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

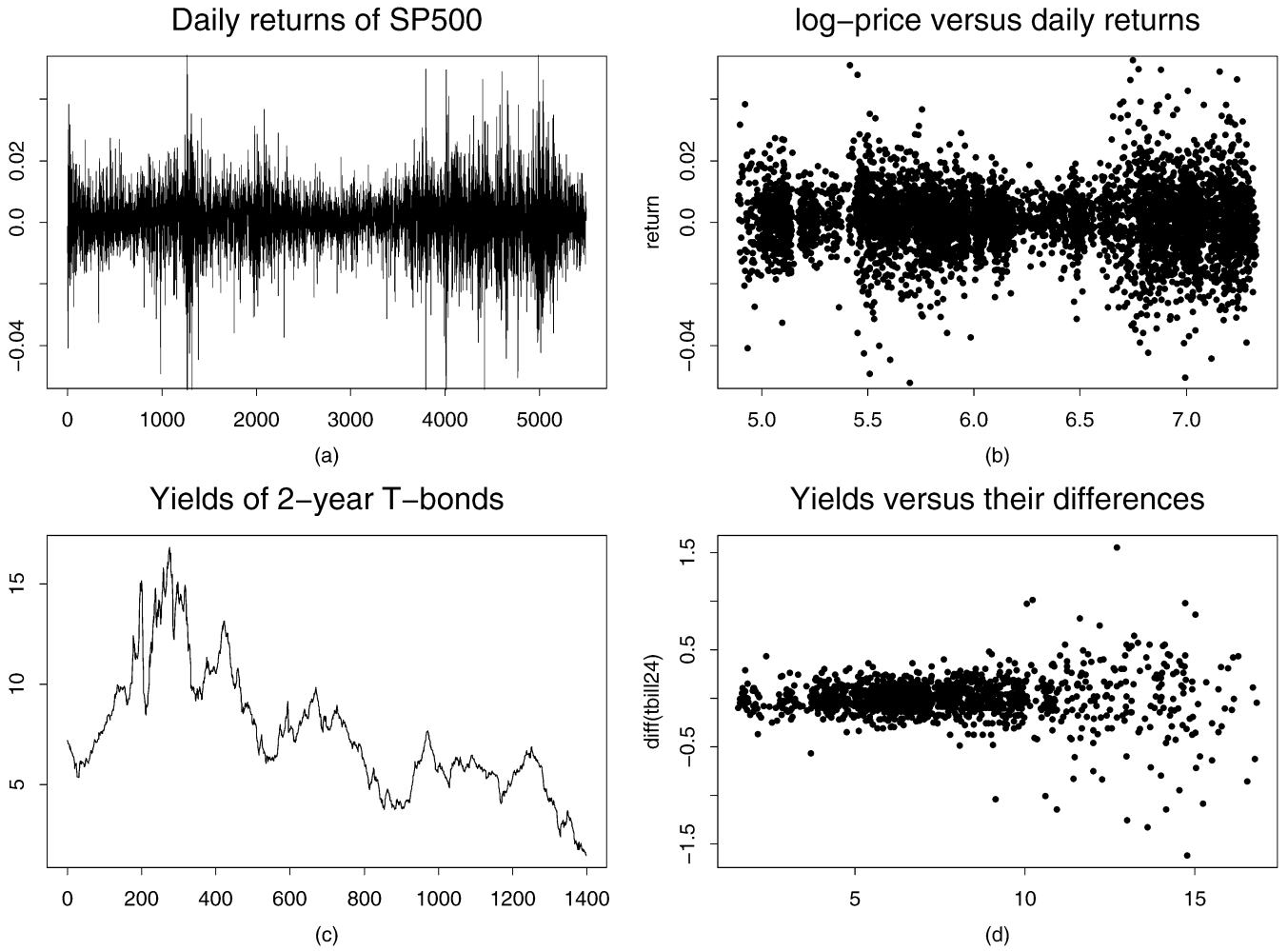


FIG. 1. (a) *Daily log-returns of the Standard and Poor's* 500 *index from October 21, 1980 to July 29, 2004.* (b) *Scatterplot of the returns against logarithm of the index (price level).* (c) *Interest rates of two-year U.S. Treasury notes from June 4, 1976 to March 7, 2003 sampled at weekly frequency.*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

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

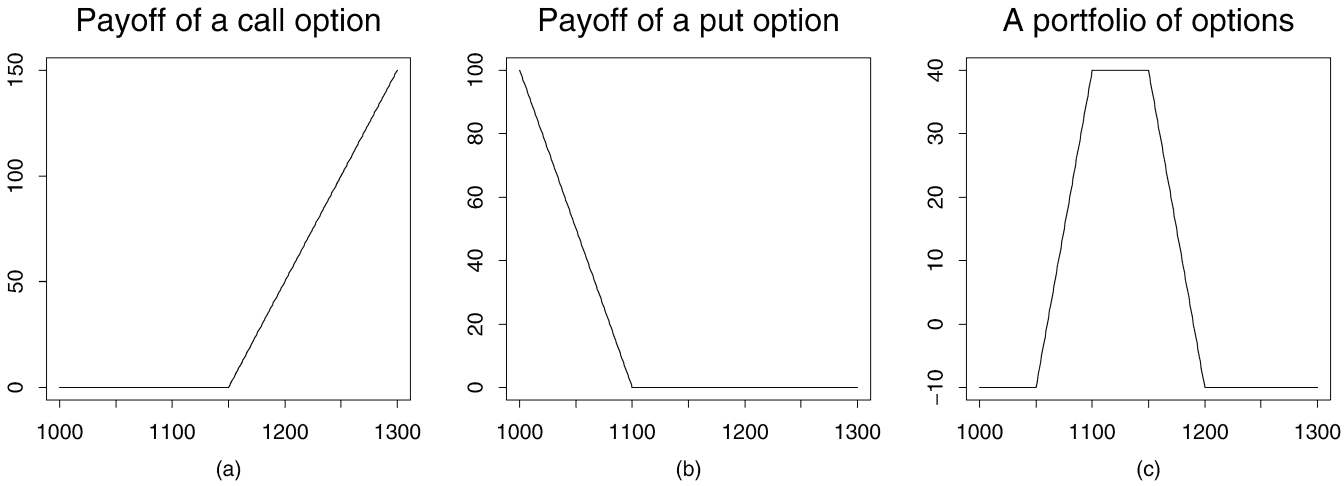


FIG. 2. (a) *Payoff of a call option*. (b) *Payoff of a put option*. (c) *Payoff of a portfolio of four options with different strike prices and different* (*long and short*) *positions*.

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.

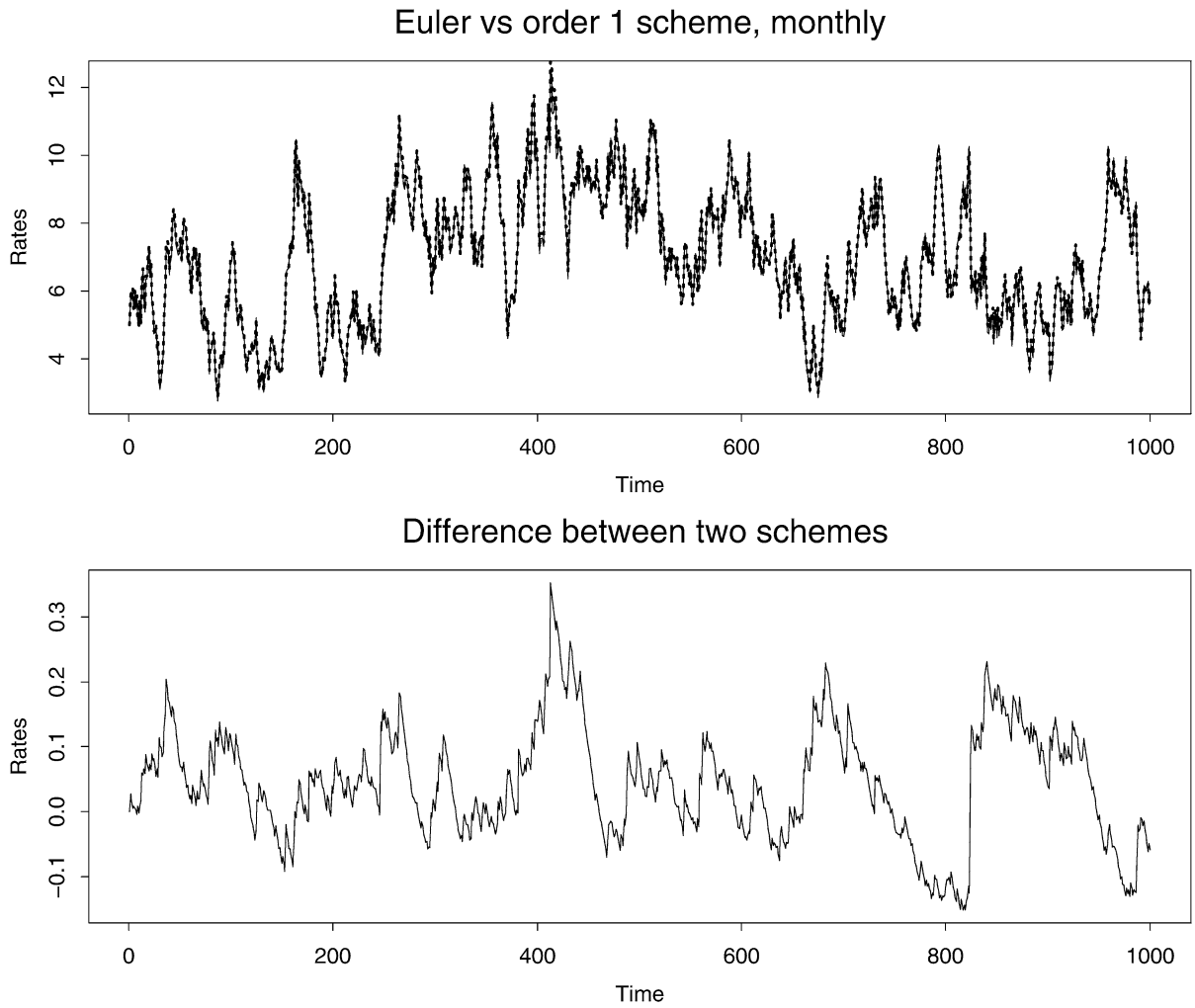


FIG. 3. *Simulated trajectories* (*multiplied by* 100) *using the Euler approximation and the strong order-one approximation for a CIR model. Top panel*: *solid curve corresponds to the Euler approximation and the dashed curve is based on the order-one approximation. Bottom panel*: *the difference between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

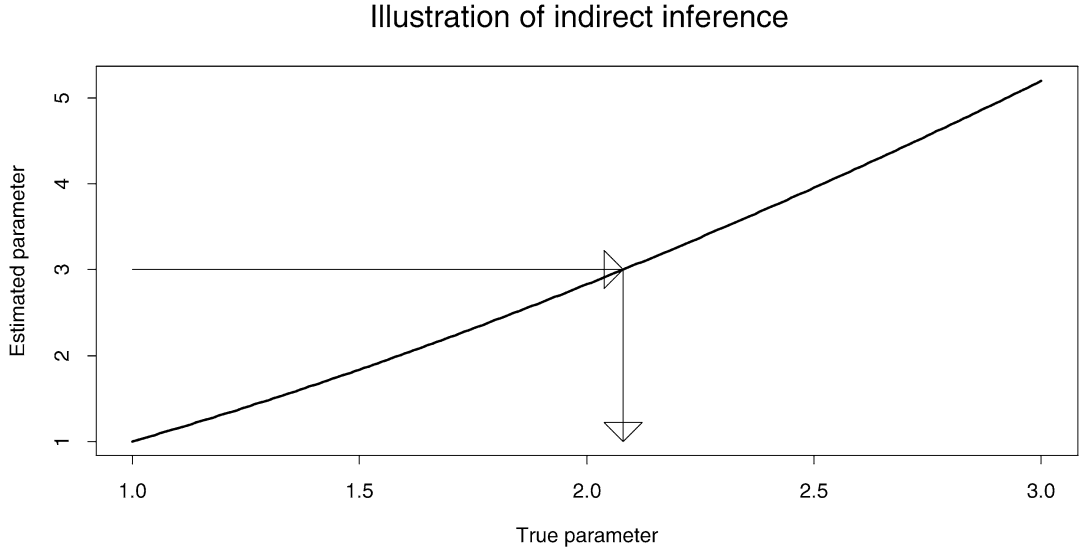


FIG. 4. *The idea of indirect inference. For each given true θ, one obtains an estimate using the Euler approximation and the simulated data. This gives a calibration curve as shown. Now for a given estimate $\hat{\theta}_0 = 3$ based on the Euler approximation and real data, one finds the calibrated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

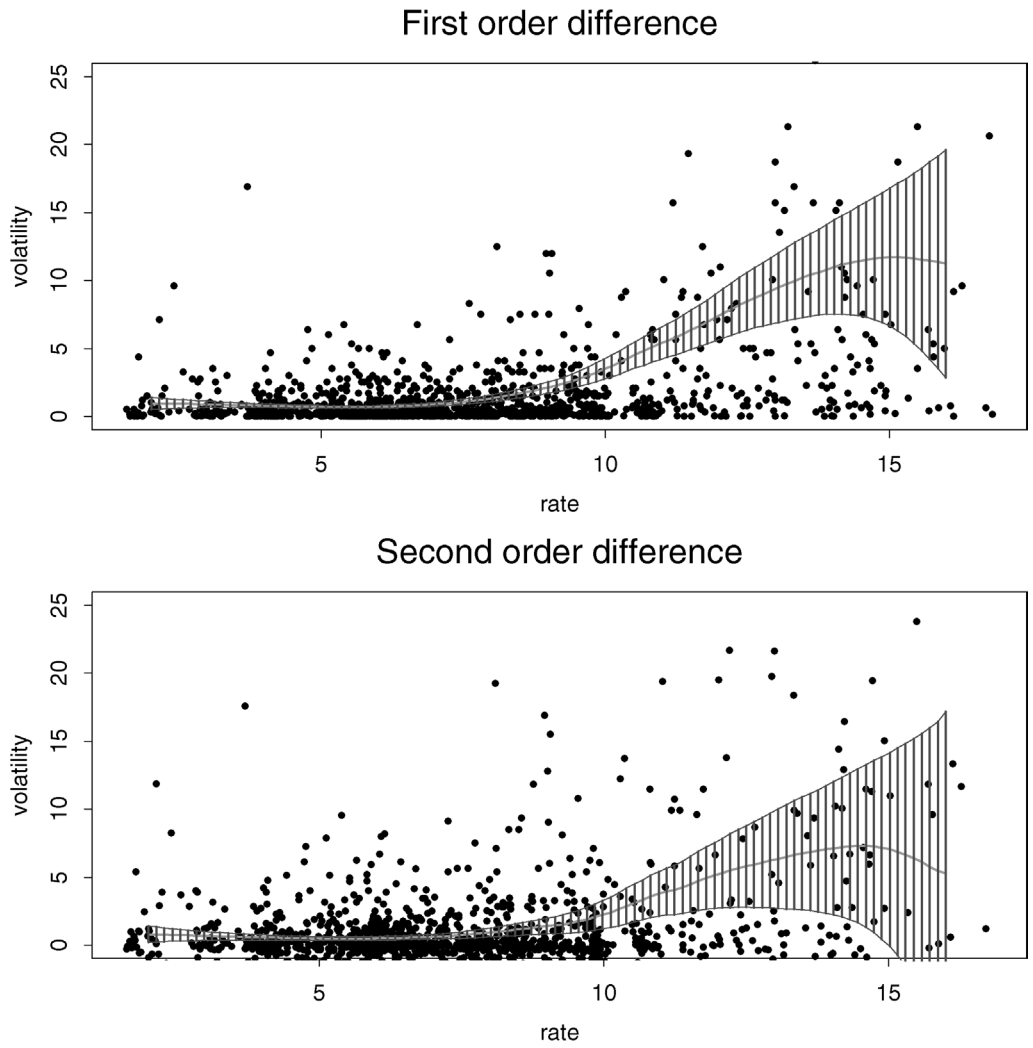


FIG. 5. *Nonparametric estimates of volatility based on order one and two differences. The bars represent two standard deviations above and below the estimated volatility. Top panel*: *order one fit. Bottom panel*: *order two fit.*

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-

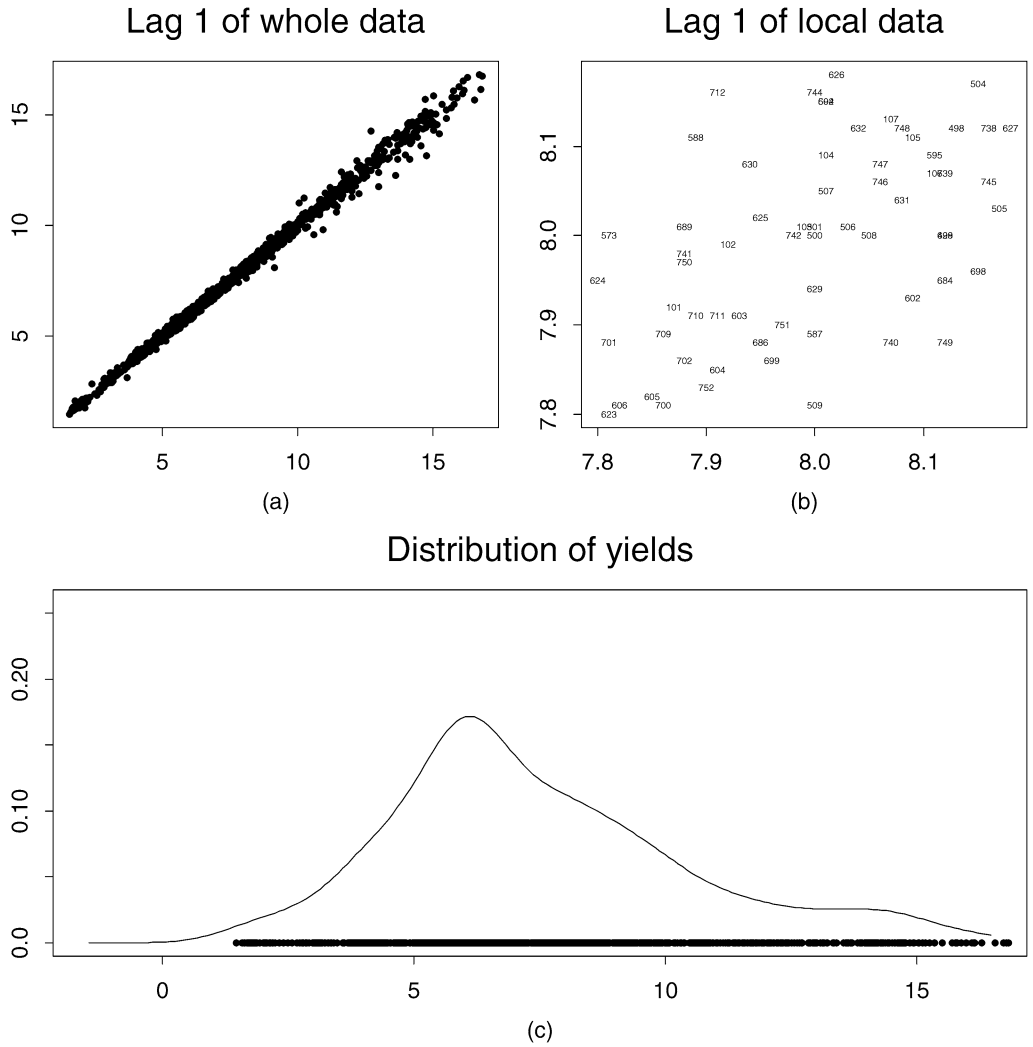


FIG. 6. (a) *Lag* 1 *scatterplot of the two-year Treasury note data.* (b) *Lag* 1 *scatterplot of those data falling in the neighborhood* $8\% \pm 0.2\%$*—the points are represented by the times of the observed data. The numbers in the scatterplot show the indices of the data falling in the neighborhood.*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

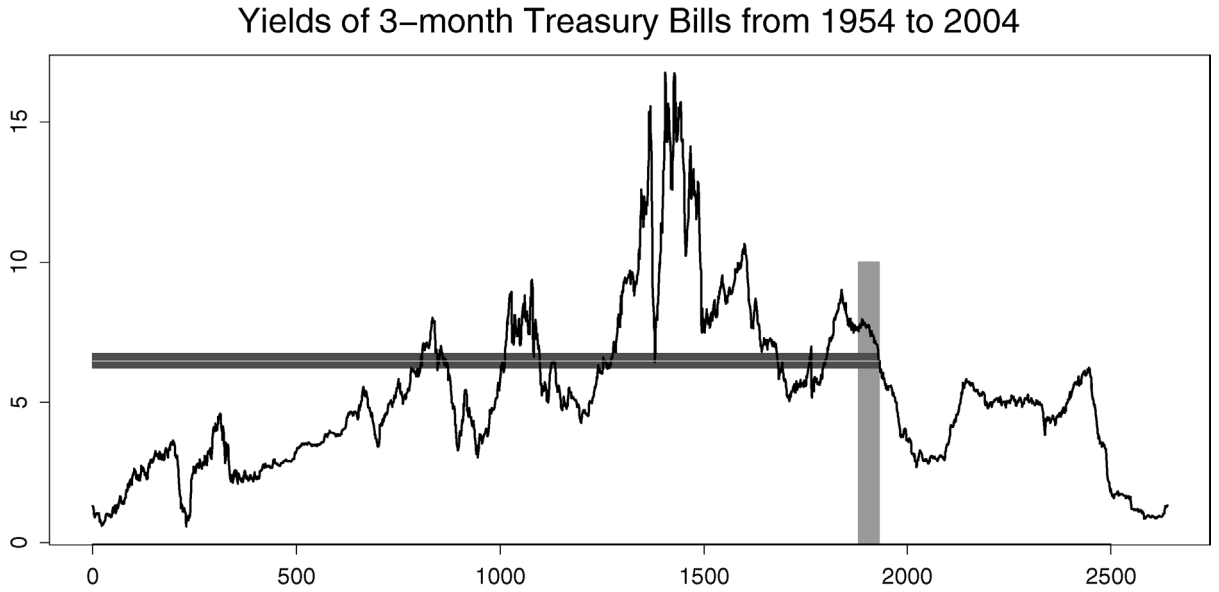


FIG. 7. *Illustration of time- and state-domain smoothing using the yields of three-month Treasury bills. The state-domain smoothing is localized in the horizontal bars, while the time-domain smoothing is concentrated in the vertical bars.*

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.

Statistical Science
2005, Vol. 20, No. 4, 338–343
DOI 10.1214/088342305000000430

Comment: A Selective Overview of Nonparametric Methods in Financial Econometrics

Peter C. B. Phillips and Jun Yu

Abstract. These comments concentrate on two issues arising from Fan's overview. The first concerns the importance of finite sample estimation bias relative to the specification and discretization biases that are emphasized in Fan's discussion. Past research and simulations given here both reveal that finite sample effects can be more important than the other two effects when judged from either statistical or economic viewpoints. Second, we draw attention to a very different nonparametric technique that is based on computing an empirical version of the quadratic variation process. This technique is not mentioned by Fan but has many advantages and has accordingly attracted much recent attention in financial econometrics and empirical applications.

Key words and phrases: Nonparametric method, continuous time models, financial time series, jackknife, realized volatility.

1. INTRODUCTION

In recent years there has been increased interest in using nonparametric methods to deal with various aspects of financial data. The paper by Fan gives an overview of some nonparametric techniques that have been used in the financial econometric literature, focusing on estimation and inference for diffusion models in continuous time and estimation of state price and transition density functions.

Continuous time specifications have been heavily used in recent work, partly because of the analytic convenience of stochastic calculus in mathematical finance and partly because of the availability of high-frequency data sets for many financial series. While the early work in continuous-time finance began in the 1970s with the work of Merton [29] and Black and Scholes [16], economists have been looking at the econometric problems of fitting continuous time systems for much longer. The idea of statistically fitting diffusion models and continuously distributed lagged dependencies with discretely observed data has a long history dating back to some original work in econometrics by Koopmans [27] and subsequent work by Phillips [31], Bergstrom [14], Sims [35], Phillips [32] and Sargan [34]. Bartlett and Rajalakshman [13] and Bartlett [12] are two references in the early statistical literature on fitting linear diffusions. Bergstrom [15] provides a short history of some of this early work. Also, the history of mathematical finance and stochastic integration prior to 1970 has recently been overviewed in an interesting historical review by Jarrow and Protter [24].

Our comments on Fan's paper will concentrate on two issues that relate in important ways to the paper's focus on misspecification and discretization bias and the role of nonparametric methods in empirical finance. The first issue deals with the finite sample effects of various estimation methods and their implications for asset pricing. A good deal of recent attention in the econometric literature has focused on the benefits of full maximum likelihood (ML) estimation of diffusions and mechanisms for avoiding discretization bias in the construction of the likelihood. However, many of the

Peter C. B. Phillips is Professor, Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, Connecticut 06520-8281, USA (e-mail: peter.phillips@yale.edu). Jun Yu is Associate Professor, School of Economics and Social Science, Singapore Management University, 90 Stamford Road, Singapore 178903 (e-mail: yujun@smu.edu.sg).

problems of estimating dynamic models that are well known in discrete time series, such as the bias in ML estimation, also manifest in the estimation of continuous time systems and affect subsequent use of these estimates, for instance in derivative pricing. In consequence, a relevant concern is the relative importance of the estimation and discretization biases. As we will show below, the former often dominates the latter even when the sample size is large (at least 500 monthly observations, say). Moreover, it turns out that correction for the finite sample estimation bias continues to be more important when the diffusion component of the model is itself misspecified. Such corrections appear to be particularly important in models that are nonstationary or nearly nonstationary.

The second issue we discuss deals with a very different nonparametric technique, which is not discussed by Fan, but which has recently attracted much attention in financial econometrics and empirical applications. This method involves the use of quadratic variation measures of realized volatility using ultra high frequency financial data. Like other nonparametric methods, empirical quadratic variation techniques also have to deal with statistical bias, which in the present case arises from the presence of microstructure noise. The field of research on this topic in econometrics is now very active.

2. FINITE SAMPLE EFFECTS

In his overview of diffusion equation estimation, Fan discusses two sources of bias, one arising from the discretization process and the second from misspecification. We review these two bias effects and then discuss the bias that comes from finite sample estimation effects.

The attractions of Itô calculus have made it particularly easy to work with stochastic differential equations driven by Brownian motion. Diffusion processes in particular have been used widely in finance to model asset prices, including stock prices, interest rates and exchange rates. Despite their mathematical attractability, diffusion processes present some formidable challenges for econometric estimation. The primary reason for the difficulty is that sample data, even very high-frequency data, are always discrete and for many popular nonlinear diffusion models the transition density of the discrete sample does not have a closed form expression, as noted by Fan. The problem is specific to nonlinear diffusions, as consistent methods for estimating exact discrete models corresponding to linear systems of diffusions have been available since Phillips [32]. A simple approach discussed in the paper is to use the Euler approximation scheme to discretize the model, a process which naturally creates some discretization bias. This discretization bias can lead to erroneous financial pricing and investment decisions. In consequence, the issue of discretization has attracted a lot of attention in the literature and many methods have been proposed to reduce the bias that it causes. Examples are Pedersen [30], Kessler [26], Durham and Gallant [18], Aït-Sahalia [2, 3] and Elerian, Chib and Shephard [19], among many others.

Next, many diffusion models in practical use are specified in a way that makes them mathematically convenient. These specifications are typically not derived from any underlying economic theory and are therefore likely to be misspecified. Potential misspecifications, like discretization, can lead to erroneous financial decisions. Accordingly, specification bias has attracted a great deal of attention in the literature and has helped to motivate the use of functional estimation techniques that treat the drift and diffusion coefficients nonparametrically. Important contributions include Aït-Sahalia [1], Stanton [36], Bandi and Phillips [5] and Hong and Li [21].

While we agree that both discretization and specification bias are important issues, finite sample estimation bias can be of equal or even greater importance for financial decision making, as noted by Phillips and Yu [33] in the context of pricing bonds and bond options. The strong effect of the finite sample estimation bias in this context can be explained as follows. In continuous time specifications, the prices of bonds and bond options depend crucially on the mean reversion parameter in the associated interest rate diffusion equation. This parameter is well known to be subject to estimation bias when standard methods like ML are used. The bias is comparable to, but generally has larger magnitude than, the usual bias that appears in time series autoregression. As the parameter is often very close to zero in empirical applications (corresponding to near martingale behavior and an autoregressive root near unity in discrete time), the estimation bias can be substantial even in very large samples.

To reduce the finite sample estimation bias in parameter estimation as well as the consequential bias that arises in asset pricing, Phillips and Yu [33] proposed the use of jackknife techniques. Suppose a sample of n observations is available and that this sample is decomposed into m consecutive sub-samples each with ℓ observations ($n = m \times \ell$). The jackknife estimator of a

parameter θ in the model is defined by

$$\hat{\theta}_{\text{jack}} = \frac{m}{m-1}\hat{\theta}_n - \frac{\sum_{i=1}^{m}\hat{\theta}_{\ell i}}{m^2-m}, \tag{2.1}$$

where $\hat{\theta}_n$ and $\hat{\theta}_{\ell i}$ are the extreme estimates of θ based on the entire sample and the i'th sub-sample, respectively. The parameter θ can be a coefficient in the diffusion process, such as the mean reversion parameter, or a much more complex function of the parameters of the diffusion process and the data, such as an asset price or derivative price. Typically, the full sample extreme estimator has bias of order $O(n^{-1})$, whereas under mild conditions the bias in the jackknife estimate is of order $O(n^{-2})$.

The following simulation illustrates these various bias effects and compares their magnitudes. In the experiment, the true generating process is assumed to be the following commonly used model (CIR hereafter) of short term interest rates due to Cox, Ingersoll and Ross [17]:

$$dr(t) = \kappa\big(\mu - r(t)\big)\,dt + \sigma r^{1/2}(t)\,dB(t). \tag{2.2}$$

The transition density of the CIR model is known to be $ce^{-u-v}(v/u)^{q/2}I_q(2(uv)^{1/2})$ and the marginal density is $w_1^{w_2}r^{w_2-1}e^{-w_1 r}/\Gamma(w_2)$, where $c = 2\kappa/(\sigma^2(1-e^{-\kappa\Delta}))$, $u = cr(t)e^{-\kappa\Delta}$, $v = cr(t+\Delta)$, $q = 2\kappa\mu/\sigma^2 - 1$, $w_1 = 2\kappa/\sigma^2$, $w_2 = 2\kappa\mu/\sigma^2$, Δ is the sampling frequency, and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q. The transition density together with the marginal density can be used for simulation purposes as well as to obtain the exact ML estimator of θ $(= (\kappa, \mu, \sigma)')$. In the simulation, we use this model to price a discount bond, which is a three-year bond with a face value of \$1 and initial interest rate of 5%, and a one-year European call option on a three-year discount bond which has a face value of \$100 and a strike price of \$87. The reader is referred to [33] for further details.

In addition to exact ML estimation, we may discretize the CIR model via the Euler method and estimate the discretized model using (quasi-) ML. The Euler scheme leads to the discretization

$$\begin{aligned} r(t+\Delta) &= \kappa\mu\Delta + (1-\kappa\Delta)r(t) \\ &\quad + \sigma N\big(0, \Delta r(t)\big). \end{aligned} \tag{2.3}$$

One thousand samples, each with 600 monthly observations (i.e., $\Delta = 1/12$), are simulated from the true model (2.2) with $(\kappa, \mu, \sigma)'$ being set at $(0.1, 0.08, 0.02)'$, which are settings that are realistic in many financial applications. To investigate the effects of discretization bias, we estimate model (2.3) by the (quasi-) ML approach. To investigate the finite sample estimation bias effects, we estimate model (2.2) based on the true transition density. To examine the effects of bias reduction in estimation, we apply the jackknife method (with $m = 3$) to the mean reversion parameter κ, the bond price and the bond option price.

To examine the effects of specification bias, we fit each simulated sequence from the true model to the misspecified Vasicek model [37] to obtain the exact ML estimates of κ, the bond price and the option price from this misspecified model. The Vasicek model is given by the simple linear diffusion

$$dr(t) = \kappa\big(\mu - r(t)\big)\,dt + \sigma\,dB(t). \tag{2.4}$$

We use this model to price the same bond and bond option. Vasicek [37] derived the expression for bond prices and Jamshidian [23] gave the corresponding formula for bond option prices. The transition density for the Vasicek model is

$$\begin{aligned} &r(t+\Delta)|r(t) \\ &\sim N\big(\mu(1-e^{-\kappa\Delta}) \\ &\qquad + e^{-\kappa\Delta}r_t, \sigma^2(1-e^{-2\kappa\Delta})/(2\kappa)\big). \end{aligned} \tag{2.5}$$

This transition density is utilized to obtain the exact ML estimates of κ, the bond price and the bond option price, all under the mistaken presumption that the misspecified model (2.4) is correctly specified.

Table 1 reports the means and root mean square errors (RMSEs) for all these cases. It is clear that the finite sample estimation bias is more substantial than

TABLE 1

Finite sample properties of ML and jackknife estimates of κ, bond price and option price for the (true) CIR model using a (correctly specified) fitted CIR model and a (misspecified) fitted Vasicek model (sample size $n = 600$)

Parameter		κ	**Bond price**	**Option price**
True value		**0.1**	**0.8503**	**2.3920**
Exact ML	Mean	0.1845	0.8438	1.8085
of CIR	RMSE	0.1319	0.0103	0.9052
Euler ML	Mean	0.1905	0.8433	1.7693
of CIR	RMSE	0.1397	0.0111	0.9668
Jackknife ($m = 3$)	Mean	0.0911	0.8488	2.1473
of CIR	RMSE	0.1205	0.0094	0.8704
ML of Vasicek	Mean	0.1746	0.8444	1.8837
(misspecified)	RMSE	0.1175	0.0088	0.7637
Jackknife ($m = 2$) of	Mean	0.0977	0.8488	2.2483
Vasicek (misspecified)	RMSE	0.1628	0.0120	1.0289

the discretization bias and the specification bias for all three quantities, at least in this experiment. In particular, κ is estimated by the exact ML method with 84.5% upward bias, which contributes toward the -0.76% bias in the bond price and the -24.39% bias in the option price. Relative to the finite sample bias, the bias in κ due to the discretization is almost negligible since the total bias in κ changes from 84.5% to 90.5%. (The increase in the total bias indicates that the discretization bias effect is in the same direction as that of the estimation bias.) The total bias changes from -0.76% to -0.82% in the bond price and from -24.39% to -26.03% in the option price. These changes are marginal. Similarly, relative to the finite sample bias, the bias in κ due to misspecification of the drift function is almost negligible since the total bias changes from 84.5% to 74.6%. (The decrease in the total bias indicates that the misspecification bias effect is in the opposite direction to that of the estimation bias.) The total bias changes from -0.76% to -0.69% in the bond price and from -24.39% to -21.25% in the option price. Once again, these changes are marginal. When the jackknife method is applied to the correctly specified model, the estimation bias is greatly reduced in all cases (from 84.5% to -8.9% for κ; from -0.76% to -0.18% for the bond price; and from -24.39% to -10.23% for the option price).

Even more remarkably, when the jackknife method is applied to the incorrectly specified model (see the final row of Table 1), the estimation bias is also greatly reduced in all cases (from 84.5% to -2.3% for κ; from -0.76% to -0.18% for the bond price; and from -24.39% to -6.01% for the option price). These figures reveal that dealing with estimation bias can be much more important than ensuring correct specification in diffusion equation estimation, suggesting that general econometric treatment of the diffusion through nonparametric methods may not address the major source of bias effects on financial decision making.

Although the estimation bias is not completely removed by the jackknife method, the bias reduction is clearly substantial and the RMSE of the jackknife estimate is smaller in all cases than that of exact ML. In sum, it is apparent from Table 1 that the finite sample estimation bias is larger in magnitude than either of the biases due to discretization and misspecification and correcting this bias is therefore a matter of importance in empirical work on which financial decisions depend.

Although this demonstration of the relative importance of finite sample estimation bias in relation to discretization bias and specification bias is conducted in a parametric context, similar results can be expected for some nonparametric models. For example, in the semiparametric model examined in [1], the diffusion function is nonparametrically specified and the drift function is linear, so that the mean reversion parameter is estimated parametrically as in the above example. In such cases, we can expect substantial finite sample estimation bias to persist and to have important practical implications in financial pricing applications.

3. REALIZED VOLATILITY

As noted in Fan's overview, many models used in financial econometrics for modeling asset prices and interest rates have the fully functional scalar differential form

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dB_t, \tag{3.1}$$

where both drift and diffusion functions are nonparametric and where the equation is driven by Brownian motion increments dB_t. For models such as (3.1), we have $(dX_t)^2 = \sigma^2(X_t)\,dt$ a.s. and hence the quadratic variation of X_t is

$$[X]_T = \int_0^T (dX_t)^2\,dt = \int_0^T \sigma^2(X_t)\,dt, \tag{3.2}$$

where $\int_0^T \sigma^2(X_t)\,dt$ is the accumulated or integrated volatility of X. Were X_t observed continuously, $[X]_T$ and, hence, integrated volatility, would also be observed. For discretely recorded data, estimation of (3.2) is an important practical problem. This can be accomplished by direct nonparametric methods using an empirical estimate of the quadratic variation that is called realized volatility. The idea has been discussed for some time, an early reference being Maheswaran and Sims [28], and it has recently attracted a good deal of attention in the econometric literature now that very high frequency data has become available for empirical use. Recent contributions to the subject are reviewed in [4] and [8].

Suppose X_t is recorded discretely at equispaced points $(\Delta, 2\Delta, \ldots, n_\Delta\Delta(\equiv T))$ over the time interval $[0, T]$. Then, $[X]_T$ can be consistently estimated by the realized volatility of X_t defined by

$$[X_\Delta]_T = \sum_{i=2}^{n_\Delta} \big(X_{i\Delta} - X_{(i-1)\Delta}\big)^2, \tag{3.3}$$

as $\Delta \to 0$, as is well known. In fact, any construction of realized volatility based on an empirical grid of observations where the maximum grid size tends to zero will produce a consistent estimate. It follows that the

integrated volatility can be consistently estimated by this nonparametric approach, regardless of the form of $\mu(X_t)$ and $\sigma(X_t)$. The approach has received a great deal of attention in the recent volatility literature and serves as a powerful alternative to the methods discussed by Fan, especially when ultra-high frequency data are available.

While this approach is seemingly straightforward, it is not without difficulties. First, in order for the approach to be useful in empirical research, it is necessary to estimate the precision of the realized volatility estimates. Important contributions on the central limit theory of these empirical quadratic variation estimates by Jacod [22] and Barndorff-Nielson and Shephard [10, 11] has facilitated the construction of suitable methods of inference. Second, in practical applications, realized volatility measures such as (3.3) are usually contaminated by microstructure noise bias, especially at ultra high frequencies and tick-by-tick data. Noise sources arise from various market frictions and discontinuities in trading behavior that prevent the full operation of efficient financial markets. Recent work on this subject (e.g., [8, 9, 21, 38]) has developed various methods, including nonparametric kernel techniques, for reducing the effects of microstructure noise bias.

4. ADDITIONAL ISSUES

Given efficient market theory, there is good reason to expect that diffusion models like (3.1) may have nonstationary characteristics. Similar comments apply to term structure models and yield curves. In such cases, nonparametric estimation methods lead to the estimation of the local time (or sojourn time) of the corresponding stochastic process and functionals of this quantity, rather than a stationary probability density. Moreover, rates of convergence in such cases become path dependent and the limit theory for nonparametric estimates of the drift and diffusion functions in (3.1) is mixed normal. Asymptotics of this type require an enlarging time span of data as well as increasing in-fill within each discrete interval as $n \to \infty$. An overview of this literature and its implications for financial data applications is given in [6]. Nonparametric estimates of yield curves in multifactor term structure models are studied in [25].

Not all models in finance are driven by Brownian motion. In some cases, one can expect noise to have to have some memory and, accordingly, models such as (3.1) have now been extended to accommodate fractional Brownian motion increments. The stochastic calculus of fractional Brownian motion, which is not a semi-martingale, is not as friendly as that of Brownian motion and requires new constructs, involving Wick products and versions of the Stratonovich integral. Moreover, certain quantities, such as quadratic variation, that have proved useful in the recent empirical literature may no longer exist and must be replaced by different forms of variation, although the idea of volatility is still present. Developing a statistical theory of inference to address these issues in financial econometric models is presenting new challenges.

ACKNOWLEDGMENTS

Peter C. B. Phillips gratefully acknowledges visiting support from the School of Economics and Social Science at Singapore Management University. Support was provided by NSF Grant SES-04-142254. Jun Yu gratefully acknowledges financial support from the Wharton-SMU Research Center at Singapore Management University.

REFERENCES

[1] AÏT-SAHALIA, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica* **64** 527–560.

[2] AÏT-SAHALIA, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *J. Finance* **54** 1361–1395.

[3] AÏT-SAHALIA, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **70** 223–262.

[4] ANDERSEN, T. G., BOLLERSLEV, T. and DIEBOLD, F. X. (2005). Parametric and nonparametric volatility measurement. In *Handbook of Financial Econometrics* (Y. Aït-Sahalia and L. P. Hansen, eds.). North-Holland, Amsterdam. To appear.

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

[6] BANDI, F. M. and PHILLIPS, P. C. B. (2005). Nonstationary continuous-time processes. In *Handbook of Financial Econometrics* (Y. Aït-Sahalia and L. P. Hansen, eds.). North-Holland, Amsterdam. To appear.

[7] BANDI, F. M. and RUSSELL, J. (2005). Volatility. In *Handbook of Financial Engineering* (J. R. Birge and V. Linetsky, eds.). To appear.

[8] BANDI, F. M. and RUSSELL, J. (2005). Microstructure noise, realized volatility and optimal sampling. Working paper, Graduate School of Business, Univ. Chicago.

[9] BARNDORFF-NIELSEN, O., HANSEN, P., LUNDE, A. and SHEPHARD, N. (2005). Regular and modified kernel-based estimators of integrated variance: The case with independent noise. Working paper, Nuffield College.

[10] BARNDORFF-NIELSEN, O. and SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 253–280.

[11] BARNDORFF-NIELSEN, O. and SHEPHARD, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72** 885–925.

[12] BARTLETT, M. S. (1955). *An Introduction to Stochastic Processes.* Cambridge Univ. Press.

[13] BARTLETT, M. S. and RAJALAKSHMAN, D. V. (1953). Goodness of fit tests for simultaneous autoregressive series. *J. Roy. Statist. Soc. Ser. B* **15** 107–124.

[14] BERGSTROM, A. (1966). Nonrecursive models as discrete approximations to systems of stochastic differential equations. *Econometrica* **34** 173–182.

[15] BERGSTROM, A. (1988). The history of continuous-time econometric models. *Econometric Theory* **4** 365–383.

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

[17] COX, J., INGERSOLL, J. and ROSS, S. (1985). A theory of the term structure of interest rates. *Econometrica* **53** 385–407.

[18] DURHAM, G. and GALLANT, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with discussion). *J. Bus. Econom. Statist.* **20** 297–338.

[19] ELERIAN, O., CHIB, S. and SHEPHARD, N. (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica* **69** 959–993.

[20] HANSEN, P. and LUNDE, A. (2006). An unbiased measure of realized volatility. *J. Bus. Econom. Statist.* To appear.

[21] 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.

[22] JACOD, J. (1994). Limit of random measures associated with the increments of a Brownian semimartingale. Working paper, Univ. P. and M. Curie, Paris.

[23] JAMSHIDIAN, F. (1989). An exact bond option formula. *J. Finance* **44** 205–209.

[24] JARROW, R. and PROTTER, P. (2004). A short history of stochastic integration and mathematical finance: The early years, 1880–1970. In *A Festschrift for Herman Rubin* (A. DasGupta, ed.) 75–91. IMS, Beachwood, OH.

[25] JEFFREY, A., KRISTENSEN, D., LINTON, O., NGUYEN, T. and PHILLIPS, P. C. B. (2004). Nonparametric estimation of a multifactor Heath–Jarrow–Morton model: An integrated approach. *J. Financial Econometrics* **2** 251–289.

[26] KESSLER, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24** 211–229.

[27] KOOPMANS, T., ed. (1950). *Statistical Inference in Dynamic Economic Models*. Wiley, New York.

[28] MAHESWARAN, S. and SIMS, C. A. (1993). Empirical implications of arbitrage-free asset markets. In *Models, Methods and Applications of Econometrics* (P. C. B. Phillips, ed.) 301–316. Blackwell, Cambridge, MA.

[29] MERTON, R. (1973). Theory of rational option pricing. *Bell J. Econom. and Management Sci.* **4** 141–183.

[30] PEDERSEN, A. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.* **22** 55–71.

[31] PHILLIPS, A. W. (1959). The estimation of parameters in systems of stochastic differential equations. *Biometrika* **46** 67–76.

[32] PHILLIPS, P. C. B. (1972). The structural estimation of a stochastic differential equation system. *Econometrica* **40** 1021–1041.

[33] PHILLIPS, P. C. B. and YU, J. (2005). Jackknifing bond option prices. *Review of Financial Studies* **18** 707–742.

[34] SARGAN, J. D. (1974). Some discrete approximations to continuous time stochastic models. *J. Roy. Statist. Soc. Ser. B* **36** 74–90.

[35] SIMS, C. (1971). Discrete approximations to continuous time distributed lags in econometrics. *Econometrica* **39** 545–563.

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

[37] VASICEK, O. (1977). An equilibrium characterization of the term structure. *J. Financial Economics* **5** 177–188.

[38] ZHANG, L., MYKLAND, P. and AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411.

Statistical Science
2005, Vol. 20, No. 4, 344–346
DOI 10.1214/088342305000000449

Comment: A Selective Overview of Nonparametric Methods in Financial Econometrics

Michael Sørensen

1. INTRODUCTION

Professor Fan should be congratulated for his review that convincingly demonstrates the usefulness of nonparametric techniques to financial econometric problems. He is mainly concerned with financial models given by stochastic differential equations, that is, diffusion processes. I will therefore complement his selective review by discussing some important problems and useful methods for diffusion models that he has not covered. My concern will mainly, but not solely, be with parametric techniques. A recent comprehensive survey of parametric inference for discretely sampled diffusion models can be found in [19].

2. GAUSSIAN LIKELIHOOD FUNCTIONS

In his brief review of parametric methods, Professor Fan mentions the Gaussian approximate likelihood function based on the Euler scheme and states that this method has some bias when the time between observations Δ is large. This is actually a very serious problem. As an example, consider a model with a linear drift of the form $\mu(x) = -\beta(x-\alpha)$ ($\beta > 0$). The estimator $\hat{\beta}_n$ of β obtained from the Gaussian approximate likelihood based on the Euler scheme converges to

$$(1 - e^{-\beta_0\Delta})\Delta^{-1}$$

as the number of observations n tends to infinity. Here β_0 denotes the true parameter value. The limiting value of $\Delta\hat{\beta}_n$ is always smaller than one, and the limit of $\hat{\beta}_n$ is always smaller than Δ^{-1}. Thus the asymptotic bias can be huge if Δ is large. A simulation study in [3] demonstrates that also for finite sample sizes an enormous bias can occur. When $\Delta\beta_0$ is small so that $(1 - e^{-\beta_0\Delta})\Delta^{-1} \approx \beta_0$, the asymptotic bias is negligible. The problem is, however, that if we use the approximate likelihood function based on the Euler scheme, there is no way we can know whether $\Delta\beta_0$ is small or large because $\Delta\hat{\beta}_n$ will always tend to be small. I suspect that the nonparametric methods outlined in Sections 3.2 and 3.5 might suffer from a similar shortcoming as they are based on the same type of approximation as the Euler scheme.

A simple solution to this problem is to use an approximate likelihood function where the transition density is replaced by a normal distribution with mean equal to the exact conditional expectation $F(x,\theta) = E_\theta(X_\Delta | X_0 = x)$ and with the variance equal to the exact conditional variance $\Phi(x;\theta) = \mathrm{Var}_\theta(X_\Delta | X_0 = x)$. Here θ is the (typically multivariate) parameter to be estimated. This approach is exactly the same as using quadratic martingale estimating functions; see [3] and [20]. The estimators obtained from quadratic martingale estimating functions have the same nice properties for high frequency observations (small Δ) as the estimators based on the Euler likelihood, but they are consistent for any value of Δ and can thus be used whether or not Δ is small. In most cases there is no explicit expression for the functions $F(x,\theta)$ and $\Phi(x;\theta)$, so often they must be determined by simulation. This requires, however, only a modest amount of computation and is not a problem in practice. If a completely explicit likelihood is preferred, one can approximate $F(x,\theta)$ and $\Phi(x;\theta)$ by expansions of a higher order than those used in the Euler scheme; see [16].

The nonparametric method in Section 3.5 could probably be improved in a similar way by using in (27) and (28) the functions $F(x,\theta)$ and $\Phi(x;\theta)$ (or the higher-order expansions in [16]) instead of the first-order approximation used in the Euler scheme.

3. MARTINGALE ESTIMATING FUNCTIONS

More generally, martingale estimating functions provide a simple and versatile technique for estimation in discretely sampled parametric stochastic differential equation models that works whether or not Δ is small.

Michael Sørensen is Professor and Head, Department of Applied Mathematics and Statistics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark (e-mail: michael@math.ku.dk).

An estimator is obtained by solving the equation

$$\sum_{i=1}^{n}\sum_{j=1}^{N} a_j(X_{(i-1)\Delta};\theta) \cdot \left[f_j(X_{i\Delta}) - H_\Delta^\theta f_j(X_{(i-1)\Delta})\right] = 0,$$

where H_Δ is the transition operator, and where the function a has the same dimension as the parameter θ. First suitable functions f_j are chosen, and then the weight functions a_j are determined so that an optimal estimating function in the sense of Godambe and Heyde [9] is obtained; see also [10]. Optimal estimating functions are approximations to the non-explicit score function. Usually $H_\Delta^\theta f_j$ must be determined by a modest amount of simulation, but Kessler and Sørensen [17] demonstrated how completely explicit estimating functions can be obtained if the functions f_j are eigenfunctions of the operator L_θ (called the generator); see also [18] for details on how to explicitly find the optimal weight functions. A review of the theory of estimating functions for discretely sampled diffusion-type models can be found in [1].

For martingale estimating functions large sample results concerning estimators can be obtained via martingale limit theory. Under weak conditions, estimators are consistent, and optimal estimating functions tend to work well when the functions f_j are chosen reasonably, that is, such that a good approximation to the score function can be obtained. At low sampling frequencies the estimators are, however, usually not efficient. The behavior of the estimators at high sampling frequencies can be investigated by considering an asymptotic scenario where the time between observations Δ_n is assumed to go to zero, as the sample size n tends to infinity, sufficiently slowly that the time horizon over which observations are made, $n\Delta_n$, tends to infinity. It is well known that in this situation estimators of parameters appearing in the diffusion coefficient may converge at a suboptimal rate, $1/\sqrt{n\Delta_n}$. The reason is that there is a lot of information about the diffusion coefficient in the fine structure of diffusion trajectories, which some estimators do not capture. Recently Sørensen [22] has given conditions ensuring that a martingale estimating function provides estimators that are rate-optimal (rate $1/\sqrt{n}$) and efficient in the high-frequency asymptotic scenario. Optimal martingale estimating functions satisfy these conditions. Quadratic martingale estimating functions are always rate-optimal, and if they are obtained from Gaussian approximate likelihood functions they are efficient too. These results are closely related to the theory of small Δ-optimality developed in [13] and [14].

4. NON-MARKOVIAN OBSERVATIONS

There are several situations in which observations from a diffusion process are non-Markovian. Most prominently this happens if a function of lower dimension of a multivariate diffusion is observed. An example is the stochastic volatility model that plays an important role as a model of financial time series since it is well known that a simple one-dimensional diffusion often cannot capture all the salient features of such data. Another example is given by the sums of diffusions proposed by Bibby, Skovgaard and Sørensen [2] as models of phenomena with more than one time scale. Other situations where diffusion data are non-Markovian are in the presence of measurement error, or when only integrals of the diffusion over time-intervals are observed; see [4]. The latter is, for instance, the case when climate data from ice cores are analyzed by means of a diffusion model. When the data are non-Markovian, it is usually not possible to find a tractable martingale estimating function, but an alternative is provided by the prediction-based estimating functions proposed in [21], which can be interpreted as approximations to martingale estimating functions.

Asymptotic results for estimators based on non-Markovian data are usually based on the assumption that the underlying diffusion process is strongly mixing. The condition ensuring exponential ρ-mixing cited in Section 2.2 is not easy to check for concrete diffusion models. A condition on the drift and diffusion coefficient that is easy to verify and that implies exponential ρ-mixing and α-mixing was given by Genon-Catalot, Jeantheau and Larédo [6].

5. NONPARAMETRIC METHODS

Let me conclude by drawing attention to some relatively early work on nonparametric methods for discretely sampled diffusion models. Wavelet methods for estimating the diffusion coefficient of a time-dependent model were proposed by Genon-Catalot, Larédo and Picard [7]. The first estimator of the diffusion coefficient mentioned in Section 3.2 was first proposed by Florens-Zmirou [5]. She considered a high frequency asymptotic scenario with fixed time span, that is, with $n\Delta_n$ constant, and proved that the asymptotic distribution of her estimator is a mixture of normal distributions where the mixing distribution is the distribution of the local time of the diffusion. If a data-dependent normalization of the estimator is used, an asymptotic normal distribution is obtained. In a series

of important papers, Marc Hoffmann has studied optimal rates of convergence of nonparametric estimators of the drift and diffusion coefficient under the three asymptotic scenarios usually considered for diffusion models including optimal estimators; see [8, 11, 12]. Other estimators of the diffusion coefficient were proposed by Soulier [23] and Jacod [15].

REFERENCES

[1] BIBBY, B. M., JACOBSEN, M. and SØRENSEN, M. (2005). Estimating functions for discretely sampled diffusion-type models. In *Handbook of Financial Econometrics* (Y. Aït-Sahalia and L. P. Hansen, eds.). North-Holland, Amsterdam. To appear.

[2] BIBBY, B. M., SKOVGAARD, I. M. and SØRENSEN, M. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli* **11** 191–220.

[3] BIBBY, B. M. and SØRENSEN, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1** 17–39.

[4] DITLEVSEN, S. and SØRENSEN, M. (2004). Inference for observations of integrated diffusion processes. *Scand. J. Statist.* **31** 417–429.

[5] FLORENS-ZMIROU, D. (1993). On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.* **30** 790–804.

[6] GENON-CATALOT, V., JEANTHEAU, T. and LARÉDO, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6** 1051–1079.

[7] GENON-CATALOT, V., LARÉDO, C. and PICARD, D. (1992). Nonparametric estimation of the diffusion coefficient by wavelet methods. *Scand. J. Statist.* **19** 317–335.

[8] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253.

[9] GODAMBE, V. P. and HEYDE, C. C. (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **55** 231–244.

[10] HEYDE, C. C. (1997). *Quasi-Likelihood and Its Application*. Springer, New York.

[11] HOFFMANN, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* **79** 135–163.

[12] HOFFMANN, M. (1999). L_p estimation of the diffusion coefficient. *Bernoulli* **5** 447–481.

[13] JACOBSEN, M. (2001). Discretely observed diffusions: Classes of estimating functions and small Δ-optimality. *Scand. J. Statist.* **28** 123–149.

[14] JACOBSEN, M. (2002). Optimality and small Δ-optimality of martingale estimating functions. *Bernoulli* **8** 643–668.

[15] JACOD, J. (2000). Nonparametric kernel estimation of the coefficient of a diffusion. *Scand. J. Statist.* **27** 83–96.

[16] KESSLER, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24** 211–229.

[17] KESSLER, M. and SØRENSEN, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **5** 299–314.

[18] LARSEN, K. S. and SØRENSEN, M. (2005). A diffusion model for exchange rates in a target zone. *Math. Finance*. To appear.

[19] SØRENSEN, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: A survey. *Internat. Statist. Rev.* **72** 337–354.

[20] SØRENSEN, M. (1997). Estimating functions for discretely observed diffusions: A review. In *Selected Proceedings of the Symposium on Estimating Functions* (I. V. Basawa, V. P. Godambe and R. L. Taylor, eds.) 305–325. IMS, Hayward, CA.

[21] SØRENSEN, M. (2000). Prediction-based estimating functions. *Econom. J.* **3** 123–147.

[22] SØRENSEN, M. (2005). Efficient martingale estimating functions for discretely sampled ergodic diffusions. Preprint, Dept. Appl. Math. and Statistics, Univ. Copenhagen.

[23] SOULIER, P. (1998). Nonparametric estimation of the diffusion coefficient of a diffusion process. *Stochastic Anal. Appl.* **16** 185–200.

Statistical Science
2005, Vol. 20, No. 4, 347–350
DOI 10.1214/088342305000000458

Comment: A Selective Overview of Nonparametric Methods in Financial Econometrics

Per A. Mykland and Lan Zhang

We would like to congratulate Jianqing Fan for an excellent and well-written survey of some of the literature in this area. We will here focus on some of the issues which are at the research frontiers in financial econometrics but are not covered in the survey. Most importantly, we consider the estimation of actual volatility. Related to this is the realization that financial data is actually observed with error (typically called *market microstructure*), and that one needs to consider a *hidden semimartingale model*. This has implications for the Markov models discussed above.

For reasons of space, we have not included references to all the relevant work by the authors that are cited, but we have tried to include at least one reference to each of the main contributors to the realized volatility area.

1. THE ESTIMATION OF ACTUAL VOLATILITY: THE IDEAL CASE

The paper discusses the estimation of Markovian systems, models where the drift and volatility coefficients are functions of time t or state x. There is, however, scope for considering more complicated systems. An important tool in this respect is the direct estimation of volatility based on high-frequency data. One considers a system of, say, log securities prices, which follows:

$$dX_t = \mu_t\,dt + \sigma_t\,dB_t, \tag{1}$$

where B_t is a standard Brownian motion. Typically, μ_t, the drift coefficient, and σ_t^2, the instantaneous variance (or volatility) of the returns process X_t, will be stochastic processes, but these processes can depend on the past in ways that need not be specified, and can be substantially more complex than a Markov model. This is known as an *Itô process*.

A main quantity of econometric interest is to obtain time series of the form $\Xi_i = \int_{T_i^-}^{T_i^+} \sigma_t^2\,dt$, $i = 1, 2, \ldots$. Here T_i^- and T_i^+ can, for example, be the beginning and the end of day number i. Ξ_i is variously known as the *integrated variance* (*or volatility*) or *quadratic variation* of the process X. The reason why one can hope to obtain this series is as follows. If $T_i^- = t_0 < t_1 < \cdots < t_n = T_i^+$ spans day number i, define the *realized volatility* by

$$\hat{\Xi}_i = \sum_{j=0}^{n-1} (X_{t_{j+1}} - X_{t_j})^2. \tag{2}$$

Then stochastic calculus tells us that

$$\Xi_i = \lim_{\max |t_{j+1} - t_j| \to 0} \hat{\Xi}_i. \tag{3}$$

In the presence of high frequency financial data, in many cases with transactions as often as every few seconds, one can, therefore, hope to almost *observe* Ξ_i. One can then either fit a model to the series of $\hat{\Xi}_i$, or one can use it directly for portfolio management (as in [12]), options hedging (as in [29]), or to test goodness of fit [31].

There are too many references to the relationship (3) to name them all, but some excellent treatments can be found in [27], Section 1.5; [26], Theorem I.4.47 on page 52; and [33], Theorem II-22 on page 66. An early econometric discussion of this relationship can be found in [2].

To make it even more intriguing, recent work both from the probabilistic and econometric sides gives the mixed normal distribution of the error in the approximation in (3). References include [6, 25, 31]. The random variance of the normal error is $2\frac{T_i^+ - T_i^-}{n}$.

Per A. Mykland is Professor, Department of Statistics, The University of Chicago, Chicago, Illinois 60637, USA (e-mail: mykland@galton.uchicago.edu). Lan Zhang is Assistant Professor, Department of Finance, University of Illinois at Chicago, Chicago, Illinois 60607, and Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: lzhang@stat.cmu.edu).

$\int_{T_i^-}^{T_i^+} \sigma_t^4 \, dH(t)$, where H is the *quadratic variation of time*. $H(t) = t$ in the case where the t_i are equidistant.

Further econometric literature includes, in particular, [3, 4, 8, 9, 14, 18, 32]. Problems that are attached to the estimation of covariations between two processes are discussed in [22]. Estimating σ_t^2 at each point t goes back to [13]; see also [30], but this has not caught on quite as much in econometric applications.

2. THE PRESENCE OF MEASUREMENT ERROR

The theory described above runs into a problem with real data. For illustration, consider how the realized volatility depends on sampling frequency for the stock (and day) considered in Figure 1. The estimator does not converge as the observation points t_i become dense in the interval of this one day, but rather seems to take off to infinity. This phenomenon was originally documented in [2]. For transaction data, this picture is repeated for most liquid securities [19, 37].

In other words, the model (1) is wrong. What can one do about this? A lot of people immediately think that the problem is due to jumps, but that is not the case. The limit in (3) exists even when there are jumps. The requirement for (3) to exist is that the process X be a semimartingale (we again cite Theorem I.4.47 of [26]), which includes both Itô processes and jumps.

The inconsistency between the empirical results where the realized volatility diverges with finer sampling, and the semimartingale theory which dictates the convergence of the realized volatility, poses a problem, since financial processes are usually assumed to be semimartingales. Otherwise, somewhat loosely speaking, there would be arbitrage opportunities in the financial markets. For rigorous statements, see, in particular, [11]. The semimartingaleness of financial processes, therefore, is almost a matter of theology in most of finance, and yet, because of Figure 1 and similar graphs for other stocks, we have to abandon it.

Our alternative model is that there is measurement error in the observation. At transaction number i, instead of seeing X_{t_i} from model (1) or, more generally, from a semimartingale, one observes

$$Y_{t_i} = X_{t_i} + \varepsilon_i. \tag{4}$$

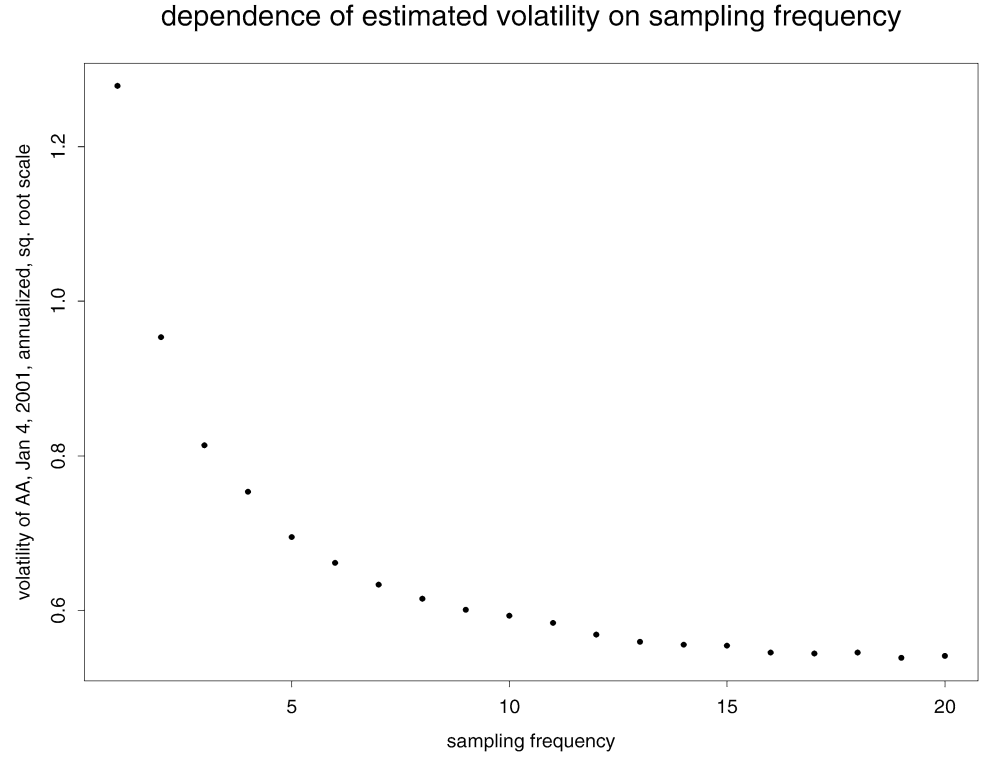


FIG. 1. *Plot of realized volatility for Alcoa Aluminum for January 4, 2001. The data is from the TAQ database. There are* 2011 *transactions on that day, on average one every* 13.365 *seconds. The most frequently sampled volatility uses all the data, and this is denoted as* "*frequency* $= 1$." "*Frequency* $= 2$" *corresponds to taking every second sampling point. Because this gives rise to two estimators of volatility, we have averaged the two. And so on for* "*frequency* $= k$" *up to* 20. *The plot corresponds to the average realized volatility discussed in* [37]. *Volatilities are given on an annualized and square root scale.*

We call this the *hidden semimartingale model*. The rationale is (depending on your subject matter) either that a transaction is a measurement of the underlying price X_{t_i}, and of course there is error, or that it is due to *market microstructure*, as documented by, among others, Roll [34], Glosten [15], Glosten and Harris [16], Brown [7], Harris [20] and Hasbrouck [21]. See [1] for a discussion of this.

A natural model for the error is that it is either i.i.d. or a stationary process, as considered by Zhou [38], Gloter and Jacod [17], Zhang, Mykland and Aït-Sahalia [37], Bandi and Russell [5], Zhang [36], Aït-Sahalia, Mykland and Zhang [1] and Hansen and Lunde [19].

Under quite loose conditions, this alternative model is consistent with the plot in Figure 1. Instead of (3), one gets that the realized volatility becomes $nE(\varepsilon_1 - \varepsilon_0)^2 + O_p(n^{-1/2})$. In the early literature (as cited in the previous section), the problem is usually taken care of by (*sic*) reducing n. A variety of approaches that improve on this are documented in [37], to which we refer for an in depth discussion. As demonstrated by Zhang [36], the true volatility Ξ_i can be consistently estimated at rate $O_p(n^{-1/4})$, as opposed to $O_p(n^{-1/2})$ when there is no error. This is not as slow as it seems, since n is quite large for liquid securities.

An alternative description of the error is that it arises due to rounding (financial instruments are, after all, traded on a price grid). Research in this direction has been done by Delattre and Jacod [10] and by Zeng [35]. To first order, the rounding and additive error models are similar, as documented by Delattre and Jacod [10]; see also [28].

It is awkward that these models imply the existence of arbitrage. The size of the error, however, is so small that it is hard to take economic advantage of them, and this, presumably, is why such deviations can persist.

3. IMPLICATIONS FOR MARKOV MODELS

We now return to the subject to Jianqing Fan's overview, namely the Markov case. It is clear that the model without observation error is not consistent with the data. This may not be a problem when working with, say, daily data, but would pose problems when using high-frequency (intraday) observations. It is presumably quite straightforward to extend the methods discussed in the paper to the case of observation error, and it would be interesting to see the results. The same applies to similar studies on Markov models by the "French school," such as Hoffmann [23] and Jacod [24].

ACKNOWLEDGMENT

We gratefully acknowledge the support of the National Science Foundation under Grant DMS-02-04639.

REFERENCES

[1] AÏT-SAHALIA, Y., MYKLAND, P. A. and ZHANG, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* **18** 351–416.

[2] ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2000). Great realizations. *Risk* **13** 105–108.

[3] ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2001). The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.* **96** 42–55.

[4] ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2003). Modeling and forecasting realized volatility. *Econometrica* **71** 579–625.

[5] BANDI, F. M. and RUSSELL, J. R. (2005). Microstructure noise, realized volatility and optimal sampling. Technical report, Graduate School of Business, Univ. Chicago.

[6] BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 253–280.

[7] BROWN, S. J. (1990). Estimating volatility. In *Financial Options*: *From Theory to Practice* (S. Figlewski, W. Silber and M. Subrahmanyam, eds.) 516–537. Business One-Irwin, Homewood, IL.

[8] CHERNOV, M. and GHYSELS, E. (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *J. Financial Economics* **56** 407–458.

[9] DACOROGNA, M. M., GENÇAY, R., MÜLLER, U., OLSEN, R. B. and PICTET, O. V. (2001). *An Introduction to High-Frequency Finance*. Academic Press, San Diego.

[10] DELATTRE, S. and JACOD, J. (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli* **3** 1–28.

[11] DELBAEN, F. and SCHACHERMAYER, W. (1995). The existence of absolutely continuous local martingale measures. *Ann. Appl. Probab.* **5** 926–945.

[12] FLEMING, J., KIRBY, C. and OSTDIEK, B. (2001). The economic value of volatility timing. *J. Finance* **56** 329–352.

[13] FOSTER, D. and NELSON, D. (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica* **64** 139–174.

[14] GALLANT, A. R., HSU, C.-T. and TAUCHEN, G. (1999). Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *Review of Economics and Statistics* **81** 617–631.

[15] GLOSTEN, L. R. (1987). Components of the bid-ask spread and the statistical properties of transaction prices. *J. Finance* **42** 1293–1307.

[16] GLOSTEN, L. R. and HARRIS, L. E. (1988). Estimating the components of the bid/ask spread. *J. Financial Economics* **21** 123–142.

[17] GLOTER, A. and JACOD, J. (2000). Diffusions with measurement errors: I—Local asymptotic normality and II—Optimal estimators. Technical report, Univ. de Paris VI.
[18] GONCALVES, S. and MEDDAHI, N. (2005). Bootstrapping realized volatility. Technical report, Univ. de Montréal.
[19] HANSEN, P. R. and LUNDE, A. (2006). Realized variance and market microstructure noise. *J. Bus. Econom. Statist.* To appear.
[20] HARRIS, L. (1990). Statistical properties of the Roll serial covariance bid/ask spread estimator. *J. Finance* **45** 579–590.
[21] HASBROUCK, J. (1993). Assessing the quality of a security market: A new approach to transaction-cost measurement. *Review of Financial Studies* **6** 191–212.
[22] HAYASHI, T. and YOSHIDA, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **11** 359–379.
[23] HOFFMANN, M. (1999). L_p estimation of the diffusion coefficient. *Bernoulli* **5** 447–481.
[24] JACOD, J. (2000). Nonparametric kernel estimation of the coefficient of a diffusion. *Scand. J. Statist.* **27** 83–96.
[25] JACOD, J. and PROTTER, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307.
[26] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. Springer, New York.
[27] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, New York.
[28] KOLASSA, J. and MCCULLAGH, P. (1990). Edgeworth series for lattice distributions. *Ann. Statist.* **18** 981–985.
[29] MYKLAND, P. A. (2003). Financial options and statistical prediction initervals. *Ann. Statist.* **31** 1413–1438.
[30] MYKLAND, P. A. and ZHANG, L. (2001). Inference for volatility type objects and implications for hedging. Technical report, Dept. Statistics, Carnegie Mellon Univ.
[31] MYKLAND, P. A. and ZHANG, L. (2002). ANOVA for diffusions. Technical report, Dept. Statistics, Univ. Chicago.
[32] OOMEN, R. (2004). Properties of realized variance for a pure jump process: Calendar time sampling versus business time sampling. Technical report, Warwick Business School, Univ. Warwick.
[33] PROTTER, P. (2004). *Stochastic Integration and Differential Equations*: *A New Approach*, 2nd ed. Springer, New York.
[34] ROLL, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *J. Finance* **39** 1127–1139.
[35] ZENG, Y. (2003). A partially-observed model for micromovement of asset process with Bayes estimation via filtering. *Math. Finance* **13** 411–444.
[36] ZHANG, L. (2004). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Technical report, Dept. Statistics, Carnegie Mellon Univ.
[37] ZHANG, L., MYKLAND, P. A. and AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411.
[38] ZHOU, B. (1996). High-frequency data and volatility in foreign-exchange rates. *J. Bus. Econom. Statist.* **14** 45–52.

Statistical Science
2005, Vol. 20, No. 4, 351–357
DOI 10.1214/088342305000000421

Rejoinder: A Selective Overview of Nonparametric Methods in Financial Econometrics

Jianqing Fan

I am very grateful to the Executive Editor, Edward George, for organizing this stimulating discussion. I would like to take this opportunity to thank Professors Peter Phillips, Jun Yu, Michael Sørensen, Per Mykland and Lan Zhang for their insightful and stimulating comments, touching both practical, methodological and theoretical aspects of financial econometrics and their applications in asset pricing, portfolio allocation and risk management. They have made valuable contributions to the understanding of various financial econometric problems.

The last two decades have witnessed an explosion of developments of data-analytic techniques in statistical modeling and analysis of complex systems. At the same time, statistical techniques have been widely employed to confront various complex problems arising from financial and economic activities. While the discipline has grown rapidly over the last two decades and has rich and challenging statistical problems, the number of statisticians involved in studying financial econometric problems is still limited. In comparison with statisticians working on problems in biological sciences and medicine, the group working on financial and econometric problems is dismally small. It is my hope that this article will provide statisticians with quick access to some important and interesting problems in financial econometrics and to catalyze the romance between statistics and finance. A similar effort was made by Cai and Hong [12], where various aspects of nonparametric methods in continuous-time finance are reviewed. It is my intention to connect financial econometric problems as closely to statistical problems as possible so that familiar statistical tools can be employed. With this in mind, I sometimes oversimplify the problems and techniques so that key features can be highlighted.

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).

I am fully aware that financial econometrics has grown into a vast discipline itself and that it is impossible for me to provide an overview within a reasonable length. Therefore, I greatly appreciate what all discussants have done to expand the scope of discussion and provide additional references. They have also posed open statistical problems for handling nonstationary and/or non-Markovian data with or without market noise. In addition, statistical issues on various versions of capital asset pricing models and their related stochastic discount models [15, 19], the efficient market hypothesis [44] and risk management [17, 45] have barely been discussed. These reflect the vibrant intersection of the interfaces between statistics and finance. I will make some further efforts in outlining econometric problems where statistics plays an important role after brief response to the issues raised by the discussants.

1. BIASES IN STATISTICAL ESTIMATION

The contributions by Professors Phillips, Yu and Sørensen address the bias issues on the estimation of parameters in diffusion processes. Professors Phillips and Yu further translate the bias of diffusion parameter estimation into those of pricing errors of bonds and bond derivatives. Their results are very illuminating and illustrate the importance of estimation bias in financial asset pricing. Their results can be understood as follows. Suppose that the price of a financial asset depends on certain parameters θ (the speed of the reversion κ in their illustrative example). Let us denote it by $p(\theta)$, which can be in one case the price of a bond and in another case the prices of derivatives of a bond. The value of the asset is now estimated by $p(\hat{\theta})$ with $\hat{\theta}$ being estimated from empirical data. When $\hat{\theta}$ is overestimated (say), which shifts the whole distribution of $\hat{\theta}$ to the left, the distribution of $p(\hat{\theta})$ will also be shifted, depending on the sensitivity of p to θ. The sensitivity is much larger for bond derivatives when κ is close to zero (see Figure 2 of [46]), and hence the pricing errors are much larger. On the other hand, as the distribution

of κ is shifted to the left, from Figure 2 of [46], both prices of bonds and their derivatives get smaller and so does the variance of pricing errors. Simulation studies in [46] suggest that these two effects cancel each other out in terms of mean square error.

I agree with Phillips and Yu's observation that discretization is not the main source of biases for many reasonable financial applications. Finite-sample estimation bias can be more severe. This partially answers the question raised by Professor Sørensen. On the other hand, his comments give theoretical insights into the bias due to discretization. For financial applications (such as modeling short-term rates) when the data are collected at monthly frequency, the bias $\{1-\exp(-\kappa\Delta)\}/\Delta-\kappa=-0.0019$ and -0.00042, respectively, for $\kappa=0.21459$ used in Figure 3 of [34] and for $\kappa=0.1$ used in the discussion by Phillips and Yu. For weekly data, using the parameter $\kappa=0.0446$ cited in [14], the discretization bias is merely 9.2×10^{-5}.

For other types of applications, such as climatology, Professor Sørensen is right that the bias due to discretization can sometimes be substantial. It is both theoretically elegant and practically viable to have methods that work well for all situations. The quasi-maximum likelihood methods and their modifications discussed by Professor Sørensen are attractive alternatives. As he pointed out, analytical solutions are rare and computation algorithms are required. This increases the chance of numerical instability in practical implementations. The problem can be attenuated with the estimates based on the Euler approximation as an initial value. The martingale method is a generalization of his quasi-maximum likelihood estimator, which aims at improving efficiency by suitable choice of weighting functions a_j. However, unless the conditional density has multiplicative score functions, the estimation equations will not be efficient. This explains the observation made by Professor Sørensen that the methods based on martingale estimating functions are usually not efficient for low frequency data. The above discussion tends to suggest that when the Euler approximation is reasonable, the resulting estimates tend to have smaller variances.

In addition to the discretization bias and finite sample estimation bias, there is model specification bias. This can be serious in many applications. In the example given by Professors Phillips and Yu, the modeling errors do not have any serious adverse effects on pricing bonds and their derivatives. However, we should be wary of generalizing this statement. Indeed, for the model parameters given in the discussion by Phillips and Yu, the transition density of the CIR model has a noncentral χ^2-distributions with degrees of freedom 80, which is close to the normal transition density given by the Vasicek model. Therefore, the model is not very seriously misspecified.

Nonparametric methods reduce model specification errors by either global modeling such as spline methods or local approximations. This reduces significantly the possibility of specification errors. Since nonparametric methods are somewhat crude and often used as model diagnostic and exploration tools, simple and quick methods serve many practical purposes. For example, in time domain smoothing, the bandwidth h is always an order of magnitude larger than the sampling frequency Δ. Therefore, the approximation errors due to discretization are really negligible. Similarly, for many realistic problems, the function approximation errors can be an order of magnitude larger than discretization errors. Hence, discretization errors are often not a main source of errors in nonparametric inference.

2. HIGH-FREQUENCY DATA

Professors Mykland, Zhang, Phillips and Jun address statistical issues for high-frequency data. I greatly appreciate their insightful comments and their elaborations on the importance and applications of the subject. Thanks to the advances in modern trading technology, the availability of high-frequency data over the last decade has significantly increased. Research in this area has advanced very rapidly lately. I would like to thank Professors Mykland and Zhang for their comprehensive overview on this active research area.

With high-frequency data, discretization errors have significantly been reduced. Nonparametric methods become even more important for this type of large sample problem. The connections between the realized volatility and the time-inhomogeneous model can simply be made as follows. Consider a subfamily of models of (8) in [34],

$$dX_t=\alpha_t\,dt+\sigma_t\,dW_t.$$

For high-frequency data the sampling interval is very small. For the sampling frequency of a minute, $\Delta=1/(252*24*60)\approx2.756\times10^{-6}$. Hence, standardized residuals in Section 2.5 of [34] become $E_t=\Delta^{-1/2}(X_{t+\Delta}-X_t)$ and the local constant estimate of the spot volatility reduces to

$$\hat\sigma^2_{j\Delta}=\sum_{i=-\infty}^{j-1}w_{j-i}E^2_{i\Delta},$$

where $\{w_i\}$ are the weights induced by a kernel function satisfying $\sum_{i=1}^{\infty} w_i = 1$. Now, for the weights with a bounded support, the quadratic variation of the process or integrated volatility $\int_t^T \sigma_t^2\, dt$ is naturally estimated by $\Delta \sum_{i=t/\Delta}^{T/\Delta-1} \hat{\sigma}_i^2$, which is simply

$$\sum_{i=t/\Delta}^{T/\Delta-1} \{X_{i\Delta} - X_{(i-1)\Delta}\}^2.$$

This shows that our nonparametric estimation of the integrated volatility for high-frequency data is indeed the same as the realized volatility.

As suggested by Professors Mykland, Zhang, Phillips and Yu, the applications of realized volatilities are not without difficulties. Market microstructure noises emerge at such a fine frequency of observation and market prices can contain multiple jumps due to the flux of information during a trading session. Figure 1 in the discussion by Mykland and Zhang demonstrates convincingly the existence of the market microstructure noise. Aït-Sahalia, Mykland and Zhang [1] and Zhang, Mykland and Aït-Sahalia [50] give comprehensive accounts of this under the assumption that the observed prices are the true ones contaminated with random noise of market microstructure: $Y_t = X_t + \varepsilon_t$. However, they do not take into account that the price processes $\{X_t\}$ may contain jumps in addition to random noises. An effort in this direction has been made recently by Fan and Wang [38] using wavelet techniques.

3. ESTIMATING COVARIANCE MATRICES

Covariance matrices play an important role in risk management and asset allocation. They are featured prominently in many financial econometrics problems. For example, the smallest and largest eigenvalues are related to the minimum and the maximum of the volatility of portfolios and their corresponding eigenvectors are related to portfolio allocation. See [40] for applications of covariance matrices to portfolio selection and [43] for their applications to other scientific problems. There are a couple of approaches to these kinds of problems, depending on the size of the covariance matrices. I hereby give a brief overview and address some of the open challenges.

The simplest estimate of a covariance matrix is probably the sample covariance matrix of the log-returns of p assets over a period of n days prior to the current time t. This is indeed a nonparametric estimation of the covariance matrix localizing in time and has been studied in multivariate analysis when p is finite and the underlying model is correct, that is, the covariance matrix remains the same in the n days prior to time t. See, for example, [26, 27, 47]. However, the impact of the biases in nonparametric methods on the estimation of eigenvalues and eigenvectors has not yet been thoroughly investigated.

The sample covariance matrices can be augmented by using the information from the state domain, which is an extension of the method discussed in Section 3.6 of [34] and allows us to use the historical information. This is particularly useful for estimating the covariance matrices of bonds with different maturities. Useful parametric models such as affine models have been popularly used in interest rate modeling. See, for example, [20, 24, 23]. Nonparametric methods provide useful alternatives to estimating the covariance matrices and to validating parametric models. A naive extension involves high-dimensional smoothing in the state domain. But this can be avoided by localizing only on the yields of a few bonds with intermediate length of maturity.

Another class of techniques is to use a form of GARCH model [28] to estimate covariance matrices. As noted in [30], the number of parameters grows rapidly with the dimensionality p. Various efforts have been made to reduce the complexity of the models. These include constant conditional correlation multivariate GARCH models [10], vectorized multivariate GARCH models [11], dynamic conditional correlation models [29, 31], orthogonal GARCH models [2], generalized orthogonal GARCH models [48] and conditionally uncorrelated component models [37]. For a survey, see [8].

In portfolio allocation and risk management, the number of stocks p can be well in the order of hundreds, which is typically in the same order as the sample size n. The sample covariance matrix may not be a good estimator of the population one. The estimated variance of a portfolio based on the sample covariance may far exceed the true one. The estimation errors can accumulate quickly when p grows with n. Indeed, Johnstone [43] shows that the largest eigenvalue of the covariance matrix is far larger than the population one. There are many studies on the behavior of random matrices when the dimensionality p grows with n. See, for example, [5, 22, 21, 49]. For a survey, see [4].

Estimating covariance matrices for large p is intrinsically challenging. For example, when $p = 200$, there are more than 20,000 free parameters. Yet, the

available sample size is usually in the order of hundreds or a few thousand. Longer time series (larger n) will increase modeling biases. Without imposing structures on the covariance matrices, they are hard to estimate. Thanks to the multi-factor models (see Chapter 6 of [13]), if a few factors can capture completely the cross-sectional risks, the number of parameters can be significantly reduced. For example, using the Fama–French three-factor models [32, 33], there are $4p$ instead of $p(p+1)/2$ parameters. Natural questions arise with this structured estimate of the covariance matrix, how large p can be such that the estimation error in the covariance matrix is negligible in asset allocation and risk management. The problems of this kind are interesting and remain open.

Another possible approach to the estimation of covariance matrices is to use a model selection approach. First of all, according to Chapter 3 of [39], the Cholesky decomposition admits nice autoregressive interpretation. We may reasonably assume that the elements in the Cholesky decomposition of the covariance matrix are sparse. Hence, the penalized likelihood method [3, 35, 42] can be employed to select and estimate nonsparse elements. The sampling property of such a method remains unknown. Its impact on portfolio allocation and risk management needs to be studied.

4. STATISTICS IN DERIVATIVE PRICING

Over last three decades, option pricing has witnessed an explosion of new models that extend the original work of Black and Scholes [9]. Empirically pricing financial derivatives is innately related to statistical regression problems. This is well documented in papers such as [6, 7, 15, 16, 25, 41]. See also a brief review given by Cai and Hong [12]. For a given stochastic model with given structural parameters under the risk-neutral measure, the prices of European options can be determined, which are simply the discounted expected payoffs under the risk-neutral measure. Bakshi, Cao and Chen [6] give the analytic formulas of option prices for five commonly used stochastic models, including the stochastic-volatility random-jump model. They then estimate the risk-neutral parameters by minimizing the discrepancies between the observed prices and the theoretical ones. With estimated risk-neutral parameters, option prices with different characteristics can be evaluated. They conduct a comprehensive study of the relative merits of competing option pricing models by computing pricing errors for new options. Dumas, Fleming and Whaley [25] model the implied volatility function by a quadratic function of the strike price and time to maturity and determine these parameters by minimizing pricing errors. Based on the analytic formula of Bakshi, Cao and Chen [6] for option price under the stochastic volatility models, Chernov and Ghysels [16] estimate the risk neutral parameters by integrating information from both historical data and risk-neutral data implied by observed option prices. Instead of using continuous-time diffusion models, Heston and Nandi [41] assume that the stock prices under the risk-neutral world follow a GARCH model and derive a closed form for European options. They determine the structural parameters by minimizing the discrepancy between the empirical and theoretical option prices. Barone-Adesi, Engle and Mancini [7] estimate risk-neutral parameters by integrating the information from both historical data and option prices. Christoffersen and Jakobs [18] expand the flexility of the model by introducing long- and short-run volatility components.

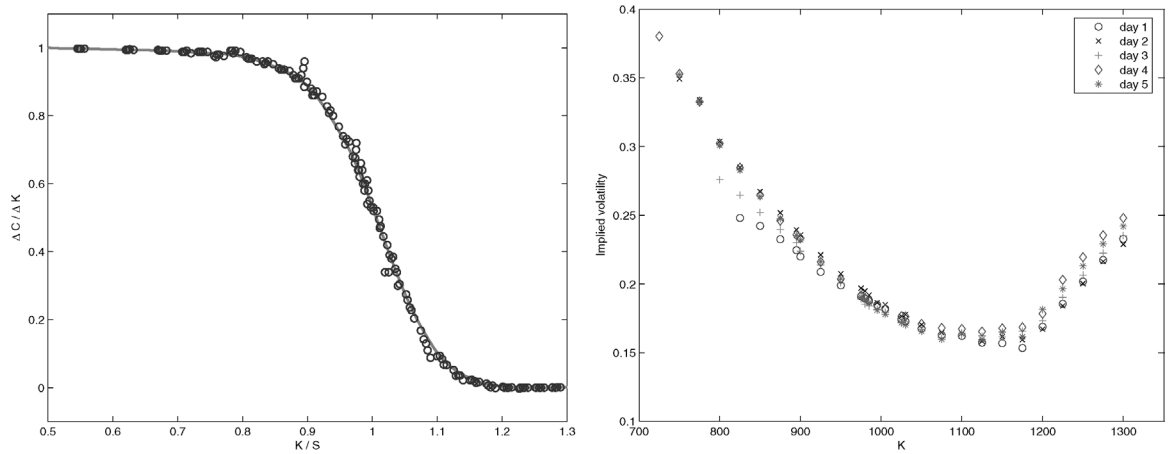
The above approaches can be summarized as follows. Using the notation in Section 4.1 of [34], the theoretical option price with option characteristics $(S_i, K_i, T_i, r_i, \delta_i)$ is governed by a parametric form $C(S_i, K_i, T_i, r_i, \delta_i, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of structural parameters of the stock price dynamics under the risk-neutral measure. The form depends on the underlying parameters of the stochastic model. This can be in one case a stochastic volatility model and in another case a GARCH model. The parameters are then determined by minimizing

$$\sum_{i=1}^{n}\{C_i - C(S_i, K_i, T_i, r_i, \delta_i, \boldsymbol{\theta})\}^2$$

or similar discrepancy measures. The success of a method depends critically on the correctness of model assumptions under the risk-neutral measure. Since these assumptions are not on the physical measure, they are hard to verify. This is why so many parametric models have been introduced. Their efforts can be regarded as searching an appropriate parametric form $C(\cdot;\boldsymbol{\theta})$ to better fit the option data. Nonparametric methods in Section 4.1 provide a viable alternative for this purpose. They can be combined with parametric approaches to improve the accuracy of pricing.

As an illustration, let us consider the options with fixed $(S_i, T_i, r_i, \delta_i)$ so that their prices are only a function of K or equivalently a function of the moneyness $m = K/S$,

$$C = \exp(-rT)\int_K^\infty (x-K) f^*(x)\,dx.$$



(a) (b)

FIG. 1. (a) *Scatterplot of the response variable computed based on option prices with consecutive strike price against the moneyness.* (b) *The implied volatilities of the options during the period July 7–11, 2003.*

Denoting $D = \exp(rT)C/S$ and letting $\bar{F}^*(x) = 1 - F^*(x) = \int_x^\infty f^*(y)\,dy$ be the survival function, then by integration by parts,

$$D = -S^{-1}\int_K^\infty (x-K)\,d\bar{F}^*(x) = S^{-1}\int_K^\infty \bar{F}^*(x)\,dx.$$

By a change of variable, we have

$$D = \int_m^\infty \bar{F}(u)\,du,$$

where $F(u) = F^*(Su)$ is the state price distribution in the normalized scale (the stock price is normalized to \$1). Let us write explicitly $D(m)$ to stress the dependence of discounted option price on the moneyness m. Then

$$\frac{D(m_1)-D(m_2)}{m_2-m_1} = (m_2-m_1)^{-1}\int_{m_1}^{m_2}\bar{F}(u)\,du = \bar{F}\left(\frac{m_2+m_1}{2}\right) + O\big((m_2-m_1)^2\big).$$

Assume that the moneyness $m_i = K_i/S_t$ has already been ordered for N_t options with strike prices $\{K_i, i = 1,\ldots,N_t\}$ traded at time t. Let $x_i = (m_i + m_{i+1})/2 = (K_i + K_{i+1})/(2S)$ and y_i be the observed value of $\frac{D(m_i)-D(m_{i+1})}{m_{i+1}-m_i}$, namely,

$$y_i = \exp(r_t T_t)\{C_i - C_{i+1}\}/\{K_{i+1}-K_i\}, \qquad i = 1,\ldots,N_t - 1,$$

where r_t, T_t and S_t are, respectively, the risk-free interest rate, time to maturity and spot stock price at time t, and C_{i+1} and C_i are the option prices at time t associated with strike prices K_{i+1} and K_i. Then, estimating the state price distribution becomes a familiar nonparametric regression problem,

$$y_i \approx \bar{F}(x_i) + \varepsilon_i.$$

In the above equation, the dependence on t is suppressed. Figure 1(a) shows the scatterplot of the pairs (x_i, y_i) based on the closing call option prices (average of bid-ask prices) of the Standard and Poor's 500 index with maturity of $T_t = 75 - t$ days on the week of July 7 to July 11, 2003 ($t = 0,\ldots,4$). The implied volatility curve is given in Figure 1(b). It is not a constant and provides stark evidence against the Black–Scholes formula.

The waterfall shape of the regression curve is very clear. The naive applications of nonparametric techniques will incur large approximation biases resulting in systematic pricing errors. One possible improvement is to use a parametric method such as the ad-hoc Black–Scholes model of Dumas, Fleming and Whaley [25] to estimate the main shape of the regression function and then use a nonparametric method to estimate the difference. This kind of idea has been investigated by Fan and Mancini [36]. When we aggregate the data in the week of July 7 to July 11, 2003, the times to maturity T_t vary slightly. Semiparametric techniques can be used to adjust for this effect. Similarly to many practical problems, we always have side information available that can be incorporated into modeling and analysis of the data. This reinforces the claim that pricing financial derivatives is fundamentally a statistical problem where statisticians can play an important role.

ACKNOWLEDGMENTS

This research was supported in part by NSF Grant DMS-03-55179. The author acknowledges gratefully various discussions with Dr. Loriano Mancini and would like to thank him for providing Figure 1 presented in this discussion.

REFERENCES

[1] AÏT-SAHALIA, Y., MYKLAND, P. and ZHANG, L. (2005). How often to sample a continuous-time process in the presence of maket microstructure noise. *Review of Financial Studies* **18** 351–416.

[2] ALEXANDER, C. (2001). Orthogonal GARCH. In *Mastering Risk* **2** (C. Alexander, ed.) 21–38. Financial Times Prentice Hall, London.

[3] ANTONIADIS, A. and FAN, J. (2001). Regularization of wavelet approximations (with discussion). *J. Amer. Statist. Assoc.* **96** 939–967.

[4] BAI, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices. A review (with discussion). *Statist. Sinica* **9** 611–677.

[5] BAI, Z. D. and SILVERSTEIN, J. W. (1999). Exact separation of eigenvalues of large-dimensional sample covariance matrices. *Ann. Probab.* **27** 1536–1555.

[6] BAKSHI, G., CAO, C. and CHEN, Z. (1997). Empirical performance of alternative option pricing models. *J. Finance* **52** 2003–2049.

[7] BARONE-ADESI, G., ENGLE, R. F. and MANCINI, L. (2004). GARCH options in incomplete markets. NCCR-FinRisk Working Paper 155, Univ. Zurich.

[8] BAUWENS, L., LAURENT, S. and ROMBOUTS, J. V. K. (2003). Multivariate GARCH models: A survey. CORE Discussion Paper 2003/31.

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

[10] BOLLERSLEV, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics* **72** 498–505.

[11] BOLLERSLEV, T., ENGLE, R. and WOOLDRIDGE, J. (1988). A capital asset pricing model with time-varying covariances. *J. Political Economy* **96** 116–131.

[12] 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.

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

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

[15] CHEN, X. and LUDVIGSON, S. (2005). Land of Addicts? An empirical investigation of habit-based asset pricing model. Unpublished manuscript.

[16] CHERNOV, M. and GHYSELS, E. (2000). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *J. Financial Economics* **56** 407–458.

[17] CHRISTOFFERSEN, P. F. (2003). *Elements of Financial Risk Management*. Academic Press, San Diego.

[18] CHRISTOFFERSEN, P. F. and JAKOBS, K. (2005). Option valuation with long-run and short-run volatility components. Unpublished manuscript.

[19] COCHRANE, J. H. (2001). *Asset Pricing*. Princeton Univ. Press.

[20] DAI, Q. and SINGLETON, K. (2000). Specification analysis of affine term structure models. *J. Finance* **55** 1943–1978.

[21] DETTE, H. (2002). Strong approximation of eigenvalues of large dimensional Wishart matrices by roots of generalized Laguerre polynomials. *J. Approx. Theory* **118** 290–304.

[22] DIACONIS, P. and SHAHSHAHANI, M. (1994). On the eigenvalues of random matrices. *J. Appl. Probab.* **31A** 49–61.

[23] DUFFEE, G. R. (2002). Term premia and interest rate forecasts in affine models. *J. Finance* **57** 405–443.

[24] DUFFIE, D. and KAN, R. (1996). A yield-factor model of interest rates. *Math. Finance* **6** 379–406.

[25] DUMAS, B., FLEMING, J. and WHALEY, R. E. (1998). Implied volatility functions: Empirical tests. *J. Finance* **53** 2059–2106.

[26] EATON, M. L. and TYLER, D. E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Ann. Statist.* **19** 260–271.

[27] EATON, M. L. and TYLER, D. (1994). The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis. *J. Multivariate Anal.* **50** 238–264.

[28] ENGLE, R. F., ed. (1995). *ARCH, Selected Readings*. Oxford Univ. Press.

[29] ENGLE, R. F. (2002). Dynamic conditional correlation—a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econom. Statist.* **20** 339–350.

[30] ENGLE, R. F. and KRONER, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* **11** 122–150.

[31] ENGLE, R. F. and SHEPPARD, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Preprint.

[32] FAMA, E. and FRENCH, K. (1992). The cross-section of expected stock returns. *J. Finance* **47** 427–465.

[33] FAMA, E. and FRENCH, K. (1993). Common risk factors in the returns on stocks and bonds. *J. Financial Economics* **33** 3–56.

[34] FAN, J. (2005). A selective overview of nonparametric methods in financial econometrics. *Statist. Sci.* **20** 317–337.

[35] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360.

[36] FAN, J. and MANCINI, L. (2005). Derivative pricing using semiparametric techniques. Unpublished manuscript.

[37] FAN, J., WANG, M. and YAO, Q. (2005). Modelling multivariate volatilities via conditionally uncorrelated components. Unpublished manuscript.

[38] FAN, J. and WANG, Y. (2005). Multiscale jump and volatility analysis for high-frequency financial data. Unpublished manuscript.

[39] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series*: *Nonparametric and Parametric Methods*. Springer, New York.

[40] GOLDFARB, D. and IYENGAR, G. (2003). Robust portfolio selection problems. *Math. Oper. Res.* **28** 1–37.

[41] HESTON, S. L. and NANDI, S. (2000). A closed-form GARCH option valuation model. *Review of Financial Studies* **13** 585–625.

[42] HUANG, J. Z., LIU, N. and POURAHMADI, M. (2005). Covariance selection and estimation via penalized normal likelihood. Unpublished manuscript.

[43] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327.

[44] LO, A. (2000). Finance: A selective survey. *J. Amer. Statist. Assoc.* **95** 629–635.

[45] MOFFATT, H. K. (2002). *Risk Management*: *Value at Risk and Beyond*. Cambridge Univ. Press.

[46] PHILLIPS, P. C. B. and YU, J. (2005). Jackknifing bond option prices. *Review of Financial Studies* **18** 707–742.

[47] TYLER, D. E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9** 725–736.

[48] VAN DER WEIDE, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model. *J. Appl. Econometrics* **17** 549–564.

[49] WIEAND, K. (2002). Eigenvalue distributions of random unitary matrices. *Probab. Theory Related Fields* **123** 202–224.

[50] ZHANG, L., MYKLAND, P. and AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411.