X_T - K)_+$, one obtains the Black–Scholes [23] option pricing formula

$$P_0 = x_0\Phi(d_1) - K\exp(-rT)\Phi(d_2), \tag{18}$$

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

2.4 Simulation of Stochastic Models

Simulation methods provide useful tools for the valuation of financial derivatives and other financial instruments when the analytical formula (14) is hard to obtain. For example, if the price under the "risk-neutral" measure is (16), the analytical formula for pricing derivatives is usually not analytically tractable and simulation methods offer viable alternatives (together with variance reduction techniques) to evaluate it. They also provide useful tools for assessing performance of statistical methods and statistical inferences.

The simplest method is perhaps the Euler scheme. The SDE (7) is approximated as

$$X_{t+\Delta} = X_t + \mu(t, X_t)\Delta + \sigma(t, X_t)\Delta^{1/2}\varepsilon_t, \tag{19}$$

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 (19) to obtain a sequence of simulated data $\{X_{j\Delta}, j = 1, 2, \ldots\}$. The approximation error can be reduced if one uses a smaller step size Δ/M for a given integer M to first obtain a more detailed sequence $\{X_{j\Delta/M}, j = 1, 2, \ldots\}$ and then one takes the subsequence $\{X_{j\Delta}, j = 1, 2, \ldots\}$. For example, to simulate daily prices of a stock, one can simulate hourly data first and then take the daily closing prices. Since the step size Δ/M is smaller, the approximation (19) 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. [87]. The higher-order approximations can be obtained by the Itô–Taylor expansion (see [100], page 242). In particular, a strong order-one approximation is given by

$$\begin{aligned} X_{t+\Delta} &= X_t + \mu(t, X_t)\Delta + \sigma(t, X_t)\Delta^{1/2}\varepsilon_t \\ &\quad + \tfrac{1}{2}\sigma(t, X_t)\sigma_x'(t, X_t)\Delta\{\varepsilon_t^2 - 1\}, \end{aligned} \tag{20}$$

where $\sigma_x'(t, x)$ is the partial derivative function with respect to x. This method can be combined with a smaller step size method in the last paragraph. For the time-homogeneous model (1), an alternative form, without evaluating the derivative function, is given in (3.14) of [87].

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 (11). In the latter case, the generated sequence is stationary.

There are only a few processes where exact simulation is possible. For GBM, one can generate the sequence from the explicit solution (17), where the Brownian motion can be simulated from independent Gaussian increments. The conditional density of Vasicek's model (3) is Gaussian with mean $\alpha + (x_0 - \alpha)\rho$ and variance $\sigma_\Delta^2 = \sigma^2(1 - \rho^2)/(2\kappa)$ as indicated by (4). Generate X_0 from the invariant density $N(\alpha, \sigma^2/(2\kappa))$. With X_0, generate X_Δ from the normal distribution with mean $\alpha + (X_0 - \alpha)\exp(-\kappa\Delta)$ and variance σ_Δ^2. With X_Δ, we generate $X_{2\Delta}$ from mean $\alpha + (X_\Delta - \alpha)\exp(-\kappa\Delta)$ and variance σ_Δ^2. Repeat this process until we obtain the desired length of the process.

For the CIR model (2), provided that $q = 2\kappa\alpha/\sigma^2 - 1 \geq 0$ (a sufficient condition for $X_t \geq 0$), the transition density is determined by the fact that given $X_t = x_0$, $2cX_{t+\Delta}$ has a noncentral χ^2 distribution with degrees of freedom $2q + 2$ and noncentrality parameter $2u$, where $c = 2\kappa/\{\sigma^2(1 - \exp(-\kappa\Delta))\}$, $u = cx_0\exp(k\Delta)$. The invariant density is the Gamma distribution with shape parameter $q + 1$ and scale parameter $\sigma^2/(2\kappa)$.

As an illustration, we consider the CIR model (7) with parameters $\kappa = 0.21459$, $\alpha = 0.08571$, $\sigma = 0.07830$ and $\Delta = 1/12$. The model parameters are taken from [30]. We simulated 1000 monthly data values using both the Euler scheme (19) and the strong order-one approximation (20) with the same random shocks. Figure 3 depicts one of their trajectories. The difference is negligible. This is in line with the observations made by Stanton [104] that as long as data are sampled monthly or more frequently, the errors introduced by using the Euler approximation are very small for stochastic dynamics that are similar to the CIR model.

3. ESTIMATION OF RETURN AND VOLATILITY FUNCTIONS

There is a large literature on the estimation of the return and volatility functions. Early references include [93] and [94]. 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.

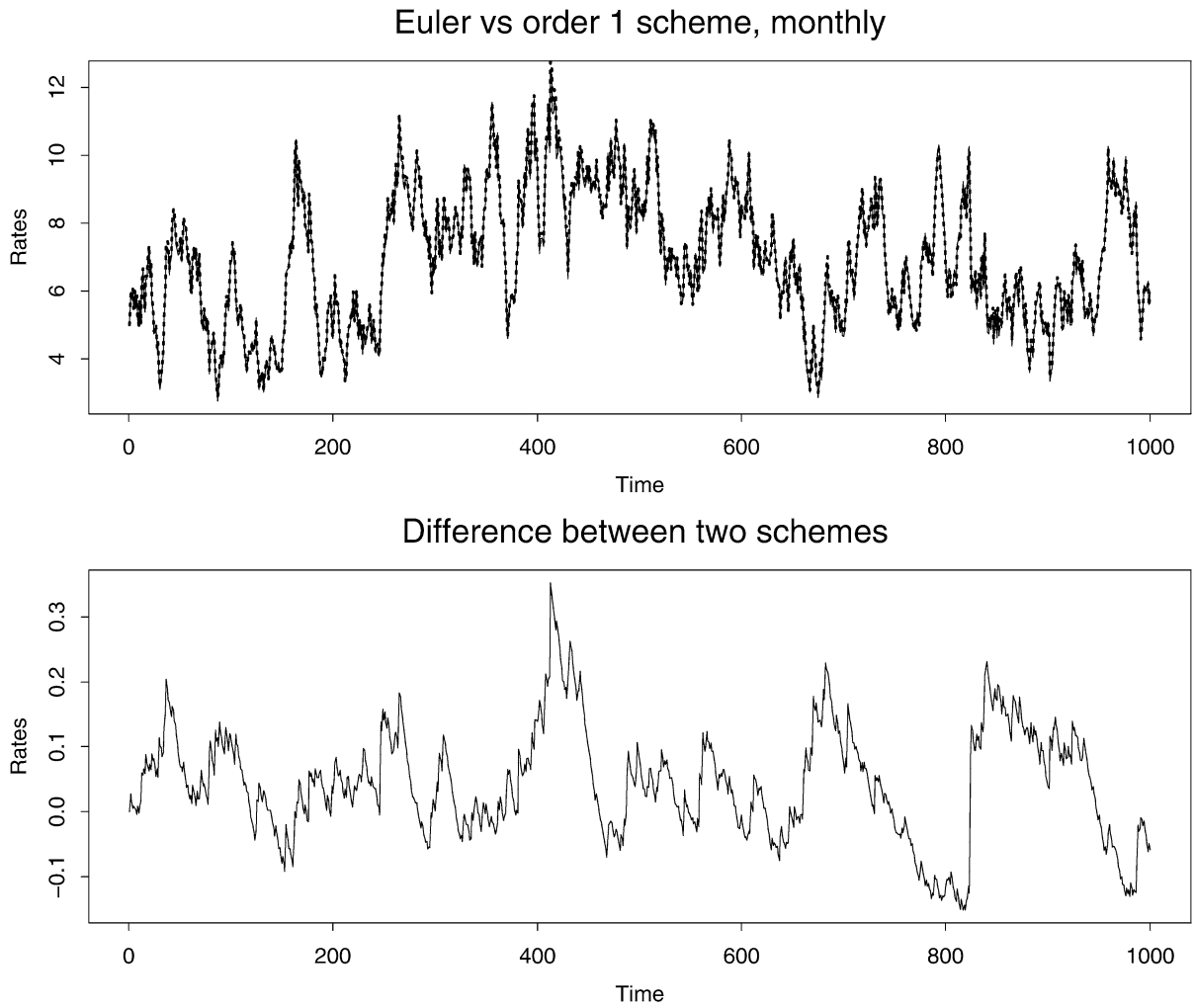

FIG. 3. *Simulated trajectories* (*multiplied by* 100) *using the Euler approximation and the strong order-one approximation for a CIR model. Top panel*: *solid curve corresponds to the Euler approximation and the dashed curve is based on the order-one approximation. Bottom panel*: *the difference between the order-one scheme and the Euler scheme.*

3.1 Methods of Estimation

We first outline several methods of estimation for parametric models. The idea can be extended to nonparametric models. Suppose that we have a sample $\{X_{i\Delta}, i = 0, \ldots, n\}$ from model (5). Then, the likelihood function, under the stationary condition, is

$$\log f(X_0) + \sum_{i=1}^{n} \log p_\Delta\big(X_{i\Delta} | X_{(i-1)\Delta}\big). \tag{21}$$

If the functions μ and σ are parameterized and the explicit form of the transition density is available, one can apply the maximum likelihood method. However, the explicit form of the transition density is not available for many simple models such as the CLKS model (6). Even for the CIR model (2), its maximum likelihood estimator is very difficult to find, as the transition density involves the modified Bessel function of the first kind.

One simple technique is to rely on the Euler approximation scheme (19). Then proceed as if the data come from the Gaussian location and scale model. This method works well when Δ is small, but can create some biases when Δ is large. However, the bias can be reduced by the following calibration idea, called indirect inference by Gouriéroux et al. [61]. The idea works as follows. Suppose that the functions μ and σ have been parameterized with unknown parameters θ. Use the Euler approximation (19) and the maximum likelihood method to obtain an estimate $\hat{\theta}_0$. For each given parameter θ around $\hat{\theta}_0$, simulate data from (5) and apply the crude method to obtain an estimate $\hat{\theta}_1(\theta)$ which depends on θ. Since we simulated the data with the true

parameter θ, the function $\hat{\theta}_1(\theta)$ tells us how to calibrate the estimate. See Figure 4. Calibrate the estimate via $\hat{\theta}_1^{-1}(\hat{\theta}_0)$, which improves the bias of the estimate. One drawback of this method is that it is intensive in computation and the calibration cannot easily be done when the dimensionality of parameters θ is high.

Another method for bias reduction is to approximate the transition density in (21) by a higher order approximation, and to then maximize the approximated likelihood function. Such a scheme has been introduced by Aït-Sahalia [4, 5], who derives the expansion of the transition density around a normal density function using Hermite polynomials. The intuition behind such an expansion is that the diffusion process $X_{t+\Delta} - X_t$ in (5) can be regarded as sum of many independent increments with a very small step size and hence the Edgeworth expansion can be obtained for the distribution of $X_{t+\Delta} - X_t$ given X_t. See also [43].

An "exact" approach is to use the method of moments. If the process X_t is stationary as in the interest-rate models, the moment conditions can easily be derived by observing

$$E\left\{\lim_{\Delta\to 0}\Delta^{-1}E[g(X_{t+\Delta})-g(X_t)|X_t]\right\} = \lim_{\Delta\to 0}\Delta^{-1}E[g(X_{t+\Delta})-g(X_t)]=0$$

for any function g satisfying the regularity condition that the limit and the expectation are exchangeable. The right-hand side is the expectation of $dg(X_t)$. By Itô's formula (10), the above equation reduces to

$$E[g'(X_t)\mu(X_t)+g''(X_t)\sigma^2(X_t)/2]=0. \tag{22}$$

For example, if $g(x)=\exp(-ax)$ for some given $a>0$, then

$$E\exp(-aX_t)\{\mu(X_t)-a\sigma^2(X_t)/2\}=0.$$

This can produce an arbitrary number of equations by choosing different a's. If the functions μ and σ are parameterized, the number of moment conditions can be more than the number of equations. One way to efficiently use this is the generalized method of moments introduced by Hansen [65], minimizing a quadratic form of the discrepancies between the empirical and the theoretical moments, a generalization of the classical method of moments which solves the moment equations. The weighting matrix in the quadratic form can be chosen to optimize the performance of the resulting estimator. To improve the efficiency of the estimate, a large system of moments is needed. Thus, the generalized method of moments needs a large system of nonlinear equations which can be expensive in computation. Further, the moment equations (22) use only the marginal information of the process. Hence, the procedure is not efficient. For example, in the CKLS model (6), σ and κ are estimable via (22) only through σ^2/κ.

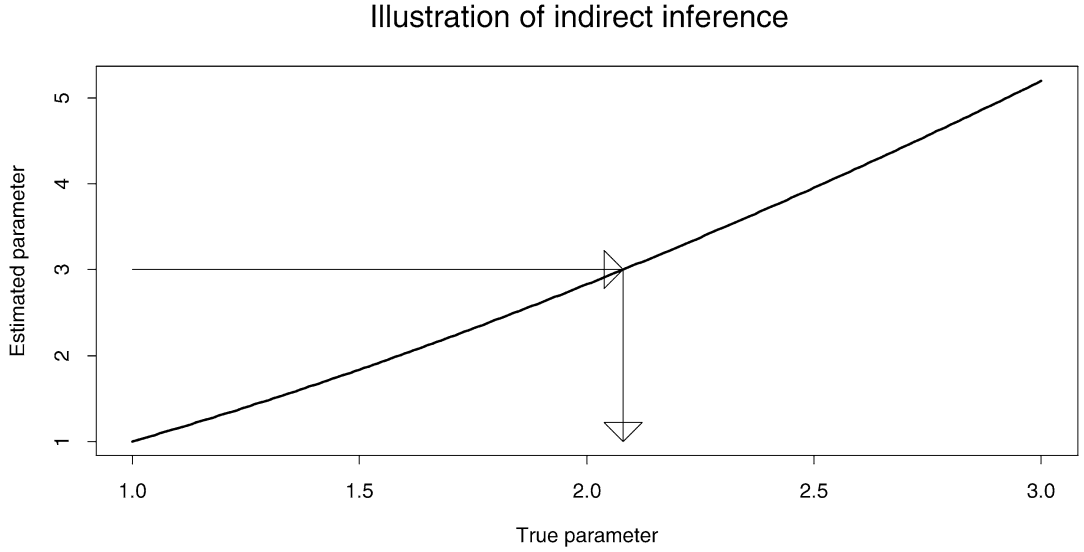

FIG. 4. *The idea of indirect inference. For each given true θ, one obtains an estimate using the Euler approximation and the simulated data. This gives a calibration curve as shown. Now for a given estimate $\hat{\theta}_0 = 3$ based on the Euler approximation and real data, one finds the calibrated estimate $\hat{\theta}_1^{-1}(3) = 2.080$.*

3.2 Time-Homogeneous Model

The Euler approximation can easily be used to estimate the drift and diffusion nonparametrically. 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

$$E(Y_{i\Delta}|X_{i\Delta}) = \mu(X_{i\Delta}) + O(\Delta)$$

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 [104] applies kernel regression [102, 107] 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

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

and

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

where $K_h(u) = h^{-1}K(u/h)$ is a rescaled kernel. The consistency and asymptotic normality of the estimator are studied in [15]. Fan and Yao [49] apply the local linear technique (Section 6.3 in [50]) to estimate the return and volatility functions, under a slightly different setup. The local linear estimator [46] is given by

$$\hat{\mu}(x) = \sum_{i=0}^{n-1} K_n(X_{i\Delta} - x, x) Y_{i\Delta}, \tag{23}$$

where

$$K_n(u, x) = K_h(u) \frac{S_{n,2}(x) - u S_{n,1}(x)}{S_{n,2}(x) S_{n,0}(x) - S_{n,1}(x)^2}, \tag{24}$$

with $S_{n,j}(x) = \sum_{i=0}^{n-1} K_h(X_{i\Delta} - x)(X_{i\Delta} - x)^j$, is the equivalent kernel induced by the local linear fit. In contrast to the kernel method, the local linear weights depend on both X_i and x. In particular, they satisfy

$$\sum_{i=1}^{n-1} K_n(X_{i\Delta} - x, x) = 1$$

and

$$\sum_{i=1}^{n-1} K_n(X_{i\Delta} - x, x)(X_{i\Delta} - x) = 0.$$

These are the key properties for the bias reduction of the local linear method as demonstrated in [46]. Further, Fan and Yao [49] use the squared residuals

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

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

Stanton [104] derives a higher-order approximation scheme up to order three in an effort to reduce biases. He suggests that higher-order approximations must outperform lower-order approximations. To verify such a claim, Fan and Zhang [53] derived the following order k approximation scheme:

$$\begin{aligned} E(Y^*_{i\Delta}|X_{i\Delta}) &= \mu(X_{i\Delta}) + O(\Delta^k), \\ E(Z^*_{i\Delta}|X_{i\Delta}) &= \sigma^2(X_{i\Delta}) + O(\Delta^k), \end{aligned} \tag{25}$$

where

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

and

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

and the coefficients $a_{k,j} = (-1)^{j+1}\binom{k}{j}/j$ are chosen to make the approximation error in (25) of order Δ^k. For example, the second approximation is

$$1.5(X_{t+\Delta} - X_t) - 0.5(X_{t+2\Delta} - X_{t+\Delta}).$$

By using the independent increments of Brownian motion, its variance is $1.5^2 + 0.5^2 = 2.5$ times as large as that of the first-order difference. Indeed, Fan and Zhang [53] show that while higher-order approximations give better approximation errors, we have to pay a huge premium for variance inflation,

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

where the variance inflation factors $V_1(k)$ and $V_2(k)$ are explicitly given by Fan and Zhang [53]. Table 1 shows some of the numerical results for the variance inflation factor.

The above theoretical results have also been verified via empirical simulations in [53]. The problem is no monopoly for nonparametric fitting—it is shared by

TABLE 1
Variance inflation factors by using higher-order differences

	Order k				
	1	**2**	**3**	**4**	**5**
$V_1(k)$	1.00	2.50	4.83	9.25	18.95
$V_2(k)$	1.00	3.00	8.00	21.66	61.50

the parametric methods. Therefore, the methods based on higher-order differences should seldomly be used unless the sampling interval is very wide (e.g., quarterly data). It remains open whether it is possible to estimate nonparametrically the return and the volatility functions without seriously inflating the variance with other higher-order approximation schemes.

As an illustration, we take the yields of the two-year Treasury notes depicted in Figure 1. Figure 5 presents nonparametrically estimated volatility functions, based on order $k = 1$ and $k = 2$ approximations. The local linear fit is employed with the Epanechnikov kernel and bandwidth $h = 0.35$. It is evident that the order two approximation has higher variance than the order one approximation. In fact, the magnitude of variance inflation is in line with the theoretical result: the increase of the standard deviation is $\sqrt{3}$ from order one to order two approximation.

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, [12, 27, 39, 58, 59, 85, 109] and references therein. Arapis

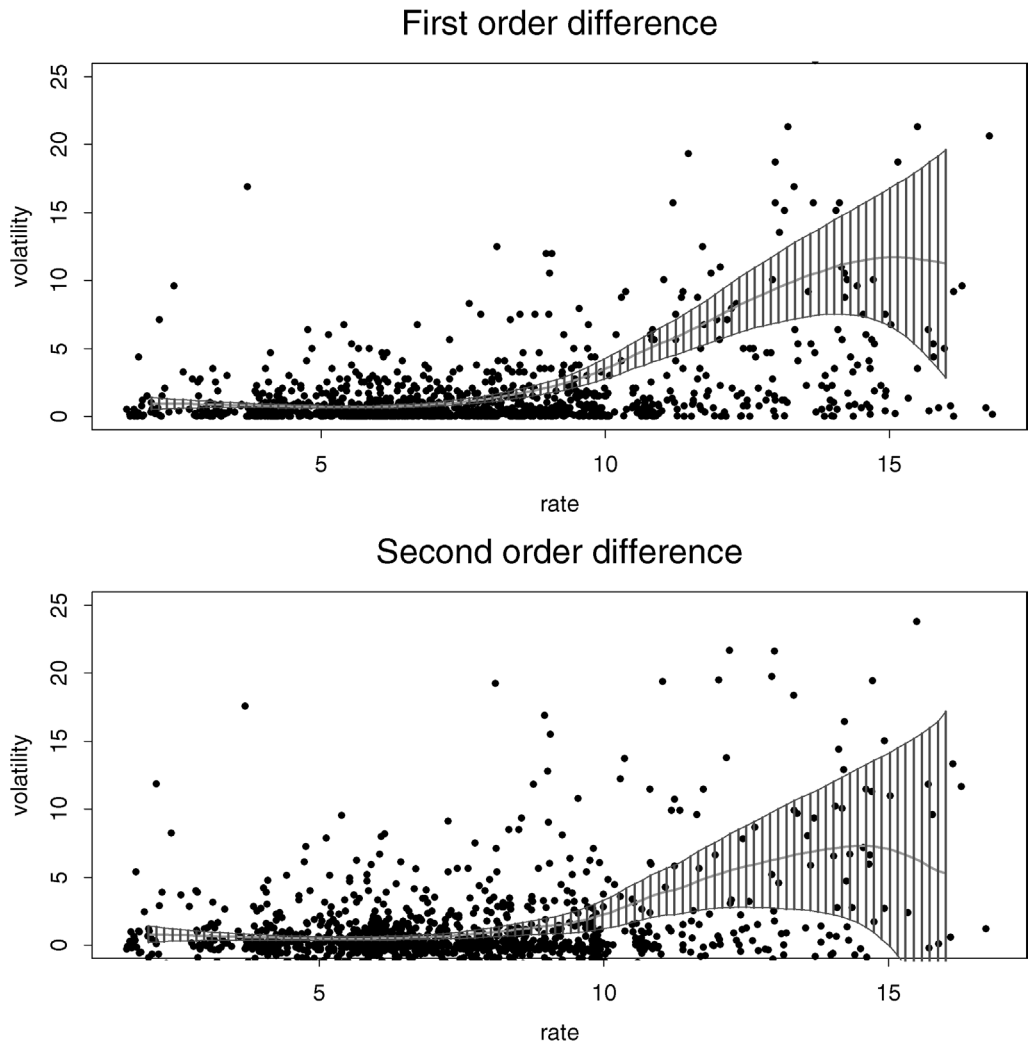

FIG. 5. *Nonparametric estimates of volatility based on order one and two differences. The bars represent two standard deviations above and below the estimated volatility. Top panel*: *order one fit. Bottom panel*: *order two fit.*

and Gao [11] investigate the mean integrated square error of several methods for estimating the drift and diffusion and compare their performances. Aït-Sahalia and Mykland [9, 10] study the effects of random and discrete sampling when estimating continuous-time diffusions. Bandi and Nguyen [14] investigate small sample behavior of nonparametric diffusion estimators. Thorough study of nonparametric estimation of conditional variance functions can be found in [62, 69, 91, 99]. In particular, Section 8.7 of [50] gives various methods for estimating the conditional variance function. Wang [108] studies the relationship between diffusion and GARCH models.

3.3 Model Validation

Stanton [104] applies his kernel estimator to a Treasury bill data set and observes a nonlinear return function in his nonparametric estimate, particularly in the region where the interest rate is high (over 14%, say). This leads him to postulate the hypothesis that the return functions of short-term rates are nonlinear. Chapman and Pearson [30] study the finite sample properties of Stanton's estimator. By applying his procedure to the CIR model, they find that Stanton's procedure produces spurious nonlinearity, due to the boundary effect and the mean reversion.

Can we apply a formal statistics test to Stanton's hypothesis? The null hypothesis can simply be formulated: the drift is of a linear form as in model (6). What is the alternative hypothesis? For such a problem our alternative model is usually vague. Hence, it is natural to assume that the drift is a nonlinear smooth function. This becomes a testing problem with a parametric null hypothesis versus a nonparametric alternative hypothesis. There is a large body of literature on this. The basic idea is to compute a discrepancy measure between the parametric estimates and nonparametric estimates and to reject the parametric hypothesis when the discrepancy is large. See, for example, the book by Hart [73].

In an effort to derive a generally applicable principle, Fan et al. [54] propose the generalized likelihood ratio (GLR) tests for parametric-versus-nonparametric or nonparametric-versus-parametric hypotheses. The basic idea is to replace the maximum likelihood under a nonparametric hypothesis (which usually does not exist) by the likelihood under good nonparametric estimates. Section 9.3 of [50] gives details on the implementation of the GLR tests, including estimating P-values, bias reduction and bandwidth selection. The method has been successfully employed by Fan and Zhang [53] for checking whether the return and volatility functions possess certain parametric forms.

Another viable approach of model validation is to base it on the transition density. One can check whether the nonparametrically estimated transition density is significantly different from the parametrically estimated one. Section 4.3 provides some additional details. Another approach, proposed by Hong and Li [77], uses the fact that under the null hypothesis the random variables $\{Z_i\}$ are a sequence of i.i.d. uniform random variables where $Z_i = P(X_{i\Delta}|X_{(i-1)\Delta}, \theta)$ and $P(y|x, \theta)$ is the transition distribution function. They propose to detect the departure from the null hypothesis by comparing the kernel-estimated bivariate density of $\{(Z_i, Z_{i+1})\}$ with that of the uniform distribution on the unit square. The transition-density-based approaches appear more elegant as they check simultaneously the forms of drift and diffusion. However, the transition density does often not admit an analytic form and the tests can be computationally intensive.

3.4 Fixed Sampling Interval

For practical analysis of financial data, it is hard to determine whether the sampling interval tends to zero. The key determination is whether the approximation errors for small "Δ" are negligible. It is ideal when a method is applicable whether or not "Δ" is small. This kind of method is possible, as demonstrated below.

The simplest problem to illustrate the idea is the kernel density estimation of the invariant density of the stationary process $\{X_t\}$. For the given sample $\{X_{t\Delta}\}$, the kernel density estimate for the invariant density is

$$\hat{f}(x) = n^{-1} \sum_{i=1}^{n} K_h(X_{i\Delta} - x), \tag{26}$$

based on the discrete data $\{X_{i\Delta}, i = 1, \ldots, n\}$. This method is valid for all Δ. It gives a consistent estimate of f as long as the time horizon is long: $n\Delta \to \infty$. We will refer to this kind of nonparametric method as state-domain smoothing, as the procedure localizes in the state variable X_t. Various properties, including consistency and asymptotic normality, of the kernel estimator (26) are studied by Bandi [13] and Bandi and Phillips [15]. Bandi [13] also uses the estimator (26), which is the same as the local time of the process spending at a point x except for a scaling constant, as a descriptive tool for potentially nonstationary diffusion processes.

Why can the state-domain smoothing methods be employed as if the data were independent? This is due

to the fact that localizing in the state domain weakens the correlation structure and that nonparametric estimates use essentially only local data. Hence many results on nonparametric estimators for independent data continue to hold for dependent data as long as their mixing coefficients decay sufficiently fast. As mentioned at the end of Section 2.2, geometric mixing and mixing are equivalent for time-homogeneous diffusion processes. Hence, the mixing coefficients decay usually sufficiently fast for theoretical investigation.

The localizing and whitening can be understood graphically in Figure 6. Figure 6(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 $8\% \pm 0.2\%$. In fact, as detailed in Figure 6(b), the indices of the data that fall in the local window are quite far apart. This in turn implies the weak dependence for the data in the local window, that is, "whitening by windowing." See Section 5.4 of [50] and Hart [72] for further details. The effect of dependence structure on kernel density estimation was thoroughly studied by Claeskens and Hall [35].

The diffusion function can also be consistently estimated when Δ is fixed. In pricing the derivatives of interest rates, Aït-Sahalia [2] assumes $\mu(x) = k(\alpha - x)$. Using the kernel density estimator $\hat{f}$ and estimated κ and α from a least-squares method, he applied (11) to estimate $\sigma(\cdot): \hat{\sigma}^2(x) = 2\int_0^x \hat{\mu}(u)\hat{f}(u)\,du/\hat{f}(x)$. He further established the asymptotic normality of such an estimator. Gao and King [56] propose tests of diffusion models based on the discrepancy between the paramet-

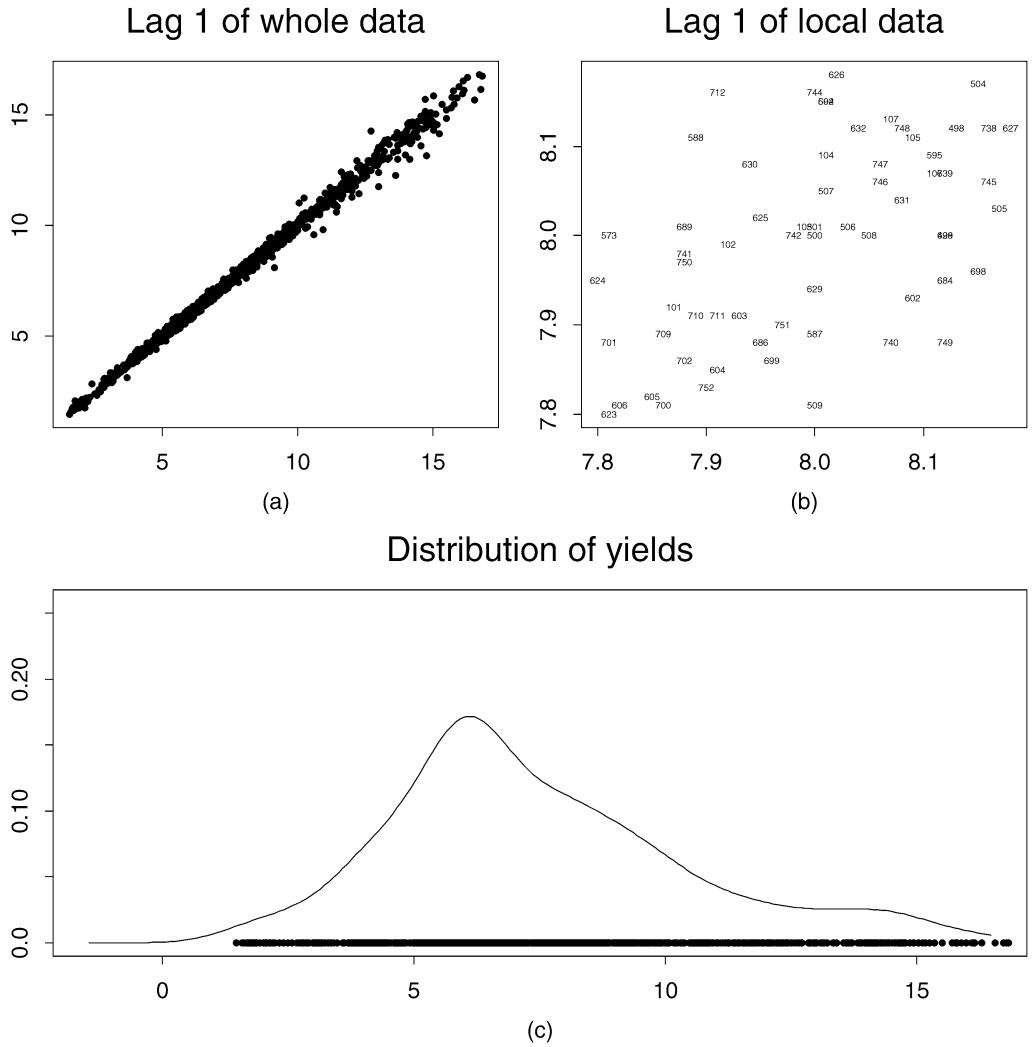

FIG. 6. (a) *Lag* 1 *scatterplot of the two-year Treasury note data.* (b) *Lag* 1 *scatterplot of those data falling in the neighborhood* $8\% \pm 0.2\%$*—the points are represented by the times of the observed data. The numbers in the scatterplot show the indices of the data falling in the neighborhood.* (c) *Kernel density estimate of the invariant density.*

ric and nonparametric estimates of the invariant density.

The Aït-Sahalia method [2] easily illustrates that the volatility function can be consistently estimated for fixed Δ. However, we do not expect that it is efficient. Indeed, we use only the marginal information of the data. As shown in (21), almost all information is contained in the transition density $p_\Delta(\cdot|\cdot)$. The transition density can be estimated as in Section 4.2 below whether Δ is small or large. Since the transition density and drift and volatility are in one-to-one correspondence for the diffusion process (5), the drift and diffusion functions can be consistently estimated via inverting the relationship between the transition density and the drift and diffusion functions.

There is no simple formula for expressing the drift and diffusion in terms of the transition density. The inversion is frequently carried out via a spectral analysis of the operator $H_\Delta = \exp(\Delta L)$, where the infinitesimal operator L is defined as

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

It has the property

$$Lg(x) = \lim_{\Delta \to 0} \Delta^{-1} [E\{g(X_{t+\Delta})|X_t = x\} - g(x)]$$

by Itô's formula (10). The operator H_Δ is the transition operator in that [see also (12)]

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

The works of Hansen and Scheinkman [66], Hansen, Scheinkman and Touzi [67] and Kessler and Sørensen [86] consist of the following idea. The first step is to estimate the transition operator H_Δ from the data. From the transition operator, one can identify the infinitesimal operator L and hence the functions $\mu(\cdot)$ and $\sigma(\cdot)$. More precisely, let λ_1 be the largest negative eigenvalue of the operator L with eigenfunction $\xi_1(x)$. Then $L\xi_1 = \lambda_1 \xi_1$, or equivalently, $\sigma^2 \xi_1'' + 2\mu\xi_1' = 2\lambda_1 \xi_1$. This gives one equation of μ and σ. Another equation can be obtained via (11): $(\sigma^2 f)' - 2\mu f = 0$. Solving these two equations we obtain

$$\sigma^2(x) = 2\lambda_1 \int_0^x \xi_1(y) f(y)\, dy / [f(x)\xi_1(x)]$$

and another explicit expression for $\mu(x)$. Using semigroup theory ([44], Theorem IV.3.7), ξ_1 is also an eigenfunction of H_Δ with eigenvalue $\exp(\Delta\lambda_1)$. Hence, the proposal is to estimate the invariant density f and the transition density $p_\Delta(y|x)$, which implies the values of λ_1 and ξ_1. Gobet [58] derives the optimal rate of convergence for such a scheme, using a wavelet basis. In particular, [58] shows 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.

3.5 Time-Dependent Model

The time-dependent model (8) was introduced to accommodate the possibility of economic changes over time. The coefficient functions in (8) are assumed to be slowly time-varying and smooth. Nonparametric techniques can be applied to estimate these coefficient functions. The basic idea is to localizing in time, resulting in a time-domain smoothing.

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 (19), 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 the minimizers of the local regression (27). Fan et al. [48] 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 are no significant differences between nonparametric fitting with one-sided and two-sided kernels. We opt for local constant approximations instead of local linear approximations in (27), 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. To appreciate this, for constant functions α_1 and α_2 a large bandwidth will be chosen to reduce the variance in the estimation. This is in essence fitting a global linear regression by (27). If the local linear approximations are used, since no variable selection procedures have been incorporated in the local fitting (27), the slopes of the local linear approximations will not be estimated as zero and hence artificial linear trends will be created for the estimated coefficients.

The coefficient functions in the volatility can be estimated by the local approximated likelihood method. Let

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

be the normalized residuals. Then

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

The conditional log-likelihood of $\hat{E}_t$ given X_t can easily be obtained by the approximation (28). Using local constant approximations and incorporating the kernel weight, we obtain the local approximated likelihood at each time point and estimates of the functions $\beta_0(\cdot)$ and $\beta_1(\cdot)$ at that time point. This type of local approximated-likelihood method is related to the generalized method of moments of Hansen [65] and the ideas of Florens-Zmirou [55] and Genon-Catalot and Jacod [57].

Since the coefficient functions in both return and volatility functions are estimated using only historical data, their bandwidths can be selected based on a form of the average prediction error. See Fan et al. [48] 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)]

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

but we do not continue in 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.

The question arises naturally if the coefficients in the model (8) 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. [48] proposed a formal test for this kind of problem.

The coefficient functions in the semiparametric model (9) can also be estimated by using the profile approximated-likelihood method. For each given β_1, one can easily estimate $\beta_0(\cdot)$ via the approximation (28), resulting in an estimate $\hat{\beta}_0(\cdot;\beta_1)$. Regarding the nonparametric function $\beta_0(\cdot)$ as being parameterized by $\hat{\beta}_0(\cdot;\beta_1)$, model (28) with $\beta_1(t) \equiv \beta_1$ becomes a "synthesized" parametric model with unknown β_1. The parameter β_1 can be estimated by the maximum (approximated) likelihood method. Note that β_1 is estimated by using all the data points, while $\hat{\beta}_0(t) = \hat{\beta}_0(t;\hat{\beta}_1)$ is obtained by using only the local data points. See [48] for details.

For other nonparametric methods of estimating volatility in time inhomogeneous models, see Härdle, Herwartz and Spokoiny [68] and Mercurio and Spokoiny [89]. Their methods are based on model (8) with $\alpha_1(t) = \beta_1(t) = 0$.

3.6 State-Domain Versus Time-Domain Smoothing

So far, we have introduced both state- and time-domain smoothing. The former relies on the structural invariability implied by the stationarity assumption and depends predominantly on the (remote) historical data. The latter uses the continuity of underlying parameters and concentrates basically on the recent data. This is illustrated in Figure 7 using the yields of the three-month Treasury bills from January 8, 1954 to July 16,

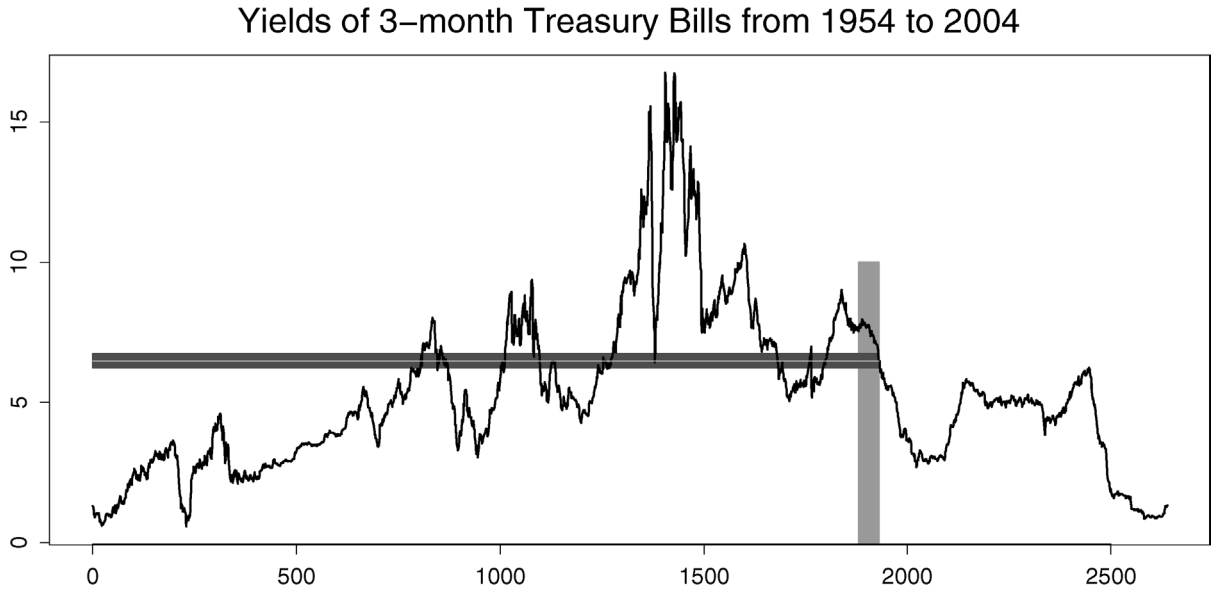

FIG. 7. *Illustration of time- and state-domain smoothing using the yields of three-month Treasury bills. The state-domain smoothing is localized in the horizontal bars, while the time-domain smoothing is concentrated in the vertical bars.*

2004 sampled at weekly frequency. On December 28, 1990, the interest rate was about 6.48%. To estimate the drift and diffusion around $x = 6.48$, the state-domain smoothing focuses on the dynamics where interest rates are around 6.48%, the horizontal bar with interest rates falling in $6.48\% \pm 0.25\%$. The estimated volatility is basically the sample standard deviation of the differences $\{X_{i\Delta} - X_{(i-1)\Delta}\}$ within this horizontal bar. On the other hand, the time-domain smoothing focuses predominantly on the recent history, say one year, as illustrated in the figure. The time-domain estimate of volatility is basically a sample standard deviation within the vertical bar.

For a given time series, it is hard to say which estimate is better. This depends on the underlying stochastic processes and also on the time when the forecast is made. If the underlying process is continuous and stationary, such as model (5), both methods are applicable. For example, standing at December 28, 1990, one can forecast the volatility by using the sample standard deviation in either the horizontal bar or the vertical bar. However, the estimated precision depends on the local data. Since the sample variance is basically linear in the squared differences $\{Z_{i\Delta}^2\}$, the standard errors of both estimates can be assessed and used to guide the forecasting.

For stationary diffusion processes, it is possible to integrate both the time-domain and state-domain estimates. Note that the historical data (with interest rates in $6.48\% \pm 0.25\%$) are far apart in time from the data used in the time-domain smoothing (vertical bar), except the last segment, which can be ignored in the state-domain fitting. The next-to-last segment with interest rates in $6.48\% \pm 0.25\%$ is May 11 to July 20, 1988, 123 weeks prior to the last segment. Hence, these two estimates are nearly independent. The integrated estimate is a linear combination of these two nearly independent estimates. The weights can easily be chosen to minimize the variance of the integrated estimator, by using the assessed standard errors of the state- and time-domain estimators. The optimal weights are proportional to the variances of the two estimators, which depend on time t. This forms a dynamically integrated predictor for volatility estimation, as the optimal weights change over time.

3.7 Continuously Observed Data

At the theoretical level, one may also examine the problem of estimating the drift and diffusion functions assuming the whole process is observable up to time T. Let us assume again that the observed process $\{X_t\}$ follows the SDE (5). 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 (11), estimating the drift function $\mu(x)$ is equivalent to estimating the invariant density f. In fact,

$$\mu(x) = [\sigma^2(x) f(x)]'/[2f(x)]. \tag{29}$$

The invariant density f can easily be estimated by kernel density estimation. When $\Delta \to 0$, the summation in (26) converges to

$$\hat{f}(x) = T^{-1} \int_0^T K_h(X_t - x)\, dt. \tag{30}$$

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 (29). Such an approach has been employed by Kutoyants [88] and Dalalyan and Kutoyants [40, 41]. They established the 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 rates $T^{-1/2}$, $T^{-2s/(2s+1)}$ and $T^{-2s/(2s+1)}$, respectively, where s is the degree of smoothness of μ. These are the optimal rates of convergence.

An alternative approach is to estimate the drift function directly from (23). By letting $\Delta \to 0$, one can easily obtain a local linear regression estimator for continuously observed data, which admits a similar form to (23) and (30). This is the approach that Spokoiny [103] used. He showed that this estimator attains the optimal rate of convergence and established further a data-driven bandwidth such that the local linear estimator attains adaptive minimax rates.

4. ESTIMATION OF STATE PRICE DENSITIES AND TRANSITION DENSITIES

The state price density (SPD) is the probability density of the value of an asset under the risk-neutral world (14) (see [38]) or equivalent martingale measure [71]. It is directly related to the pricing of financial derivatives. It is the transition density of X_T given X_0 under the equivalent martingale Q. The SPD does not depend on the payoff function and hence it can be used to evaluate other illiquid derivatives, once it is estimated from more liquid derivatives. On the other hand, 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 the State Price Density

For some specific models, the state price density can be formed explicitly. For example, for the GBM (1) with a constant risk-free rate r, according to (17), 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

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

See (14) (we have changed the notation from P_0 to C to emphasize the price of the European call option). Hence,

$$f^*(K) = \exp\Big(\int_0^T r_s\,ds\Big)\frac{\partial^2 C}{\partial K^2}. \tag{31}$$

This was observed by Breeden and Litzenberger [25]. 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 using formula (14).

In general, the price of a 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 dividend 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 unless we assume the Black–Scholes model. Based on historical data $\{(C_i, S_i, K_i, T_i, r_i, \delta_i), i = 1, \ldots, n\}$, where C_i is the ith traded-option price with associated characteristics $(S_i, K_i, T_i, r_i, \delta_i)$, Aït-Sahalia and Lo [7] fit the 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, the five-dimensional nonparametric function cannot be estimated well with practical range of sample sizes. Aït-Sahalia and Lo [7] 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 five to four. By assuming further that the option-pricing function is homogeneous of degree one in F and K, namely,

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

they reduced the dimensionality to three. Aït-Sahalia and Lo [7] imposed a semiparametric form on the pricing formula,

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

where $C_{\mathrm{BS}}(F, K, T, r, \sigma)$ is the Black–Scholes pricing formula given in (18) and $\sigma(F, K, T)$ is the implied volatility, computed by inverting the Black–Scholes formula. Thus, the problem becomes one of nonparametrically estimating the implied volatility function $\sigma(F, K, T)$. This is 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 two. This is one of the options in [4].

The state price density f^* is nonnegative and hence the function C should be convex in the strike price K. Aït-Sahalia and Duarte [6] propose to estimate the option price under the convexity constraint using a local linear estimator. See also [70] 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 and for statistical estimation and inference. 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 would be estimated using the methods to be introduced.

Assume that we have a sample $\{X_{i\Delta}, i = 0, \ldots, n\}$ from model (5). The "double-kernel" method of Fan, Yao and Tong [51] is to observe that

$$E\big\{W_{h_2}(X_{i\Delta} - y)|X_{(i-1)\Delta} = x\big\} \approx p_\Delta(y|x) \quad \text{as } h_2 \to 0, \tag{32}$$

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 $W_{h_2}(X_{i\Delta} - y)$ on $X_{(i-1)\Delta}$. An application of the local linear estimator (23) yields

$$\hat{p}_\Delta(y|x) = \sum_{i=1}^n K_n\big(X_{(i-1)\Delta} - x, x\big) \cdot W_{h_2}(X_{i\Delta} - y), \tag{33}$$

where the equivalent kernel $K_n(u, x)$ was defined in (24). Fan, Yao and Tong [51] establish the asymptotic normality of such an estimator under stationarity and ρ-mixing conditions [necessarily decaying at geometric rate for SDE (5)], which gives explicitly the asymptotic bias and variance of the estimator. See also Section 6.5 of [50]. The cross-validation idea of Rudemo [98] and Bowman [24] can be extended to select bandwidths for estimating conditional densities. See [52, 63].

The transition distribution can be estimated by integrating the estimator (33) over y. By letting $h_2 \to 0$, the estimator is the regression of the indicator $I(X_{i\Delta} < y)$ on $X_{(i-1)\Delta}$. Alternative estimators can be obtained by an application of the local logistic regression and adjusted Nadaraya–Watson method of Hall et al. [64].

Early references on the estimation of the transition distributions and densities include [96, 97] and [95].

4.3 Inferences Based on Transition Densities

With the estimated transition density, one can now verify whether parametric models such as (1)–(3), (6) are consistent with the observed data. Let $p_{\Delta,\theta}(y|x)$ be the transition density under a parametric diffusion model. For example, for the CIR model (2), the parameter $\theta = (\kappa, \alpha, \sigma)$. As in (21), 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. [54], the GLR test for the null hypothesis $H_0 : p_\Delta(y|x) = p_{\Delta,\theta}(y|x)$ is

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

where $\hat{p}$ is a nonparametric estimate of the transition density. Since the transition density cannot 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 $\|\hat{p}_\Delta - p_{\Delta,\hat{\theta}}\|^2$ and $\|\hat{P}_\Delta - P_{\Delta,\hat{\theta}}\|^2$ for an appropriate norm $\|\cdot\|$, where $\hat{P}_\Delta$ and $P_{\Delta,\hat{\theta}}$ are the estimates of the cumulative transition distributions under respectively the parametric and nonparametric models. The transition density-based methods depend on two bandwidths and are harder to implement. Indeed, their null distributions are harder to determine than those based on the transition distribution methods. In comparison with the invariant density-based approach of Arapis and Gao [11], it is consistent against a much larger family of alternatives.

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 use the distance between $\hat{p}_{2\Delta}(y|x)$ and $\int_{-\infty}^{+\infty} \hat{p}_\Delta(y|z)\hat{p}_\Delta(z|x)\,dz$ as a test statistic.

The transition density can also be used for parameter estimation. One possible approach is to find the parameter which minimizes the distance $\|\hat{P}_\Delta - P_{\Delta,\theta}\|$. In this case, the bandwidth should be chosen to optimize the performance for estimating θ. The approach is applicable whether or not $\Delta \to 0$.

5. CONCLUDING REMARKS

Enormous efforts in financial econometrics have been made in modeling the dynamics of stock prices and bond yields. There are directly related to pricing derivative securities, proprietary trading and portfolio management. Various parametric models have been proposed to facilitate mathematical derivations. They have risks that misspecifications of models lead to erroneous pricing and hedging strategies. Nonparametric models provide a powerful and flexible treatment. They aim at reducing modeling biases by increasing somewhat the variances of resulting estimators. They provide an elegant method for validating or suggesting a family of parametric models.

The versatility of nonparametric techniques in financial econometrics has been demonstrated in this paper. They are applicable to various aspects of diffusion models: drift, diffusion, transition densities and even state price densities. They allow us to examine whether the stochastic dynamics for stocks and bonds are time varying and whether famous parametric models are consistent with empirical financial data. They permit us to price illiquid or nontraded derivatives from liquid derivatives.

The applications of nonparametric techniques in financial econometrics are far wider than what has been presented. There are several areas where nonparametric methods have played a pivotal role. One example

is to test various versions of capital asset pricing models (CAPM) and their related stochastic discount models [36]. See, for example, the research manuscript by Chen and Ludvigson [34] in this direction. Another important class of models are stochastic volatility models [19, 101], where nonparametric methods can be also applied. The nonparametric techniques have been prominently featured in the RiskMetrics of J. P. Morgan. It can be employed to forecast the risks of portfolios. See, for example, [8, 32, 33, 47, 82] for related nonparametric techniques on risk management.

ACKNOWLEDGMENTS

The author gratefully acknowledges various discussions with Professors Yacine Aït-Sahalia and Jia-an Yan and helpful comments of the editors and reviewers that led to significant improvement of the presentation of this paper. This research was supported in part by NSF Grant DMS-03-55179 and a direct allocation RGC grant of the Chinese University of Hong Kong.

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. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica* **64** 527–560.

[3] AÏT-SAHALIA, Y. (1996). 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. *J. Finance* **54** 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. *J. Econometrics* **116** 9–47.

[7] AÏT-SAHALIA, Y. and LO, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *J. Finance* **53** 499–547.

[8] AÏT-SAHALIA, Y. and LO, A. W. (2000). Nonparametric risk management and implied risk aversion. *J. Econometrics* **94** 9–51.

[9] AÏT-SAHALIA, Y. and MYKLAND, P. (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* **71** 483–549.

[10] AÏT-SAHALIA, Y. and MYKLAND, P. (2004). Estimators of diffusions with randomly spaced discrete observations: A general theory. *Ann. Statist.* **32** 2186–2222.

[11] ARAPIS, M. and GAO, J. (2004). Nonparametric kernel estimation and testing in continuous-time financial econometrics. Unpublished manuscript.

[12] ARFI, M. (1998). Non-parametric variance estimation from ergodic samples. *Scand. J. Statist.* **25** 225–234.

[13] BANDI, F. (2002). Short-term interest rate dynamics: A spatial approach. *J. Financial Economics* **65** 73–110.

[14] BANDI, F. and NGUYEN, T. (1999). Fully nonparametric estimators for diffusions: A small sample analysis. Unpublished manuscript.

[15] BANDI, F. and PHILLIPS, P. C. B. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica* **71** 241–283.

[16] BANON, G. (1977). Estimation non paramétrique de densité de probabilité pour les processus de Markov. Thése, Univ. Paul Sabatier de Toulouse, France.

[17] BANON, G. (1978). Nonparametric identification for diffusion processes. *SIAM J. Control Optim.* **16** 380–395.

[18] BANON, G. and NGUYEN, H. T. (1981). Recursive estimation in diffusion models. *SIAM J. Control Optim.* **19** 676–685.

[19] BARNDOFF-NIELSEN, O. E. and SHEPHARD, N. (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** 167–241.

[20] BINGHAM, N. H. and KIESEL, R. (1998). *Risk-Neutral Valuation*: *Pricing and Hedging of Financial Derivatives*. Springer, New York.

[21] 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**(1) 33–39.

[22] BLACK, F. and KARASINSKI, P. (1991). Bond and option pricing when short rates are lognormal. *Financial Analysts Journal* **47**(4) 52–59.

[23] BLACK, F. and SCHOLES, M. (1973). The pricing of options and corporate liabilities. *J. Political Economy* **81** 637–654.

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

[25] BREEDEN, D. and LITZENBERGER, R. H. (1978). Prices of state-contingent claims implicit in option prices. *J. Business* **51** 621–651.

[26] CAI, Z., FAN, J. and YAO, Q. (2000). Functional-coefficient regression models for nonlinear time series. *J. Amer. Statist. Assoc.* **95** 941–956.

[27] CAI, Z. and HONG, Y. (2003). Nonparametric methods in continuous-time finance: A selective review. In *Recent Advances and Trends in Nonparametric Statistics* (M. G. Akritas and D. N. Politis, eds.) 283–302. North-Holland, Amsterdam.

[28] CAMPBELL, J. Y., LO, A. W. and MACKINLAY, A. C. (1997). *The Econometrics of Financial Markets*. Princeton Univ. Press.

[29] CHAN, K. C., KAROLYI, G. A., LONGSTAFF, F. A. and SANDERS, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. *J. Finance* **47** 1209–1227.

[30] CHAPMAN, D. A. and PEARSON, N. D. (2000). Is the short rate drift actually nonlinear? *J. Finance* **55** 355–388.

[31] CHEN, R. and TSAY, R. S. (1993). Functional-coefficient autoregressive models. *J. Amer. Statist. Assoc.* **88** 298–308.

[32] CHEN, S. X. (2005). Nonparametric estimation of expected shortfall. *Econometric Theory*. To appear.
[33] CHEN, S. X. and TANG, C. Y. (2005). Nonparametric inference of value-at-risk for dependent financial returns. *J. Financial Econometrics* **3** 227–255.
[34] CHEN, X. and LUDVIGSON, S. (2003). Land of Addicts? An empirical investigation of habit-based asset pricing model. Unpublished manuscript.
[35] CLAESKENS, G. and HALL, P. (2002). Effect of dependence on stochastic measures of accuracy of density estimators. *Ann. Statist.* **30** 431–454.
[36] COCHRANE, J. H. (2001). *Asset Pricing*. Princeton Univ. Press.
[37] COX, J. C., INGERSOLL, J. E. and ROSS, S. A. (1985). A theory of the term structure of interest rates. *Econometrica* **53** 385–407.
[38] COX, J. C. and ROSS, S. (1976). The valuation of options for alternative stochastic processes. *J. Financial Economics* **3** 145–166.
[39] DACUNHA-CASTELLE, D. and FLORENS, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **19** 263–284.
[40] DALALYAN, A. S. and KUTOYANTS, Y. A. (2002). Asymptotically efficient trend coefficient estimation for ergodic diffusion. *Math. Methods Statist.* **11** 402–427.
[41] DALALYAN, A. S. and KUTOYANTS, Y. A. (2003). Asymptotically efficient estimation of the derivative of the invariant density. *Stat. Inference Stoch. Process.* **6** 89–107.
[42] DUFFIE, D. (2001). *Dynamic Asset Pricing Theory*, 3rd ed. Princeton Univ. Press.
[43] EGOROV, A. V., LI, H. and XU, Y. (2003). Maximum likelihood estimation of time-inhomogeneous diffusions. *J. Econometrics* **114** 107–139.
[44] ENGEL, K.-J. and NAGEL, R. (2000). *One-Parameter Semigroups for Linear Evolution Equations*. Springer, Berlin.
[45] ENGELBERT, H. J. and SCHMIDT, W. (1984). On one-dimensional stochastic differential equations with generalized drift. *Stochastic Differential Systems. Lecture Notes in Control and Inform. Sci.* **69** 143–155. Springer, Berlin.
[46] FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004.
[47] FAN, J. and GU, J. (2003). Semiparametric estimation of value-at-risk. *Econom. J.* **6** 261–290.
[48] FAN, J., JIANG, J., ZHANG, C. and ZHOU, Z. (2003). Time-dependent diffusion models for term structure dynamics. *Statist. Sinica* **13** 965–992.
[49] FAN, J. and YAO, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85** 645–660.
[50] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series*: *Nonparametric and Parametric Methods*. Springer, New York.
[51] FAN, J., YAO, Q. and TONG, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* **83** 189–206.
[52] FAN, J. and YIM, T. H. (2004). A crossvalidation method for estimating conditional densities. *Biometrika* **91** 819–834.
[53] FAN, J. and ZHANG, C. (2003). A re-examination of diffusion estimators with applications to financial model validation. *J. Amer. Statist. Assoc.* **98** 118–134.
[54] FAN, J., ZHANG, C. and ZHANG, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** 153–193.
[55] FLORENS-ZMIROU, D. (1993). On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.* **30** 790–804.
[56] GAO, J. and KING, M. (2004). Adaptive testing in continuous-time diffusion models. *Econometric Theory* **20** 844–882.
[57] GENON-CATALOT, V. and JACOD, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **29** 119–151.
[58] GOBET, E. (2002). LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.* **38** 711–737.
[59] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253.
[60] GOURIÉROUX, C. and JASIAK, J. (2001). *Financial Econometrics*: *Problems, Models, and Methods*. Princeton Univ. Press.
[61] GOURIÉROUX, C., MONFORT, A. and RENAULT, E. (1993). Indirect inference. *J. Appl. Econometrics* **8** suppl. S85–S118.
[62] HALL, P. and CARROLL, R. J. (1989). Variance function estimation in regression: The effect of estimating the mean. *J. Roy. Statist. Soc. Ser. B* **51** 3–14.
[63] HALL, P., RACINE, J. and LI, Q. (2004). Cross-validation and the estimation of conditional probability densities. *J. Amer. Statist. Assoc.* **99** 1015–1026.
[64] HALL, P., WOLFF, R. C. L. and YAO, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.* **94** 154–163.
[65] HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054.
[66] HANSEN, L. P. and SCHEINKMAN, J. A. (1995). Back to the future: Generating moment implications for continuous-time Markov processes. *Econometrica* **63** 767–804.
[67] HANSEN, L. P., SCHEINKMAN, J. A. and TOUZI, N. (1998). Spectral methods for identifying scalar diffusions. *J. Econometrics* **86** 1–32.
[68] HÄRDLE, W., HERWARTZ, H. and SPOKOINY, V. (2003). Time inhomogeneous multiple volatility modelling. *J. Financial Econometrics* **1** 55–95.
[69] HÄRDLE, W. and TSYBAKOV, A. B. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econometrics* **81** 223–242.
[70] 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.
[71] HARRISON, J. M. and KREPS, D. (1979). Martingales and arbitrage in multiperiod securities markets. *J. Econom. Theory* **2** 381–408.
[72] HART, J. D. (1996). Some automated methods of smoothing time-dependent data. *Nonparametr. Statist.* **6** 115–142.

[73] HART, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer, New York.
[74] HASTIE, T. J. and TIBSHIRANI, R. J. (1993). Varying-coefficient models (with discussion). *J. Roy. Statist. Soc. Ser. B.* **55** 757–796.
[75] HO, T. S. Y. and LEE, S.-B. (1986). Term structure movements and pricing interest rate contingent claims. *J. Finance* **41** 1011–1029.
[76] HONG, Y. and LEE, T.-H. (2003). Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models. *Review of Economics and Statistics* **85** 1048–1062.
[77] HONG, Y. and LI, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies* **18** 37–84.
[78] HULL, J. (2003). *Options, Futures, and Other Derivatives*, 5th ed. Prentice Hall, Upper Saddle River, NJ.
[79] HULL, J. and WHITE, A. (1990). Pricing interest-rate-derivative securities. *Review of Financial Studies* **3** 573–592.
[80] ITÔ, K. (1942). Differential equations determining Markov processes. *Zenkoku Shijo Sugaku Danwakai* **244** 1352–1400. (In Japanese.)
[81] ITÔ, K. (1946). On a stochastic integral equation. *Proc. Japan Acad.* **22** 32–35.
[82] JORION, P. (2000). *Value at Risk*: *The New Benchmark for Managing Financial Risk*, 2nd ed. McGraw–Hill, New York.
[83] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York.
[84] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, New York.
[85] KESSLER, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24** 211–229.
[86] KESSLER, M. and SØRENSEN, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **5** 299–314.
[87] KLOEDEN, P. E., PLATEN, E., SCHURZ, H. and SØRENSEN, M. (1996). On effects of discretization on estimators of drift parameters for diffusion processes. *J. Appl. Probab.* **33** 1061–1076.
[88] KUTOYANTS, Y. A. (1998). Efficient density estimation for ergodic diffusion processes. *Stat. Inference Stoch. Process.* **1** 131–155.
[89] MERCURIO, D. and SPOKOINY, V. (2004). Statistical inference for time-inhomogeneous volatility models. *Ann. Statist.* **32** 577–602.
[90] MERTON, R. (1973). Theory of rational option pricing. *Bell J. Econom. Management Sci.* **4** 141–183.
[91] MÜLLER, H.-G. and STADTMÜLLER, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15** 610–625.
[92] OSBORNE, M. F. M. (1959). Brownian motion in the stock market. *Operations Res.* **7** 145–173.
[93] PHAM, D. T. (1981). Nonparametric estimation of the drift coefficient in the diffusion equation. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 61–73.
[94] PRAKASA RAO, B. L. S. (1985). Estimation of the drift for diffusion process. *Statistics* **16** 263–275.
[95] ROSENBLATT, M. (1970). Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 199–213. Cambridge Univ. Press.
[96] ROUSSAS, G. G. (1969). Nonparametric estimation in Markov processes. *Ann. Inst. Statist. Math.* **21** 73–87.
[97] ROUSSAS, G. G. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.* **40** 1386–1400.
[98] RUDEMO, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.* **9** 65–78.
[99] RUPPERT, D., WAND, M. P., HOLST, U. and HÖSSJER, O. (1997). Local polynomial variance function estimation. *Technometrics* **39** 262–273.
[100] SCHURZ, H. (2000). Numerical analysis of stochastic differential equations without tears. In *Handbook of Stochastic Analysis and Applications* (D. Kannan and V. Lakshmikantham, eds.) 237–359. Dekker, New York.
[101] SHEPHARD, N., ed. (2005). *Stochastic Volatility*: *Selected Readings*. Oxford Univ. Press.
[102] SIMONOFF, J. S. (1996). *Smoothing Methods in Statistics*. Springer, New York.
[103] SPOKOINY, V. (2000). Adaptive drift estimation for nonparametric diffusion model. *Ann. Statist.* **28** 815–836.
[104] STANTON, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *J. Finance* **52** 1973–2002.
[105] STEELE, J. M. (2001). *Stochastic Calculus and Financial Applications*. Springer, New York.
[106] VASICEK, O. A. (1977). An equilibrium characterization of the term structure. *J. Financial Economics* **5** 177–188.
[107] WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.
[108] WANG, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *Ann. Statist.* **30** 754–783.
[109] YOSHIDA, N. (1992). Estimation for diffusion processes from discrete observations. *J. Multivariate Anal.* **41** 220–242.