Nonparametric Methods in Continuous-Time Finance: A Selective Review

Zongwu Cai
Department of Mathematics
University of North Carolina
Charlotte, NC 28223
E-mail: zcai@uncc.edu

Yongmiao Hong
Department of Economics
Cornell University
Ithaca, NY 14853
E-mail: yh20@cornell.edu

March 17, 2003

Abstract

This paper gives a selective review on the recent developments of nonparametric methods in continuous-time finance, particularly in the areas of nonparametric estimation of diffusion processes, nonparametric testing of parametric diffusion models, and nonparametric pricing of derivatives. For each financial context, the paper discusses the suitable statistical concepts, models, and modeling procedures, as well as some of their applications to financial data. Their relative strengths and weakness are discussed. Much theoretical and empirical research is needed in this area, and more importantly, the paper points to several aspects that deserve further investigation.

Key Words: Continuous time model, derivative pricing, jump process, kernel smoothing, nonparametric test, non-stationarity, options, stochastic discount factor, time-dependent model.

The authors thank Haitao Li for his valuable and helpful comments, suggestions and discussions. Cai's research was supported, in part, by the National Science Foundation grant DMS-0072400 and funds provided by the University of North Carolina at Charlotte and Sonderforschungsbereich 373, Berlin, Germany. Hong's research was supported, in part, by the National Science Foundation grant SES-0111769.

1 Introduction

Nonparametric methods have become a core area in statistics (see, e.g., Härdle, 1990; Fan and Gijbels, 1996) in the last two decades and have been used successfully to various fields such as economics and finance (see, e.g., Pagan and Ullah, 1999; Mittelhammer, Judge and Miller, 2000) due to their advantage of requiring little prior information on the data generating process. Recently, nonparametric techniques have been proved to be the most attractive way to conduct research and gain economic intuition in certain core areas in finance, such as asset and derivative pricing, term structure theory, portfolio choice, and risk management, particularly, in modeling continuous-time models.

Finance is characterized by time and uncertainty. Continuous-time modelling has been a basic analytic tool in modern finance since the seminar papers by Black and Scholes (1973) and Merton (1973). The rationale behind it is that most of time, news arrives at financial markets in a continuous manner. More importantly, derivative pricing in theoretical finance is generally much more convenient and elegant in a continuous-time framework than through binomial or other discrete approximations. However, statistical analysis based on continuous-time financial models has just emerged as a field in less than a decade. This is apparently due to the difficulty of estimating and testing continuous-time models using discretely observed data. The purpose of this survey is to review some recent developments of nonparametric methods in continuous-time finance, and particularly in the areas of nonparametric estimation and testing of diffusion models, and derivative pricing. Financial time series data have some distinct important stylized facts, such as persistent volatility clustering, heavy tails, strong serial dependence, and occasionally sudden but large jumps. In addition, financial modelling is often closely embedded in a financial theoretical framework. These features suggest that standard statistical theory may not be readily applicable to financial time series. This is a promising and fruitful area for both financial economists and statisticians to interact each other.

Section 2 introduces various continuous-time diffusion processes and nonparametric estimation methods for diffusion processes. Section 3 reviews the estimation and testing of a parametric diffusion model using nonparametric methods. Section 4 discusses nonparametric estimation of derivative pricing, particularly the estimation of risk neutral density functions. Section 5 concludes.

2 Diffusions and Nonparametric Estimation

2.1 Models

Modeling the dynamics of interest rates, stock prices, foreign exchange rates, and macroeconomic factors, *inter alia*, is one of the most important topics in asset pricing studies. The instantaneous risk-free interest rate or the so-called short rate is, for example, the state variable that determines the evolution of the yield curve in an important class of term structure models, such as Vasicek (1977) and Cox, Ingersoll and Ross (1985, CIR). It is of fundamental importance for pricing fixed-income securities. Many theoretical models have been developed in mathematical finance to describe the short rate movement.[1]

In the theoretical term structure literature, the short rate or the underlying process of interest, $\{X_t,\, t \geq 0\}$, is often modelled as a time-homogeneous diffusion process, or stochastic differential equation,

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dB_t, \tag{2.1}$$

where $\{B_t,\, t \geq 0\}$ is a standard Brownian motion. The functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ are respectively the drift (or instantaneous mean) and the diffusion (or instantaneous variance) of the process, which determine the dynamics of the short rate.

There are two approaches to modeling $\mu(\cdot)$ and $\sigma(\cdot)$. The first is a parametric approach, which assumes some parametric forms $\mu(\cdot, \theta)$ and $\sigma(\cdot, \theta)$, and estimates the unknown model parameters θ. Most existing models in the literature assume that the interest rate exhibits mean-reversion and that the drift $\mu(\cdot)$ is a linear function of the interest rate level. It is also often assumed that the diffusion $\sigma(\cdot)$ takes the form of $\sigma\,|X_t|^\gamma$, where γ measures the sensitivity of interest rate volatility to the interest rate level. This specification, in modelling interest rate dynamics, captures the so-called "level effect"; i.e., the higher the interest rate level, the larger the volatility. With $\gamma = 0$ and 0.5, the model (2.1) reduces to the well-known Vasicek and CIR models, respectively. The forms of $\mu(\cdot, \theta)$ and $\sigma(\cdot, \theta)$ are typically chosen due to theoretical convenience. They may not be consistent with the data generating process.

The second approach is a nonparametric one, which does not assume any restrictive functional form for $\mu(\cdot)$ and $\sigma(\cdot)$ beyond regularity conditions. In the last few years, great progress has been made in estimating and testing continuous-time models for the short term interest rate using nonparametric methods.[2] Despite many studies, empirical analysis on the

[1] Other theoretical models are studied by Brennan and Schwartz (1979), Constantinides (1992), Courtadon (1982), Cox, Ingersoll and Ross (1980), Dothan (1978), Duffie and Kan (1996), Longstaff and Schwartz (1992), Marsh and Rosenfield (1983), and Merton (1973). Heath, Jarrow and Morton (1992) consider another important class of term structure models which use the forward rate as the underlying state variable.

[2] Empirical studies on the short rate include Ait-Sahalia (1996a, b), Andersen and Lund (1997), Ang and Bekaert (1998), Brenner, Harjes and Kroner (1996), Brown and Dybvig (1986), Chan, Karolyi, Longstaff

functional forms of the drift and diffusion is still not conclusive. For example, recent studies by Ait-Sahalia (1996b) and Stanton (1997) using nonparametric methods, overwhelmingly reject all linear drift models for the short rate. They find that the drift of the short rate is a nonlinear function of the interest rate level. Both studies show that for the lower and middle ranges of the interest rate, the drift is almost zero, i.e., the interest rate behaves like a random walk. But the short rate exhibits strong mean-reversion when the interest rate level is high. These findings lead to the development of nonlinear term structure models such as those of Ahn and Gao (1999).

However, the evidence of nonlinear drift has been challenged by Pritsker (1998) and Chapman and Pearson (2000), who find that the nonparametric methods of Ait-Sahalia (1996b) and Stanton (1997) have severe finite sample problems, especially near the extreme observations. The finite sample problems with nonparametric methods cast doubt on the evidence of nonlinear drift. On the other hand, the findings in Ait-Sahalia (1996b) and Stanton (1997) that the drift is nearly flat for the middle range of the interest rate are not much affected by the small sample bias. Chapman and Pearson (2000) point out that this is a puzzling fact, since "there are strong theoretical reasons to believe that short rate cannot exhibit the asymptotically explosive behavior implied by a random walk model." They conclude that "time series methods alone are not capable of producing evidence of nonlinearity in the drift." Recently, Fan and Zhang (2001) fit a nonparametric model using local linear technique and apply the generalized likelihood ratio test of Fan, Zhang and Zhang (2001) to test whether the drift is linear. They support Chapman and Pearson's (2000) conclusion. However, the generalized likelihood ratio test is developed by Fan, Zhang and Zhang (2001) for the iid samples but it is unknown whether it is valid for financial time series contexts, which is warranted for a further investigation. Interest rate data are well-known for persistent serial dependence. Pritsker (1998) uses Vasicek's (1977) model of interest rates to investigate the performance of a nonparametric density estimation in finite samples. He finds that asymptotic theory gives poor approximation even for a rather large sample size.

Controversies also exist on the diffusion $\sigma(\cdot)$. The specification of $\sigma(\cdot)$ is important, because it affects derivative pricing. Chan, Karolyi, Longstaff and Sanders (1992) show that in a single factor model of the short rate, γ roughly equals to 1.5 and all the models with $\gamma \leq 1$ are rejected. Ait-Sahalia (1996b) finds that γ is close to 1, Stanton (1997) finds that in his simiparametric model γ is about 1.5, and Conley, Hansen, Luttmer and Scheikman (1997) show that their estimate of γ is between 1.5 and 2. However, Bliss and Smith (1998)

and Sanders (1992), Chapman and Pearson (2000), Chapman, Long and Pearson (1999), Conley, Hansen, Luttmer and Scheinkman (1997), Gray (1996), and Stanton (1997).

argue that the result that γ equals 1.5 depends on whether the data between October 1979 to September 1982 are included.

2.2 Nonparametric estimation

Under some regularity conditions (Jiang and Knight, 1997; Bandi and Nguyen, 2000a), the diffusion process in (2.1) is a one-dimensional, regular, strong Morkov process with continuous sample paths and time-invariant stationary transition density. The drift and diffusion are respectively the first two moments of the infinitesimal conditional distribution of X_t :

$$\mu(X_t) = \lim_{\Delta \to 0} E\left\{ \frac{X_{t+\Delta} - X_t}{\Delta} \,\Big|\, X_t \right\} \tag{2.2}$$

and

$$\sigma^2(X_t) = \lim_{\Delta \to 0} E\left\{ \frac{(X_{t+\Delta} - X_t)^2}{\Delta} \,\Big|\, X_t \right\}. \tag{2.3}$$

See, e.g., Øksendal (1985) and Karatzas and Shreve (1988). The drift describes the movement of X_t due to time changes, whereas the diffusion term measures the magnitude of random fluctuations around the drift.

Using the Dynkin operator (see, e.g., Øksendal, 1985; Karatzas and Shreve, 1988), Stanton (1997) shows that the first order approximation

$$\mu(X_t)^{(1)} = \frac{1}{\Delta}\, E\left\{X_{t+\Delta} - X_t \,|\, X_t\right\} + O(\Delta),$$

the second order approximation

$$\mu(X_t)^{(2)} = \frac{1}{2\,\Delta}\left[4\, E\left\{X_{t+\Delta} - X_t \,|\, X_t\right\} - E\left\{X_{t+2\,\Delta} - X_t \,|\, X_t\right\}\right] + O(\Delta^2),$$

and the third order approximation

$$\begin{aligned}\mu(X_t)^{(3)} \;&=\; \frac{1}{6\,\Delta}\left[18\, E\left\{X_{t+\Delta} - X_t \,|\, X_t\right\} - 9\, E\left\{X_{t+2\,\Delta} - X_t \,|\, X_t\right\}\right.\\ &\quad \left. +2\, E\left\{X_{t+3\,\Delta} - X_t \,|\, X_t\right\}\right] + O(\Delta^3),\end{aligned}$$

etc. Fan and Zhang (2001) derive higher-order approximations. Similar formulas hold for the diffusion (Stanton, 1997). Bandi and Nguyen (2000a) argue that approximations to the drift and diffusion of any order display the same rate of convergence and limiting variance, so that asymptotic argument in conjunction with computational issues suggest simply using the first order approximations in practice. As indicated by Stanton (1997, p.1982), the higher the order of the approximations, the faster they will converge to the true drift and diffusion. However, as noted by Bandi and Nguyen (2000a) and Fan and Zhang (2001), higher order approximations can be detrimental to the efficiency of the estimation procedure in finite samples. In fact, the variance grows nearly exponentially fast as the order increases and they are much more volatile than their lower order counterparts. For more discussions, see Bandi (2000), Bandi and Nguyen (2000a), and Fan and Zhang (2001).

Now suppose we observe X_t at $t = \tau\Delta$, $\tau = 1, \ldots, n$, in a fixed time interval $[0,\ T]$ with T. Denote the random sample as $\{X_{\tau\Delta}\}_{\tau=1}^{n}$. Then from (2.2) and (2.3) that the first order approximations to $\mu(\cdot)$ and $\sigma(\cdot)$ lead to

$$\mu(X_{\tau\Delta}) \approx \frac{1}{\Delta}\, E[X_{(\tau+1)\Delta} - X_{\tau\Delta} |\, X_{\tau\Delta}] \quad \text{and} \quad \sigma^2(X_{\tau\Delta}) \approx \frac{1}{\Delta}\, E[(X_{(\tau+1)\Delta} - X_{\tau\Delta})^2 \,|\, X_{\tau\Delta}]$$

for all $1 \le \tau \le n-1$. This becomes a classical nonparametric regression problem.

There are many nonparametric approaches to estimating conditional expectations. Most existing nonparametric methods in finance dwell mainly on the Nadaraya-Watson (NW) kernel estimator due to its simplicity. According to Ait-Sahalia (1996a, b), Stanton (1997), Jiang and Knight (1997), and Chapman and Pearson (2000), the NW estimators of $\mu(x)$ and $\sigma^2(x)$ are given, respectively, by

$$\widehat{\mu}(x) = \frac{1}{\Delta}\, \frac{\sum_{\tau=1}^{n-1} (X_{(\tau+1)\Delta} - X_{\tau\Delta})\; K_h(x - X_{\tau\Delta})}{\sum_{\tau=1}^{n-1} K_h(x - X_{\tau\Delta})},$$

and

$$\widehat{\sigma}^2(x) = \frac{1}{\Delta}\, \frac{\sum_{\tau=1}^{n-1} (X_{(\tau+1)\Delta} - X_{\tau\Delta})^2\; K_h(x - X_{\tau\Delta})}{\sum_{\tau=1}^{n-1} K_h(x - X_{\tau\Delta})}, \tag{2.4}$$

where $K_h(u) = K(u/h)/h$, $h = h_n > 0$ is the bandwidth with $h \to 0$ and $n\,h \to \infty$ as $n \to \infty$, and $K(\cdot) : \mathbb{R} \to \mathbb{R}$ is a standard kernel. Jiang and Knight (1997) suggest first using (2.4) to estimate $\sigma^2(\cdot)$. Observing that the drift

$$\mu(X_t) = \frac{1}{2\, \pi(X_t)}\, \frac{\partial[\sigma^2(X_t)\, \pi(X_t)]}{\partial X_t},$$

where $\pi(\cdot)$ is the stationary density of $\{X_t\}$ (see, e.g., Ait-Sahalia, 1996a; Jiang and Knight, 1997; Stanton, 1997; Bandi and Nguyen, 2000a), Jiang and Knight (1997) suggest estimating $\mu(\cdot)$ by

$$\widehat{\mu}(x) = \frac{1}{2\, \widehat{\pi}(x)}\, \frac{\partial \left\{\widehat{\sigma}^2(x)\, \widehat{\pi}(x)\right\}}{\partial x},$$

where $\widehat{\pi}(\cdot)$ is a consistent estimator of $\pi(\cdot)$, say, the classical kernel density estimator. The reason of doing so is based on the fact that in (2.1) the drift is of order dt and the diffusion is of order $\sqrt{dt}$, as $(d\,B_t)^2 = dt + O((dt)^2)$. That is, the diffusion has lower order than the drift for infinitesimal changes in time, and the local-time dynamics of the sampling path reflects more of the diffusion than those of the drift term. Therefore, when Δ is very small, identification becomes much easier for the diffusion term than the drift term.

It is well known that the NW estimator suffers from some disadvantages such as larger bias, boundary effects, and inferior minimax efficiency (see, e.g., Fan and Gijbels, 1996). To overcome these drawbacks, Fan and Zhang (2001) suggest using the local linear technique, for $k = 1$ and 2,

$$\sum_{\tau=1}^{n-1} \left\{ \Delta^{-1} (X_{(\tau+1)\Delta} - X_{\tau\Delta})^k - \beta_0 - \beta_1 (x - X_{\tau\Delta}) \right\}^2 K_h(x - X_{\tau\Delta}), \tag{2.5}$$

which gives the local linear estimate of $\mu(\cdot)$ for $k = 1$ and $\sigma^2(\cdot)$ for $k = 2$. However, the local linear estimator of the diffusion $\sigma(\cdot)$ cannot be always nonnegative in finite samples. To attenuate this disadvantage of local polynomial method, a weighted NW method due to Cai (2001) can be used to estimate $\sigma(\cdot)$ although the method needs further verification.

The asymptotic theory can be found in Jiang and Knight (1997) and Bandi and Nguyen (2000a) for the NW estimator and in Fan and Zhang (2001) for the local linear estimator. To implement kernel estimates, the bandwidth(s) must be chosen. In the iid setting, there are theoretically optimal bandwidth selections. There are no such results for diffusion processes available although there are many theoretic and empirical studies in the literature.

One crucial assumption in the above development is the stationarity of $\{X_t\}$. However, it might not hold for real financial time series data. If $\{X_t\}$ is not stationary, Bandi and Phillips (2003) propose using the following estimators to estimate $\mu(x)$ and $\sigma^2(x)$,

$$\begin{aligned} \widehat{\mu}(x) &= \frac{\sum_{\tau=1}^{n} K_h(x - X_{\tau\Delta})\, \widetilde{\mu}(X_{\tau\Delta})}{\sum_{\tau=1}^{n} K_h(x - X_{\tau\Delta})}, \qquad \text{and} \\ \widehat{\sigma}^2(x) &= \frac{\sum_{\tau=1}^{n} K_h(x - X_{\tau\Delta})\, \widetilde{\sigma}^2(X_{\tau\Delta})}{\sum_{\tau=1}^{n} K_h(x - X_{\tau\Delta})}, \end{aligned}$$

where

$$\begin{aligned} \widetilde{\mu}(x) &= \frac{1}{\Delta} \frac{\sum_{\tau=1}^{n-1} I(|X_{\tau\Delta} - x| \le b)\, (X_{(\tau+1)\Delta} - X_{\tau\Delta})}{\sum_{\tau=1}^{n} I(|X_{\tau\Delta} - x| \le b)}, \qquad \text{and} \\ \widetilde{\sigma}^2(x) &= \frac{1}{\Delta} \frac{\sum_{\tau=1}^{n-1} I(|X_{\tau\Delta} - x| \le b)\, (X_{(\tau+1)\Delta} - X_{\tau\Delta})^2}{\sum_{\tau=1}^{n} I(|X_{\tau\Delta} - x| \le b)}. \end{aligned}$$

See, also Bandi and Nguyen (2000a). Here, $b = b_n > 0$ is a bandwidth-like smoothing parameter that depends on the time span and on the sample size, which is called the spatial bandwidth in Bandi and Phillips (2003). This modeling approach is termed as the *chronological local time* estimation. Bandi and Philips's approach can deal well with the situation that the series is not stationary. The reader is referred to the papers by Bandi and Phillips (2003) and Bandi and Nguyen (2000a) for more discussions and asymptotic theory.

Bandi and Philips's (2003) estimator can be viewed as a two-step smoothing method: The first step defines straight sample analogs to the values that drift and diffusion take at the sampled points. Indeed, this step uses the smoothing technique (a linear estimator with same weights) to obtain the raw estimates of the two functions $\widetilde{\mu}(x)$ and $\widetilde{\sigma}^2(x)$, respectively. To implement this estimator, an empirical and theoretical study on the selection of two bandwidths b and h is needed.

2.3 Time-dependent diffusion models

The time-homogeneous diffusion models in (2.1) have certain limitations. For example, they cannot capture the time effect. A variety of time-dependent diffusion models have been proposed in the literature. A time-dependent diffusion process is

$$d\,X_t = \mu(X_t, t)\,dt + \sigma(X_t, t)\,dB_t. \tag{2.6}$$

Examples of (2.6) include Ho and Lee (HL) (1986), Hull and White (HW) (1990), Black, Derman and Toy (BDT) (1990), and Black and Karasinski (BK) (1991), among others. They consider respectively the following models:

$$\begin{aligned}
\text{HL:} \quad & d\,X_t &=& \quad \mu(t)\,dt + \sigma(t)\,d\,B_t, \\
\text{HW:} \quad & d\,X_t &=& \quad [\alpha_0 + \alpha_1(t)\,X_t]\,dt + \sigma(t)\,X_t^k\,d\,B_t, \qquad k = 0 \;\text{ or }\; 0.5 \\
\text{BDT:} \quad & d\,X_t &=& \quad [\alpha_1(t)\,X_t + \alpha_2(t)\,X_t\,\log(X_t)]\,dt + \sigma(t)\,X_t\,d\,B_t, \\
\text{BK:} \quad & d\,X_t &=& \quad [\alpha_1(t)\,X_t + \alpha_2(t)\,X_t\,\log(X_t)]\,dt + \sigma(t)\,X_t\,d\,B_t,
\end{aligned}$$

where $\alpha_2(t) = \sigma'(t)/\sigma(t)$. Similar to (2.2) and (2.3), one haw

$$\mu(X_t, t) = \lim_{\Delta\to 0} \frac{E\{X_{t+\Delta} - X_t \mid X_t\}}{\Delta} \quad \text{and} \quad \sigma^2(X_t, t) = \lim_{\Delta\to 0} \frac{E\left\{(X_{t+\Delta} - X_t)^2 \mid X_t\right\}}{\Delta},$$

which provide a regression form for estimating $\mu(\cdot, t)$ and $\sigma^2(\cdot, t)$.

Fan, Jiang, Zhang and Zhou (2001) consider the following time-varying coefficients single factor model

$$d\,X_t = [\alpha_0(t) + \alpha_1(t)\,X_t]\,dt + \beta_0(t)\,X_t^{\beta_1(t)}\,d\,B_t, \tag{2.7}$$

and use the local linear technique in (2.5) to estimate the coefficient functions $\{\alpha_j(\cdot)\}$ and $\{\beta_j(\cdot)\}$. Since the coefficients depend on time, $\{X_t\}$ might not be stationary. The asymptotic properties of the resulting estimators are still unknown. The aforementioned models are a special case of the following more general time-varying coefficient multi-factor diffusion models

$$d\,X_t = \mu(X_t,\,t)\,dt + \sigma(X_t,\,t)\,d\,B_t, \tag{2.8}$$

where

$$\mu(X_t,\,t) = \alpha_0(t) + \alpha_1(t)\,g(X_t), \quad \text{and} \quad (\sigma(X_t,\,t)\sigma(X_t,\,t)')_{ij} = \beta_{0,ij}(t) + \beta_{1,ij}(t)'\,h_{ij}(X_t),$$

and $g(\cdot)$ and $\{h_{ij}(\cdot)\}$ are known functions. This is the time-dependent version of the multi-factor affine models studied in Duffie, Pan and Singleton (2000). It allows time-varying coefficients in multi-factor affine models. A further theoretical and empirical study of the time-varying coefficient multi-factor diffusion model in (2.8) is warranted.

2.4 Jump diffusion models

There has been a vast literature on the study of diffusion models with jumps; see, for example, Pan (1997), Duffie and Pan (2001), Bollerslev and Zhou (1999), Eraker, Johannes and Polson (1999), Bates (2000), Duffie, Pan and Singleton (2000), Johannes (2000), Liu, Longstaff and Pan (2002), Zhou (2001), Singleton (2001), Perron (2001), Chernov, Gallant, Ghysels and Tauchen (2002). The main purpose of adding jumps into diffusion models or stochastic volatility diffusion models is to accommodate impact of sudden and large shocks to financial markets, such as macroeconomic announcements, the Asian and Russian finance crisis, an unusually large unemployment announcement, and a dramatic interest rate cut by the Federal Reserve. For more discussions on why it is necessary to add jumps into diffusion models, see, for example, Lobo (1999), Bollerslev and Zhou (1999), Liu, Longstaff and Pan (2002), and Johannes (2000), among others. Jumps can capture the heavy tail behavior of the distribution of the underlying process.

For the expositional purpose, we only consider a single factor diffusion model with jump:

$$d\,X_t = \mu(X_t)\,dt + \sigma(X_t)\,dB_t + dJ_t, \tag{2.9}$$

where J_t is a compensated jump process (zero conditional mean) with arrival rate $\lambda_t = \lambda(X_t) \geq 0$, which is an instantaneous intensity function, and the jump size, ξ, has a time-invariant distribution $\Pi(\cdot)$ with mean zero. There are several studies on specification of J_t. For example, $J_t = \xi\,P_t$, where P_t is a Poisson process with an intensity $\lambda(X_t)$ or a binomial distribution with probability $\lambda(X_t)$, and $\Pi(\cdot)$ can be either normal or uniform. If $\lambda_t(\cdot) = 0$ or $E(\xi^2) = 0$, the jump-diffusion model in (2.9) becomes the diffusion model in (2.1). More generally, Chernov, Gallant, Ghysels and Tauchen (2002) consider a Lévy process for $\{J_t\}$.

In practice, $\lambda(\cdot)$ might be assumed to have a particular form. For example, Chernov, Gallant, Ghysels and Tauchen (2002) consider three different types of special forms, each having the appealing feature of yielding analytic option pricing formula for European type contracts written on the stock price index. There are some open issues for the jump-diffusion model: (i) jumps are not observed and it is not possible to say surely if they exist; (ii) if they exist, a natural question arises how to estimate a jump time τ, which is defined to be the discontinuous time at which $X_{\tau+} \neq X_{\tau-}$, and the jump size ξ, which is $\xi = X_{\tau+} - X_{\tau-}$. Wavelet methods may be potentially useful here.

Similar to (2.2) and (2.3), the first two conditional moments are given by

$$\mu_1(X_t) = \lim_{\Delta \downarrow 0} \frac{E[\{X_{t+\Delta} - X_t\} \mid X_t]}{\Delta} = \mu(X_t) + \lambda(X_t)\, E(\xi),$$

and

$$\mu_2(X_t) = \lim_{\Delta \downarrow 0} \frac{E\left[\{X_{t+\Delta} - X_t\}^2 \mid X_t\right]}{\Delta} = \sigma^2(X_t) + \lambda(X_t)\, E(\xi^2).$$

This implies that the first two moments are the same as those for a diffusion model by using a new drift coefficient $\widetilde{\mu}(X_t) = \mu(X_t) + \lambda(X_t)\, E(\xi)$ and a new diffusion coefficient $\widetilde{\sigma}^2(x) = \sigma^2(x) + \lambda(x)\, E(\xi^2)$. However, the fundamental difference between a diffusion model and a diffusion model with jumps relies on higher order moments. Using the infinitesimal generator (Øksendal, 1985; Karatzas and Shreve, 1988) of X_t, we can compute, $j > 2$,

$$\mu_j(X_t) = \lim_{\Delta \to 0} \frac{E[(X_{t+\Delta} - X_t)^j \mid X_t]}{\Delta} = \lambda(X_t)\, E(\xi^j).$$

See Duffie, Pan and Singleton (2000) and Johannes (2000) for details. Obviouslyt, jumps provide a simple and intuitive mechanism for capturing the heavy tail behavior of interest rates. In particular, the conditional skewness and kurtosis are given by

$$\begin{aligned} s(X_t) &\equiv \frac{\lambda(X_t)\, E(\xi^3)}{\left[\sigma^2(X_t) + \lambda(X_t)\, E(\xi^2)\right]^{3/2}}, \\ k(X_t) &\equiv \frac{\lambda(X_t)\, E(\xi^4)}{\left[\sigma^2(X_t) + \lambda(X_t)\, E(\xi^2)\right]^{2}}. \end{aligned}$$

Note that $s(X_t) = 0$ if ξ is symmetric. By assuming $\xi \sim N(0, \sigma_\xi^2)$, Johannes (2000) uses the conditional kurtosis to measure the departures for the treasury bill data from normality and concludes that interest rates exchanges are extremely non-normal.

The NW estimation of $\mu_j(\cdot)$ is considered by Johannes (2000) and Bandi and Nguyen (2000b). Moreover, Bandi and Nguyen (2000b) provide a general asymptotic theory for the resulting estimators. Further, by specifying a particular form of $\Pi(\lambda) = \Pi_0(\lambda, \theta)$, say, $\xi \sim N(0,\ \sigma_\xi^2)$, Bandi and Nguyen (2000b) propose consistent estimators of $\lambda(\cdot)$, σ_ξ^2, and $\sigma^2(\cdot)$ and derive their asymptotic properties.

A natural question arises how to measure the departures from a pure diffusion model statistically. That is to test the model (2.9) against the model (2.1). It is equivalent to checking whether $\lambda(\cdot) \equiv 0$ or $\xi = 0$. Instead of using the conditional skewness or kurtosis, a test statistic can be constructed based on the higher order conditional moments. For example, one can construct the following nonparametric test statistics

$$T_1 = \int \widehat{\mu}_4(x)\, w(x)\, dx, \quad \text{or} \quad T_2 = \int \widehat{\mu}_3^2(x)\, w(x)\, dx,$$

where $w(\cdot)$ is a weighting function. This is under investigation theoretically and empirically.

More generally, for given a discrete sample of a diffusion process, can one tell whether the underlying model that gave rise to the data was a diffusion, or should jumps be allowed into the model? To answer this question, Ait-Sahalia (2002b) proposes an approach to identifying the sufficient and necessary restriction on the transition densities of diffusions, at the sampling interval of the observed data. This restriction characterizes the continuity of the unobservable continuous sample path of the underlying process and is valid for every sampling interval including long ones. Let $\{X_t,\, t \geq 0\}$ be a Markovian process taking values in $D \subseteq \Re$. Let $p(\Delta,\, y \,|\, x)$ denote the transition density function of the process over interval length Δ, that is, the conditional density of $X_{t+\Delta} = y$ given $X_t = x$, and it is assumed that the transition densities are time homogenous. Ait-Sahalia (2002b) shows that if the transition density $p(\Delta,\, y \,|\, x)$ is strictly positive and twice-continuously differentiable on $D \times D$ and the following condition

$$\frac{\partial^2}{\partial x\, \partial y} \ln p(\Delta,\, y \,|\, x) > 0 \quad \text{for all} \quad \Delta > 0 \quad \text{and} \quad (x,\, y) \in D \times D,$$

which is the so called "diffusion criterion" in Ait-Sahalia (2002b), is satisfied, then, the underlying process is a diffusion. From a discretely sampled time series $\{X_{\tau\Delta}\}$, once could test nonparametrically the hypothesis that the data were generated by a continuous-time diffusion $\{X_t\}$. That is to test nonparametrically the null hypothesis

$$\mathbb{H}_0 : \frac{\partial^2}{\partial x\, \partial y} \ln p(\Delta,\, y \,|\, x) > 0 \quad \text{for all} \quad x,\, y$$

versus the alternative

$$\mathbb{H}_A : \frac{\partial^2}{\partial x\, \partial y} \ln p(\Delta,\, y \,|\, x) \leq 0 \quad \text{for some} \quad x,\, y.$$

One could construct a test statistic based on checking whether the above "diffusion criterion" holds for a nonparametric estimator of $p(\Delta,\, y \,|\, x)$. This topic is still open.

2.5 Time-dependent jump diffusion models

Duffie, Pan and Singleton (2000) consider the time-varying coefficient intensity

$$\lambda(X_t,\, t) = \lambda_0(t) + \lambda_1(t)\, X_t,$$

and Chernov, Gallant, Ghysels and Tauchen (2002) consider a more general stochastic volatility model with the stochastic intensity,

$$\lambda(\xi_0,\, X_t,\, t) = \lambda_0(\xi_0,\, t) + \lambda_1(\xi_0,\, t)\, X_t,$$

where ξ_0 is the size of the previous jump. This specification yields a class of jump Lévy measures which combine the features of jump intensities depending on, say volatility, as well as the size of the previous jump. Johannes, Kumar and Polson (1999) also propose a class of jump diffusion processes with a jump intensity depending on the past jump time and the absolute return. Moreover, as pointed out by Chernov, Gallant, Ghysels and Tauchen (2002), another potentially very useful specification of the intensity function would include the past duration, i.e., the time since the last jump, say $\tau(t)$, which is the time that has elapsed between the last jump and t where $\tau(t)$ is a continuous function of t, such as

$$\lambda(\xi_0,\, X_t,\, \tau,\, t) = \{\lambda_0(t) + \lambda_1(t)\, X_t\}\, \lambda\{\tau(t)\}\, \exp\{G(\xi_0)\}, \tag{2.10}$$

which can accommodate the increasing, decreasing or hump-shaped hazard functions of the size of the previous jump, and the duration dependence of jump intensities. However, to the best of our knowledge, there have not been any attempt in the literature to discuss the estimation and test of the intensity function $\lambda(\cdot)$ nonparametrically in the above settings.

A natural question arises is how to generalize the model (2.9) economically and statistically to a more general time-dependent jump diffusion model

$$d\, X_t = \mu(X_t,\, t)\, dt + \sigma(X_t,\, t)\, dB_t + dJ_t$$

with the time-dependent intensity function $\lambda(\xi_0,\, X_t,\, \tau,\, t)$ without any specified form or with some nonparametric structure, say, like (2.10). Clearly, they include the aforementioned models as a special case, which are studied by Duffie, Pan and Singleton (2000), Johannes, Kumar and Polson (1999), and Chernov, Gallant, Ghysels and Tauchen (2002), among others. This is still an open problem.

3 Nonparametric Estimation and Testing of Parametric Diffusions

3.1 Nonparametric estimation of parametric diffusion models

As is well-known, derivative pricing in mathematical finance is generally much more tractable in a continuous-time modelling framework than through binomial or other discrete approximations. In the empirical literature, however, it is an usual practice to abandon continuous-time modeling when estimating derivative pricing models. This is mainly due to the difficulty that the transition density for most continuous-time models with discrete observations has no closed form and therefore the maximum likelihood estimation (MLE) is infeasible.

One major focus of the continuous-time literature is on developing econometric methods to estimate continuous-time models using discretely-sampled data.[3] This is largely motivated by the fact that using the discrete version of a continuous-time model can result in inconsistent parameter estimates (Lo, 1988). Available estimation procedures include the MLE method of Lo (1988), the simulated methods of moments of Duffie and Singleton (1993) and Gourieroux, Monfort and Renault (1993), the generalized method of moments (GMM) of Hansen and Scheinkman (1995), the efficient method of moments (EMM) of Gallant and Tauchen (1996), the Markov chain Monte Carlo (MCMC) of Jacquier, Polson and Rossi (1994), Eraker (1998) and Jones (1998), and the methods based on the empirical characteristic function of Jiang and Knight (2001) and Singleton (2001).

Below we focus on nonparametric estimation of a parametric continuous-time model

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

where $\mu(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ are known functions, and θ is unknown parameter vector in an open bounded parameter space Θ. Ait-Sahalia (1996b) proposes a minimum distance estimator

$$\widehat{\theta} = \arg\min_{\theta\in\Theta}\; n^{-1}\; \sum_{\tau=1}^{n} \left[\widehat{\pi}_0(X_{\tau\Delta}) - \pi(X_{\tau\Delta},\theta)\right]^2, \tag{3.2}$$

where

$$\widehat{\pi}_0(x) = n^{-1}\sum_{\tau=1}^{n} K_h(x - X_{\tau\Delta})$$

is a kernel estimator for the stationary density of X_t, and

$$\pi\,(x,\theta) = \frac{c(\theta)}{\sigma^2(x,\theta)}\exp\left\{\int_{x_0^*}^{x}\frac{2\,\mu(u,\theta)}{\sigma^2(u,\theta)}du\right\}, \tag{3.3}$$

is the marginal density estimator implied by the diffusion model, where the standardization factor $c(\theta)$ ensures that $\pi(\cdot,\theta)$ integrates to 1 for every $\theta\in\Theta$, and x_0^* is the lower bound of the support of X_t. Because the marginal density cannot capture the full dynamics of the diffusion process, one can expect that $\widehat{\theta}$ will not be asymptotically most efficient, although it is root-n consistent for θ_0.

Let $p_x(\Delta, x\,|\,x_0,\theta)$ be the conditional density function of $X_{\tau\Delta} = x$ given $X_{(\tau-1)\Delta} = x_0$ induced by model (3.1). The log-likelihood function of the model for the sample is

$$l_n(\theta) = \sum_{\tau=1}^{n} \ln p_x(\Delta, X_{\tau\Delta}\,|\,X_{(\tau-1)\Delta},\theta).$$

[3] Sundaresan (2001) states that "perhaps the most significant development in the continuous-time field during the last decade has been the innovations in econometric theory and in the estimation techniques for models in continuous time." For other reviews of the recent literature, see, e.g., Melino (1994), Tauchen (1997, 2001), and Campbell, Lo and MacKinlay (1997).

The MLE estimator that maximizes $l_n(\theta)$ would be asymptotically most efficient if the conditional density $p_x(\Delta, x|x_0, \theta)$ has a closed form. Unfortunately, except for some simple models, $p_x(\Delta, x|x_0, \theta)$ usually does not have a closed form.

Using the Hermite polynomial series, Ait-Sahalia (2002a) proposes a closed form sequence $\{p_x^{(J)}(\Delta, x|x_0, \theta)\}$ to approximate $p_x(\Delta, x|x_0, \theta)$, and then obtains an estimator $\widehat{\theta}_n^{(J)}$ that maximizes the approximated model likelihood. The estimator $\widehat{\theta}_n^{(J)}$ enjoys the same asymptotic efficiency as the (infeasible) MLE as $J = J_n \to \infty$. More specifically, Ait-Sahalia (2002a) first considers a transformed process

$$Y_t \equiv \gamma(X_t, \theta) = \int_{-\infty}^{X_t} \frac{1}{\sigma(u, \theta)} du.$$

This transformed process obeys the following diffusion

$$dY_t = \mu_y(Y_t, \theta)dt + dB_t,$$

where

$$\mu_y(y, \theta) = \frac{\mu[\gamma^{-1}(y, \theta), \theta]}{\sigma[\gamma^{-1}(y, \theta), \theta]} - \frac{1}{2}\frac{\partial \sigma[\gamma^{-1}(y, \theta), \theta]}{\partial x}.$$

The transform $X \to Y$ ensures that the tail of the transition density $p_y(\Delta, y|y_0, \theta)$ of Y_t will generally vanish exponentially fast so that Hermite series approximations will converge. However, $p_y(\Delta, y|y_0, \theta)$ may get peaked at y_0 when the sample frequency Δ gets smaller. To avoid this, Ait-Sahalia (2002a) considers a further transform

$$Z_t = \Delta^{-1/2}(Y_t - y_0)$$

and then approximates the transition density of Z_t by the Hermite polynomials:

$$p_z^{(J)}(z \mid y_0, \theta) = \phi(z) \sum_{j=0}^{J} \eta_z^{(j)}(y_0, \theta) H_j(z),$$

where $\phi(\cdot)$ is the $N(0,\ 1)$ density, and $\{H_j(z)\}$ is the Hermite polynomial series. The coefficients $\{\eta_z^{(j)}(y_0, \theta\}$ are specific conditional moments of process Z_t, and can be explicitly computed using the Monte Carlo method or using a higher Taylor series expansion in Δ.

The approximated transition density of X_t is then given as follows:

$$\begin{aligned} p_x(x \mid x_0, \theta) &= \sigma(x, \theta)^{-1} p_y(\gamma(x, \theta) \mid \gamma(x, \theta), \theta) \\ &= \Delta^{-1/2} p_z(\Delta^{-1/2}(\gamma(x, \theta) - \gamma(x_0, \theta)) \mid \gamma(x_0, \theta), \theta). \end{aligned}$$

Under suitable regularity conditions, particularly when $J = J_n \to \infty$ as $n \to \infty$, the estimator

$$\widehat{\theta}_n = \arg\min_{\theta \in \Theta} \sum_{\tau=1}^{n} \ln p_x^{(J)}(X_{\tau\Delta}|X_{(\tau-1)\Delta}, \theta)$$

will be asymptotically equivalent to the infeasible MLE. Ait-Sahalia (1999) applies this method to estimate a variety of diffusion models for spot interest rates, and finds that $J = 2$ or 3 already gives accurate approximation for most financial diffusion models. Egorov, Li and Xu (2003) extend this approach to stationary time-inhomogeneous diffusion models. Ait-Sahalia (2002c) extend this method to general multivariate diffusion models and Ait-Sahalia and Kimmel (2002) to affine multi-factor term structure models.

In a rather general continuous-time setup which allows for stationary multi-factor diffusion models with partially observable state variables, Gallant and Tauchen (1996) propose an efficient method of moment estimator that also enjoys the asymptotic efficiency as the MLE. The basic idea of EMM is to first use a Hermite-polynomial based semi-nonparametric (SNP) density estimator to approximate the transition density of the observed state variables. This is called the auxiliary model and its score is called the score generator, which has expectation zero under the model-implied distribution when the parametric model is correctly specified. Then, given a parameter setting for the multi-factor model, one may use simulation to evaluate the expectation of the score under the stationary density of the model and compute a chi-square criterion function. A nonlinear optimizer is used to find the parameter values that minimize the proposed criterion.

Specifically, suppose $\{X_t\}$ is a stationary possibly vector-valued process such that the true conditional density function $p_0(\Delta, X_{\tau\Delta}|\, X_{s\Delta},\, s \le \tau - 1) = p_0(\Delta, X_{\tau\Delta}|Y_{\tau\Delta})$, where $Y_{\tau\Delta} \equiv (X_{(\tau-1)\Delta}, \ldots, X_{(\tau-d)\Delta})'$ for some fixed integer $d \ge 0$. This is a Markovian process of order d. To check the adequacy of a parametric model in (3.1), Gallant and Tauchen (1996) propose to check whether the following moment condition holds:

$$M(\beta_n, \theta) \equiv \int \frac{\partial \log f(\Delta, x, y; \beta_n)}{\partial \beta_n} p(\Delta, x, y; \theta) dx dy = 0, \quad \text{if} \quad \theta = \theta_0 \in \Theta, \tag{3.4}$$

where $p(\Delta, x, y; \theta)$ is the model-implied joint density for $(X_{\tau\Delta}, Y'_{\tau\Delta})'$, θ_0 is the unknown true parameter value, and $f(\Delta, x, y; \beta_n)$ is an auxiliary model for the conditional density of $(X_{\tau\Delta}, Y'_{\tau\Delta})'$. Note that β_n is the parameter vector in the SNP density model $f(\Delta, x, y; \beta_n)$ and generally does not nest the parametric parameter θ. By allowing the dimension of β_n to grow with the sample size n, the SNP density $f(\Delta, x, y; \beta_n)$ will eventually span the true density $p_0(\Delta, x, y)$ of $(X_{\tau\Delta}, Y'_{\tau\Delta})'$, and thus is free of model misspecification asymptotically. Gallant and Tauchen (1996) use a Hermite polynomial approximation for $f(\Delta, x, y; \beta_n)$, with the dimension of β_n determined by such model selection criteria as BIC. The integration in (3.4) can be computed by simulating a large number of realizations under the distribution of the parametric model $p(\Delta, x, y; \theta)$.

The efficient method of moment estimator is defined as follows:

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} M(\widehat{\beta}_n, \theta)' \, \widehat{I}^{-1}(\theta) \, M(\widehat{\beta}_n, \theta),$$

where $\widehat{\beta}$ is the quasi-MLE for β_n, the coefficients in the Hermite polynomial expansion of the SNP density model $f(x, y, \beta_n)$ and the matrix $\widehat{I}(\theta)$ is an estimate of the asymptotic variance of $\sqrt{n}\partial M_n(\widehat{\beta}_n, \theta)/\partial\theta$ (Gallant and Tauchen, 2001). This estimator $\widehat{\theta}$ is asymptotically as efficient as the (infeasible) MLE.

The EMM has been applied widely in financial applications. See, for example, Anderson and Lund (1997), Dai and Singleton (2000), Ahn, Dittmar and Gallant (2002) for interest rate applications, Liu (2000), Anderson, Berzoni and Lund (2002), Chernov, Gallant, Ghysels and Tauchen (2001) for estimating stochastic volatility models for stock prices with such complications as long memory and jumps, Chung and Tauchen (2001) for estimating and testing target zero models of exchange rates, Jiang and van der Sluis (2000) for price option pricing, and Valderamma (2001) for a macroeconomic application. It would be interesting to compare the EMM method and Ait-Sahalia's (2002a) approximate MLE in finite sample performance.

3.2 Nonparametric testing of diffusion models

In financial applications, most continuous-time models are parametric. It is important to test whether a parametric diffusion model adequately captures the dynamics of the underlying process. Model misspecification generally renders inconsistent estimators of model parameters and their variance-covariance matrix, leading to misleading conclusions in inference and hypothesis testing. More importantly, a misspecified model can yield large errors in hedging, pricing and risk management.

Unlike the vast literature of estimation of parametric diffusion models, there are relatively few test procedures for parametric diffusion models using discrete observations. Suppose $\{X_t\}$ follows a continuous-time diffusion process in (2.6). Often it is assumed that the drift and diffusion $\mu(\cdot, t)$ and $\sigma(\cdot, t)$ have some parametric forms $\mu(\cdot, t, \theta)$ and $\sigma(\cdot, t, \theta)$, where $\theta \in \Theta$. We say that models $\mu(\cdot, t, \theta)$ and $\sigma(\cdot, t, \theta)$ are correctly specified for the drift and diffusion $\mu(\cdot, t)$ and $\sigma(\cdot, t)$ respectively if

$$\mathbb{H}_0 : P\left[\mu(X_t, t, \theta_0) = \mu(X_t, t), \sigma(X_t, t, \theta_0) = \sigma(X_t, t)\right] = 1 \quad \text{for some} \quad \theta_0 \in \Theta. \tag{3.5}$$

As noted earlier, warious methods have been developed to estimate θ_0, taking (3.5) as given. However, these methods generally cannot deliver consistent parameter estimates if $\mu(\cdot, t, \theta)$ or $\sigma(\cdot, t, \theta)$ is misspecified in the sense that

$$\mathbb{H}_A : P\left[\mu(X_t, t, \theta) = \mu(X_t, t), \sigma(X_t, t, \theta) = \sigma(X_t, t)\right] < 1 \quad \text{for all} \quad \theta \in \Theta. \tag{3.6}$$

Under $\mathbb{H}_A$ of (3.6), there exists no parameter value $\theta \in \Theta$ such that the drift model $\mu(\cdot, t, \theta)$ and the diffusion model $\sigma(\cdot, t, \theta)$ coincide with the true drift $\mu(\cdot, t)$ and the true diffusion $\sigma(\cdot, t)$ respectively.

There is a growing interest in testing whether a continuous-time model is correctly specified using a discrete sample $\{X_{\tau\Delta}\}_{\tau=1}^{n}$. Ait-Sahalia (1996b) observes that for a stationary time-homogeneous diffusion process in (3.1), a pair of drift and diffusion models $\mu(\cdot,\theta)$ and $\sigma(\cdot,\theta)$ uniquely determines the stationary density $\pi(\cdot,\theta)$ in (3.3). Ait-Sahalia (1996b) compares a parametric marginal density estimator $\pi(\cdot,\widehat{\theta})$ with a nonparametric density estimator $\widehat{\pi}_0(\cdot)$ via the quadratic form

$$M \equiv \int_{x_0^*}^{x_1^*} \left[\widehat{\pi}_0(x) - \pi(x,\widehat{\theta})\right]^2 \widehat{\pi}_0(x)dx, \tag{3.7}$$

where x_1^* is the upper bound for X_t, $\widehat{\theta}$ is the minimum distance estimator given by (3.2). The M statistic, after demeaning and scaling, is asymptotically normal under $\mathbb{H}_0$.

The M test makes no restrictive assumptions on the data generating process and can detect a wide range of alternatives. This appealing power property is not shared by parametric approaches such as generalized method of moment tests (e.g., Conley *et al.* 1997). The latter has optimal power against certain alternatives (depending on the choice of moment functions) but may be completely silent against other alternatives. In an application to Euro-dollar interest rates, Ait-Sahalia (1996b) rejects all existing one-factor linear drift models using asymptotic theory and finds that "the principal source of rejection of existing models is the strong nonlinearity of the drift," which is further supported by Stanton (1997).

However, several limitations of this test may hinder its empirical applicability. First, as Ait-Sahalia (1996b) has pointed out, the marginal density cannot capture the full dynamics of $\{X_t\}$. It cannot distinguish two diffusion models that have the same marginal density but different transition densities.[4] Second, subject to some regularity conditions, the asymptotic distribution of the quadratic form M in (3.7) remains the same whether the sample $\{X_{\tau\Delta}\}_{\tau=1}^{n}$ is iid or highly persistently dependent (Ait-Sahalia, 1996b). This convenient asymptotic property unfortunately results in a substantial discrepancy between the asymptotic and finite sample distributions, particularly when data display persistent dependence (Pritsker, 1998). This discrepancy and the slow convergence of kernel estimators are the main reasons identified by Pritsker (1998) for the poor finite sample performance of the M test. They cast some doubt on the applicability of first order asymptotic theory of nonparametric methods in finance, since persistent serial dependence is a stylized fact for interest rates and many other high frequency financial data. Third, a kernel density estimator produces biased estimates near the boundaries of the data (e.g., Härdle, 1990 and Fan and Gijbels, 1996). In the present context, the boundary bias can generate spurious nonlinear drifts, giving misleading conclusions on the dynamics of $\{X_t\}$.

[4] A simple example is the Vasicek model, where if we vary the speed of mean reversion and the scale of diffusion in the same proportion, the marginal density will remain unchanged, but the transition density will be different.

Recently, Hong and Li (2002) have developed a nonparametric test for the model in (2.6) using the transition density, which can capture the full dynamics of $\{X_t\}$ in (3.1). Let $p_x(x, t \,|\, x_0, s)$ be the true transition density of the diffusion process X_t; that is, the conditional density of $X_t = x$ given $X_s = x_0$, $s < t$. For a given pair of drift and diffusion models $\mu(\cdot, t, \theta)$ and $\sigma(\cdot, t, \theta)$, a certain family of transition densities $\{p(x, t \,|\, x_0, s, \theta)\}$ is characterized. When (and only when) $\mathbb{H}_0$ in (3.5) holds, there exists some $\theta_0 \in \Theta$ such that $p(x, t \,|\, x_0, s, \theta_0) = p_0(x, t \,|\, x_0, s)$ almost everywhere for all $t > s$. Hence, the hypotheses of interest $\mathbb{H}_0$ in (3.5) versus $\mathbb{H}_A$ in (3.6) can be equivalently written as follows:

$$\mathbb{H}_0 : p(x, t \,|\, y, s, \theta_0) = p_0(x, t \,|\, y, s) \;\text{ almost everywhere for some }\; \theta_0 \in \Theta \tag{3.8}$$

versus the alternative hypothesis

$$\mathbb{H}_A : p(x, t \,|\, y, s, \theta) \neq p_0(x, t \,|\, y, s) \;\text{ for some }\; t > s \;\text{ and for all }\; \theta \in \Theta. \tag{3.9}$$

A natural approach to testing $\mathbb{H}_0$ in (3.8) versus $\mathbb{H}_A$ in (3.9) would be to compare a model transition density estimator $p(x, t \,|\, x_0, s, \widehat{\theta})$ with a nonparametric transition density estimator, say $\widehat{p}_0(x, t \,|\, x_0, s)$. Instead of comparing $p(x, t \,|\, x_0, s, \widehat{\theta})$ and $\widehat{p}_0(x, t \,|\, x_0, s)$ directly, Hong and Li (2002) first transform $\{X_{\tau\Delta}\}_{\tau=1}^{n}$ via a probability integral transform. Define a discrete transformed sequence

$$Z_\tau(\theta) \equiv \int_{-\infty}^{X_{\tau\Delta}} p\left[x, \tau\Delta | X_{(\tau-1)\Delta}, (\tau-1)\Delta, \theta\right] dx, \quad \tau = 1, \ldots, n. \tag{3.10}$$

Under (and only under) $\mathbb{H}_0$ in (3.8) there exists some $\theta_0 \in \Theta$ such that $p[x, \tau\Delta | X_{(\tau-1)\Delta}, (\tau - 1)\Delta, \boldsymbol{\theta}_0] = p_0[x, \tau\Delta | X_{(\tau-1)\Delta}, (\tau-1)\Delta]$ almost surely for all $\Delta > 0$. Consequently, the transformed series $\{Z_\tau \equiv Z_\tau(\theta_0)\}_{\tau=1}^{n}$ is iid $U[0,1]$ under $\mathbb{H}_0$ in (3.8). This result is first proven, in a simpler context, by Rosenblatt (1952), and is more recently used to evaluate out-of-sample density forecasts (e.g., Diebold, Gunther and Tay, 1998) in a discrete-time context. Intuitively, we may call $\{Z_\tau(\theta)\}$ "generalized residuals" of the model $p(x, t \,|\, y, s, \boldsymbol{\theta})$.

To test $\mathbb{H}_0$ in (3.8), Hong and Li (2002) check whether $\{Z_\tau\}_{\tau=1}^{n}$ is both iid and $U[0,1]$. They compare a kernel estimator $\widehat{g}_j(z_1, z_2)$ defined in (3.11) below for the joint density of $\{Z_\tau, Z_{\tau-j}\}$ with unity, the product of two $U[0,1]$ densities. This approach has at least three advantages. First, since there is no serial dependence in $\{Z_\tau\}$ under $\mathbb{H}_0$ in (3.8), nonparametric joint density estimators are expected to perform much better in finite samples. In particular, the finite sample distribution of the resulting tests is expected to be robust to persistent dependence in data. Second, there is no asymptotic bias for nonparametric density estimators under $\mathbb{H}_0$ in (3.8). Third, no matter whether $\{X_t\}$ is time-inhomogeneous or even nonstationary, $\{Z_\tau\}$ is always iid $U[0,1]$ under correct model specification.

Hong and Li (2002) employ the kernel joint density estimator,

$$\widehat{g}_j(z_1, z_2) \equiv (n-j)^{-1} \sum_{\tau=j+1}^{n} K_h(z_1, \widehat{Z}_\tau) K_h(z_2, \widehat{Z}_{\tau-j}), \qquad j > 0, \tag{3.11}$$

where $\widehat{Z}_\tau = Z_\tau(\widehat{\theta})$, $\widehat{\theta}$ is any $\sqrt{n}$-consistent estimator for θ_0, and for $x \in [0,1]$,

$$K_h(x,y) \equiv \begin{cases} h^{-1}k\left(\frac{x-y}{h}\right) / \int_{-(x/h)}^{1} k(u)du, & \text{if } \; x \in [0,h), \\ h^{-1}k\left(\frac{x-y}{h}\right), & \text{if } \; x \in [h, 1-h], \\ h^{-1}k\left(\frac{x-y}{h}\right) / \int_{-1}^{(1-x)/h} k(u)du, & \text{if } \; x \in (1-h, 1] \end{cases}$$

is the kernel with boundary correction (Rice, 1986) and $k(\cdot)$ is a standard kernel. This avoids the boundary bias problem, and has some advantages over some alternative methods such as trimming and the use of the jackknife kernel.[5] To avoid the boundary bias problem, one might apply other kernel smoothing methods such as local polynomial (Fan and Gijbels, 1996) or weighted NW (Cai, 2001).

Hong and Li's (2002) test statistic is

$$\widehat{Q}(j) \equiv \left[(n-j)h \int_0^1 \int_0^1 \left[\widehat{g}_j(z_1, z_2) - 1\right]^2 dz_1 dz_2 - A_h^0\right] / V_0^{1/2},$$

where A_h^0 and V_0 are non-stochastic centering and scale factors that depends on h and $k(\cdot)$.

In a simulation experiment mimicking the dynamics of U.S. interest rates via the Vasicek model, Hong and Li (2002) find that $\widehat{Q}(j)$ has rather reasonable sizes for $n = 500$ (i.e., about two years of daily data). This is a rather substantial improvement over Ait-Sahalia's (1996b) test, in lights of Pritsker's (1998) simulation evidence. Moreover, $\widehat{Q}(j)$ has better power than the marginal density test. Hong and Li (2002) find extremely strong evidence against a variety of existing one-factor diffusion models for the spot interest rate and affine models for interest rate term structures.

Because the transition density of a continuous-time model generally has no closed form, the probability integral transform $\{Z_\tau(\theta)\}$ in (3.10) is difficult to compute. However, one can approximate the model transition density using the simulation methods developed by (e.g.)

[5] One could simply ignore the data in the boundary regions and only use the data in the interior region. Such a trimming procedure is simple, but in the present context, it would lead to the loss of significant amount of information. If $h = sn^{-\frac{1}{5}}$ where $s^2 = \text{var}(X_t)$, for example, then about 23, 20and 10of a uniformly distributed sample will fall into the boundary regions when $n = 100, 500$ and $5,000$ respectively. For financial time series, one may be particularly interested in the tail distribution of the underlying process, which is exactly contained in (and only in) the boundary regions!

Another solution is to use a kernel that adapts to the boundary regions and can effectively eliminate the boundary bias. One example is the so-called jackknife kernel, as used in Chapman and Pearson (2000). In the present context, the jackknife kernel, however, has some undesired features in finite samples. For example, it may generate negative density estimates in the boundary regions because the jackknife kernel can be negative in these regions. It also induces a relatively large variance for the kernel estimates in the boundary regions, adversely affecting the power of the test in finite samples.

Pedersen (1995), Brandt and Santa-Clara (2001), and Elerian, Chib and Shephard (2001). Alternatively, we can use Ait-Sahalia's (2002a) Hermite expansion method to construct a closed-form approximation of the model transition density.

When a misspecified model is rejected, one may like to explore what are possible sources for the rejection. For example, is the rejection due to misspecification in the drift, such as the ignorance of mean shifts or jumps? Is it due to the ignorance of GARCH effects or stochastic volatility? Or is it due to the ignorance of asymmetric behaviors (e.g., leverage effects)? Hong and Li (2002) consider to examine the autocorrelations in the various powers of $\{Z_\tau\}$, which are very informative about how well a model fits various dynamic aspects of the underlying process (e.g., conditional mean, variance, skewness, kurtosis, ARCH-in-mean effect, and leverage effect).

Gallant and Tauchen (1996) also propose an EMM-based minimum chi-square specification test for stationary continuous-time models. They examine the simulation-based expectation of an auxiliary SNP score function under the model distribution, which is zero under correct model specification. The greatest appeal of the EMM approach is that it applies to a wide range of stationary continuous-time processes, including both one-factor and multi-factor diffusion processes with partially observable state variables (e.g., stochastic volatility models). In addition to the minimum chi-square test for generic model misspecifications, the EMM approach also provides a class of individual t-statistics that are informative in revealing possible sources of model misspecification. This is perhaps the most appealing strength of the EMM approach.

Another feature of the EMM tests is that all EMM test statistics avoid estimating long-run variance-covariances, thus resulting in reasonable finite sample size performance (cf. Anderson, Chung and Sorensen, 1999). In practice, however, it may not be easy to find an adequate SNP density model for financial time series. For example, Anderson and Lund (1997) find that an AR(1)-EGARCH model with a number of Hermite polynomials adequately captures the full dynamics of daily S&P 500 return series, using a BIC criterion. However, Hong and Lee (2003) find that there still exists strong evidence on serial dependence in the standardized residuals of the model, indicating that the auxiliary SNP model is inadequate. This affects the validity of the EMM tests, because their asymptotic variance estimators have exploit the correct specification of the SNP density model. [6]

There has been also an interest in separately testing the drift model and the diffusion model in (3.1). For example, it has been controversial whether the drift of interest rates is

[6] Chen, Gao and Li (2001) consider kernel-based simultaneous specification testing for both mean and variance models in a discrete-time setup with dependent observations. The empirical likelihood principle is used to construct the test statistic. They apply the test to check adequacy of a discrete version of a continuous-time diffusion model.

linear. To test the linearity of the drift term, one can write it as a functional coefficient form (Cai, Fan and Yao, 2000) $\mu\left(X_t\right) = \alpha_0(X_t) + \alpha_1\left(X_t\right) X_t$. Then, the null hypothesis is $\mathbb{H}_0:\ \alpha_0(\cdot) \equiv \alpha_0$ and $\alpha_1(\cdot) \equiv \alpha_1$. Fan and Zhang (2001) apply the generalized likelihood ratio test developed by Fan, Zhang and Zhang (2001). They find that $\mathbb{H}_0$ is not rejected for the short-term interest rates. It is noted that the generalized likelihood ratio test is developed for the iid samples but it is still unknown whether it is valid for a time series context. Fan and Zhang (2001) and Fan, Jiang, Zhang and Zhou (2001) conjecture that it would hold based on their simulations. On the other hand, Chen, Härdle and Kleinow (2001) consider an empirical likelihood goodness-of-ft test for time series regression model, and they apply the test to test a discretized drift model of a diffusion process.

There has been also interest in testing the diffusion model $\sigma(\cdot, \theta)$. The motivation comes from the fact that derivative pricing with an underlying equity process only depends on the diffusion $\sigma(\cdot)$, which is one of the most important features of (3.1) for derivative pricing. Kleisten (2002) recently proposes a nonparametric test for a diffusion model $\sigma(\cdot)$. More specifically, Kleisten (2002) compares a nonparametric diffusion estimator $\widehat{\sigma}^2(\cdot)$ with a parametric diffusion estimator $\sigma^2(\cdot, \theta)$ via an asymptotically χ^2 test statistic

$$\widehat{T}_k = \sum_{l=1}^{k} \left[\widehat{T}(x_l)\right]^2,$$

where

$$\widehat{T}(x) = [nh\widehat{\pi}(x)]^{1/2} \left[\frac{\widehat{\sigma}^2(x)}{\tilde{\sigma}^2(x, \widehat{\theta})} - 1\right],$$

$\widehat{\theta}$ is an $\sqrt{n}$-consistent estimator for θ_0 and

$$\tilde{\sigma}^2(x, \theta) = \frac{1}{nh\widehat{\pi}(x)} \sum_{t=1}^{n} \sigma^2(x, \widehat{\theta}) K_h[(x - X_t)/h]$$

is a smooth version of $\sigma^2(x, \theta)$. The use of $\tilde{\sigma}^2(x, \widehat{\theta})$ instead of $\sigma^2(x, \widehat{\theta})$ directly reduces the kernel estimation bias in $\widehat{T}(x)$, thus allowing the use of the optimal bandwidth h for $\widehat{\sigma}^2(x)$. This device is also used in Härdle and Mammen (1993) in testing a parametric regression model. Kleinow (2002) finds that the empirical level of $\widehat{T}_k$ is too large relative to the significance level in finite samples and then proposes a modified test statistic using the empirical likelihood approach, which endogenously studentizes conditional heteroscedasticity. As expected, the empirical level of the modified test improves in finite samples, though not necessarily for the power of the test.

Finally, Fan, Jiang, Zhang and Zhang (2001) test whether the coefficients in the time-varying coefficient single factor diffusion model of (2.7) are really time-varying. Specially, they apply the generalized likelihood ratio test to check whether some or all of $\{\alpha_j(\cdot)\}$ and $\{\beta_j(\cdot)\}$ are constant.

4 Derivative Pricing and Risk Neural Density Estimation

4.1 Risk neutral density

In modern finance, the pricing of contingent claims is important given the phenomenal growth in turnover and volume of financial derivatives over the past decades. Derivative pricing formulas are highly nonlinear even when they are available in closed form. Nonparametric techniques are expected to be very useful in this area. In a standard dynamic exchange economy, the equilibrium price of a security at date t with a single liquidating payoff $Y(C_T)$ at date T that is a function of aggregate consumption C_T is given by

$$P_t = E_t\left[Y(C_T)M_{t,T}\right], \tag{4.1}$$

where the conditional expectation is taken with respect to the information set available to the representative economic agent at time t, $M_{t,T} = \delta^{T-1}U'(C_T)/U'(C_t)$, the so-called stochastic discount factor, is the marginal rate of substitution between dates t and T, δ is the rate of time preference, and $U(\cdot)$ is the utility function of the economic agent. This is the stochastic Euler equation, or the first order condition of the intertemporal utility maximization of the economic agent with suitable budget constraints (e.g., Cochrane, 1996, 2001). It holds for all securities, including assets and various derivatives. All capital asset pricing models (CAPM) and derivative pricing models can be embedded in this unified framework — each model can be viewed as a specific specification of $M_{t,T}$. See Cochrane (1996, 2001) for an excellent discussion.

There have been some parametric tests for CAMP models (e.g., Hansen and Janaganan, 1997). To our knowledge, there is only one nonparametric test for CAMP models based on the kernel method (Wang, 2002). Also, all the tests for CAMP models are formulated in terms of discrete time frameworks. Below, we focus on nonparametric derivative pricing.

Assuming that the conditional distribution of future consumption C_T has a density representation $f_t(\cdot)$, then the conditional expectation can be expressed as

$$E_t\left[Y(C_T)M_{t,T}\right] = \exp(-\tau\, r_t) \int Y(C_T) f_t^*(C_T)dC_T = \exp(-\tau\, r_t)\; E_t^*\left[Y(C_t)\right],$$

where r_t is the riskfree interest rate, $\tau = T - t$, and

$$f_t^*(C_T) = \frac{M_{t,T}f_t(C_T)}{\int M_{t,T}f_t(C_T)dC_T}$$

is called the risk neutral density (RND) function. This function is also called the risk-neural pricing probability (Cox and Ross, 1976), or equivalent martingale measure (Harrison and Kreps, 1979), or the state-price density. It contains rich information on the pricing and hedging of risky assets in an economy, and can be used to price other assets, or to recover the information about the market preferences and asset price dynamics (Bahra 1997, Jackwerth 1999). Obviously, the RND function differs from $f_t(C_T)$, the physical density function of C_T conditional on the information available at time t.

4.2 Nonparametric pricing

In order to calculate an option price from (4.1), one has to make some assumption on the data generating process of the underlying asset, $\{P_t\}$. For example, Black and Scholes (1973) assume that the underlying asset follows a geometric Brwonian motion:

$$dP_t = \mu P_t dt + \sigma P_t \, dB_t, \tag{4.2}$$

where μ and σ are two constants. Applying Ito's lemma, one can show immediately that P_τ follows a lognormal distribution with parameter $(\mu - \frac{1}{2}\sigma^2)\tau$ and $\sigma\sqrt{\tau}$. Using a no-arbitrage argument, Black and Scholes (1973) show that options can be priced if investors are risk neutral by setting the expected rate of return in the underlying asset, μ, equal to the risk-free interest rate, r. Specifically, the European call option price is

$$\pi(K_t, P_t, r, \tau) = P_t \, \Phi(d_{1t}) - e^{-r_t \tau} \, K_t \, \Phi(d_{2t}), \tag{4.3}$$

where K_t is the strike price, $\Phi(\cdot)$ is the standard normal cumulative distribution function,

$$d_{1t} = \frac{\ln(P_t/K_t) + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}, \quad \text{and} \quad d_{2t} = \frac{\ln(P_t/K_t) + (r - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}} = d_{1t} - \sigma\sqrt{\tau}.$$

In (4.3), the only parameter that is not observable a time t is σ. This parameter, when multiplied with $\sqrt{\tau}$, is the underlying asset return volatility over the remaining life of the option. An knowledge of σ can be inferred from the prices of options traded in the markets: given an observed option price, one can solve an appropriate option pricing model for σ which is essentially a market estimate of the future volatility of the underlying asset returns. This estimate of σ is known as "implied volatility".

The most important implication of Black-Scholes option pricing is that when the option is correctly priced, the implied volatility σ^2 should be the same across all exercise prices of options on the same underlying asset and with the same maturity date. However, the implied volatility observed in the market is usually a convex function of exercise price, which is often referred to as the "volatility smile". This indicates that market participants make more complicated assumptions than the geometric Bownian motion for the dynamics of the underlying asset. In particular, the convexity of "volatility smile" indicates the degree to which the market RND function has a heavier tail than a lognormal density. A great deal of effort has been made to use alternative models for the underlying asset to smooth out the volatility smile and so to achieve higher accuracy in pricing and hedging.

A more general approach to derivative pricing is to estimate the RND function directly from the observed option prices and then use it to price derivatives or to extract market information. To obtain better estimation of the RND function, several econometric techniques have been introduced. These methods are all based on the following fundamental relation between option prices and RNDs: Suppose $G_t = G(K_t, P_t, r_t, \tau)$ is the option pricing formula. Then there is a close relation between the second derivative of G_t with respect to the strike price K_t and the RND function:

$$\frac{\partial^2 G_t}{\partial K_t^2} = \exp(-\tau\, r_t)\; f_t^*(P_T). \tag{4.4}$$

This is first shown by Breeden and Litzenberger (1978) in a time-state preference framework.

Most commonly used estimation methods for RNDs are various parametric approaches. One of them is to assume that the underlying asset follows a parametric diffusion process, from which one can obtain the option pricing formula by a no-arbitrage argument, and then obtain the RND function from (4.4) (see, e.g., Bates 1991, 2000, Anagnou, Bedendo, Hodges and Tompkins 2001). Another parametric approach is to directly impose some form for the RND function and then estimate unknown parameters by minimizing the distance between the observed option prices and those generated by the assumed RND function (e.g., Jackwerth and Rubinstein, 1996, Melick and Thomas, 1997, Rubinstein, 1994). A third parametric approach is to assume a parametric form for the call pricing function or the implied volatility smile curve and then apply (4.4) to get the RND function (Bates 1991, Jarrow and Tudd, 1982, Longstaff, 1992, 1995, Shimko, 1993).

The aforementioned parametric approaches all impose certain restrictive assumptions, directly or indirectly, on the data generating process as well as the stochastic discount factor in some cases. The obtained RND function is not robust to the violation of these restrictions. To avoid this drawback, Ait-Sahalia and Lo (1998) use a nonparametric method to extract the RND function from option prices.

Given observed call option prices $\{G_t, K_t, \tau\}$, the price of the underlying asset $\{P_t\}$, and the risk free rate of interest $\{r_t\}$, Ait-Sahalia and Lo (1998) construct a kernel-estimator for $E(G_t | P_t, K_t, \tau, r_t)$. Under standard regularity conditions, Ait-Sahalia and Lo (1998) show that the RND estimator is consistent and asymptotically normal and they provide explicit expressions for the asymptotic variance of the estimator.

Armed with the RND estimator, Ait-Sahalia and Lo (1998) apply it to the pricing and delta-hedging of S&P 500 call and put options using daily data obtained from the Chicago Board Options Exchange for the sample period from January 4, 1993 to December 31, 1993. The RND estimator exhibits negative skewness and excess kurtosis, a common feature of historical stock returns. Unlike many parametric option pricing models, the RND-generated option pricing formula is capable of capturing persistent "volatility smiles" and other empirical features of market prices. Ait-Sahalia and Lo (2000) use a nonparametrical RND estimator to compute the economic value at risk, that is, the value at risk of the RND function.

The artificial neural network (ANN) has received much attention in economics and finance over the last decade. Hutchinson, Lo and Poggio (1994), Anders, Korn and Schmitt (1998) and Hanke (1999) have successfully applied the ANN models to estimate pricing formulas of finnacial derivatives. In particular, Hutchinson, Lo and Poggio (1994) use the ANN to address the following question: If option prices are truly determined by the Black-Scholes formula exactly, can ANN "learn" the Black-Scholes formula? In other words, can the Black-Scholes formula be estimated nonparametrically via learning networks with a sufficient degree of accuracy to be of practical use? Hutchinson, Lo and Poggio (1994) perform Monte Carlo simulation experiments in which various ANNs are trained on artificially generated Black-Scholes formula and then compare to the Black-Scholes formula both analytically and in out-of-sample hedging experiments. They begin by simulating a two-year sample of daily stock prices, and creating a across-section of options each day according to the rules used by the Chicago Broad Options Exchange with prices given by the Black-Scholes formula. They find that, even with training sets of only six months of daily data, learning network pricing formulas can approximate the Black-Scholes formula with reasonable accuracy. The nonlinear models obtained from neutral networks yield estimates option prices and deltas that are difficult to distinguish visually from the true Black-Scholes values.

There are several directions of further research on nonparametric estimation and testing of RNDs for derivative pricing. First, how to evaluate the quality of a RND function estimated from option prices? In other words, how to judge how well an estimated RND function reflects the market expected uncertainty of the underlying asset? Because the RND function differs from the physical probability density function of the underlying asset, the valuation of the RND function is rather challenging. The method developed by Hong and Li (2002) cannot be applied directly. One possible way to evaluate the RND function is to assume a certain family of utility functions for the representative investor, as in Rubinstein (1994) and Anagnou, Bedendo, Hodges and Tompkins (2001). Based on this assumption, one can obtain the stochastic discount factor and then the physical probability density function, to which

Hong and Li's (2002) test can be applied. However, the utility function of the economic agent is not observable. Thus, when the test delivers a rejection, it may be due to either misspecification of the utility function or misspecification of the data generating process, or both. More fundamentally, it is not clear whether the economy can be a proxy by an representative agent.

A practical issue in recovering the RND function is the limitation of option prices data with certain common characterizations. In other words, the sample size of option price data could be small in many applications. As a result, nonparametric methods should be carefully developed to fit the problems at hand.

Most econometric techniques to estimate the RND function is restricted to European options, while many of the more liquid exchange-traded options are often American. Rather complex extensions of the existing methods, including the nonparametric ones, are required in order to estimate the RND functions from the prices of American options. This is an interesting and practically important direction for further research.

5 Conclusion

Over the last several years, nonparametric continuous-time methods have become an integral part of research in financial economics. The literature is already vast and continues to grow swiftly, involving a full spread of participants for both financial economists and statisticians and engaging a wide sweep of academic journals. The field has left indelible mark on almost all core areas in finance such as asset pricing theory, consumption-portfolio selection, derivatives and risk analysis. The popularity of this field is also witnessed by the fact that the graduate students at both Master and doctoral levels in economics, finance, mathematics and statistics are expected to take courses in this discipline or alike and review the important research papers in this area to search for their own research interests, particularly dissertation topics for doctoral students. On the other hand, this area also has made an impact in the financial industry as the sophisticated nonparametric techniques can be of practical assistance in the industry. We hope that this selective review has provided the reader a perspective on this important field in finance and statistics and some open research problems.

References

Ahn, D.H., Dittmar, R.F. and Gallant, A.R. (2002), Quadratic term structure models: Theory and evidence, *The Review of Financial Studies* **15**, 243-288.

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

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

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

Ait-Sahalia, Y. (1999), Transition densities for interest rate and other nonlinear diffusions, *Journal of Finance* **54**, 1361-1395.

Ait-Sahalia, Y. (2002a), Maximum likelihood estimation of discretely sampled diffusions: A closed-form approach, *Econometrica* **70**, 223-262.

Ait-Sahalia, Y. (2002b), Telling from discrete data whether the underlying continuous-time model is a diffusion, *Journal of Finance* **57**, 2075-2112.

Ait-Sahalia, Y. (2002c), Closed-form likelihood expansions for multivariate diffusion, *Working Paper*, Department of Economics, Princeton University.

Ait-Sahalia, Y. and Kimmel, R. (2002), Estimating affine multifactor term structure models using closed-form likelihood expansions, *Working Paper*, Department of Economics, Princeton University.

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

Ait-Sahalia, Y. and Lo, A.W. (2000), Nonparametric risk management and implied risk aversion, *Journal of Econometric* **94**, 9-51.

Anagnou, I., Bedendo, M., Hodges, S. and Tompkins, R. (2001), The relation between implied and realized probability density functions, *Working Paper*, University of Technology, Vienna.

Anders, U., Korn, O. and Schmitt, C. (1998), Improving the pricing of options: A neural network approach, *Journal of Forecasting* **17**, 369-388.

Andersen, T.G., Benzoni, L. and Lund, J. (2001), Towards an empirical foundation for continuous-time equity return models, *Working Paper*, Graduate School of Management, Northwestern University.

Andersen, T.G., Chung, H.-J. and Sorensen, B.E. (1999), Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study, *Journal of Econometrics* **91**, 61-87.

Andersen, T.G. and Lund, J. (1997), Estimating continuous-time stochastic volatility models of the short-term interest rate, *Journal of Econometrics* **77**, 343-377.

Bahra, B. (1997), Implied risk-neutral probability density functions from option prices: Theory and application, *Working Paper*, Bank of England.

Bandi, F. (2000), Nonparametric fixed income pricing: theoretical issues, *Working Paper*, Graduate School of Business, The University of Chicago.

Bandi, F. and Nguyen, T.H. (2000a), Fully nonparametric estimators for diffusions: a small sample analysis, *Working Paper*, Graduate School of Business, The University of Chicago.

Bandi, F. and Nguyen, T.H. (2000b), On the functional estimation of jump-diffusion models, *Working Paper*, Graduate School of Business, The University of Chicago.

Bandi, F. and Phillips, P.C.B. (2003), Fully nonparametric estimation of scalar diffusion models, *Econometrica* **71**, 241-283.

Bates, D.S. (1991), The Crash of '87: Was it expected? The evidence from options markets, *Journal of Finance* **46**, 1009-44.

Bates, D.S. (2000), Post-'87 crash fears in the S&P 500 futures option market, *Journal of Econometrics* **94**, 181-238.

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

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

Black, F. and Scholes, M. (1973), The pricing of options and corporate liabilities, *Journal of Political Economy* **71**, 637-654.

Bliss, R.R. and Smith, D. (1998), The elasticity of interest rate volatility: Chan, Karolyi, Longstaff, and Sanders revisited, *Journal of Risk* **1**, 21-46.

Bollerslev, T. and Zhou, H. (1999), Estimating stochastic volatility diffusion using conditional moments of integrated volatility, *Working Paper*, Department of Economics, Duke University.

Brandt, M.W. and Santa-Clara, P. (2002), Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets, *Journal of Financial Economics* **63**, 161-210.

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

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

Brenner, R., Harjes, R. and Kroner, K. (1996), Another look at alternative models of the short-term interest rate, *Journal of Financial and Quantitative Analysis* **31**, 85-107.

Brown, S.J. and Dybvig, P.H. (1986), The empirical implications of the Cox, Ingersoll, Ross theory of the term structure of interest rates, *Journal of Finance* **41**, 617-630.

Cai, Z. (2001), Weighted Nadaraya-Watson regression estimation, *Statistics and Probability Letters* **51**, 307-318.

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

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.

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, *Journal of Finance* **47**, 1209-1227.

Chapman, D., Long, J. and Pearson, N. (1999), Using proxies for the short rate: When are three months like an instant, *The Review of Financial Studies* **12**, 763-807.

Chapman, D. and Pearson, N. (2000), Is the short rate drift actually nonlinear? *Journal of Finance* **55**, 355-388.

Chen, S.X., Gao, J. and Li, M. (2001), Simultaneous specification tests for the mean and variance structures of regression with applications to testing of diffusion models, *Working Paper*, Department of Statistics and Applied Probability, National University of Singapore.

Chen, S.X., Härdle, W. and Kleinow, T. (2001), An empirical likelihood goodness-of-fit test for time series, *Working paper*, Institute of Statistics and Economics, Humboldt University, Germany.

Chernov, M., Gallant, A.R., Ghysels, E. and Tauchen, G. (2002), A new class of stochastic volatility models with jumps: Theory and estimation, *Working Paper*, Department of Economics, University of North Carolina at Chapel Hill.

Chung, C.C. and Tauchen, G. (2001), Testing target zone models using efficient method of moments, *Journal of Business and Economic Statistics* **19**, 255-277.

Cochrane, J.H. (1996), A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* **104**, 572-621.

Cochrane, J.H. (2001), *Asset Pricing*, New Jersey: Princeton University Press.

Conley, T.G., Hansen, L.P., Luttmer, E.G.J. and Scheinkman, J.A. (1997), Short-term interest rates as subordinated diffusions, *Review of Financial Studies* **10**, 525-577.

Constantinides, G.M. (1992), A theory of the nominal term structure of interest rates, *Review of Financial Studies* **5**, 531-552.

Courtadon, G. (1982), A more accurate finite difference approximation for the valuation of options, *Journal of Financial and Quantitative Analysis* **17**, 697-703.

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

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

Cox, J.C. and Ross, S.A. (1976), The Volatility of option for alternative stochastic processes, *Journal of Financial Economics* **3**, 145-66.

Dai, Q. and K.J. Singleton (2000), Specification analysis of affine term structure models, *Journal of Finance* **55**, 1943-1978.

Diebold, F.X., Gunther, T. and Tay, A. (1998), Evaluating density forecasts with applications to financial risk management, *International Economic Review* **39**, 863-883.

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

Duffie, D. and Kan, R. (1996), A yield factor model of interest rate, *Mathematical Finance* **6**, 379-406.

Duffie, D. and Pan, J. (2001), Analytical value-at-risk with jumps and credit risk, *Finance and Stochastics* **5**, 15-180.

Duffie, D., Pan, J. and Singleton, K.J. (2000), Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* **68**, 1343-1376.

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

Egorov, A., Li, H. and Xu, Y. (2003), Maximum likelihood estimation of time-inhomogeneous diffusions, *Journal of Econometrics*, in press.

Elerian, O., Chib, S. and Shephard, N. (2001), Likelihood inference for discretely observed nonlinear diffusions, *Econometrica* **69**, 959-993.

Eraker, B. (1998), Markov chain Monte Carlo analysis of diffusion models with application to finance, HAE Thesis, Norwegian School of Economics and Business Administration.

Eraker, B., Johannes, M.S. and Polson, N.G. (1999), The impact of jumps in volatility and returns, *Working Paper*, Graduate School of Business, Columbia University.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modeling and its Applications*, London: Chapman and Hall.

Fan, J., Jiang, J., Zhang, C. and Zhou, Z. (2001), Time-dependent diffusion models for term structure dynamics and the stock price volatility, *Working Paper*, Department of Statistics, University of North Carolina at Chapel Hill.

Fan, J. and Zhang, C. (2001), A re-examination of diffusion estimators with applications to financial model validation, *Working Paper*, Department of Statistics, University of North Carolina at Chapel Hill.

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

Gallant, A.R. and Tauchen, G. (1996), Which moments to match? *Econometric Theory* **12**, 657-681.

Gallant, A.R. and Tauchen, G. (2001), Efficient method of moments, *Working Paper*, Department of Economics, University of North Carolina and Department of Economics, Duke University.

Gourieroux, C., Monfort, A. and Renault, E. (1993), Indirect inference, *Journal of Applied Econometrics* **8**, 85-118.

Gray, S. (1996), Modeling the conditional distribution of interest rates as a regime switching process, *Journal of Financial Economics* **42**, 27-62.

Hanke, M. (1999), Neural networks versus Black-Scholes: An empirical comparison of the pricing accuracy of two fundamentally different option pricing methods, *Journal of Computational Finance* **5**, 26-34.

Hansen, L.P. and Janaganan, R. (1997), Assessing specification errors in stochastic discount factor models, *Journal of Finance* **52**, 557-590.

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

Härdle, W. (1990), *Applied Nonparametric Regression*, New York: Cambridge University Press.

Härdle, W. and Mammen, E. (1993), Comparing nonparametric versus parametric regression fits, *The Annals of Statistics* **21**, 1926-1947.

Harrison, J.M. and Kreps, D.M. (1979), Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* **20**, 381-408.

Heath, D.C., Jarrow, R.A. and Morton, A. (1992), Bond pricing and the term structure of interest rates: A new methodology for contingent claim valuation, *Econometrica* **60**, 77-105.

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

Hong, Y. (1998), Testing for pairwise independence via the empirical distribution function, *Journal of the Royal Statistical Society, Series B* **60**, 429-453.

Hong, Y. (1999), Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach, *Journal of the American Statistical Association* **84**, 1201-1220.

Hong, Y. and Lee, T.H. (2003), Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series, *Review of Economics and Statistics*, in press.

Hong, Y. and Li, H. (2002), Nonparametric specification testing for continuous-time models with applications to interest rate term structures. Submitted to *Review of Financial Studies*.

Hong, Y. and White, H. (2000), Consistent nonparametric entropy-based testing for model specification, *Working Paper*, Departments of Economics and Department of Statistical Science, Cornell University and Department of Economics, University of California, San Diego.

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

Hutchinson, J., Lo, A.W. and Poggio, T. (1994), A nonparametric approach to pricing and hedging derivative securities via learning networks, *Journal of Finance* **49**, 851-889.

Jackwerth, J.C. (1999), Option-implied risk-neutral distributions and implied binomial trees: A literature review, *Journal of Derivative* **7**, 66-82.

Jackwerth J.C. and Rubinstein M. (1996), Recovering probability distributions from contemporary security prices, *Journal of Finance* **51**, 1611-1631.

Jacquier, E., Polson, N.G. and Rossi, P. (1994), Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics* **12**, 371-389.

Jagannathan, R. and Wang, Z. (2002), Empirical evaluation of asset pricing models: A comparison of the SDF and beta methods, *Journal of Finance* **58**, 2337-2367.

Jarrow, R. and Tudd, A. (1982), Approximate option valuation for arbitrary stochastic processes, *Journal of Financial Economics* **10**, 347-369.

Jiang, G.J. and Knight, J.L. (1997), A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model, *Econometric Theory* **13**, 615-645.

Jiang, G.J. and Knight, J.L. (2002), Estimation of continuous-time processes via the empirical characteristic function, *Journal of Business and Economic Statistics* **20**, 198-212.

Jiang, G.J. and van der Sluis, P.J. (2000), Option pricing with the efficient method of moments, in *Computational Finance* (Y.S. Abu-Mostafa, B. LeBaron, A.W. Lo, and A.S. Weigend, eds.), Cambridge: MIT Press.

Johannes, M.S. (2000), A nonparametric view of the role of jumps to interest rates, *Working Paper*, Graduate school od Business, Columbia University.

Johannes, M.S., Kumar, R. and Polson, N.G. (1999), State dependent jump models: How do US equity indices jump? *Working Paper*, Graduate School of Business, University of Chicago.

Jones, C.S. (1998), Bayesian estimation of continuous-time finance models, *Working Paper*, Simon School of Business, University of Rochester.

Kan, R. and Zhou, G. (1999), A critique of the stochastic discount factor methodology, *Journal of Finance* **54**, 1221-1248.

Karatzas, I. and Shreve, S.E. (1988), *Brownian Motion and Stochastic Calculus*, second edition, New York: Spring-Verlag.

Kleinow, T. (2002), Testing the diffusion coefficients, *Working paper*, Institute of Statistics and Economics, Humboldt University, Germany.

Liu, M. (2000), Modeling long memory in stock market volatility, *Journal of Econometrics* **99**, 139-171.

Liu, J., Longstaff, F.A. and Pan, J. (2002), Dynamic asset allocation with event risk, *Journal of Finance* **58**, 231-259.

Lo, A.W. (1988), Maximum likelihood estimation of generalized Ito processes with discretely sampled data, *Econometric Theory* **4**, 231-247.

Lo, A.W. and MacKinlay, A.C. (1988), Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies* **1**, 41-66.

Lobo, B.J. (1999), Jump risk in the U.S. stock market: Evidence using political information, *Review of Financial Economics* **8**, 149-163.

Longstaff, F.A. (1992), Multiple equilibria and tern structure models, *Journal of Financial Economics* **32**, 333-344.

Longstaff, F.A. (1995), Option pricing and the martingale restriction, *Review of Financial Studies* **8**, 1091-1124.

Longstaff, F.A. and Schwartz, E. (1992), Interest rate volatility and the term structure: A two-factor general equilibrium model, *Journal of Finance* **47**, 1259-1282.

Marsh, T. and Rosenfeld, E. (1983), Stochastic processes for interest rates and equilibrium bond prices, *Journal of Finance* **38**, 635-646.

Melick, W.R. and Thomas, C.P. (1997), Recovering an asset's implied PDF from option prices: An application to crude oil during the Gulf crisis, *Journal of Financial and Quantitative Analysis* **32**, 91-115.

Melino, A. (1994), Estimation of continuous-time models in finance, in: *Advances in Econometrics: Sixth World Congress,* Vol. 2 (C. Sims, ed.), Cambridge: Cambridge University Press.

Merton, R.C. (1973), Theory of rational option pricing, *Bell Journal of Economics and Management Science* **4**, 141-183.

Mittelhammer, R.C., Judge, G.G. and Miller, D.J. (2000), *Econometrics Foundation*, New York: Cambridge University Press.

Nagypal, E. (2001), Learning-by-doing versus selection: Can we tell them apart? *Working paper*, Department of Economics, Stanford University.

Øksendal, B. (1985), *Stochastic Differential Equations: An Introduction with Applications*, 3th edition, New York: Springer-Verlag.

Pagan, A. and Ullah, A. (1999), *Nonparametric Econometrics*, New York: Cambridge University Press.

Pan, J. (1997), Stochastic volatility with reset at jumps, *Working Paper*, School of Management, MIT.

Pedersen, A. R. (1995), A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Scandinavian Journal of Statistics* **22**, 55-71.

Perron, B. (2001), Jumps in the volatility of financial markets, *Working Paper*, Department of Economics, University of Montreal.

Pritsker, M. (1998), Nonparametric density estimation and tests of continuous time interest rate models, *Review of Financial Studies* **11**, 449-487.

Rice, J. (1986), Boundary modification for kernel regression, *Communications in Statistics* **12**, 1215-1230.

Rosenblatt, M. (1952), Remarks on a multivariate transformation, *The Annals of Mathematical Statistics* **23**, 470-472.

Rubinstein M. (1994), Implied binomial trees, *Journal of Finance* **49**, 771-818.

Shimko, D. (1993), Bounds of probability, *Risk* **6**, 33-37.

Singleton, K.J. (2001), Estimation of affine asset pricing models using the empirical characteristic function, *Journal of Econometrics* **102**, 111-141.

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

Sundaresan, S. (2001), Continuous-time methods in finance: A review and an assessment, *Journal of Finance* **55**, 1569-1622.

Tauchen, G. (1997), New minimum chi-square methods in empirical finance, in: *Advances in Econometrics: Seventh World Congress* (D.M. Kreps and K. Wallis, eds.), Cambridge: Cambridge University Press.

Tauchen, G. (2001), Notes on financial econometrics, *Journal of Econometrics* **100**, 57-64.

Valderrama, D. (2001), Can a standard real business cycle model explain the nonlinearities in U.S. national accounts data? Ph.D. Thesis, Department of Economics, Duke University.

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

Wang, K. (2002), Asset pricing with conditioning information: A new test, *Journal of Finance* **58**, 161-196.

Zhou, H. (2001), Jump-diffusion term structure and Ito conditional moment generator, *Working Paper*, Federal Reserve Board.