

Some Recent Developments in Nonparametric Finance ^{*†}

Zongwu Cai^{a,b} and Yongmiao Hong^{c,b}

^aDepartment of Mathematics & Statistics, University of North Carolina at Charlotte,
Charlotte, NC 28223, USA. E-mail: zcai@uncc.edu

^bWang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China

^cDepartment of Economics, Cornell University, Ithaca, NY 14853, USA. E-mail: yh20@cornell.edu.

Abstract

This paper gives a selective review on some recent developments of nonparametric methods in both continuous and discrete time finance, particularly in the areas of nonparametric estimation and testing of diffusion processes, nonparametric testing of parametric diffusion models, nonparametric pricing of derivatives, nonparametric estimation and hypothesis testing for nonlinear pricing kernel, and nonparametric predictability of asset returns. 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 weaknesses 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.

Keywords: Continuous time model; derivative pricing; jump process; kernel smoothing; nonparametric test; nonparametric pricing kernel; non-stationarity; options; predictability; stochastic discount factor; time-dependent model.

Forthcoming in *Advances in Econometrics*

*The authors thank two referees, Federico M. Bandi, Haitao Li and Aman Ullah for their valuable and helpful comments, suggestions and discussions. Also, the authors thank the participants at the seminars at University of Chicago, Columbia University, Academia Sinica and NYU, and the audiences at the 7th Annual Advances in Econometrics Conference (November 2008 at Louisiana State University) for their helpful comments.

†Cai's research was supported, in part, by the National Science Foundation grant DMS-0404954 and the National Science Foundation of China grant #70871003, and funds provided by the University of North Carolina at Charlotte, the Cheung Kong Scholarship from Chinese Ministry of Education, the Minjiang Scholarship from Fujian Province, China and Xiamen University. Hong thanks financial support from the Overseas Outstanding Youth Grant from the National Science Foundation of China and the Cheung Kong Scholarship from Chinese Ministry of Education and Xiamen University.

1 Introduction

Nonparametric modeling has become a core area in statistics and econometrics in the last two decades; see the books by Härdle (1990), Fan and Gijbels (1996) and Li and Racine (2007) for general statistical methodology and theory as well as applications. It has been used successfully in various fields such as economics and finance due to its advantage of requiring little prior information on the data generating process; see the books by Pagan and Ullah (1999), Mittelhammer, Judge and Miller (2000), Tsay (2005), Taylor (2005) and Li and Racine (2007) for real examples in economics and finance. Recently, nonparametric techniques have been proved to be the most attractive way of conducting research and gaining economic intuition in certain core areas in finance, such as asset and derivative pricing, term structure theory, portfolio choice, risk management, and predictability of asset returns, particularly, in modeling both continuous and discrete financial time series models; see the books by Campbell, Lo and MacKinlay (1997), Gouriéroux and Jasiak (2001), Duffie (2001), Tsay (2005) and Taylor (2005).

Finance is characterized by time and uncertainty. Modeling both continuous and discrete financial time series has been a basic analytic tool in modern finance since the seminal papers by Sharpe (1964), Fama (1970), Black and Scholes (1973) and Merton (1973). The rationale behind it is that for most of time, news arrives at financial markets in both continuous and discrete manners. 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, although it has been used for more than four decades for discrete financial time series. 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 used in both continuous and discrete time finance in recent years, and particularly in the areas of nonparametric estimation and testing of diffusion models, nonparametric derivative pricing and its tests, and predictability of asset returns based on nonparametric approaches. Financial time series data have some distinct important stylized facts, such as persistent volatility clusterings, heavy tails, strong serial dependence, and occasionally sudden but large jumps. In addition, financial modeling is often closely embedded in a financial theoretical framework. These features suggest that standard statistical theory may not be readily applicable to both continuous and discrete financial time series. This is a promising and fruitful area for both financial economists and statisticians to interact with.

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 and hypothesis testing of derivative and asset pricing, particularly the nonparametric estimation of risk neutral density functions and nonlinear pricing kernel models. Nonparametric predictability of asset returns is presented in Section 5. In Sections 2 to 5, we point out some open and interesting research problems, which might be useful for graduate students to review the important research papers in this field and to search for their own research topics, particularly dissertation topics for doctoral students. Finally, in Section 6, we highlight some important research areas that are not covered in this paper due to space limitation, say nonparametric volatility (conditional variance) and ARCH or GARCH type models and nonparametric methods in volatility for high-frequency data with/without microstructure noise. We plan to write a separate survey paper to discuss some of these omitted topics in the near future.

2 Nonparametric Diffusion Models

2.1 Diffusion 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,

$$dX_t = \mu(X_t) dt + \sigma(X_t) dB_t, \quad (2.1)$$

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. Indeed, model (2.1) can be

¹Other 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 Rosenfeld (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.

applied to many core areas in finance, such as options, derivative pricing, asset pricing, term structure of interest rates, dynamic consumption and portfolio choice, default risk, stochastic volatility, exchange rate dynamics, and others.

There are two basic approaches to identifying $\mu(\cdot)$ and $\sigma(\cdot)$. The first is a parametric approach, which assumes some parametric forms of $\mu(\cdot, \theta)$ and $\sigma(\cdot, \theta)$, and estimates the unknown model parameters, say θ . Most existing models in the literature assume that the interest rate exhibits mean-reversion and that the drift $\mu(\cdot)$ is a linear or quadratic 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. In modeling interest rate dynamics, this specification captures the so-called “level effect”; i.e., the higher the interest rate level, the larger the volatility. With $\gamma = 0$ and 0.5 , 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 wisdom or convenience. They may not be consistent with the data generating process and there may be at risk of misspecification.

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 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

²Empirical studies on the short rate include Ait-Sahalia (1996a, b), Andersen and Lund (1997), Ang and Bekaert (2002a, b), Brenner, Harjes and Kroner (1996), Brown and Dybvig (1986), Chan, Karolyi, Longstaff and Sanders (1992), Chapman and Pearson (2000), Chapman, Long and Pearson (1999), Conley, Hansen, Luttmer and Scheinkman (1997), Gray (1996), and Stanton (1997).

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. The reason is that near the extreme observations, the nonparametric estimation might not be accurate due to the sparsity of data in this region. Also, this region is close to the boundary point, so that the Nadaraya-Watson (NW) estimate suffers a boundary effect. 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, to overcome the boundary effect, Fan and Zhang (2003) fit a nonparametric model using a local linear technique and apply the generalized likelihood ratio test of Cai, Fan and Yao (2000) and 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 Cai, Fan and Yao (2000) for discrete time series and Fan, Zhang and Zhang (2001) for iid samples but it is still unknown whether it is valid for continuous 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 semiparametric 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) argue that the result that γ equals to 1.5 depends on whether the data between October 1979 to September 1982 are included. From the foregoing discussions, it seems that the value of γ may change over time.

2.2 Nonparametric Estimation

Under some regularity conditions; see Jiang and Knight (1997) and Bandi and Nguyen (2000), the diffusion process in (2.1) is a one dimensional, regular, strong Markov 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 :

$$\mu(X_t) = \lim_{\Delta \rightarrow 0} \Delta^{-1} E[Y_t | X_t], \quad \text{and} \quad \sigma^2(X_t) = \lim_{\Delta \rightarrow 0} \Delta^{-1} E[Y_t^2 | X_t], \quad (2.2)$$

where $Y_t = X_{t+\Delta} - X_t$. 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 (infinitesimal) operator; see, e.g., Øksendal (1985) and Karatzas and Shreve (1988), Stanton (1997) shows that the first order approximation

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

the second order approximation

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

and the third order approximation

$$\mu(X_t)^{(3)} = \frac{1}{6\Delta} [18 E\{Y_t | X_t\} - 9 E\{X_{t+2\Delta} - X_t | X_t\} + 2 E\{X_{t+3\Delta} - X_t | X_t\}] + O(\Delta^3),$$

etc. Fan and Zhang (2003) derive higher-order approximations. Similar formulas hold for the diffusion; see Stanton (1997). Bandi and Nguyen (2000) 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), 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 (2000) and Fan and Zhang (2003), 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 (2000) and Fan and Zhang (2003). The question arises is how to choose the order in application. As demonstrated in Fan and Zhang (2003), the first or second order may be enough in most applications.

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

$$\mu(x) \approx \frac{1}{\Delta} E[Y_\tau | X_{\tau\Delta} = x] \quad \text{and} \quad \sigma^2(x) \approx \frac{1}{\Delta} E[Y_\tau^2 | X_{\tau\Delta} = x] \quad (2.3)$$

for all $1 \leq \tau \leq n - 1$, where $Y_\tau = X_{(\tau+1)\Delta} - X_{\tau\Delta}$. Both $\mu(x)$ and $\sigma^2(x)$ become classical nonparametric regressions and a nonparametric kernel smoothing approach can be applied to estimating them.

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 for any given grid point x , respectively, by

$$\hat{\mu}(x) = \frac{1}{\Delta} \frac{\sum_{\tau=1}^{n-1} Y_\tau K_h(x - X_{\tau\Delta})}{\sum_{\tau=1}^{n-1} K_h(x - X_{\tau\Delta})}, \quad \text{and} \quad \hat{\sigma}^2(x) = \frac{1}{\Delta} \frac{\sum_{\tau=1}^{n-1} Y_\tau^2 K_h(x - X_{\tau\Delta})}{\sum_{\tau=1}^{n-1} K_h(x - X_{\tau\Delta})}, \quad (2.4)$$

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

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

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

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

where $\hat{\pi}(x)$ is a consistent estimator of $\pi(x)$, 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 $(dB_t)^2 = dt + O((dt)^2)$. 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 (2003) suggest using the local linear technique to estimate $\mu(x)$ as follows. When $X_{\tau\Delta}$ is in a neighborhood of the grid point x , by assuming that the second derivative of $\mu(\cdot)$ is continuous, $\mu(X_{\tau\Delta})$ can be approximated linearly as $\beta_0 + \beta_1(X_{\tau\Delta} - x)$, where $\beta_0 = \mu(x)$ and $\beta_1 = \mu'(x)$, the first derivative of $\mu(x)$. Then, the

locally weighted least square is given by

$$\sum_{\tau=1}^{n-1} \{\Delta^{-1}Y_{\tau} - \beta_0 - \beta_1(X_{\tau\Delta} - x)\}^2 K_h(X_{\tau\Delta} - x). \quad (2.5)$$

Minimizing the above with respect to β_0 and β_1 gives the local linear estimate of $\mu(x)$. Similarly, in view of (2.3), the local linear estimator of $\sigma^2(\cdot)$ can be obtained by changing $\Delta^{-1}Y_{\tau}$ in (2.5) into $\Delta^{-1}Y_{\tau}^2$. 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 proposed by Cai (2001) can be used to estimate $\sigma(\cdot)$. Recently, Xu and Phillips (2007) study this approach and investigate its properties.

The asymptotic theory can be found in Jiang and Knight (1997) and Bandi and Nguyen (2000) for the NW estimator and in Fan and Zhang (2003) for the local linear estimator as well as Xu and Phillips (2007) for the weighted NW 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 theoretical and empirical studies in the literature. As a rule of thumb, an easy way to choose a data-driven fashion bandwidth is to use the nonparametric version of the Akaike information criterion; see Cai and Tiwari (2000).

One crucial assumption in the foregoing 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 $\sigma^2(x)$, respectively,

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

where

$$\tilde{\mu}(x) = \frac{1}{\Delta} \frac{\sum_{\tau=1}^{n-1} I(|X_{\tau\Delta} - x| \leq b) Y_{\tau}}{\sum_{\tau=1}^n I(|X_{\tau\Delta} - x| \leq b)}, \quad \text{and} \quad \tilde{\sigma}^2(x) = \frac{1}{\Delta} \frac{\sum_{\tau=1}^{n-1} I(|X_{\tau\Delta} - x| \leq b) Y_{\tau}^2}{\sum_{\tau=1}^n I(|X_{\tau\Delta} - x| \leq b)}.$$

See, also Bandi and Nguyen (2000). 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 Phillips'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 (2000) for more discussions and asymptotic theory.

Bandi and Phillips's (2003) estimator can be viewed as a double kernel smoothing method: The first step defines straight sample analogs to the values that drift and diffusion take at the sampled points and it can be regarded as a generalization of the moving average. Indeed, this step uses the smoothing technique (a linear estimator with the same weights) to obtain the raw estimates of the two functions $\tilde{\mu}(x)$ and $\tilde{\sigma}^2(x)$, respectively. This approach is different from classical two-step method in the literature; see Cai (2002a, b). The key is to figure out how important the first is to the second step. 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, as addressed at the end of Section 2.1. A variety of time-dependent diffusion models have been proposed in the literature. A time-dependent diffusion process is formulated as

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t. \quad (2.6)$$

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:

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

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

$$\mu(X_t, t) = \lim_{\Delta \rightarrow 0} \Delta^{-1} E\{Y_t | X_t\}, \quad \text{and} \quad \sigma^2(X_t, t) = \lim_{\Delta \rightarrow 0} \Delta^{-1} E\{Y_t^2 | X_t\},$$

where $Y_t = X_{t+\Delta} - X_t$, which provide a regression form for estimating $\mu(\cdot, t)$ and $\sigma^2(\cdot, t)$.

By assuming that the drift and diffusion functions are linear in X_t with time varying coefficients, Fan, Jiang, Zhang and Zhou (2003) consider the following time-varying coefficient single factor model

$$dX_t = [\alpha_0(t) + \alpha_1(t) X_t] dt + \beta_0(t) X_t^{\beta_1(t)} dB_t, \quad (2.7)$$

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. Indeed, the aforementioned models are a special case of the following more general time-varying coefficient multi-factor diffusion model

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t, \quad (2.8)$$

where

$$\mu(X_t, t) = \alpha_0(t) + \alpha_1(t) g(X_t), \quad \text{and} \quad (\sigma(X_t, t)\sigma(X_t, t)^\top)_{ij} = \beta_{0,ij}(t) + \beta_{1,ij}(t)^\top h_{ij}(X_t),$$

and $g(\cdot)$ and $\{h_{ij}(\cdot)\}$ are known functions. This is the time-dependent version of the multi-factor affine model 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. It is interesting to point out that the estimation approaches described above are still applicable to model (2.8) but the asymptotic theory is very challenging because of the nonstationarity of unknown structure of the underlying process $\{X_t\}$.

2.4 Jump Diffusion Models

There has been a vast literature on the study of diffusion models with jumps.³ 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, the U.S. 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 (2002), Liu, Longstaff and Pan (2002), and Johannes (2004), among others. Also, Jumps can capture the heavy tail behavior of the distribution of the underlying process; see later.

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

$$dX_t = \mu(X_t) dt + \sigma(X_t) dB_t + dJ_t, \quad (2.9)$$

where J_t is a compensated jump process (zero conditional mean) with arrival rate (conditional probability) $\lambda_t = \lambda(X_t) \geq 0$, which is an instantaneous intensity function. There are several studies on specification of J_t . For example, a simple specification is to assume $J_t = \xi P_t$, where P_t is a Poisson process with an intensity $\lambda(X_t)$ or a binomial distribution

³See, to name just a few, Pan (1997), Duffie and Pan (2001), Bollerslev and Zhou (2002), Eraker, Johannes and Polson (2003), Bates (2000), Duffie, Pan and Singleton (2000), Johannes (2004), Liu, Longstaff and Pan (2002), Zhou (2001), Singleton (2001), Perron (2001), Chernov, Gallant, Ghysels and Tauchen (2003).

with probability $\lambda(X_t)$, and the jump size, ξ , has a time-invariant distribution $\Pi(\cdot)$ with mean zero. $\Pi(\cdot)$ is commonly assumed to be either normally or uniformly distributed. If $\lambda(\cdot) = 0$ 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 (2003) consider a Lévy process for J_t . A simple jump diffusion model proposed by Kou (2002) is discussed in Tsay (2005) by assuming that $J_t = \sum_{i=1}^{n_t} (L_i - 1)$, where n_t is a Poisson process with rate λ , and $\{L_i\}$ is a sequence of independent and identically distributed nonnegative random variables such that $\ln(L_i)$ has a double exponential distribution with probability density function $f(x) = \exp(-|x - \theta_1|/\theta_2)/2\theta_2$ for $0 < \theta_2 < 1$. This simple model enjoys several nice properties. The returns implied by the model are leptokurtic and asymmetric with respect to zero. In addition, the model can reproduce volatility smile and provide analytical formulas for the prices of many options.

In practice, $\lambda(\cdot)$ might be assumed to have a particular form. For example, Chernov, Gallant, Ghysels and Tauchen (2003) 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 is 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 $\xi = X_{\tau+} - X_{\tau-}$. We conjecture that a wavelet method may be potentially useful here because a wavelet approach has an ability of capturing the discontinuity and removing the contaminated noise. For detailed discussion on how to use a wavelet method in this regard, the reader is referred to the paper by Fan and Wang (2007). Indeed, Fan and Wang (2007) propose using a wavelet method to cope with both jumps in the price and market microstructure noise in the observed data to estimate both integrated volatility and jump variation from the data sampled from jump-diffusion price processes, contaminated with the market microstructure noise.

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

$$\mu_1(X_t) = \lim_{\Delta \downarrow 0} \Delta^{-1} E[Y_t | X_t] = \mu(X_t) + \lambda(X_t) E(\xi),$$

and

$$\mu_2(X_t) = \lim_{\Delta \downarrow 0} \Delta^{-1} E[Y_t^2 | X_t] = \sigma^2(X_t) + \lambda(X_t) E(\xi^2).$$

Clearly, $\mu_2(X_t)$ is much bigger than $\sigma^2(X_t)$ if there is a jump. This means that adding a jump into the model can capture the heavy tails. Also, it is easy to see that the first two moments are the same as those for a diffusion model by using a new drift coefficient

$\tilde{\mu}(X_t) = \mu(X_t) + \lambda(X_t) E(\xi)$ and a new diffusion coefficient $\tilde{\sigma}^2(x) = \sigma^2(x) + \lambda(x) E(\xi^2)$. 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, $j > 2$,

$$\mu_j(X_t) = \lim_{\Delta \rightarrow 0} \Delta^{-1} E[Y_t^j | X_t] = \lambda(X_t) E(\xi^j).$$

See Duffie, Pan and Singleton (2000) and Johannes (2004) for details. Obviously, jumps provide a simple and intuitive mechanism for capturing the heavy tail behavior of underlying process. In particular, the conditional skewness and kurtosis are respectively given by

$$s(X_t) \equiv \frac{\lambda(X_t) E(\xi^3)}{[\sigma^2(X_t) + \lambda(X_t) E(\xi^2)]^{3/2}}, \quad \text{and} \quad k(X_t) \equiv \frac{\lambda(X_t) E(\xi^4)}{[\sigma^2(X_t) + \lambda(X_t) E(\xi^2)]^2}.$$

Note that $s(X_t) = 0$ if ξ is symmetric. By assuming $\xi \sim N(0, \sigma_\xi^2)$, Johannes (2004) 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 (2004) and Bandi and Nguyen (2003). Moreover, Bandi and Nguyen (2003) provide a general asymptotic theory for the resulting estimators. Further, by specifying a particular form of $\Pi(\xi) = \Pi_0(\xi, \theta)$, say, $\xi \sim N(0, \sigma_\xi^2)$, Bandi and Nguyen (2003) propose consistent estimators of $\lambda(\cdot)$, σ_ξ^2 , and $\sigma^2(\cdot)$ and derive their asymptotic properties.

A natural question arises is how to measure the departures from a pure diffusion model statistically. That is to test model (2.9) against 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

$$T_1 = \int \hat{\mu}_4(x) w(x) dx, \quad \text{or} \quad T_2 = \int \hat{\mu}_3^2(x) w(x) dx, \quad (2.10)$$

where $w(\cdot)$ is a weighting function. The asymptotic theory for T_1 and T_2 is still unknown. It needs a further investigation theoretically and empirically. Based on a Monte Carlo simulation approach, Cai and Zhang (2008b) use the aforementioned testing statistics in an application, described as follows.

It is well known that prices fully reflect the available information in the efficient market. Thus, Cai and Zhang (2008b) consider the market information consisting of two components. The first is the anticipated information that drives market prices' daily normal fluctuation

and the second is the unanticipated information that determines prices to exceptional fluctuation, which can be characterized by a jump process. Therefore, Cai and Zhang (2008b) investigate the market information via a jump-diffusion process. The jump term in the dynamic of stock price or return rate reflects the sensitivity of unanticipated information for the related firms. This implies that the investigation of the jump parameters for firms with different sizes would help us to find the relationship between firm sizes and information sensitivity. With the nonparametric method as described above, Cai and Zhang (2008b) use the kernel estimation method, and reveal how the nonparametric estimation of the jump parameters (functions) reflect the so called information effect. Also, they test the model based on the test statistic formulated in (2.10). Due to the lack of the relevant theory of the test statistics in (2.10), Cai and Zhang (2008b) use the Monte Carlo simulation, and find that a jump diffusion process performs better to model with all market information, including anticipated and unanticipated information than the pure diffusion model. Empirically, Cai and Zhang (2008b) estimate the jump intensity and jump variance for portfolios with different firm sizes for data from both the U.S. and Chinese markets, and find some evidences that there exists information effect among different firm sizes, from which we could get valuable references for investors' decision making. Finally, using a Monte Carlo simulation method, Cai and Zhang (2008a) exam the test statistics in (2.10) to see how the discontinuity of drift or diffusion function affects the performance of the test statistics. They find that the discontinuity of drift or diffusion function has an impact on the performance of the test statistics in (2.10).

More generally, 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 $D \subseteq \mathbb{R}$. 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

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

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 $\{X_{\tau\Delta}\}$, one 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

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

versus the alternative

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

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. If the model has a specific form, say a parametric form, the diffusion criterion becomes a simple form, say, it becomes just a constraint for some parameters. Then, the testing problem becomes testing a constraint on parameters; see Ait-Sahalia (2002b) for some real applications.

2.5 Time-Dependent Jump Diffusion Models

Duffie, Pan and Singleton (2000) consider the following time-dependent jump diffusion model

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dB_t + dJ_t, \quad (2.11)$$

where J_t is a compensated jump process with the time-varying intensity $\lambda(X_t, t) = \lambda_0(t) + \lambda_1(t) X_t$, and Chernov, Gallant, Ghysels and Tauchen (2003) consider a more general stochastic volatility model with the time-varying 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 (2003), 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

$$\lambda(\xi_0, X_t, \tau, t) = \{\lambda_0(t) + \lambda_1(t) X_t\} \lambda\{\tau(t)\} \exp\{G(\xi_0)\}, \quad (2.12)$$

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 model (2.9) economically and statistically to a more general time-dependent jump diffusion model given in (2.11) with the time-dependent intensity function $\lambda(\xi_0, X_t, \tau, t)$ without any specified form or with some nonparametric structure, say, like (2.12). 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 (2003), among others. This is still an open problem.

3 Nonparametric Inferences of Parametric Diffusion Models

3.1 Nonparametric Estimation

As is well-known, derivative pricing in mathematical finance is generally much more tractable in a continuous-time modeling 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.⁴ This is largely motivated by the fact that using the discrete version of a continuous-time model can result in inconsistent parameter estimates; see Lo (1988). Available estimation procedures include the MLE method of Lo (1988), the simulated methods of moments of Duffie and Singleton (1993) and Gouriéroux, 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 (2002) and Singleton (2001).

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

$$dX_t = \mu(X_t, \theta) dt + \sigma(X_t, \theta) dB_t, \quad (3.1)$$

⁴Sundaresan (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).

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

$$\hat{\theta} = \arg \min_{\theta \in \Theta} n^{-1} \sum_{\tau=1}^n [\hat{\pi}_0(X_{\tau\Delta}) - \pi(X_{\tau\Delta}, \theta)]^2, \quad (3.2)$$

where

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

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

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

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 $\hat{\theta}$ will not be asymptotically most efficient, although it is root- n consistent for θ_0 if the parametric model is correctly specified.

Next, we introduce the approximate likelihood estimation (AMLE) approach, due to Ait-Sahalia (2002a). 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

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

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 $\hat{\theta}_n^{(J)}$ that maximizes the approximated model likelihood. The estimator $\hat{\theta}_n^{(J)}$ enjoys the same asymptotic efficiency as the (infeasible) MLE as $J = J_n \rightarrow \infty$. 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 $X \rightarrow Y$ 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 transformation as

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

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

$$p_z^{(J)}(z | z_0, \theta) = \phi(z) \sum_{j=0}^J \eta_z^{(j)}(z_0, \theta) H_j(z),$$

where $\phi(\cdot)$ is the $N(0, 1)$ density, and $\{H_j(z)\}$ is the Hermite polynomial series. The coefficients $\{\eta_z^{(j)}(z_0, \theta)\}$ 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:

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

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

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

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 (2008) extends this method to general multivariate diffusion models and Ait-Sahalia and Kimmel (2007) to affine multi-factor term structure models.

In contrast to the AMLE in Ait-Sahalia (2002a), Jiang and Knight (2006) consider a more general Markov models where the transition density is unknown. The approach Jiang and Knight (2006) propose is based on the empirical characteristic function estimation procedure

with an approximate optimal weight function. The approximate optimal weight function is obtained through an Edgeworth/Gram-Charlier expansion of the logarithmic transition density of the Markovian process. They derive the estimating equations and demonstrate that they are equivalent to the approximate maximum likelihood estimation (AMLE) as in Ait-Sahalia (2002a). However, in contrast to the common AMLE, their approach ensures the consistency of the estimator even in the presence of approximation error. When the approximation error of the optimal weight function is arbitrarily small, the estimator has MLE efficiency. For details, see Jiang and Knight (2006).

Finally, 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 $p_0(\Delta, X_{\tau\Delta} | X_{s\Delta}, s \leq \tau - 1) = p_0(\Delta, X_{\tau\Delta} | Y_{\tau\Delta})$, where $Y_{\tau\Delta} \equiv (X_{(\tau-1)\Delta}, \dots, X_{(\tau-d)\Delta})^\top$ for some fixed integer $d \geq 0$. 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:

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

where $p(\Delta, x, y; \theta)$ is the model-implied joint density for $(X_{\tau\Delta}, Y_{\tau\Delta}^\top)^\top$, θ_0 is the unknown true parameter value, and $f(\Delta, x, y; \beta_n)$ is an auxiliary model for the conditional density of $(X_{\tau\Delta}, Y_{\tau\Delta}^\top)^\top$. 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 $(X_{\tau\Delta}, Y_{\tau\Delta}^\top)^\top$, and thus it 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 a model selection criterion such as the

Bayesian information criterion (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:

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

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

The EMM has been applied widely in financial applications. See, for example, Andersen and Lund (1997), Dai and Singleton (2000), Ahn, Dittmar and Gallant (2002) for interest rate applications, Liu (2000), Andersen, Berzoni and Lund (2002), Chernov, Gallant, Ghysels and Tauchen (2003) 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 Valderrama (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 and this topic is still open.

3.2 Nonparametric Testing

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

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

As noted earlier, various 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

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

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 $\{X_{\tau\Delta}\}_{\tau=1}^n$. Next we will present some test procedures for testing the continuous-time models. 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 determine the stationary density $\pi(\cdot, \theta)$ in (3.3). Ait-Sahalia (1996b) compares a parametric marginal density estimator $\pi(\cdot, \hat{\theta})$ with a nonparametric density estimator $\hat{\pi}_0(\cdot)$ via the quadratic form

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

where x_1^* is the upper bound for X_t , $\hat{\theta}$ 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.⁵ Second, subject to some regularity conditions, the asymptotic

⁵A 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.

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 the 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 doubts 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 $\{X_t\}$.

Recently, Hong and Li (2005) 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 $p_0(x, t | x_0, s)$ 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:

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

versus the alternative hypothesis

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

Clearly, to test \mathbb{H}_0 in (3.8) versus \mathbb{H}_a in (3.9) would be to compare a model transition density estimator $p(x, t | x_0, s, \hat{\theta})$ with a nonparametric transition density estimator, say $\hat{p}_0(x, t | x_0, s)$. Instead of comparing $p(x, t | x_0, s, \hat{\theta})$ and $\hat{p}_0(x, t | x_0, s)$ directly, Hong and Li (2005) first transform $\{X_{\tau\Delta}\}_{\tau=1}^n$ via a probability integral transformation. Define a discrete transformed sequence

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

Under (and only under) \mathbb{H}_0 in (3.8) there exists some $\theta_0 \in \Theta$ such that $p[x, \tau\Delta | X_{(\tau-1)\Delta}, (\tau-1)\Delta, \theta_0] = p_0[x, \tau\Delta | X_{(\tau-1)\Delta}, (\tau-1)\Delta]$ 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 $p(x, t | y, s, \theta)$.

To test \mathbb{H}_0 in (3.8), Hong and Li (2005) 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 (2005) employ the kernel joint density estimator,

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

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.⁶ 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).

⁶One 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, 20 and 10 of 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.

Hong and Li's (2005) test statistic is

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

where A_h^0 and V_0 are non-stochastic centering and scale factors which are functions of h and $k(\cdot)$.

In a simulation experiment mimicking the dynamics of U.S. interest rates via the Vasicek model, Hong and Li (2005) 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 (2005) 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. Egorov, Hong and Li (2006) have recently extended Hong and Li (2005) to evaluate out-of-sample of density forecasts of a multivariate diffusion model possibly with jumps and partially unobservable state variables.

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.) Pedersen (1995), Brandt and Santa-Clara (2002), 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 the 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 (2005) 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 mis-specifications, 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, as is shown in Hong and Lee (2003b). For example, Andersen 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 (2003a) 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 exploited the correct specification of the SNP density model.⁷

There has also been 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 linear. To test the linearity of the drift term, one can write it as a functional coefficient form (Cai, Fan and Yao, 2000) $\mu(X_t) = \alpha_0(X_t) + \alpha_1(X_t) X_t$. Then, the null hypothesis is $\mathbb{H}_0 : \alpha_0(\cdot) \equiv \alpha_0$ and $\alpha_1(\cdot) \equiv \alpha_1$. Fan and Zhang (2003) apply the generalized likelihood ratio test developed by Cai, Fan and Yao (2000) and 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 asymptotic theory for 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. One might follow the idea from Cai, Fan and Yao (2000) to use the bootstrap or wild bootstrap method instead of the asymptotic theory for time series context. Fan and Zhang (2003) and Fan, Jiang, Zhang and Zhou (2003) conjecture that it would hold based on their simulations. On the other hand, Chen, Härdle and Kleinow (2002) consider an empirical likelihood goodness-of-fit test for time series regression model, and they apply the test to test a discrete drift model of a diffusion process.

There has also been 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

⁷Chen, Gao and Tang (2008) 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.

diffusion $\sigma(\cdot)$, which is one of the most important features of (3.1) for derivative pricing. Kleinow (2002) recently proposes a nonparametric test for a diffusion model $\sigma(\cdot)$. More specifically, Kleinow (2002) compares a nonparametric diffusion estimator $\hat{\sigma}^2(\cdot)$ with a parametric diffusion estimator $\sigma^2(\cdot, \theta)$ via an asymptotically χ^2 test statistic

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

where

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

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

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

is a smooth version of $\sigma^2(x, \theta)$. The use of $\tilde{\sigma}^2(x, \hat{\theta})$ instead of $\sigma^2(x, \hat{\theta})$ directly reduces the kernel estimation bias in $\hat{T}(x)$, thus allowing the use of the optimal bandwidth h for $\hat{\sigma}^2(x)$. This device is also used in Härdle and Mammen (1993) in testing a parametric regression model. Kleinow (2002) finds that the empirical level of \hat{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.

Furthermore, Fan, Jiang, Zhang and Zhou (2003) 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. However, the validity of the generalized likelihood ratio test for nonstationary time series is still unknown and it needs a further investigation.

Finally, Kristensen (2008) considers an estimation method for two classes of semiparametric scalar diffusion models. In the first class, the diffusion term is parameterized and the drift is left unspecified while in the second class, only the drift term is specified. Under the assumption of stationarity, the unspecified term can be identified as a function of the parametric component and the stationary density. Given a discrete sample with a fixed time distance, the parametric component is then estimated by maximizing the associated likelihood with a preliminary estimator of the unspecified term plugged in. Kristensen (2008) shows that this pseudo-MLE (PMLE) is \sqrt{n} -consistent with an asymptotically normal distribution under regularity conditions, and demonstrates how the estimator can be used in

specification testing not only of the semiparametric model itself but also of fully parametric ones. Since the likelihood function is not available on closed form, the practical implementation of the proposed estimator and tests will rely on simulated or approximate PMLE. Under regularity conditions, Kristensen (2008) verifies that the approximate/simulated version of the PMLE inherits the properties of the actual but infeasible estimator. Also, Kristensen (2007) proposes a nonparametric kernel estimator of the drift (diffusion) term in a diffusion model based on a preliminary parametric estimator of the diffusion (drift) term. Under regularity conditions, rates of convergence and asymptotic normality of the nonparametric estimators are established. Moreover, Kristensen (2007) develops misspecification tests of diffusion models based on the nonparametric estimators, and derives the asymptotic properties of the tests. Furthermore, Kristensen (2007) proposes a Markov Bootstrap method for the test statistics to improve on the finite-sample approximations.

4 Nonparametric Pricing Kernel Models

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 a 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 , which is a function of aggregate consumption C_T , is given by

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

where the conditional expectation is taken with respect to the information set available to the representative economic agent at time t , $M_{t,T} = \delta^{T-t} U'(C_T)/U'(C_t)$, the so-called stochastic discount factor (SDF), 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 (CAP) models 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 CAP models (e.g., Hansen and Janaganan, 1997). To the best of our knowledge, there are only a few nonparametric tests available in the literature for testing CAP models based on the kernel method; see Wang (2002, 2003)

and Cai, Kuan and Sun (2008), which will be elaborated in detail in Section 4.3 later. Also, all the tests for CAP models are formulated in terms of discrete time frameworks. We focus on nonparametric derivative pricing in Section 4.2 and the nonparametric asset pricing will be discussed separately in Section 4.3.

4.1 Nonparametric Risk Neutral Density

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

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

where r_t is the risk-free 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. See Taylor (2005, Chapter 16) for details about the definition and estimation methods. This function is also called the risk-neutral pricing probability (Cox and Ross, 1976), or equivalent martingale measure (Harrison and Kreps, 1979), or the state-price density (SPD). 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 Derivative 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 Brownian motion:

$$dP_t = \mu P_t dt + \sigma P_t dB_t,$$

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

$$\pi(K_t, P_t, r, \tau) = P_t \Phi(d_t) - e^{-r\tau} K_t \Phi(d_t - \sigma\sqrt{\tau}), \quad (4.2)$$

where K_t is the strike price, $\Phi(\cdot)$ is the standard normal cumulative distribution function and $d_t = \{\ln(P_t/K_t) + (r + \frac{1}{2}\sigma^2)\tau\} / (\sigma\sqrt{\tau})$. In (4.2), 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. The 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 model 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 Brownian 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:

$$\frac{\partial^2 G_t}{\partial K_t^2} = \exp(-\tau r_t) f_t^*(P_T). \quad (4.3)$$

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 (4.3) (see, e.g., Bates 1991, 2000, Anagnou, Bedendo, Hodges and Tompkins 2005). 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 (4.3) 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 nonparametric 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 financial 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 neural networks yield estimated option prices and deltas that are difficult to distinguish visually from the true Black-Scholes values.

Based on the economic theory of option pricing, the price of a call option should be a monotonically decreasing convex function of the strike price and the state price density proportional to the second derivative of the call function (see (4.3)). Hence, the SPD is a valid density function over future values of the underlying asset price and must be nonnegative and integrate to one. Therefore, Yatchew and Härdle (2006) combine shape restrictions with nonparametric regression to estimate the call price function and the SPD within a single least squares procedure. Constraints include smoothness of various order derivatives, monotonicity and convexity of the call function and integration to one of the SPD. Confidence intervals and test procedures are implemented using bootstrap methods. In addition, they apply the procedures to option data on the DAX index.

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 (2005) cannot be applied directly. One possible way of evaluating 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 (2005). Based on this assumption, one can obtain the stochastic discount factor and then the physical probability density function, to which Hong and Li's (2005) 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 regarded as a proxy by a 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 on 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.

4.3 Nonparametric Asset Pricing

The CAP model and the arbitrage asset pricing theory (APT) have been cornerstones in theoretical and empirical finance for decades. A classical CAP model usually assumes a simple and stable linear relationship between an asset's systematic risk and its expected return; see the books by Campbell, Lo and MacKinlay (1997) and Cochrane (2001) for details. However, this simple relationship assumption has been challenged and rejected by several recent studies based on empirical evidences of time variation in betas and expected returns (as well as return volatilities). As with other models, one considers the conditional CAP models or nonlinear APT with time-varying betas to characterize the time variations in betas and risk premia. In particular, Fama and French (1992, 1993, 1995) use some instrumental variables such as book-to-market equity ratio and market equity, as proxies for some unidentified risk factors to explain the time variation in returns. Whereas Ferson (1989), Harvey (1989), Ferson and Harvey (1991, 1993, 1998, 1999), Ferson and Korajczyk (1995), and Jagannathan and Wang (1996) conclude that beta and market risk premium vary over time, therefore, a static CAP model should incorporate time variations in beta in the model. Although there is a vast amount of empirical evidences on time variation in betas and risk premia, there is no theoretical guidance on how betas and risk premia vary with time or variables that represent conditioning information. Many recent studies focus on modeling the variation in betas using continuous approximation and the theoretical framework of the conditional CAP models; see Cochrane (1996), Jagannathan and Wang (1996, 2002), Wang (2002, 2003), Ang and Liu (2004) and the references therein. Recently, Ghysels (1998) discusses the problem in detail and stresses the impact of misspecification of beta risk dynamics on inference and estimation. Also, he argues that betas change through time very slowly and linear factor models like the conditional CAP model may have a tendency to overstate the time variation. Further, Ghysels (1998) shows that among several well-known time-varying beta models, a serious misspecification produces time variation in beta that is highly volatile and leads to large pricing errors. Finally, Ghysels (1998) concludes that it is

better to use the static CAP model in pricing when we do not have a proper model to capture time variation in betas correctly.

It is well documented that large pricing errors could be due to the linear approach used in a nonlinear model, and treating a non-linear relationship as a linear could lead to serious prediction problems in estimation. To overcome these problems, some nonlinear models have been considered in the recent literature. Following are some examples. Bansal, Hsieh and Viswanathan (1993) and Bansal and Viswanathan (1993) advocate the idea of a flexible stochastic discount factor model in empirical asset pricing and they focus on nonlinear arbitrage pricing theory models by assuming that the SDF is a nonlinear function of a few state variables. Further, Akdeniz, Altay-Salih and Caner (2003) test for the existence of significant evidence of non-linearity in the time series relationship of industry returns with market returns using the heteroskedasticity consistent Lagrange multiplier test of Hansen (1996) under the framework of the threshold model and they find that there exists statistically significant non-linearity in this relationship with respect to real interest rates. Wang (2002, 2003) explores a nonparametric form of the SDF model and conducted a test based on the nonparametric model. Parametric models for time-varying betas can be the most efficient if the underlying betas are correctly specified. However, a misspecification may cause serious bias, and model constraints may distort the betas in local area.

To follow the nations from Bansal, Hsieh and Viswanathan (1993), Bansal and Viswanathan (1993), Ghysels (1998) and Wang (2002, 2003), which are slightly different from those used in (4.1), a very simplified version of the SDF framework for asset pricing admits a basic pricing representation, which is a special case of model (4.1),

$$E[m_{t+1} r_{i,t+1} | \Omega_t] = 0, \tag{4.4}$$

where Ω_t denotes the information set at time t , m_{t+1} is the SDF or the pricing kernel, and $r_{i,t+1}$ is the excess return on the i -th asset or portfolio. Here $\varepsilon_{t+1} = m_{t+1} r_{i,t+1}$ is called the pricing error. In empirical finance, different models impose different constraints on the SDF. Particularly, the SDF is usually assumed to be a linear function of factors in various applications and then it becomes the well known CAP model; see Jagannathan and Wang (2002) and Wang (2003). Indeed, Jagannathan and Wang (2002) give the detailed comparison of the SDF and CAP model representations. Further, when the SDF is fully parameterized such as linear form, the general method of moments (GMM) of Hansen (1982) can be used to estimate parameters and test the model; see Campbell, Lo and MacKinlay (1997) and Cochrane (2001) for details.

Recently, Bansal, Hsieh and Viswanathan (1993) and Bansal and Viswanathan (1993) assume that m_{t+1} is a nonlinear function of a few state variables. Since the exact form of the nonlinear pricing kernel is unknown, Bansal and Viswanathan (1993) suggest using the polynomial expansion to approximate it and then apply the GMM for estimating and testing. As pointed out by Wang (2003), although this approach is intuitive and general, one of the shortcomings is that it is difficult to obtain the distribution theory and the effective assessment of finite sample performance. To overcome this difficulty, instead of considering the nonlinear pricing kernel, Ghysels (1998) focuses on the nonlinear parametric model and uses a set of moment conditions suitable for GMM estimation of parameters involved. Wang (2003) studies the nonparametric conditional CAP model and gives an explicit expression for the pricing kernel m_{t+1} ; that is, $m_{t+1} = 1 - b(Z_t) r_{p,t+1}$, where Z_t is a $k \times 1$ vector of conditioning variables from Ω_t , $b(Z_t) = E(r_{p,t+1}|Z_t)/E(r_{p,t+1}^2|Z_t)$ is an unknown function, and $r_{p,t+1}$ is the return on the market portfolio in excess of the riskless rate. Since the functional form of $b(\cdot)$ is unknown, Wang (2003) suggests estimating $b(\cdot)$ by using the Nadaraya-Watson method to two regression functions $E(r_{p,t+1}|Z_t)$ and $E(r_{p,t+1}^2|Z_t)$. Also, he conducts a simple nonparametric test about the pricing error. Indeed, his test is the well known F -test by running a multiple regression of the estimated pricing error $\hat{\varepsilon}_{t+1}$ versus a group of information variables; see (4.9) later for details. Further, Wang (2003) extends this setting to multifactor models by allowing $b(\cdot)$ to change over time; that is, $b(Z_t) = b(t)$. Finally, Bansal, Hsieh and Viswanathan (1993), Bansal and Viswanathan (1993), and Ghysels (1998) do not assume that m_{t+1} is a linear function of $r_{p,t+1}$ and instead they consider a parametric model by using the polynomial expansion.

To combine the models studied by Bansal, Hsieh and Viswanathan (1993), Bansal and Viswanathan (1993), Ghysels (1998) and Wang (2002, 2003), and some other models in the finance literature under a very general framework, Cai, Kuan and Sun (2008a) assume that the nonlinear pricing kernel has the form of $m_{t+1} = 1 - m(Z_t)r_{p,t+1}$, where $m(\cdot)$ is unspecified and they focus on the following nonparametric APT model

$$E[\{1 - m(Z_t) r_{p,t+1}\} r_{i,t+1} | \Omega_t] = 0, \quad (4.5)$$

where $m(\cdot)$ is an unknown function of Z_t which is a $k \times 1$ vector of conditioning variables from Ω_t . Indeed, (4.5) can be regarded as a moment (orthogonal) condition. The main interest of (4.5) is to identify and estimate the function $m(Z_t)$ as well as test whether the model is correctly specified.

Let I_t be a $q \times 1$ ($q \geq k$) vector of conditional variables from Ω_t , including Z_t , satisfying

the following orthogonal condition

$$E[\{1 - m(Z_t)r_{p,t+1}\} r_{i,t+1} | I_t] = 0, \quad (4.6)$$

which can be regarded as an approximation of (4.5). It follows from the orthogonality condition in (4.6) that, for any vector function $Q(V_t) \equiv Q_t$ with a dimension d_q specified later,

$$E [Q_t \{1 - m(Z_t)r_{p,t+1}\} r_{i,t+1} | I_t] = 0,$$

and its sample version is

$$\frac{1}{T} \sum_{t=1}^T Q_t \{1 - m(Z_t)r_{p,t+1}\} r_{i,t+1} = 0. \quad (4.7)$$

Therefore, Cai, Kuan and Sun (2008a) propose a new nonparametric estimation procedure to combine the orthogonality conditions given in (4.7) with the local linear fitting scheme of Fan and Gijbels (1996) to estimate the unknown function $m(\cdot)$. This nonparametric estimation approach is called by Cai, Kuan and Sun (2008a) as the nonparametric generalized method of moment (NPGMM).

For a given grid point z_0 and $\{Z_t\}$ in a neighborhood of z_0 , the orthogonality conditions in (4.7) can be approximated by the following locally weighted orthogonality conditions

$$\sum_{t=1}^T Q_t [1 - (a - b^\top (Z_t - z_0))r_{p,t+1}] r_{i,t+1} K_h(Z_t - z_0) = 0, \quad (4.8)$$

where $K_h(\cdot) = h^{-k}K(\cdot/h)$, $K(\cdot)$ is a kernel function in \mathbb{R}^k , and $h = h_n > 0$ is a bandwidth, which controls the amount of smoothing used in the estimation. (4.8) can be viewed as a generalization of the nonparametric estimation equations in Cai (2003) and the locally weighted version of (9.2.29) in Hamilton (1994, p.243). Therefore, solving the above equations leads to the NPGMM estimate of $m(z_0)$, denoted by $\widehat{m}(z_0)$, which is \widehat{a} , where $(\widehat{a}, \widehat{b})$ is the minimizer of (4.8). Cai, Kuan and Sun (2008a) discuss how to choose Q_t and derive the asymptotic properties of the proposed nonparametric estimator.

Let $\widehat{e}_{i,t+1}$ be the estimated pricing error; that is, $\widehat{e}_{i,t+1} = \widehat{m}_{t+1} r_{i,t+1}$, where $\widehat{m}_{t+1} = 1 - \widehat{m}(Z_t)r_{p,t+1}$. To test $E(e_{i,t+1} | \Omega_t) = 0$, Wang (2002, 2003) considers a simple test as follows. First, to run a multiple regression

$$\widehat{e}_{i,t+1} = V_t^\top \delta_i + v_{i,t+1}, \quad (4.9)$$

where V_t is a $q \times 1$ ($q \geq k$) vector of observed variables from Ω_t ,⁸ and then test if all the regression coefficients are zero; that is, $\mathbb{H}_0 : \delta_1 = \dots = \delta_q = 0$. By assuming that the

⁸Wang (2003) takes V_t to be Z_t in his empirical analysis.

distribution of $v_{i,t+1}$ is normal, Wang (2002, 2003) uses a conventional F-test. Also, Wang (2002) discusses two alternative test procedures. Indeed, the above model can be viewed as a linear approximation of $E[e_{i,t+1} | V_t]$. To examine the magnitude of pricing errors, Ghysels (1998) considers the mean square error (MSE) as a criterion to test if the conditional CAP model or APT model is misspecified relative to the unconditional one.

To check the misspecification of the model, Cai, Kuan and Sun (2008b) consider the testing hypothesis \mathbb{H}_0 :

$$\mathbb{H}_0 : m(\cdot) = m_0(\cdot) \quad \text{versus} \quad \mathbb{H}_a : m(\cdot) \neq m_0(\cdot), \quad (4.10)$$

where $m_0(\cdot)$ has a particular form. For example, if $m_0(\cdot) = b(\cdot)$, where $b(\cdot)$ is given in Wang (2003), this test is about testing the mean-covariance efficiency. If $m(\cdot)$ is a linear function, the test reduces to testing whether the linear pricing kernel is appropriate. Then, Cai, Kuan and Sun (2008b) construct a consistent nonparametric test based on a U-Statistics technique, described as follows. Since I_t is a $q \times 1$ ($q \geq k$) vector of observed variables from Ω_t , similar to Wang (2003), I_t is taken to be Z_t . It is clear that $E(e_{i,t+1} | Z_t) = 0$, where $e_{i,t+1} = [1 - m_0(Z_t)r_{p,t+1}]r_{i,t+1}$, if and only if $[E(e_{i,t+1} | Z_t)]^2 f(Z_t) = 0$, and if and only if $E[e_{i,t+1}E(e_{i,t+1} | Z_t)f(Z_t)] = 0$, where $f(\cdot)$ is the density of Z_t . Interestingly, the testing problem on conditional moment becomes unconditional. Obviously, the test statistic could be postulated as

$$U_T = \frac{1}{T} \sum_{t=1}^T e_{i,t+1} E(e_{i,t+1} | Z_t) f(Z_t), \quad (4.11)$$

if $e_{i,t+1}E(e_{i,t+1} | Z_t)f(Z_t)$ would be known. Since $E(e_{i,t+1} | Z_t)f(Z_t)$ is unknown, its leave-one-out Nadaraya-Watson estimator can be formulated as

$$\widehat{E}(e_{i,t+1} | Z_t)f(Z_t) = \frac{1}{T-1} \sum_{s \neq t}^T e_{i,s+1} K_h(Z_s - Z_t). \quad (4.12)$$

Plugging (4.12) into (4.11) and replacing $e_{i,t+1}$ by its estimate $\widehat{e}_{i,t+1} = \widehat{e}_t$, one obtain the test statistic, denoted by \widehat{U}_T ,

$$\widehat{U}_T = \frac{1}{T(T-1)} \sum_{s \neq t} K_h(Z_s - Z_t) \widehat{e}_s \widehat{e}_t, \quad (4.13)$$

which is indeed a second order U-statistics. Finally, Cai, Kuan and Sun (2008b) show that this nonparametric test statistic is consistent. In addition, they apply the proposed testing procedure to test if either the CAP model or the Fama and French model, in the flexible nonparametric form, can explain the momentum profit which is the value-weighted portfolio

of NYSE stocks as the market portfolio, using the dividend price ratio, the default premium, the one-month Treasury bill rate, and the excess return on the NYSE equally weighted portfolio as the conditioning variables.

5 Nonparametric Predictive Models for Asset Returns

The predictability of stock returns has been studied for the last two decades as a cornerstone research topic in economics and finance,⁹ and it is now routinely used in studies of many financial applications such as mutual fund performances, tests of the conditional capital asset pricing and optimal asset allocations.¹⁰ Tremendous empirical studies document the predictability of stock returns using various lagged financial variables, such as the log dividend-price ratio, the log earning-price ratio, the log book-to-market ratio, the dividend yield, the term spread and default premium, and the interest rates. Important questions are often asked about whether the returns are predictable and whether the predictability is stable over time. Since many of the predictive financial variables are highly persistent and even nonstationary, it is really challenging econometrically or statistically to answer these questions.

Predictability issues are generally assessed in the context of parametric predictive regression models in which rates of returns are regressed against the lagged values of stochastic explanatory variables (or state variables). Mankiw and Shapiro (1986) and Stambaugh (1986) were first to discern the econometric and statistical difficulties inherent in the estimation of predictive regressions through the structural predictive linear model as

$$y_t = \alpha_0 + \alpha_1 x_{t-1} + \varepsilon_t, \quad x_t = \rho x_{t-1} + u_t, \quad 1 \leq t \leq n \quad (5.1)$$

where y_t is the predictable variable, say excess stock return at time t , innovations $\{(\varepsilon_t, u_t)\}$ are independently and identically distributed (iid) bivariate normal $N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon u} \\ \sigma_{\varepsilon u} & \sigma_u^2 \end{pmatrix}$, and x_{t-1} is the first lag of a financial variable such as the log dividend-price ratio, which is commonly modelled by an AR(1) model as the second equation in (5.1).

There are several limitations to model (5.1) that should be seriously considered. First, note that the correlation between two innovations ε_t and u_t in (5.1) is $\phi = \sigma_{\varepsilon u} / \sigma_\varepsilon \sigma_u$, which

⁹See, for example, Fama and French (1988), Keim and Stambaugh (1986), Campbell and Shiller (1988), Cutler, Poterba, and Summers (1991), Balvers, Cosimano, and McDonald (1990), Schwert (1990), and Fama (1990), Kothari and Shanken (1997)

¹⁰See, Christopherson et al. (1998), Ferson and Schadt (1996), Ferson and Harvey (1991), Ghysels (1998), Ait-Sahalia and Brandt (2001), Barberis (2000), Brandt (1999), Campbell and Viceira (1998), and Kandel and Stambaugh (1996).

is unfortunately non-zero for many empirical applications; see, for example, see Table 4 in Campbell and Yogo (2006) and Table 1 in Torous, Valkanov, and Yan (2004) for some real applications. This creates the so called “endogeneity” (x_{t-1} and ε_t may be correlated) problem which makes modeling difficult and produces biased estimation. Another difficulty comes from the parameter ρ , which is the unknown degree of persistence of the variable x_t . That is, x_t is stationary if $|\rho| < 1$; see Viceira (1997), Amihud and Hurvich (2004), Paye and Timmermann (2006), and Dangl and Halling (2007), or it is unit root or integrated if $\rho = 1$, denoted by I(1); see Park and Hanh (1999), Chang and Martinez-Chombo (2003), and Cai, Li and Park (2008), or it is local-to-unity or nearly integrated if $\rho = 1 + c/n$ for some $c < 0$, denoted by NI(1); see, Elliott and Stock (1994), Cavanagh, Elliott, and Stock (1995), Torous, Valkanov, and Yan (2004), Campbell and Yogo (2006), Polk, Thompson, and Vuolteenaho (2006), and Rossi (2007), among others. This means that the predictive variable x_t is highly persistent, and even nonstationary, which may cause troubles for econometric modeling.

The third difficulty is the instability issue of the return predictive model. In fact, in return predictive models based on financial instruments such as the dividend and earnings yield, short interest rates, term spreads, and default premium, and so on, there have been many evidences on the instability of prediction model, particularly based on the dividend and earnings yield and the sample from the second half of the 1990s. This leads to the conclusion that the coefficients should change over time; see, for example, Viceira (1997), Lettau and Ludvigsson (2001), Goyal and Welch (2003), Paye and Timmermann (2006), Ang and Bekaert (2007), and Dangl and Halling (2007). While the aforementioned studies found evidences of instability in return predictive models, they did not provide any guideline on how the coefficients change over the time and where the return models may have changed. It is well known that if return predictive models are unstable, one can only assess the economic significance of return predictability provided it can be determined how widespread such instability changes over time and the extent to which it affects the predictability of stock returns. Therefore, all of the foregoing difficulties about the classical predictive regression models motivate us to propose a new varying-coefficient predictive regression model. The proposed model is not only interesting in its applications to finance and economics but also important in enriching the econometric theory.

As shown in Nelson and Kim (1993), because of the endogeneity, the ordinary least squares (OLS) estimate of the slope coefficient α_1 in (5.1) and its standard errors are substantially biased in finite samples if x_t is highly persistent, not really exogenous, and even nonstationary. Conventional tests based on standard t -statistics from OLS estimates tend to over-reject the null of non-predictability in Monte Carlo simulations. Some improvements

have been developed recently to deal with the bias issue. For example, the first order bias-correction estimator is proposed by Stambaugh (1999) based on Kendall (1954)'s analytical result for the bias expression of the least squares estimate of ρ , while Amihud and Hurvich (2004) propose a two-stage least squares estimator by using a linear projection of ε_t onto u_t . Finally, the conservative bias-adjusted estimator is proposed by Lewellen (2004) if ρ is very close to one for some predicting variables. Unfortunately, all of them still have not overcome the instability difficulty mentioned above. To deal with the instability problems, Paye and Timmermann (2006) analyze the excess returns on international equity indices related to state variables such as the lagged dividend yield, short interest rate, term spread and default premium, to investigate how widespread the evidence of structural breaks is and to what extent breaks affect the predictability of stock returns. Finally, Dangl and Halling (2007) consider equity return prediction model with random coefficients generated from a unit root process, related to 16 state variables.

Cai and Wang (2008a) consider a time-varying coefficient predictive regression model to allow the coefficients α_0 and α_1 in (5.1) to change over time (to be function of time), denoted by $\alpha_0(t)$ and $\alpha_1(t)$. They use a nonlinear projection of ε_t onto u_t ; that is $\varepsilon_t = \alpha_2(t) u_t + v_t$, and then model (5.1) becomes the following time-varying coefficient predictive model

$$y_t = \alpha_0(t) + \alpha_1(t)x_{t-1} + \alpha_2(t) u_t + v_t, \quad x_t = \rho x_{t-1} + u_t, \quad 1 \leq t \leq n. \quad (5.2)$$

They apply the local linear method to find the nonparametric estimates for $\alpha_j(t)$ and derive the asymptotic properties for the proposed estimator. Also, they derive the limiting distribution of the proposed nonparametric estimator, which is a mixed normal with conditional variance being a function of integrations of an Ornstein-Uhlenbeck process (mean-reverting process). They also show that the convergence rates for the intercept function (the regular rate at $(nh)^{1/2}$) and the slope function (a faster rate at $(n^2h)^{1/2}$) are totally different due to the NI(1) property of the state variable, although the asymptotic bias, coming from the local linear approximation, is the same as the stationary covariate case. Therefore, to estimate the intercept function optimally, Cai and Wang (2008a) propose a two-stage optimal estimation procedure similar to the profile likelihood method; see, e.g., Speckman (1988), Cai (2002a, b), and Cai, Li, and Park (2008), and they also show that the proposed two-stage estimator reaches indeed the optimality.

Cai and Wang (2008b) consider some consistent nonparametric tests for testing the null hypothesis of whether a parametric linear regression model is suitable or if there is no relationship between the dependent variable and predictors. Therefore, these testing problems

can be postulated as the following general testing hypothesis:

$$\mathbb{H}_0 : \alpha_j(t) = \alpha_j(t, \theta_j), \quad (5.3)$$

where $\alpha_j(t, \theta_j)$ is a known function with unknown parameter θ_j . If $\alpha_j(t, \theta_j)$ is constant, (5.3) becomes to test if model (5.1) is appropriate. If $\alpha_1(t, \theta_1) = 0$, it is to test if there exists predictability. If $\alpha_j(t, \theta_j)$ is a piecewise constant function, it is to test whether there exists any structural change. Cai and Wang (2008b) propose a nonparametric test which is a U-statistic type, similar to (4.13), and they also show that the proposed test statistic has different asymptotic behaviors depending on the stochastic properties of x_t . Specifically, Cai and Wang (2008b) address the following two scenarios: (a) x_t is non-stationary (either I(1) or NI(1)); (b) x_t contains both stationary and non-stationary components. Cai and Wang (2008a, 2008b) apply the estimation and testing procedures described above to consider the instability of predictability of some financial variables. Their test finds evidence for instability of predictability for the dividend-price and earnings-price ratios. They also find evidence for instability of predictability with the short rate and the long-short yield spread, for which the conventional test leads to valid inference.

For the linear projection used by Amihud and Hurvich (2004), it is implicitly assumed that the joint distribution of two innovations ε_t and u_t in model (5.1) is normal and this assumption might not hold for all applications. To relax this harsh assumption, Cai (2008) considers a nonlinear projection of ε_t onto x_{t-1} instead of u_t as $\varepsilon_t = \phi(x_{t-1}) + v_t$, so that $E(v_t|x_{t-1}) = 0$. Therefore, the endogeneity is removed. Then, model (5.1) becomes to the following classical regression model with nonstationary predictors

$$y_t = g(x_{t-1}) + v_t, \quad x_t = \rho x_{t-1} + u_t, \quad 1 \leq t \leq n \quad (5.4)$$

where $g(x_{t-1}) = \alpha_0 + \alpha_1 x_{t-1} + \phi(x_{t-1})$ and $E(v_t|x_{t-1}) = 0$. Now, for model (5.4), the testing predictability $\mathbb{H}_0 : \alpha_1 = 0$ for model (5.1) as in Campbell and Yogo (2006) becomes the testing hypothesis $\mathbb{H}_0 : g(x) = c$ for model (5.4), which is indeed more general. To estimate $g(\cdot)$ nonparametrically, Cai (2008) uses a local linear or local constant method and derives the limiting distribution of the nonparametric estimator when x_t is an I(1) process. It is interesting to note that the limiting distribution of the proposed nonparametric estimator is a mixed normal with a conditional variance associated with a local time of a standard Brownian motion and the convergence rate is $\sqrt{n^{1/2}h}$ instead of the conventional rate \sqrt{nh} . Furthermore, Cai (2008) proposes two test procedures. The first one is similar to the testing approach proposed in Sun, Cai and Li (2008) when x_t is integrated and the second one is to use the generalized likelihood ratio type testing procedure as in Cai, Fan and Yao

(2000) and the bootstrap. Finally, Cai (2008) applies the aforementioned estimation and testing procedures to consider the predictability of some financial instruments. The tests find some strong evidences that the predictability exists for the log dividend-price ratio, log earnings-price ratio, the short rate, and the long-short yield spread.

6 Conclusion

Over the last several years, nonparametric methods for both continuous and discrete time 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.

Finally, we would like to point out that the paper by Cai and Li (2009) gives a comprehensive survey on some recent developments in nonparametric econometrics, including nonparametric estimation and testing of regression functions with mixed discrete and continuous covariates, nonparametric estimation/testing with nonstationary data, nonparametric instrumental variable estimations, and nonparametric estimation of quantile regression models, which can be applied to financial studies. Other two promising lines of nonparametric finance are nonparametric volatility (conditional variance) and ARCH or GARCH type models and nonparametric methods in volatility for high-frequency data with/without microstructure noise. The reader interested in these areas of research should consult with the recent works, to name just a few, including Fan and Wang (2007), Long, Su and Ullah (2009), and Mishra, Su and Ullah (2009) and the references therein. Unfortunately, these topics are omitted in this paper due to too vast literature. However, we will write a separate survey paper on this important financial area, which is volatility models for both low-frequency and high-frequency data.

References

- Ahn, D.H., R.F. Dittmar and A.R. Gallant (2002). Quadratic term structure models: Theory and evidence, *The Review of Financial Studies*, **15**, 243-288.
- Ahn, D.H. and B. Gao (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. (2008). Closed-form likelihood expansions for multivariate diffusion, *Annals of Statistics*, **36**, 906-937.
- Ait-Sahalia, Y. and M. Brandt (2001). Variable selection for portfolio choice, *Journal of Finance*, **56**, 1297-1350.
- Ait-Sahalia, Y. and R. Kimmel (2007). Maximum likelihood estimation of stochastic volatility models, *Journal of Financial Economics*, **83**, 413-452.
- Ait-Sahalia, Y. and A.W. Lo (1998). Nonparametric estimation of state-price densities implicit in financial asset prices, *Journal of Finance*, **53**, 499-547.
- Ait-Sahalia, Y. and A.W. Lo (2000). Nonparametric risk management and implied risk aversion, *Journal of Econometrics*, **94**, 9-51.
- Akdeniz, L., A. Altay-Salih and M. Caner (2003). Time-varying betas help in asset pricing: The threshold CAPM, *Studies in Nonlinear Dynamics & Econometrics*, **6**, No.4, Article 1.
- Amihud, Y. and C. Hurvich (2004). Predictive regression: A reduced-bias estimation method, *Journal of Financial and Quantitative Analysis*, **39**, 813-841.
- Anagnou, I., M. Bedendo, S. Hodges and R. Tompkins (2005). The relation between implied and realized probability density functions, *Review of Futures Markets*, **11**, 41-66.
- Anders, U., O. Korn and C. Schmitt (1998). Improving the pricing of options: A neural network approach, *Journal of Forecasting*, **17**, 369-388.

- Andersen, T.G., L. Benzoni and J. Lund (2002). Towards an empirical foundation for continuous-time equity return models, *Journal of Finance*, **57**, 1239-1284.
- Andersen, T.G., H.-J. Chung and B.E. Sorensen (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 J. Lund (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate, *Journal of Econometrics*, **77**, 343-377.
- Ang, A. and G. Bekaert (2002a). Short rate nonlinearities and regime switches, *Journal of Economic Dynamics and Control*, **26**, 1243-1274.
- Ang, A. and G. Bekaert (2002b). Regime switches in interest rates, *Journal of Business and Economic Statistics*, **20**, 163-182.
- Ang, A. and G. Bekaert (2007). Stock return predictability: Is it there? *Review of Financial Studies*, **20**, 651-707.
- Ang, A. and J. Liu (2004). How to discount cashflows with time-varying expected return, *The Journal of Finance*, **59**, 2745-2783.
- Bahra, B. (1997). Implied risk-neutral probability density functions from option prices: Theory and application, *Working Paper*, Bank of England.
- Balvers, R.J., T.F. Cosimano and B. McDonald (1990). Predicting stock returns in an efficient market, *Journal of Finance*, **45**, 1109-1128.
- Bandi, F. (2000). Nonparametric fixed income pricing: theoretical issues, *Working Paper*, Graduate School of Business, The University of Chicago.
- Bandi, F. and T.H. Nguyen (2000). Fully nonparametric estimators for diffusions: a small sample analysis, *Working Paper*, Graduate School of Business, The University of Chicago.
- Bandi, F. and T.H. Nguyen (2003). On the functional estimation of jump-diffusion models, *Journal of Econometrics*, **116**, 293-328.
- Bandi, F. and P.C.B. Phillips (2003). Fully nonparametric estimation of scalar diffusion models, *Econometrica*, **71**, 241-283.
- Bansal, R., D.A. Hsieh and S. Viswanathan (1993). A new approach to international arbitrage pricing, *The Journal of Finance*, **48**, 1719-1747.
- Bansal, R. and S. Viswanathan (1993). No arbitrage and arbitrage pricing: A new approach, *The Journal of Finance*, **47**, 1231-1262.
- Barberis, N. (2000). Investing for the long run when returns are predictable, *Journal of Finance*, **55**, 225-264.

- Bates, D.S. (1991). The Crash of '87: Was it expected? The evidence from options markets, *Journal of Finance*, **46**, 1009-1044.
- Bates, D.S. (2000). Post-'87 crash fears in the S&P 500 futures option market, *Journal of Econometrics*, **94**, 181-238.
- Black, F., E. Derman and W. Toy (1990). A one-factor model of interest rates and its application to treasury bond options, *Financial Analysts Journal*, **46**, 33-39.
- Black, F. and P. Karasinski (1991). Bond and option pricing when short rates are log-normal, *Financial Analysts Journal*, **47**, 52-59.
- Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities, *Journal of Political Economy*, **71**, 637-654.
- Bliss, R.R. and D. Smith (1998). The elasticity of interest rate volatility: Chan, Karolyi, Longstaff, and Sanders revisited, *Journal of Risk*, **1**, 21-46.
- Bollerslev, T. and H. Zhou (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility, *Journal of Econometrics*, **109**, 33-65.
- Brandt, M.W. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach, *Journal of Finance*, **54**, 1609-1646.
- Brandt, M.W. and P. Santa-Clara (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 R.H. Litzenberger (1978). Prices of state contingent claims implicit in option prices, *Journal of Business*, **51**, 621-51.
- Brennan, M.J. and E. Schwartz (1979). A continuous time approach to the pricing of bonds, *Journal of Banking and Finance*, **3**, 133-155.
- Brenner, R., R. Harjes and K. Kroner (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 P.H. Dybvig (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. (2002a). Two-step likelihood estimation procedure for varying-coefficient models. *Journal of Multivariate Analysis*, **82**, 189-209.
- Cai, Z. (2002b). A two-stage approach to additive time series models. *Statistica Neerlandica*, **56**, 415-433.

- Cai, Z. (2003). Nonparametric estimation equations for time series data, *Statistics and Probability Letters*, **62**, 379-390.
- Cai, Z. (2008). Nonparametric predictive regression models for asset returns. *Working paper*, Department of Mathematics and Statistics, University of North Carolina at Charlotte.
- Cai, Z., J. Fan and Q. Yao (2000). Functional-coefficient regression models for nonlinear time series, *Journal of the American Statistical Association*, **95**, 941-956.
- Cai, Z., C.M. Kuan and L. Sun (2008a). Nonparametric pricing kernel models, *Working Paper*, Department of Mathematics and Statistics, University of North Carolina at Charlotte.
- Cai, Z., C.M. Kuan and L. Sun (2008b). Nonparametric test for pricing kernel models, *Working Paper*, Department of Mathematics and Statistics, University of North Carolina at Charlotte.
- Cai, Z. and Q. Li (2009). Some recent developments on nonparametric econometrics. Forthcoming in *Advances in Econometrics*.
- Cai, Z., Q. Li and J.Y. Park (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, **148**, 101-113.
- Cai, Z. and R.C. Tiwari (2000). Application of a local linear autoregressive model to BOD time series. *Environmetrics*, **11**, 341-350.
- Cai, Z. and Y. Wang (2008a). Instability of predictability of asset returns, *Working paper*, Department of Mathematics and Statistics, University of North Carolina at Charlotte.
- Cai, Z. and Y. Wang (2008b). Testing stability of predictability of asset returns, *Working paper*, Department of Mathematics and Statistics, University of North Carolina at Charlotte.
- Cai, Z. and L. Zhang (2008a). Testing for discontinuous diffusion models versus jump diffusion models, *Working Paper*, The Wang Yanan Institute for Studies in Economics, Xiamen University.
- Cai, Z. and L. Zhang (2008b). Information effect for different firm-sizes via the nonparametric jump-diffusion model, *Working Paper*, The Wang Yanan Institute for Studies in Economics, Xiamen University.
- Campbell, J.Y., A.W. Lo and A.C. MacKinlay (1997). *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.
- Campbell, J. and R. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies*, **1**, 195-227.
- Campbell, J.Y. and L. Viceira (1998). Consumption and portfolio decisions when expected returns are time varying, *Quarterly Journal of Economics*, **114**, 433-495.

- Campbell, J. and M. Yogo (2006). Efficient tests of stock return predictability, *Journal of Financial Economics*, **81**, 27-60.
- Cavanagh, C.L., G. Elliott and J.H. Stock (1995). Inference in models with nearly integrated regressors, *Econometric Theory*, **11**, 1131-1147.
- Chan, K.C., G.A. Karolyi, F.A. Longstaff and A.B. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance*, **47**, 1209-1227.
- Chang, Y. and E. Martinez-Chombo (2003). Electricity demand analysis using cointegration and error-correction models with time varying parameters: The Mexican case, *Working paper*, Department of Economics, Texas A&M University.
- Chapman, D., J. Long and N. Pearson (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 N. Pearson (2000). Is the short rate drift actually nonlinear? *Journal of Finance*, **55**, 355-388.
- Chen, S.X., J. Gao and C. Tang (2008). A test for model specification of diffusion processes, *Annals of Statistics*, **36**, 167-198.
- Chen, S.X., W. Härdle and T. Kleinow (2002). An empirical likelihood goodness-of-fit test for time series, In *Applied Quantitative Finance* (W. Härdle, T. Kleinow, and G. Stahl, eds.), 259-281.
- Chernov, M., A.R. Gallant, E. Ghysels and G. Tauchen (2003). Alternative Models of Stock Price Dynamics, *Journal of Econometrics*, **116**, 225-257.
- Christopherson, J.A., W. Ferson and D.A. Glassman (1998). Conditioning manager alphas on economic information: another look at the persistence of performance, *Review of Financial Studies*, **11**, 111-142.
- Chung, C.C. and G. Tauchen (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., L.P. Hansen, E.G.J. Luttmer and J.A. Scheinkman (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., J.E. Ingersoll and S.A. Ross (1980). An analysis of variable rate loan contracts, *Journal of Finance*, **35**, 389-403.
- Cox, J.C., J.E. Ingersoll and S.A. Ross (1985). A theory of the term structure of interest rates, *Econometrica*, **53**, 385-407.
- Cox, J.C. and S.A. Ross (1976). The Volatility of option for alternative stochastic processes, *Journal of Financial Economics*, **3**, 145-166.
- Cutler, D.M., J.M. Poterba and L.H. Summers (1991). Speculative dynamics, *Review of Economic Studies*, **58**, 529-546.
- Dai, Q. and K.J. Singleton (2000). Specification analysis of affine term structure models, *Journal of Finance*, **55**, 1943-1978.
- Dangl, T. and M. Halling (2007). Predictive regressions with time-varying coefficients, *Working Paper*, School of Business, University of Utah.
- Diebold, F.X., T. Gunther and A. Tay (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. (2001). *Dynamic Asset Pricing Theory*, 3th Edition, Princeton University Press, Princeton, NJ.
- Duffie, D. and R. Kan (1996). A yield factor model of interest rate, *Mathematical Finance*, **6**, 379-406.
- Duffie, D. and J. Pan (2001). Analytical value-at-risk with jumps and credit risk, *Finance and Stochastics*, **5**, 155-180.
- Duffie, D., J. Pan and K.J. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions, *Econometrica*, **68**, 1343-1376.
- Duffie, D. and K.J. Singleton (1993). Simulated moments estimation of Markov models of asset prices, *Econometrica*, **61**, 929-952.
- Egorov, A., Y. Hong and H. Li (2006). Validating forecasts of the joint probability density of bond yields: Can affine models beat random walk? *Journal of Econometrics*, **135**, 255-284.
- Egorov, A., H. Li and Y. Xu (2003). Maximum likelihood estimation of time-inhomogeneous diffusions, *Journal of Econometrics*, **114**, 107-139.
- Elliott, G. and J.H. Stock (1994). Inference in time series regression when the order of integration of a regressor is unknown, *Econometric Theory*, **10**, 672-700.

- Elerian, O., S. Chib and N. Shephard (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., M.S. Johannes and N.G. Polson (2003). The impact of jumps in volatility and returns, *Journal of Finance*, **58**, 1269-1300.
- Fama, E. (1970). Efficient capital markets: A review of theory and empirical work, *Journal of Finance*, **25**, 383-417.
- Fama, E.F. (1990). Stock returns, real returns, and economic activity, *Journal of Finance*, **45**, 1089-1108.
- Fama, E.F. and K.R. French (1988). Dividend yields and expected stock returns, *Journal of Financial Economics*, **22**, 3-26.
- Fama, E. and K.R. French (1992). The cross-section of expected stock returns, *The Journal of Finance*, **47**, 427-466.
- Fama, E. and K.R. French (1993). Common risk factors in the returns on bonds and stocks, *Journal of Financial Economics*, **33**, 3-56.
- Fama, E. and K.R. French (1995). Size and book-to-market factors in earning and returns, *The Journal of Finance*, **50**, 131-155.
- Fan, J. and I. Gijbels (1996). *Local Polynomial Modeling and its Applications*, London: Chapman and Hall.
- Fan, J., J. Jiang, C. Zhang and Z. Zhou (2003). Time-dependent diffusion models for term structure dynamics and the stock price volatility, *Statistica Sinica*, **13**, 965-992.
- Fan, J. and Y. Wang (2007). Multi-scale jump and volatility analysis for high-frequency financial data, *Journal of the American Statistical Association*, **102**, 1349-1362.
- Fan, J. and C. Zhang (2003). A re-examination of diffusion estimators with applications to financial model validation, *Journal of the American Statistical Association*, **98**, 118-134.
- Fan, J., C. Zhang and J. Zhang (2001). Generalized likelihood ratio statistics and Wilks phenomenon, *The Annals of Statistics*, **29**, 153-193.
- Ferson, W.E. (1989). Changes in expected security returns, risk and the level of interest rates, *The Journal of Finance*, **44**, 1191-1214.
- Ferson, W.E. and C.R. Harvey (1991). The variation of economic risk premiums, *Journal of Political Economy*, **99**, 385-415.

- Ferson, W.E. and C.R. Harvey (1993). The risk and predictability of international equity returns, *Journal of Financial Studies*, **6**, 527-566.
- Ferson, W.E. and C.R. Harvey (1998). Fundamental determinants of national equity market returns: A perspective on conditional asset pricing, *Journal of Banking and Finance*, **21**, 1625-1665.
- Ferson, W.E. and C.R. Harvey (1999). Conditional variables and the cross section of stock return, *The Journal of Finance*, **54**, 1325-1360.
- Ferson, W.E. and R.W. Schadt (1996). Measuring fund strategy and performance in changing economic conditions, *Journal of Finance*, **51**, 425-461.
- Ferson, W.E. and R.A. Korajczyk (1995). Do arbitrage pricing models explain the predictability of stock returns?, *Journal of Business*, **68**, 309-349.
- Hamilton, J.D. (1994). *Time Series Analysis*, Princeton: Princeton University Press.
- Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, **64**, 413-430.
- Hansen, L.P. (1982). Large sample properties of the generalized method of moments estimators, *Econometrica*, **50**, 1029-1054.
- Harvey, C.R. (1989). Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics*, **24**, 289-317.
- Gallant, A.R. and G. Tauchen (1996). Which moments to match?, *Econometric Theory*, **12**, 657-681.
- Gallant, A.R. and G. Tauchen (2001). Efficient method of moments, *Working Paper*, Department of Economics, University of North Carolina and Department of Economics, Duke University.
- Ghysels, E. (1998). On stable factor structures in the pricing of risk: Do time varying betas help or hurt?, *The Journal of Finance*, **53**, 549-573.
- Gourieroux, C., A. Monfort and E. Renault (1993). Indirect inference, *Journal of Applied Econometrics*, **8**, 85-118.
- Goyal, A. and I. Welch (2003). Predicting the equity premium with dividend ratios, *Management Science*, **49**, 639-654.
- Gray, S. (1996). Modeling the conditional distribution of interest rates as a regime switching process, *Journal of Financial Economics*, **42**, 27-62.
- Gourieroux, C. and J. Jasiak (2001). *Financial Econometrics: Problems, Models, and Methods*, Princeton University Press, Princeton, NJ.

- 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, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, **64**, 413-430.
- Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica*, **50**, 1029-1054.
- Hansen, L.P. and R. Janaganan (1997). Assessing specification errors in stochastic discount factor models, *Journal of Finance*, **52**, 557-590.
- Hansen, L.P. and J.A. Scheinkman (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 E. Mammen (1993). Comparing nonparametric versus parametric regression fits, *The Annals of Statistics*, **21**, 1926-1947.
- Harrison, J.M. and D.M. Kreps (1979). Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory*, **20**, 381-408.
- Heath, D.C., R.A. Jarrow and A. Morton (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 S.B. Lee (1986). Term structure movements and pricing interest rate contingent claims, *Journal of Finance*, **41**, 1011-1029.
- Hong, Y. and T.H. Lee (2003a). Inference and forecast of exchange rates via generalized spectrum and nonlinear time series models, *The Review of Economics and Statistics*, **85**, 1048-1062.
- Hong, Y. and T.H. Lee (2003b). Diagnostic checking for nonlinear time series models, *Econometric Theory*, **19**, 1065-1121.
- Hong, Y. and H. Li (2005). Nonparametric specification testing for continuous-time models with applications to interest rate term structures, *The Review of Financial Studies*, **18**, 37-84.
- Hull, J. and H. White (1990). Pricing interest-rate derivative securities, *Review of Financial Studies*, **3**, 573-592.
- Hutchinson, J., A.W. Lo and T. Poggio (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 M. Rubinstein (1996). Recovering probability distributions from contemporary security prices, *Journal of Finance*, **51**, 1611-1631.
- Jacquier, E., N.G. Polson and P. Rossi (1994). Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics*, **12**, 371-389.
- Jagannathan, R. and Z. Wang (1996). The conditional CAPM and the cross-section of expected returns, *The Journal of Finance*, **51**, 3-53.
- Jagannathan, R. and Z. Wang (2002). Empirical evaluation of asset pricing models: A comparison of the SDF and beta methods, *The Journal of Finance*, **57**, 2337-2367.
- Jarrow, R. and A. Tudd (1982). Approximate option valuation for arbitrary stochastic processes, *Journal of Financial Economics*, **10**, 347-369.
- Jiang, G.J. and J.L. Knight (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 J.L. Knight (2002). Estimation of continuous time processes via the empirical characteristic function, *Journal of Business and Economic Statistics*, **20**, 198-212.
- Jiang, G.J. and J.L. Knight (2006). ECF Estimation of Markov models where the transition density is unknown. Manuscript.
- Jiang, G.J. and P.J. van der Sluis (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. (2004). The economic and statistical role of jumps to interest rates, *Journal of Finance*, **59**, 227-260.
- Johannes, M.S., R. Kumar and N.G. Polson (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.
- Kendall, M.G. (1954). Note on bias in the estimation of autocorrelation, *Biometrika*, **41**, 403-404.
- Kandel, S. and R. Stambaugh (1996). On the predictability of stock returns: an asset allocation perspective, *Journal of Finance*, **51**, 385-424.
- Karatzas, I. and S.E. Shreve (1988). *Brownian Motion and Stochastic Calculus*, 2th Edition, New York: Springer-Verlag.

- Keim, D.B. and R.F. Stambaugh (1986). Predicting returns in the stock and bond markets, *Journal of Financial Economics*, **17**, 357-390.
- Kleinow, T. (2002). Testing the diffusion coefficients, *Working paper*, Institute of Statistics and Economics, Humboldt University, Germany.
- Kothari, S.P. and J. Shanken (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics*, **44**, 169-203.
- Kou, S. (2002). A jump diffusion model for option pricing, *Management Science*, **48**, 1086-1101.
- Kristensen, D. (2007). Nonparametric estimation and misspecification testing of diffusion models. Manuscript.
- Kristensen, D. (2008). Pseudo-maximum likelihood estimation in two classes of semiparametric diffusion models. Manuscript.
- Lettau, M. and S. Ludvigsson (2001). Consumption, aggregate wealth, and expected stock returns, *Journal of Finance*, **56**, 815-849.
- Lewellen, J. (2004). Predicting returns with financial ratios, *Journal of Financial Economics*, **74**, 209-235.
- Li, Q. and J. Racine (2007). *Nonparametric Econometrics: Theory and Applications*, Princeton University Press.
- Liu, M. (2000). Modeling long memory in stock market volatility, *Journal of Econometrics*, **99**, 139-171.
- Liu, J., F.A. Longstaff and J. Pan (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 A.C. MacKinlay (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.
- Long, X., L. Su and A. Ullah (2009). Estimation and forecasting of dynamic conditional covariance: A semiparametric multivariate model. Manuscript.
- 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 E. Schwartz (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model, *Journal of Finance*, **47**, 1259-1282.
- Mankiw, N.G. and M. Shapiro (1986). Do we reject too often? Small sample properties of tests of rational expectation models, *Economics Letters*, **20**, 139-145.
- Marsh, T. and E. Rosenfeld (1983). Stochastic processes for interest rates and equilibrium bond prices, *Journal of Finance*, **38**, 635-646.
- Melick, W.R. and C.P. Thomas (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.
- Mishra, S. L. Su and A. Ullah (2009). Semiparametric estimator of time series conditional variance. Manuscript.
- Mittelhammer, R.C., G.G. Judge and D.J. Miller (2000). *Econometrics Foundation*, New York: Cambridge University Press.
- Nelson, C.R. and M.J. Kim (1993). Predictable stock returns: The role of small sample bias, *Journal of Finance*, **48**, 641-661.
- Øksendal, B. (1985). *Stochastic Differential Equations: An Introduction with Applications*, 3th edition, New York: Springer-Verlag.
- Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics*, New York: Cambridge University Press.
- Pan, J. (1997). Stochastic volatility with reset at jumps, *Working Paper*, School of Management, MIT.
- Park, J.Y. and S.B. Hahn (1999). Cointegrating regressions with time varying coefficients, *Econometric Theory*, **15**, 664-703.
- Paye, B.S. and A. Timmermann (2006). Instability of return prediction models, *Journal of Empirical Finance*, **13**, 274-315.
- 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.

- Polk, C., S. Thompson and T. Vuolteenaho (2006). Cross-sectional forecasts of the equity premium, *Journal of Financial Economics*, **81**, 101-141.
- 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.
- Rossi, B. (2007). Expectation hypothesis tests and predictive regressions at long horizons, *Econometrics Journal*, **10**, 1-26.
- Rubinstein M. (1994). Implied binomial trees, *Journal of Finance*, **49**, 771-818.
- Schwert, G.W. (1990). Stock returns and real activity: A century of evidence, *Journal of Finance*, **45**, 1237-1257.
- Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance*, **19**, 425-442.
- 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.
- Speckman, P. (1988). Kernel smoothing in partially linear models, *Journal of the Royal Statistical Society, Series B*, **50**, 413-436.
- Sun, Y., Z. Cai and Q. Li (2008). Consistent nonparametric test on parametric smooth coefficient model with nonstationary data, Submitted to *Journal of Econometrics*.
- Sundaresan, S. (2001). Continuous-time methods in finance: A review and an assessment, *Journal of Finance*, **55**, 1569-1622.
- Stambaugh, R. (1986). Bias in regressions with lagged stochastic regressors, *Working Paper*, University of Chicago.
- Stambaugh, R. (1999). Predictive regressions, *Journal of Financial Economics*, **54**, 375-421.
- Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance*, **52**, 1973-2002.
- Torous, W., R. Valkanov and S. Yan (2004). On predicting stock returns with nearly integrated explanatory variables, *Journal of Business*, **77**, 937-966.

- 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.
- Taylor, S. (2005). *Asset Price Dynamics, Volatility, And Prediction*, Princeton University Press, Princeton, NJ.
- Tsay, R.S. (2005). *Analysis of Financial Time Series*, 2th Edition, John Wiley & Sons, New York.
- 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.
- Viceira, L.M. (1997). Testing for structural change in the predictability of asset returns, Manuscript, Harvard University.
- Wang, K. (2002). Nonparametric tests of conditional mean-variance efficiency of a benchmark portfolio, *Journal of Empirical Finance*, **9**, 133-169.
- Wang, K.Q. (2003). Asset pricing with conditioning information: A new test, *The Journal of Finance*, **58**, 161-196.
- Yatchew, A. and W. Härdle (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap, *Journal of Econometrics*, **133**, 579-599.
- Xu, K. and P.B.C. Phillips (2007). Tilted nonparametric estimation of volatility functions, Cowles Foundation Discussion Paper No. 1612R, Department of Economics, Yale University.
- Zhou, H. (2001). Jump-diffusion term structure and Ito conditional moment generator, *Working Paper*, Federal Reserve Board.