Is the Drift of the Interest Rate Process Linear? A New Approach and Evidence

Yongmiao Hong
Department of Economics, Cornell University & WISE, Xiamen University
Yoon-Jin Lee
Department of Economics, Indiana University
Zhaogang Song
Department of Economics, Cornell University

April 2009

Yongmiao Hong, Department of Economics, Uris Hall, Cornell University, Ithaca, NY 14850, USA and Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, Xiamen 361005, Fujian, China. Email: yh20@cornell.edu; and Yoon-Jin Lee, Department of Economics, Indiana University, Wylie Hall, Bloomington, IN 47405, USA, Email: lee243@indiana.edu; Zhaogang Song, Department of Economics, Uris Hall, Cornell University, Ithaca, NY 14850, USA, Email: zs43@cornell.edu. Hong thanks supports from China National Science Foundation and the Cheung Kong Scholarship of Chinese ministry of Education and Xiamen University, and Lee thanks for financial support from Indiana University.

ABSTRACT

Continuous-time models are important for investigating interest rate term structure and pricing fixed income derivatives. Economic theory often provides little guidance on the choice of the form of continuous-time models, and existing one-factor and multi-factor continuous-time interest rate models often assume a linear drift, among other things. Some studies, based smoothed nonparametric kernel estimation, suggest that the drift of the interest rate process is nonlinear, particularly at high interest rate levels. However, this has been doubted as an artifact of smoothed nonparametric estimation in comparison with highly persistent interest rate data. Whether the drift of the interest rate process is linear or nonlinear remains an unsolved issue in the literature.

In this paper, we take a new approach to re-address this important issue by first considering a general continuous-time regression for the interest rate process and then testing it via a generalized spectral derivative approach of Hong and Lee (2005) which is tailored to the continuous-time setting. Our method avoids the undesirable features of smoothed nonparametric estimation for highly persistent financial time series data. Unlike the existing approaches to testing linearity in drift, we allow for stochastic volatility and jumps, which have been well documented for the interest rate process in the literature. An empirically realistic simulation study shows that the generalized spectral derivative provides reliable inference in finite samples for continuous-time models. Based on the widely used 7-day Eurodollar rates, we document strong evidence that the interest rate process has a nonlinear drift and such evidence is robust to the presence of level effect, stochastic volatility, jumps, and different methods of drift parameter estimation. We further document that such popular nonlinear drift models as Aït-Sahalia's (1996a) nonlinear drift model and Ahn and Gao's (1991) Inverse-Feller drift model can capture some nonlinear drift dynamics of the short rate, and Aït-Sahalia's nonlinear model outperforms Ahn and Gao's nonlinear drift model due to its flexibility to capture asymmetric mean-reverting feature. However, they are still firmly rejected, indicating room for further improving the modelling of the drift function of the interest rate.

Key Words: Continuous-Time Model, Diffusion Function, Drift Function, Interest Rate, Jumps, Robust Testing, Stochastic Volatility.

JEL NO: C12, C14

1. Introduction

Continuous-time diffusion models have been commonly used in many financial applications, such as in valuing and hedging the huge institutional holdings of fixed income securities and derivatives. Most continuous-time interest rate models involve specifications of a drift function, a diffusion function and a jump function. Economic theory often provides little guidance on specification of these functions, and existing continuous-time interest models usually employ somewhat arbitrary convenient functional forms. In particular, the most commonly used specification of drift, in both univariate (one-factor) and multivariate (multi-factor) modelling setups, is a linear drift (e.g., Vasicek (1977), Cox-Ingersoll-Ross (1985, CIR), Pearson and Sun (1994), Brennan and Schwartz (1979), Courtadon (1982), Black and Karasinski (1991), Chan, Karolyi, Longstaff and Sanders (1992), Duffie and Kan (1996), Andersen and Lund (1997), Duffie, Pan and Singleton (2000)). While the volatility may be estimated relatively accurately using high-frequency observations of the short term interest rate, it is well known that the short rate's high persistence makes the identification of the true shape of the drift function particularly difficult. For example, Jones (1999), who considers the Bayesian analysis of diffusion processes, concludes that "despite the fact that over five thousand observations of daily data are available, the data sample is effectively small. Precise statements about the shape of the drift function ... are impossible to make." For the purpose of pricing interest-rate derivative securities, it is important to identify precisely the instantaneous variance (i.e., diffusion function), since option prices depend on the shape of the diffusion function of the underlying spot interest rate. In addition, optimal hedging strategies for risk-averse investors also depend critically on the level of term structure volatility. Thus, many studies have focused on estimation of the diffusion function while imposing a parametric drift model. For example, in some studies (e.g., Aït-Sahalia (1996)), to estimate the diffusion function nonparametrically, which is characterized as a function of the drift and marginal density, a parametric specification of drift (linear drift) is assumed. If the linear drift model is misspecified, diffusion estimation will not be consistent (see, Lo and Wang (1995) for more discussion). Arapis and Gao (2006) show that imposing a parametric linear drift affects the estimation of the diffusion function. They find that the drift function may have a greater effect on pricing derivatives than what is viewed in the literature. Adequacy of a drift model is also important for understanding the drift dynamics (e.g., mean-reverting speed) and forecasting the short term interest rates. In summary, correct specification of the drift function is essential for consistent estimation of the diffusion function, correct derivatives pricing, and optimal forecasts of the short rate, among other things.

There has been an unresolved debate regarding nonlinearity of drift of the interest rate in the literature. Aït-Sahalia (1996) and Stanton (1997) use smoothed nonparametric kernel methods to estimate the drift and diffusion functions of the short rate. They find evidence of nonlinearity in drift. Specifically, the estimated drift function is highly nonlinear, especially for large values of the interest rate. Aït-Sahalia (1996a) finds nonlinear mean-reversion of the spot interest rate—around its mean, where the drift is essentially zero, the spot rate behaves like a random walk, reverting toward the mean strongly when far away from the mean (for very high or very low rates). He concludes that "the linearity of the drift imposed in the literature appears to be the main source of misspecification." On the other hand, Stanton (1997) finds little mean reversion for all rates below 15% but the estimated drift drops sharply and becomes negative as the interest rate increases beyond 15%. Similar results are obtained by Conley, Hansen, Luttmer, and Scheinkman (1997), who estimate a drift function that is nonzero only for rates below 3% or above 11%. Jiang and Knight (1997) also find a similar pattern of nonlinear mean reversion for Canadian interest rates.

It is well-known that there exists a boundary bias problem for smoothed nonparametric kernel estimation, due to the asymmetric coverage of data in the tails of a distribution (e.g., Härdle (1990), Hong and Li (2005)). This boundary bias may be particularly relevant, because high values of interest rates lie in the right-tail and such observations are sparse. More importantly, Chapman and Pearson (2000), using a striking simulation study, argue that the evidence of nonlinearity documented above may be spurious due to the nature of smoothed nonparametric kernel estimation in combination with highly persistent interest rate data. They find that the typical estimated drift function displays spurious nonlinearities at high levels of interest rates even though the true drift is linear, due to the finite sample bias of the truncation of a distribution. Such a finite sample bias still exists even if the boundary bias of the smoothed kernel estimator is corrected. The simulation study of Chapman and Pearson (2000) indicates that Aït-Sahalia (1996) and Stanton's (1997) results may not be capable of providing convincing evidence of a nonlinear drift, and the smoothed nonparametric methods used there cannot provide a reliable reference for highly persistent interest rate data. In addition the finite sample bias of the truncation of a distribution and the boundary bias problem, the shapes of these nonparametric estimators also depend on the choice of the bandwidth parameter, which is a delicate business — oversmoothed estimates tend to suggest a linear drift, whereas undersmoothed estimates are particularly susceptible to the truncation bias and correlated residual bias, thus resulting in a spurious nonlinear drift.

Jones (2003) also investigates linearity of drift using a Bayesian approach (MCMC) and argues that nonlinearity may be an outcome of special priors. Before looking at the data, a flat prior holder expects to conclude in favor of the existence of nonlinear drift even when it is not a true feature of the data. As in the autoregressive model, the flat prior therefore represents an informative prior belief that the model is stationary, and the flat prior in this case corresponds to a belief that the drift is nonlinear. Thus, the results suggest that the finding of a nonlinear drift highly depends on the choice of the sampling frequency, the type of prior, flat or Jeffreys prior, and the prior belief about whether interest rates are stationary. Durham (2002), using a simulated MLE, also finds that interest rate drift nonlinearity is more associated with noisy interest rate data, and a constant drift model is adequate. He suggests that the apparent transitory component not currently captured by the model motivates the adoption of a stochastic mean model of interest rates. Li, Pearson and Poteshman (2004) fail to reject a linear drift model but they argue that this cannot be used as evidence against nonlinearity. Recently, using panel data, Takamiazawa (2008) and Sam and Jiang (2007) find nonlinearity at high levels of the interest rate but the mean reversion is weaker than that in Stanton (1997) and Jiang (1998). Using interest rate data on the short end of the term structure, Takamiazawa (2008) finds that it is difficult to find strong evidence supporting nonlinear physical drift from a statistical perspective, but nonlinear risk-neutral drift is strongly supported with the time series and cross-sectional dimensions of data. Park (2008), using a novel approach based on martingale regression and time change, supports the hypothesis of linear drift for the short rate. In conclusion, the literature has not drawn a decisive conclusion regarding the nonlinearity in drift of interest rates yet. Different conclusions are mainly due to the use of different econometric methods as well as the use of different interest rate data. From an econometric perspective, when a test fails to reject linearity in drift, it might be due to low or little power of the test against certain alternatives of nonlinear drift (i.e., Type II error); on the other hand, when a test rejects linearity of drift, it might be due to the overrejection of the test when the null hypothesis of linear drift actually holds (i.e., Type I errors). It is important to use a test that can provide a reliable inference (i.e., does not overreject a correct null hypothesis) in finite samples and has good power against a vast range of nonlinear drift alternatives.

Most research in the interest rate literature is more concerned about estimation of continuous-time diffusion models rather than direct testing. They usually estimate a flexible parametric form of drift and check if some parameters associated with a specific nonlinear drift are zero (see, e.g., Chan *et al.* (1992), Chapman and Pearson (2000), Takamizawa (2008)). This approach essentially considers a specific type of nonlinear drift, and may overlook many important nonlinear drift alternatives. Alternatively, other researchers nonparametrically estimate the drift function and informally check if it is linear by examining the shape of the graph by eyes (see, e.g., Pritsker (1997), Chapman and Pearson (2000), Stanton (1997), Sam and Jiang (2007)). Strictly speaking, these studies may not be viewed as a formal, rigorous statistical rejection of the linearity of drift.

There have been a few studies on testing continuous-time diffusion models for the interest rate. Some develop formal tests to check the drift and diffusion specifications jointly by using the marginal density or transitional density (e.g., Aït-Sahalia (1996a), Hong and Li (2005), Gao and King (2004), Aït-Sahalia, Fan and Peng (2008)). These tests are not suitable to test the drift model because they cannot separate the sources of misspecification upon rejection. Others focus on the diffusion function only (e.g., Li (2007), Corradi and White (1999)). For those who test the drift function in a semiparametric setup (Kristensen (2008a), Gao and Casas (2008)), a parametric form of the diffusion function is assumed and thus these tests are subject to misspecification of the diffusion function. Kristensen (2008b) and Bandi and Phillips (2007) suggest nonparametric methods which could be used to test the drift specification but they do not pursue it. Bandi and Phillips (2007) use nonparametric smoothing of drift, which is expected to be subject to the truncation bias as pointed out in Chapman and Pearson (2000). On the other hand, Kristensen (2008b) compares a direct nonparametric transition density estimator and a diffusion-based semiparametric transition density. The latter usually does not have a closed form and is therefore difficult to be applied in practice.

In this paper, we will use time series interest rate data to examine the issue of linearity of drift by employing a new approach that avoids the undesired features of smoothed nonparametric estimation as pointed out by Chapman and Pearson (1999). The approach is made possible by first considering a general continuous-time regression for the interest rate and then testing it via a generalized spectral derivative approach developed in Hong and Lee (2005), which is tailored to the continuous-time setting in this paper. The generalized spectral derivative checks the martingale property of the discrete-time regression error of the drift model while being robust to a general error structure of unknown form. We allow but are not restricted to the class of one-factor diffusion process (with the diffusion function unspecified). We allow the presence of latent stochastic volatility and jumps. This is particularly appealing given the overwhelming evidence of stochastic volatility (e.g., Anderson and Lund (1997)) and jumps (e.g., Das (2002) and Johannes (2004)) in interest rates. Our method uses an empirical generalized autocovariance function which does not require smoothing at each lag. Thus, it is free of the boundary bias, and the truncation bias of the smoothed nonparametric estimation when applied to highly persistent interest rate data. Indeed, an empirically realistic simulation study confirms that the generalized spectral derivative test used in this paper provides a reliable finite sample reference for continuous-time models. Based on a widely used sample on the 7-day Eurodollar rates June 1, 1973, to February 25, 1995, we document striking evidence of a nonlinear drift with asymmetric mean-reverting features. Such popular nonlinear drift models as Aït-Sahalia's (1996a) nonlinear drift model and Ahn and Gao's (1991) Inverse-Feller drift model can capture some nonlinear drift dynamics of the short rate, and Aït-Sahalia's nonlinear model outperforms Ahn and Gao's nonlinear drift model due to its flexibility to capture asymmetric mean-reverting feature. However, these popular nonlinear drift models are still firmly rejected, indicating room for further improving the modelling of the drift function of the interest rate.

Section 2 introduces a general continuous-time regression framework for the spot interest rate process and the hypotheses of interest on the drift. In Section 3, we describe how the generalized spectral derivative test of Hong and Lee (2005) can be tailored to the continuous-time setting and how it can be used to test linearity of drift using a discretely observed sample. A simulation study shows that the test provides a reliable finite sample reference tailored to a continuous-time model that mimics the high persistence of interest rate data. In Section 4, we apply the generalized spectral derivative test to a widely used sample of 7-day Eurodollar rates. Section 5 concludes the paper.

2. Martingale Characterization of Drift Specification

2.1 Continuous-time Regression

Consider a general continuous time regression

$$dY_t = \bar{\mu}(Y_t)dt + dU_t, \tag{2.1}$$

where $\{Y_t\}$ is a stochastic process, $\{\mathcal{F}_t\}$ is a filtration to which $\{Y_t\}$ is adapted, $\bar{\mu}(Y_t)$ is the instantaneous conditional mean, namely $\bar{\mu}(Y_t) \equiv \lim_{\delta \to 0} E\left[\frac{Y_t - Y_{t+\delta}}{\delta}\middle| \mathcal{F}_t\right]$, and $\{U_t\}$ is a martingale process with respect to the filtration $\{\mathcal{F}_t\}$ so that dU_t is a martingale difference sequence (*m.d.s.*) with $E(dU_t|\mathcal{F}_t) = 0$. Following Park (2008), we may call (2.1) a martingale regression in continuous-time, and dU_t is a continuous-time regression error.[1] Equation (2.1) can be viewed as the continuous-time decomposition of a physical system into a signal and a noise.

Integrating the process in (2.1), we obtain the process

$$U_t = (Y_t - Y_0) - \int_0^t \bar{\mu}(Y_s)ds, \tag{2.2}$$

which is a martingale. One important implication is that the discrete-time process

$$\varepsilon_{t,\Delta} \equiv U_t - U_{t-\Delta}$$

is a *m.d.s.* for any given sampling frequency $\Delta > 0$ (e.g., $\Delta = \frac{1}{250}$ for the daily sampling frequency). The process $\varepsilon_{t,\Delta}$ may be viewed as the true discrete-time regression error of (2.1). Note that there is no discretization bias for the discrete-time regression error $\varepsilon_{t,\Delta}$ because no discretization approximation is made here.

We will allow U_t to be a general martingale process. When $\{U_t\}$ has *a.s.* continuous sample paths, we can write $dU_t = \sigma_t dW_t$, where $\{\sigma_t\}$ is adapted to $\{\mathcal{F}_t\}$ and $\{W_t\}$ is the standard Brownian motion with respect to $\{\mathcal{F}_t\}$. Clearly, (2.1) includes the class of such one-factor diffusion processes as a special case, where Y_t is defined as the time-homogeneous Markov process that solves a stochastic differential equation (SDE) of the form

$$dY_t = \mu(Y_t)dt + \sigma(Y_t)dW_t, \tag{2.3}$$

where $\mu(Y_t)$ is the drift function and $\sigma(Y_t)$ is the diffusion function. Here, the drift function $\mu(Y_t)$ coincides with an instantaneous conditional mean $\bar{\mu}(Y_t)$ for all $0 < t < \infty$ (see Gao (Ch.5, 2007)).

[1] Park (2008) actually considers a more general setup where $\bar{\mu}(\cdot)$ is a a function of X_t and X_t may be a different process from Y_t. The test described below applies to this more general setup.

Stanton (1997) refers to this as the first order approximation to $\mu(Y_t)$. Thus, the drift function $\mu(Y_t)$ generally represents the rate of instantaneous change in the conditional mean of $\{Y_t\}$. Similarly, the squared diffusion function $\sigma^2(Y_t)$ is the instantaneous conditional variance. The specification of $\sigma(Y_t)$ is totally unrestricted under (2.1). Examples of $\sigma(Y_t)$ include the interest rate models of Vasicek (1977), Brennan and Schwartz (1979, 1982), CIR, and Chan *et al.* (1992). The main difference among these models lies in their functional forms for $\mu(Y_t)$ and $\sigma(Y_t)$ in (2.3). See, e.g., Aït-Sahalia (1996a, Table I) for various specifications of the spot interest rate.

We allow the error process U_t in (2.1) to have a more general structure than in the class of one-factor diffusion processes in (2.3). For example, we allow stochastic volatility (SV). An example of SV models (e.g., Andersen and Lund (1997, 2004)) is

$$\begin{aligned} dY_t &= \kappa_1(\mu - Y_t)dt + \sigma_t r_t^{\gamma} dW_{1,t}, \quad \gamma > 0, \\ d\log\sigma_t^2 &= \kappa_2(\alpha - \log\sigma_t^2)dt + \xi dW_{2,t}, \end{aligned}$$

where $W_{1,t}$ and $W_{2,t}$ are independent standard Brownian motion processes. In this case, the diffusion form $\sigma(\cdot)$ is a latent process with a different deriving force from $W_{1,t}$, but we still have $E(dU_t|\mathcal{F}_t) = 0$. This provides us a test procedure for the drift model which is robust to a general form of volatility structure.

While the 'normal' vibrations or smooth variation in the changes in Y_t are modeled by a standard Brownian motion or Levy process, the 'abnormal' vibrations may be due to the arrival of important information that has more than a marginal effect on changes in Y_t. Such important information usually arrives only at discrete points in time, so this jump component is most appropriately modeled with a counting process (discrete-time process) reflecting the non-marginal impact of the information. In (2.1), we do not rule out the possibility of a jump component, since $\{U_t\}$ is a general martingale process which does not require the continuity of the sample path.[2] Specifically, we consider the following jump diffusion process

$$dY_t = \mu(Y_t)dt + \sigma(Y_t)dW_t + JdN_t, \tag{2.4}$$

where J is a jump size which follows some distribution, and $\{N_t\}$ is a homogeneous Poisson process with arrival rate λ independent of $\{W_t\}$. Since the Poisson process N_t is a counting process increasing with time and is not a martingale, with the presence of jumps, we need to transform it into a martingale by substracting a proper "mean" (Neftci (1996), p.115-116, Neftci (2000), p.179). A compensated

[2] Note that the semi-martingale process can be written as $Y_t = Y_0 + \int_0^t \mu(Y_s)ds + \int_0^t \sigma(Y_s)ds + jumps$ (see, Revuz and Yor (2005), for more discussion).

Poisson process, $N_t^* = N_t - \lambda t$, is a martingale. Thus, (2.4) can be written as follows:

$$dY_t = \bar{\mu}(Y_t)dt + \sigma(Y_t)dW_t + JdN_t^*, \tag{2.5}$$

where $\bar{\mu}(Y_t) = \mu(Y_t) - \lambda k$, $k = E(J)$. Here, $dU_t = \sigma(Y_t)dW_t + J_t dN_t^*$ and we have $E[dU_t|\mathcal{F}_t] = 0$. Note that the drift $\mu(Y_t)$ does not coincide with the instantaneous conditional mean $\bar{\mu}(Y_t)$, but they only differ by a constant. This implies, among many other things, that $\bar{\mu}(Y_t)$ is linear if and only if $\mu(Y_t)$ is linear.

2.2 Hypotheses of Interest

In financial applications, a parametric form $\mu_\theta(Y_t)$ is often used to approximate the drift function $\mu(Y_t)$, where $\mu_\theta(\cdot)$ is a known functional form, and θ is an unknown parameter vector. The most commonly used specification of drift in prior research on the interest rate modelling is a linear mean-reverting drift

$$\mu_\theta(Y_t) = \alpha(\beta - Y_t), \tag{2.6}$$

where α is the long-run equilibrium level of the interest rate, β is the speed of reversion at which the short rate returns to the long-run mean. This force is proportional to the deviation of the interest rate from its mean.

Our main interest in this paper is to check whether the drift of the interest rate process is linear under a general error structure for U_t, particularly allowing for SV and jumps. The null hypothesis of interest is

$$\mathbb{H}_0 : \mu(Y_t) = \alpha_0(\beta_0 - Y_t) \text{ for some unknown } \theta_0 = (\alpha_0, \beta_0)' \in \Theta \subset \mathbb{R}^2,$$

and the alternative hypothesis is

$$\mathbb{H}_A : \mu(Y_t) \neq \alpha(\beta - Y_t) \text{ for all } \theta = (\alpha, \beta)' \in \Theta.$$

Under the alternative hypothesis $\mathbb{H}_A$, there exists neglected nonlinearity in drift of the interest rate Y_t. There are vast possibilities of departure from linearity of drift and practitioners usually have no information about the true nonlinear alternative. Therefore, it is highly desirable to employ a test that has power against a vast class of alternatives. The generalized spectral derivative test of Hong and Lee (2005) has such an appealing feature. This test is developed for discrete-time conditional mean model and we now show that it is also applicable to test $\mathbb{H}_0$ in a continuous-time setup.

We first assume that there is no jump component. Define the model error process $U_t(\theta)$ of the linear drift model as follows:

$$dU_t(\theta) \equiv dY_t - \mu_\theta(Y_t)dt = dY_t - \alpha(\beta - Y_t)dt. \tag{2.7}$$

Then, under and only under $\mathbb{H}_0$, the model-implied error process $U_t(\theta)$ coincides with the true martingale process U_t *a.s.* As a result, $dU_t(\theta_0)$ has the *m.d.s.* property that $E[dU_t(\theta_0)|\mathcal{F}_t] = 0$ *a.s.* for some $\theta_0 \in \Theta$ if and only if $\mathbb{H}_0$ holds. Thus, to test $\mathbb{H}_0$, we can check whether $E[dU_t(\theta_0)|\mathcal{F}_t] = 0$ *a.s.* using a discretely observed sample $\{Y_{t\Delta}\}_{t=1}^n$, where Δ is a fixed sampling interval and n is the sample size. In our empirical study, we will use the same daily data ($\Delta = \frac{1}{250}$) of the 7-day Eurodollar rates as used in Aït-Sahalia (1996) and many others. We note that (2.7) implies that for any $\Delta > 0$, the discrete-time error of the linear drift model

$$\varepsilon_{t,\Delta}(\theta) = U_t(\theta) - U_{t-\Delta}(\theta) = (Y_t - Y_{t-\Delta}) - \alpha \int_{t-\Delta}^{t} (\beta - Y_s)ds \tag{2.8}$$

is a *m.d.s.* with respect to $\mathcal{F}_t$ when $\theta = \theta_0$ under and only under $\mathbb{H}_0$. The process $\varepsilon_{t,\Delta}(\theta)$ can be viewed as a discrete-time model error term of the continuous-time linear drift model in (2.6). One can check the linearity in drift by checking the *m.d.s.* property of $\{\varepsilon_{t,\Delta}(\theta_0)\}$. We note that it is not trivial to test the *m.d.s.* property for $\varepsilon_{t,\Delta}(\theta_0)$, because the departures from *m.d.s.* can be nonlinear and subtle, and the conditioning information set $\mathcal{F}_{t-\Delta}$ is of infinite dimension or grows with time t. Also, $\varepsilon_{t,\Delta}(\theta)$ usually displays conditional heteroskedasticity, as will arise when there exists a level effect (e.g., the CIR model). Therefore, it is important to use a robust test to conditional heteroskedasticity of unknown form.

When a jump component exists as in (2.4), $\varepsilon_{t,\Delta}(\theta_0)$ is no longer a *m.d.s.* under $\mathbb{H}_0$ because the jump in (2.4) only makes a difference in the intercept. However, $\varepsilon_{t,\Delta}(\theta_0) - E[\varepsilon_{t,\Delta}(\theta_0)]$ is a *m.d.s.* In other words, the presence of the jump in (2.4) does not alter the *m.d.s.* property of $\varepsilon_{t,\Delta}(\theta_0)$ after demeaning. As long as the unconditional mean of $\varepsilon_{t,\Delta}(\theta)$ is properly controlled, a test for $\mathbb{H}_0$ is then made robust to the presence of jumps in (2.4).

There are several methods to test linearity in drift. One approach is based on the smoothed nonparametric estimation of the drift function and then compares it with a linear drift model. As pointed out by Chapman and Pearson (2000), this approach suffers from the finite sample bias of the truncation of a distribution due to high persistence in interest rate data, besides the boundary bias. It also requires the sampling interval Δ shrinks to zero, and the use of high-frequency data may suffer from market microstructure noise. Park (2008) proposes a novel approach to estimating and testing the drift model in continuous-time based on a time change from calender time to volatility time (quadratic variation clock), which makes the transformed regression error become an independent Brownian motion. This approach leaves the diffusion function totally unspecified, but it assumes that the error process U_t has *a.s.* continuous-time sample paths, as required for the time change.

Jump components are therefore ruled out. Moreover, Park's (2008) approach requires using higher frequency data to estimate realized volatilities. This implies that one also need intraday interest rate data available when using daily data (i.e., $\Delta = \frac{1}{250}$).

In this paper, we apply the generalized spectral derivative test of Hong and Lee (2005) to the continuous-time regression framework and test linearity of drift for the short rate. This approach has several advantages over the existing methods. First, we do not use smoothed nonparametric estimation of the drift function, thus avoiding the finite sample truncation bias and boundary bias as pointed out by Chapman and Pearson (2000). Second, we allow but do not restrict the interest rate to follow a one-factor diffusion process (with a diffusion function of unknown form). We also allow the presence of stochastic volatility and jumps for interest rates. Third, the generalized spectral derivative test can check a vast range of linear and nonlinear alternatives. It checks many lags as well, thus providing a sensible solution to the difficult "curse of dimensionality" problem associated with the *m.d.s.* property. In the next section, we show how Hong and Lee's (2005) test can be tailored to the continuous-time setup and how it can be used to test linearity in drift for interest rates.

3. Generalized Spectral Derivative Test

3.1 Generalized Spectral Derivative Test

To show how generalized spectral derivative test of Hong and Lee (2005) can be applied to the continuous-time framework to check linearity of drift using a discretely sampled data $\{Y_{t\Delta}\}_{t=1}^{n}$, we first introduce its basic idea. For notational convenience, we put the discrete-time model error $\varepsilon_{t,\Delta} \equiv \varepsilon_{t,\Delta}(\theta^*)$, where θ^* is the probability limit of some estimator $\hat{\theta}$ (say) for θ_0. Suppose $\{\varepsilon_{t,\Delta}\}$ is a strictly stationary process with marginal characteristic function $\varphi_\Delta(u) \equiv E[\exp(\mathbf{i}u\varepsilon_{t,\Delta})]$ and pairwise joint characteristic function $\varphi_{j,\Delta}(u,v) \equiv E[\exp(\mathbf{i}u\varepsilon_{t,\Delta} + \mathbf{i}v\varepsilon_{t-|j|\Delta,\Delta})]$, where $\mathbf{i} \equiv \sqrt{-1}$, $u, v \in \mathbb{R}$ and $j = 0, \pm 1, \cdots$ is a lag/lead order. The basic idea of the generalized spectrum is to consider the spectrum of the transformed series $\{\exp(iu\varepsilon_{t,\Delta})\}$. It is defined as

$$f_\Delta(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_{j,\Delta}(u,v) \exp(-\mathbf{i}j\omega), \quad \omega \in [-\pi, \pi], \tag{3.1}$$

where ω is the frequency, $\sigma_{j,\Delta}(u,v)$ is the covariance function of the transformed series:

$$\sigma_{j,\Delta}(u,v) \equiv \operatorname{cov}[\exp(\mathbf{i}u\varepsilon_{t,\Delta}), \exp(\mathbf{i}v\varepsilon_{t-|j|\Delta,\Delta})] = \varphi_{j,\Delta}(u,v) - \varphi_\Delta(u)\varphi_\Delta(v), \quad j = 0, \pm 1, \cdots .$$

Since the generalized covariance $\sigma_{j,\Delta}(u,v)$ is the difference between the joint characteristic function $\varphi_{j,\Delta}(u,v)$ and the product of marginal characteristic functions $\varphi_\Delta(u)$ and $\varphi_\Delta(v)$, $\sigma_{j,\Delta}(u,v) = 0$ for all u, v if and only if $\varepsilon_{t,\Delta}$ and $\varepsilon_{t-|j|\Delta,\Delta}$ are independent. As a result, the generalized spectrum $f_\Delta(\omega, u, v)$ can capture any type of pairwise serial dependence in $\{\varepsilon_{t,\Delta}\}$, i.e., dependence between

$\varepsilon_{t,\Delta}$ and $\varepsilon_{t-|j|\Delta,\Delta}$ for any nonzero lag j, including that with zero autocorrelation. In other words, it can capture serial dependence not only in mean but also in higher order conditional moments of $\varepsilon_{t,\Delta}$, over all lags. Thus, $f_\Delta(\omega,u,v)$ is not applicable to test the hypothesis of linearity in drift, which only implies that $\varepsilon_{t,\Delta}$ is a *m.d.s.* but is silent about the dynamics of its higher order conditional moments. However, just as the characteristic function can be differentiated to generate various moments of $\{\varepsilon_{t,\Delta}\}$, the generalized spectrum $f_\Delta(\omega,u,v)$ can be differentiated to capture serial dependence in various conditional moments of $\varepsilon_{t,\Delta}$. To focus (and only focus) on serial dependence in conditional mean (i.e., departures from *m.d.s.*) of $\varepsilon_{t,\Delta}$, one can consider the following generalized spectral derivative

$$f_\Delta^{(0,1,0)}(\omega,0,v) \equiv \left.\frac{\partial f_\Delta(\omega,u,v)}{\partial u}\right|_{u=0} = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_{j,\Delta}^{(1,0)}(0,v)\exp(-\mathbf{i}j\omega), \quad \omega\in[-\pi,\pi], \tag{3.2}$$

where the generalized covariance derivative

$$\sigma_{j,\Delta}^{(1,0)}(0,v) \equiv \left.\frac{\partial \sigma_{j,\Delta}(u,v)}{\partial u}\right|_{u=0} = \text{cov}[\mathbf{i}\varepsilon_{t,\Delta}, \exp(\mathbf{i}v\varepsilon_{t-|j|\Delta,\Delta})].$$

The derivative measure $\sigma_{j,\Delta}^{(1,0)}(0,v)$ checks whether the autoregression function $E(\varepsilon_{t,\Delta}|\varepsilon_{t-j\Delta,\Delta})$ at lag j is zero. Under appropriate conditions, $\sigma_{j,\Delta}^{(1,0)}(0,v)=0$ for all $v\in\mathbb{R}$ if and only if $E(\varepsilon_{t,\Delta}|\varepsilon_{t-j\Delta,\Delta}) = E(\varepsilon_{t,\Delta})$. This can be seen intuitively if we take a Taylor series expansion of the exponential function $\exp(\mathbf{i}v\varepsilon_{t-\Delta})$ in a small neighborhood of v around zero:

$$\sigma_{j,\Delta}^{(1,0)}(0,v) = \sum_{l=0}^{\infty}\frac{\mathbf{i}(\mathbf{i}v)^l}{l!}\text{cov}(\varepsilon_{t,\Delta}, \varepsilon_{t-|j|,\Delta}^l).$$

The generalized covariance derivative $\sigma_{j,\Delta}^{(1,0)}(0,v)$ essentially checks covariances between $\varepsilon_{t,\Delta}$ and all possible powers of $\varepsilon_{t-|j|,\Delta}$. Because the hypothesis of linearity of drift implies $E(\varepsilon_{t,\Delta}|\mathcal{F}_{t-\Delta}) = E(\varepsilon_{t,\Delta})$ *a.s.*, $f_\Delta^{(0,1,0)}(\omega,0,v)$ becomes a "flat spectrum", namely,

$$f_{\Delta,0}^{(0,1,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\sigma_{0,\Delta}^{(1,0)}(0,v), \quad \omega\in[-\pi,\pi].$$

Thus, one can test linearity of drift by checking whether $f_\Delta^{(0,1,0)}(\omega,0,v)$ is a "flat spectrum". The autoregression function $E(\varepsilon_{t,\Delta}|\varepsilon_{t-j\Delta,\Delta})$ can equivalently $\sigma_{j,\Delta}^{(1,0)}(0,v)$ can capture a vast range of linear and nonlinear serial dependence in mean, particularly those with zero autocorrelation.[3] An additional advantage of using $f_\Delta^{(0,1,0)}(\omega,0,v)$ is that it can check many lag orders without suffering from the "curse of dimensionality" problem.[4]

[3] An example include a nonlinear MA process $\varepsilon_{t,\Delta} = \alpha z_{t-\Delta,\Delta} z_{t-2\Delta,\Delta} + z_{t,\Delta}$, where $z_{t,\Delta}$ is $i.i.d.(0,1)$.

[4] It should be noted that the *m.d.s.* hypothesis that $E(\varepsilon_{t,\Delta}|\mathcal{F}_t) = E(\varepsilon_{t,\Delta})$ *a.s.* is not exactly same as the hypothesis that $E(\varepsilon_{t,\Delta}|\varepsilon_{t-j\Delta,\Delta}) = E(\varepsilon_{t,\Delta})$ *a.s.* for all $j>0$. The former implies the latter but not vice versa (i.e., the latter is a

Although $\sigma_{j,\Delta}^{(1,0)}(0,v)$ and $E(\varepsilon_{t,\Delta}|\varepsilon_{t-j\Delta,\Delta})$ are equivalent measures for serial dependence in mean, $\sigma_{j,\Delta}^{(1,0)}(0,v)$ can be estimated consistently by its sample analogue (see (3.7) below), thus avoiding smoothed nonparametric estimation for $E(\varepsilon_{t,\Delta}|\varepsilon_{t-j\Delta,\Delta})$ which, when applied to highly persistent interest rate data, would suffer from the finite sample bias and the boundary bias problem as pointed out by Chapman and Pearson (2000). This is another advantages of using the frequency domain approach.

In the present context, $\varepsilon_{t,\Delta}$ is not observed and has to be estimated. Suppose we have a discretely observed random sample $\{Y_{t\Delta}\}_{t=1}^{n}$ of size n which is used to estimate model $\mu_\theta(Y_s)$. We then obtain the estimated discrete-time model residual

$$\hat{\varepsilon}_{t,\Delta} = Y_t - Y_{t-\Delta} - \int_{t-\Delta}^{t} \mu_{\hat{\theta}}(Y_s)ds, \tag{3.3}$$

where $\hat{\theta}$ is any $\sqrt{n}$-consistent parameter estimator for θ_0 which need not be asymptotically most efficient. For the class of one-factor diffusion models with unspecified diffusion functions, one example of $\hat{\theta}$ is OLS and Aït-Sahalia's (1996) feasible GLS. The OLS estimator can be also used; it is consistent but is less efficient than Aït-Sahalia's (1996) feasible GLS when the interest rate is a diffusion process with unknown diffusion function. One can also use Park's (2008) martingale-based minimum distance estimator if data sampled at frequency higher than Δ are available. In this paper, we use Song's (2008) conditional GMM estimator which is robust to a general error structure and is applicable even in the nonlinear drift case. As we will see, the generalized spectral derivative test is robust to different consistent estimation methods.

In calculating the estimated discrete-time model residual $\hat{\varepsilon}_{t,\Delta}(\theta)$, one needs to compute the Riemann integral $\int_{t-\Delta}^{t} \mu_{\hat{\theta}}(Y_s)ds$. When the sampling frequency Δ is small, similarly to Phillips and Yu (2008), one can approximate the integral $\int_{t-\Delta}^{t} \mu_{\theta}(Y_s)ds$ by using the average of $\mu_{\hat{\theta}}(\cdot)$ over higher frequency observations between time period $t-\Delta$ and time period t. For example, if Δ is a daily sampling frequency ($\Delta = \frac{1}{250}$), which is small, we can use the following approximation:

$$\int_{t-\Delta}^{t} \mu_\theta(Y_s)ds = \frac{\Delta}{2}\left[\mu_\theta(Y_t) + \mu_\theta(Y_{t-\Delta})\right] + O_P(\Delta^2). \tag{3.4}$$

This avoids the use of intraday data and greatly simplifies the calculation of $\hat{\varepsilon}_{t,\Delta}$. The discretization error is of order Δ^2, which is small for daily data in practice (see Das (2002) for relevant discussion in a different context).

With the estimated model residuals $\{\hat{\varepsilon}_{t,\Delta}\}$, we can estimate $f_{\Delta}^{(0,1,0)}(\omega,0,v)$ by a kernel spectral

necessary but not a sufficient condition). This is the price we have to pay for dealing with the difficulty of the "curse of dimensionality" of testing the $m.d.s.$ property of $\{\varepsilon_{t,\Delta}\}$. Nevertheless, the examples for which $E(\varepsilon_{t,\Delta}|\varepsilon_{t-j\Delta,\Delta}) = E(\varepsilon_{t,\Delta})$ $a.s.$ for all $j>0$ and $E(\varepsilon_{t,\Delta}|\mathcal{F}_{t-\Delta}) \neq E(\varepsilon_{t,\Delta})$ may be rare in practice and are pathological.

estimator

$$\hat{f}^{(0,1,0)}_{\Delta}(\omega,0,v) \equiv \frac{1}{2\pi}\sum_{j=1-n}^{n-1} k(j/p)(1-|j|/n)^{1/2}\hat{\sigma}^{(1,0)}_{j,\Delta}(0,v)e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi,\pi], \tag{3.5}$$

where $\hat{\sigma}^{(1,0)}_{j,\Delta}(0,v) = \frac{\partial}{\partial u}\hat{\sigma}_{j,\Delta}(u,v)|_{u=0}$, $\hat{\sigma}_{j,\Delta}(u,v) = \hat{\varphi}_{j,\Delta}(u,v) - \hat{\varphi}_{j,\Delta}(u,0)\hat{\varphi}_{j,\Delta}(0,v)$, and

$$\hat{\varphi}_{j,\Delta}(u,v) = \frac{1}{n-|j|}\sum_{t=|j|+1}^{n} \exp(\mathbf{i}u\hat{\varepsilon}_{t,\Delta} + \mathbf{i}v\hat{\varepsilon}_{t-|j|\Delta,\Delta}).$$

Here, $p \equiv p(n)$ is a lag order, and $k(\cdot)$ is a symmetric kernel function for lag orders. Examples of $k(\cdot)$ include the Bartlett, Daniell, Parzen and Quadratic spectral kernels (e.g., Priestley 1981, p.442). The popular Newey and West's (1987) long-run variance-covariance estimator is based on the Bartlett kernel $k(z) = (1-|z|)\mathbf{1}(|z| \le 1)$, where $\mathbf{1}(\cdot)$ is the indicator function. The factor $(1-|j|/n)^{1/2}$ is a finite-sample correction. It could be replaced by unity. Under certain regularity conditions, the estimator $\hat{f}^{(0,1,0)}_{\Delta}(\omega,0,v)$ is consistent for $f^{(0,1,0)}_{\Delta}(\omega,0,v)$. We note that lag order p is a smoothing parameter but it is different from a bandwidth used to estimate the drift function $\mu(Y_t)$ directly. In our simulation and empirical study, we will consider a data-driven choice of p and conduct sensitivity check on the impact of the choice of p. The results are quite robust to the choice of p. We emphasize that the spectral (frequency-domain) estimator $\hat{f}^{(0,1,0)}_{\Delta}(\omega,0,v)$ does not suffer from the boundary bias and finite sample bias due to the truncation of a distribution associated with smoothed time-domain nonparametric estimation for $\mu(Y_t)$ or $E(\varepsilon_{\Delta,t}|\varepsilon_{\Delta,t-\Delta})$.

To estimate the flat generalized spectral derivative $f^{(0,1,0)}_{0,\Delta}(\omega,0,v) \equiv \frac{1}{2\pi}\sigma^{(1,0)}_{0,\Delta}(0,v)$, we use the restricted spectral estimator

$$\hat{f}^{(0,1,0)}_{0,\Delta}(\omega,0,v) \equiv \frac{1}{2\pi}\hat{\sigma}^{(1,0)}_{0,\Delta}(0,v), \quad \omega \in [-\pi,\pi]. \tag{3.6}$$

The unrestricted spectral estimator $\hat{f}^{(0,1,0)}_{\Delta}(\omega,0,v)$ and the restricted spectral estimator $\hat{f}^{(0,1,0)}_{0,\Delta}(\omega,0,v)$ will converge to the same limit as $n \to \infty$ when the drift is linear. If they converge to different limits, there exists evidence of a nonlinear drift. Hong and Lee (2005) propose a test for the *m.d.s.* property of $\{\varepsilon_{t,\Delta}\}$ by comparing $f^{(0,1,0)}_{\Delta}(\omega,0,v)$ with the "flat spectrum" estimator $\hat{f}^{(0,1,0)}_{0,\Delta}(\omega,0,v)$ via the quadratic norm. The test statistic is

$$M_{0,\Delta}(p) \equiv \left[\sum_{j=1}^{n-1} k^2(j/p)(n-j)\int \left|\hat{\sigma}^{(1,0)}_{j,\Delta}(0,v)\right|^2 d\mathcal{W}(v) - \hat{C}_{\Delta}(p)\right] / \sqrt{\hat{D}_{\Delta}(p)}, \tag{3.7}$$

where $\mathcal{W}(v)$ is a nondecreasing, positive, right-continuous function that weights sets of v about zero equally. An example of $\mathcal{W}(v)$ is the N(0,1) CDF, which is used in the simulation study and empirical study below. Both $\hat{C}_\Delta(p)$ and $\hat{D}_\Delta(p)$ are centering and scaling factors that have taken into account the impact of conditional heteroskedasticity and other time-varying higher order moments of unknown form. As a consequence, $M_{0,\Delta}(p)$ is robust to conditional heteroskedasticity and higher order time-varying moments of unknown form. This is particularly appealing for the present study on linearity of drift of the short rate, because the discrete-time error $\{\varepsilon_{t,\Delta}\}$ generally display conditional heteroskedasticity and time-varying conditional higher order moments even when $\{\varepsilon_{t,\Delta}\}$ is a *m.d.s.,* as is the case when Y_t follows a CIR process where level effect exists or when a latent stochastic volatility is present. Following Hong and Lee (2005), we have $M_{0,\Delta}(p) \to N(0,1)$ in distribution when the drift is linear. Under nonlinear drift, $M_{0,\Delta}(p)$ will generally diverge to infinity as $n \to \infty$, giving the test its power.

We emphasize that the approach to testing the drift model $\mu_\theta(\cdot)$ by using the generalized spectral derivative test to check the *m.d.s.* property of the discrete-time model error $\varepsilon_{t,\Delta}(\theta)$ is not restricted to the hypothesis of linearity of drift. It is a generally applicable approach to testing a drift model (whether linear or nonlinear) in the general continuous-time regression framework in (2.1). In this paper, in addition to the linear drift model, we also apply the proposed test to check some popular nonlinear drift models for interest rates (see Section 4 for details). Moreover, it allows for a general diffusion process with the diffusion function of unknown form stochastic volatility and jumps, as often encountered in high-frequency financial data.

3.2 Simulation Study

The generalized spectral derivative test $M_{0,\Delta}(p)$ is originally proposed to test discrete-time conditional mean models. It is unclear whether it will performs well in finite samples for such highly persistent time series data as the short-term interest rates in a continuous-time setup. We now conduct an empirically realistic simulation study in a continuous-time setup. Following Chapman and Pearson (2000), we first consider a CIR process:

$$dY_t = \alpha(\beta - Y_t)dt + \sigma\sqrt{Y_t}dW_t, \tag{3.8}$$

where β is the long-run mean of Y_t, α determines the speed at which the process reverts to the long-run mean, and the instantaneous variance $\sigma^2(Y_t)$ of the process increases at the rate proportional to Y_t. Here, the drift of this process is linear, allowing us to examine the size of the test $M_{0,\Delta}(p)$. The first two unconditional moments of the square root process are $E(Y_t) = \beta$, $\text{var}(Y_t) = \beta\sigma^2/2\alpha$, and $\text{corr}(Y_{t+\Delta}, Y_t) = \exp(-\alpha\Delta)$. We choose the length of time between observations of the diffusion process $\Delta = 1/250$, corresponding to daily observations. The choice of α, the parameter that determines the persistence of the process, is important, since we are particularly interested in the impact of dependence persistence in $\{Y_t\}$ on the size of the test. Like Chapman and Pearson (2000), we set α to 0.21459 or 0.858837, respectively. The first value implies an (monthly) autocorrelation of the short rate of 0.982, which is consistent with the upper end of the range of parameter estimates based on the U.S. interest rate data. The second choice of α implies a first order (monthly) autocorrelation coefficient that is equal to that of the Eurodollar data used in Aït-Sahalia (1996a,b). The choice of β is not particularly important, and we set it to 0.085711 in order to be consistent with the Aït-Sahalia's (1996a,b) 7-day Eurodollar rate data set. Given (α, β), the value of σ is chosen such that the variance of Y_t in (3.8) equal to the sample variance in the Aït-Sahalia data set. This implies the choices of σ are equal to 0.07830 and 0.15660 respectively, where the ordering is consistent with the ordering of the α values. Thus, we consider two cases of both low and high levels of dependence persistence, with $\left(\alpha, \sigma^2\right) = (0.85837, 0.002185)$ and $(0.214592, 0.000546)$ respectively. [5]

We generate data using the Milstein's scheme, i.e.,

$$Y_{(\tau+1)\delta} = Y_{\tau\delta} + \alpha(\beta - Y_{\tau\delta})\delta + \sigma\sqrt{Y_{\tau\delta}}\sqrt{\delta}e_{\tau\delta} + \frac{1}{2}\sigma^2 Y_{\tau\delta}\delta\left(e_{\tau\delta}^2 - 1\right) \tag{3.9}$$

where $e_{\tau\delta} \sim i.i.d.N(0,1)$ and $\delta = \Delta/M$, where $\Delta = 1/250$. We set $M = 5$; namely, we generate 5 observations each day and then save the fifth one as the daily observation. The initial value of Y_t can be set equal to the average interest rate level of the data set in Aït-Sahalia (1996a,b). The discrete observations of sample size n are generated over a time period $[0, T]$, where $n = T/\Delta$.

We also consider to generate data from a stochastic volatility model (e.g., Andersen and Lund (1997, 2004)) and a jump-diffusion model (e.g., Duffie, Pan and Singleton (2000)). It has been recognized in the literature that stochastic volatility helps improving the short-rate modeling. For example, Andersen and Lund (1997) estimate the two factor stochastic volatility model and found that the stochastic volatility factor vastly improves goodness of fit. On the other hand, there has been strong evidence that jumps play an important role in capturing the dynamics of interest rates. Das (2001) and Johannes (2004) analyze the role of jumps in continuous-time short interest rate models, and the results show that jumps are economically and statistically important.

To examine the robustness of the $M_{0,\Delta}(p)$ test in (3.7) to the presence of stochastic volatility, we consider the stochastic volatility model studied in Andersen and Lund (1997, 2004), which is an extension of the CKLS model (Chan *et al.*, (1992)):

$$\begin{aligned} dY_t &= \alpha_1(\beta_1 - X_t)dt + \sigma_t Y_t^{\gamma} dW_{1t}, \\ d\log\sigma_t^2 &= \alpha_2(\beta_2 - \log\sigma_t^2)dt + \xi dW_{2t}, \end{aligned}$$

where W_{1t} and W_{2t} are two standard Brownian Motions. The specification implies mean reversion of the interest rate level as well as the stochastic (log-) volatility. The parameter values are taken

[5] We also considered the Vasicek process, i.e., $dX_t = \alpha(\beta - X_t)dt + \sigma dW_t, \alpha > 0$. Here, the diffusion function $\sigma(X_t)$ is a constant and the transition density is conditionally normal with mean $X_t\exp(-\alpha\Delta)$ and variance $\sigma^2\left[1 - \exp(-2\alpha\Delta)\right]/2\alpha$. The simulation results on the finite sample performance of $M_{0,\Delta}(p)$ were similar to those of the CIR model, thus not reported here.

from Andersen and Lund's (2004) estimation results for weekly data, with $(\alpha_1, \beta_1, \gamma, \alpha_2, \beta_2, \xi) = (0.1633, 0.0595, 0.5443, 1.0397, -6.3599, 1.2719)$.

To examine the impact of jump components, we consider an affine jump-diffusion model as considered in Duffie, Pan and Singleton (2000):

$$dY_t = \alpha(\beta - Y_t)dt + \sigma\sqrt{Y_t}dW_t + JdN_t, \tag{3.10}$$

where J is the random jump size which follows a $N\left(\mu_J, \sigma_J^2\right)$ distribution, and N_t is a Poisson process with arrival intensity λ, which denotes the number of jumps per year. The diffusion and jump processes are independent of each other, and are also independent of jump size J. We assume that the coefficients are bounded and sufficient regularity conditions are satisfied so that a unique, strong solution to (3.10) exists (see Protter (2005) for details on the regularity conditions). The parameter values are taken from the estimation results of Das (2002) for daily data, with $(\alpha, \beta, \sigma, \mu_J, \sigma_J) = (0.8542, 0.0330, 0.0173, 0.0004, 0.0058)$ except that λ is set to 5 which corresponds to five jumps a year (Carrasco, Chernov, Florens and Ghysels (2002)).

As discussed earlier, a $\sqrt{n}$-consistent parameter estimator for the linear drift parameter vector $\theta_0 = (\alpha_0, \beta_0)'$ is needed to implement the generalized spectral derivative test $M_{0,\Delta}(p)$. Under the null hypothesis of linear drift, only the drift function is specified. Therefore, estimation methods for a fully parametric diffusion model cannot be used, including the marginal density method in Aït-Sahalia (1996a), the approximated MLE method in Aït-Sahalia (2002b), the GMM method in Hansen and Scheinkman (1995), the simulated MLE method in Pedersen (1995) and the EMM method of Gallant and Tauchen (1996). We could use the estimation methods in Aït-Sahalia (1996), Kristensen (2008b), and Park (2008). Alternatively, OLS, as considered in Chan, Karolyi, Longstaff, and Sanders (1992), Chapman and Pearson (2000) and Fan and Zhang (2003), can also be used. When Y_t follows the diffusion process in (2.3), we have

$$E\left[Y_{(\tau+1)\Delta}|Y_{\tau\Delta}\right] = \beta + e^{-\alpha\Delta}(Y_{\tau\Delta} - \beta), \quad \tau = 1, \cdots, n, \tag{3.11}$$

under a linear drift. The OLS estimator does not suffer from any discretized bias.[6] It is consistent although not most efficient.

Park's (2008) estimator is based on a random time change which requires consistent estimation of the quadratic variation (integrated volatility). This requires using intraday interest rates when Δ is the daily sampling frequency. Kristensen's (2008b) estimator does not require intraday data, but it involves comparison between a direct nonparametric transition density estimator and a diffusion-based semiparametric transition density. Since the transition density generally does not have a closed-

[6] As discussed in Aït-Sahalia (1996b) and Kristensen (2008b), the conditional mean $E\left[X_{(\tau+1)\Delta}|X_{\tau\Delta}\right]$ would generally involve the diffusion term when the drift term is nonlinear.

form, simulation methods have to be used in Kristensen's (2008b) method. This is computationally burdensome in practice. Moreover, like Ait-Sahalia's (1996b) feasible GLS estimator, both Park (2008) and Kristensen (2008b) assume a one-factor diffusion model with an unspecified diffusion function. It is restrictive in our present setup.

To estimate the drift parameters consistently under a general error structure that allows for stochastic volatility and jumps, we use Song's (2008) GMM estimator based on conditional moment restriction that $E[\varepsilon_{t,\Delta}(\theta_0)|\mathcal{F}_{t-\Delta}] = E[\varepsilon_{t,\Delta}(\theta_0)]$. Specifically, Song's (2008) estimator is given as follows:

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \frac{1}{n-1}\sum_{l=2}^{n}\left\{\frac{1}{n-1}\sum_{\tau=2}^{n}\left[\varepsilon_{\tau\Delta,\Delta}(\theta) - \bar{\varepsilon}_{\Delta}(\theta)\right]\mathbf{1}\left(Y_{(\tau-1)\Delta} \le X_{l\Delta}\right)\right\}^2, \tag{3.12}$$

where $\bar{\varepsilon}_{\Delta}(\theta) = \frac{1}{n}\sum_{\tau=1}^{n}\varepsilon_{\tau\Delta,\Delta}(\theta)$ This estimator exploits the *m.d.s.* property of $\varepsilon_{t,\Delta}$ under $\mathbb{H}_0$ and is robust to a general error structure in U_t.

We examine the performance of the generalized spectral derivative test $M_{0,\Delta}(p)$ for the sample size $n = 250, 500$ and 1000 respectively. These sample sizes correspond to 1 to 4 years of daily data. For each sample size n, we generate 1000 data sets under each DGP, using the GAUSS Window Version random number generator. To compute $M_{0,\Delta}(p)$, we need to choose a kernel function $k(\cdot)$ and the lag order p. We use the Daniell kernel and Parzen kernel, which give similar results. For space, we only report results based on the Parzen kernel. For the choice of p, we use a data-driven plug-in method, which delivers an automatic data-driven choice of lag order $\hat{p}_0$ that asymptotically minimizes the integrated mean squared error of the nonparametric generalized spectral density derivative estimator $\hat{f}_{\Delta}^{(0,1,0)}(\omega, 0, v)$. This plug-in method involves a preliminary generalized spectral density estimator, which in turn requires the choice of a preliminary kernel function $\bar{k}(\cdot)$ and a preliminary lag order $\bar{p}$. We use the Bartlet kernel for $\bar{k}(\cdot)$ and choose $\bar{p} = 2, 4, 6, 8$ and 10 respectively, to examine the sensitivity of the performance of the $M_{0,\Delta}(p)$ test to the choice of $\bar{p}$.

We first examine the sizes of $M_{0,\Delta}(p)$ under the CIR process at both 10% and 5% significance levels, using the asymptotic $N(0,1)$ approximation. The results are summarized in Table 1. Under the CIR model, there exists level effect, a form of conditional heteroskedasticity. We estimate the drift parameter $\theta = (\alpha, \beta)'$ using the OLS estimator and the conditional GMM estimator in (3.12) respectively. From Table 1, we find that the sizes of $M_{0,\Delta}(p)$ are reasonable at both 10% and 5% levels for all sample sizes n, no matter whether OLS or the conditional GMM estimator is used. The size performances are similar no matter whether data is highly persistent. They are not sensitive to the choice of the preliminary lag order $\bar{p}$.

Next, we examine the sizes of $M_{0,\Delta}(p)$ under the stochastic volatility model with linear drift, as reported in Table 2. Since the OLS estimator may not be consistent in the presence of stochastic volatility, we have to rely on the results based on Song's (2008) conditional GMM estimator, which is consistent in the presence of stochastic volatility. Nevertheless, we still include the results based on the OLS estimation for comparison. For the conditional GMM estimation, the sizes of $M_{0,\Delta}(p)$

are reasonable at both 10% and 5% levels, for all sample sizes and all choices of preliminary lag order $\bar{p}$. This suggests that the test $M_{0,\Delta}(p)$ is robust to the presence of stochastic volatility and to the choice of $\bar{p}$. Although the OLS estimator may not be consistent, the sizes of $M_{0,\Delta}(p)$ based on the OLS estimation are also reasonable for all sample sizes and all choices of preliminary lag orders.

Finally, we examine the sizes of $M_{0,\Delta}(p)$ under the affine jump diffusion model, reported in Table 3. Like under stochastic volatility, the conditional GMM estimator in (3.12) is still consistent in the presence of jumps, while the OLS estimator is generally inconsistent. As a result, we have to rely on the results based on the conditional GMM estimator. As can be seen from Table 3, the $M_{0,\Delta}(p)$ test is severely underrejected (i.e., the rejection rates are smaller than the significant levels) when the sample size n is small. The underrejection is possibly due to the effect of the jumps on the asymptotic variance estimator $\hat{D}_{\Delta}(p)$ of the test statistic $M_{0,\Delta}(p)$, which takes into account the impact of all conditional moments of $\varepsilon_{t,\Delta}$. Jumps have large impact on higher order moments. Fortunately, the sizes of $M_{0,\Delta}(p)$ improve as n increases. For the OLS estimation, the underrejection of $M_{0,\Delta}(p)$ is more severe than for the conditional GMM estimation. This is apparently due to inconsistent estimation of OLS.

In summary, the generalized spectral derivative test $M_{0,\Delta}(p)$ has reasonable sizes in finite samples even for highly persistent data, and the sizes of $M_{0,\Delta}(p)$ are robust to conditional heteroskedasticity (level effect) and stochastic volatility. The test $M_{0,\Delta}(p)$ displays severe underrejection in small samples when there exist jumps, but the sizes improve as the sample size n increases and when a consistent estimator is used. For all scenarios, the sizes of $M_{0,\Delta}(p)$ are not sensitive to the choice of preliminary lag orders.

4. Evidence of Nonlinear Drift

4.1 Interest Rate Data

To compare with the results and conclusions of the existing literature, we use the same 7-day Eurodollar rate series, bid-ask midpoint, as used in Aït-Sahalia (1996a,b). It is used to proxy for the U.S. short-term interest rate. This daily series, with 5573 observations, covers the period from June 1, 1973, to February 25, 1995. Weekends and holidays have not been treated specifically (Monday is taken as the next day after Friday). Whereas weekend effects have been documented for stock prices (e.g., French and Roll (1986)), there does not seem to be a conclusive weekend effect in money market instruments. Aït-Sahalia (1996a,b) stated that choosing a 7-day rate, such as the 7-day Eurodollar, as the underlying factor for pricing derivatives is a necessary compromise between (i) literally taking an "instantaneous" rate that does not already embed the result of a pricing operator, and (ii) avoiding some of the spurious microstructure effects associated with overnight rates. For example, the second Wednesday settlement effect in the federal funds market creates a spike in the raw federal funds data that has to be smoothed (see Hamilton (1996)). More generally short-term supply and demand effects in overnight markets can create excess volatility at the short end of the yield curve that is often irrelevant for the rest of the curve.

4.2 Empirical Evidence of Nonlinear Drift

To implement the test for nonlinear drift, we estimate the linear drift parameters $\theta = (\alpha, \beta)'$ based on the data on 7-day Eurodollar rates, using both the OLS estimator and the conditional GMM estimator in (3.12). We first compute the generalized spectral derivative test statistic $M_{0,\Delta}(p)$, and report the test statistics in column 1 of Table 4. The values of the test statistic $M_{0,\Delta}(p)$ are quite large compared to the asymptotic $N(0,1)$ critical values. They essentially imply a zero P-value. Thus, there exists rather strong evidence of nonlinearity in drift for 7-day Eurodollar interest rates. To assess the sensitivity of the results to the choice of preliminary lag order $\bar{p}$, we consider $\bar{p} = 2, 5, 10$, respectively. The results are robust to the choice of $\bar{p}$. For all lag orders $\bar{p}$, we have zero P-values for $M_{0,\Delta}(p)$. It is noted that the test results are robust to different parameter estimation methods. We also note that our evidence on nonlinearity of drift differs from the finding of linear drift of Park (2008), who also uses the continuous-time regression framework in (2.1) together with an additional time change device which makes the transformed regression errors follow a Brownian motion. Park (2008) does not reject the null hypothesis of linear drift. A possible reason is that Park (2008) uses a test which only checks the normality of the transformed estimated regression errors and does not check serial dependence in the transformed errors. This may lead to lower power if misspecification of the drift dynamics shows up more as a form of serial dependence in the transformed errors.

Our finding of nonlinear drift has some important implications. First, since we avoid the direct smoothed nonparametric kernel estimation for the drift function, our result is free of the finite sample bias due to truncation of the distribution and the boundary problem of smoothed kernel estimation. Thus, our evidence of nonlinear drift is not a spurious result, particularly given that the generalized spectral derivative test $M_{0,\Delta}(p)$ has reliable finite sample performance as revealed in the empirically realistic simulation study of Section 3. Second, a vast literature has documented evidence against Markovian one-factor diffusion modeling for the interest rate, and non-Markov multifactor models such as stochastic volatility can better capture the dynamics of the interest rate. For example, Anderson and Lund (1997) find that an extended bivariate CIR diffusion model with stochastic volatility can capture most features of the interest rate. Since our finding of nonlinear drift is robust to the presence of stochastic volatility, it implies that the introduction of a nonlinear drift to Andersen and Lund's (1997, 2004) SV model may further improve goodness of fit to the interest rate data. Third, Das (2002) documents that in modelling daily Fed fund rates, the inclusion of jumps diminishes the extent of nonlinearity in drift, and renders the drift of the interest rate more or less linear. Because we allow for the existence of jumps and the presence of jumps does not lead to overrejection of $M_{0,\Delta}(p)$ (see the simulation study in Section 3), our finding of nonlinearity in drift is not likely due to the presence of jumps. Jones (2003) and Takamizawa (2008) also point out that the nonlinearity of drift may be exaggerated when transitory shocks are not taken into account. Transitory shocks are very short-lived and so are similar to jumps to certain extent. Our test statistic and so our findings are also robust to such transitory shocks.

To gauge possible reasons for the rejection of linearity in drift and possible patterns of the drift, we now conduct a variety of tests based on the derivatives of the measure $\sigma_j^{(1,0)}(0,v)$ with respect to v at zero. Specifically, we use the first four derivatives

$$\sigma_j^{(1,l)}(0,v) = \frac{\partial^l}{\partial v^l}\sigma^{(1,0)}(0,v)\bigg|_{v=0} = \mathbf{i}^{l+1}Cov(\varepsilon_{t,\Delta}, \varepsilon_{t-|j|,\Delta}^l), \quad l = 1,2,3,4.$$

Intuitively, the choice of $l = 1$ checks (linear) autocorrelation in $\{\varepsilon_{t,\Delta}\}$ over various lags, the choice of $l = 2$ checks the ARCH-in-mean type effect in $\{\varepsilon_{t,\Delta}\}$ (see Engle, Lilien and Robins (1987)), the choice of $l = 3$ checks the skewness-in-mean effects in $\{\varepsilon_{t,\Delta}\}$, and the choice of $l = 4$ checks the kurtosis-in-mean effects in $\{\varepsilon_{t,\Delta}\}$. The test statistics for these derivative measures $\sigma_j^{(1,l)}(0,v)$, denoted as $M_{l,\Delta}(p)$, can be constructed in the same way as $M_{0,\Delta}(p)$ is constructed, and they are all asymptotically $N(0,1)$ under the null hypothesis of linear drift.

Table 4 shows that except for the choice of $l = 4$, the test statistics for $l = 1,2,3$ are overwhelmingly significant, essentially with zero P-values. The (linear) autocorrelation test $M_{1,\Delta}(p)$ implies that the interest rate process is not a linear drift process, because the significance in autocorrelation of $\{\varepsilon_{t,\Delta}\}$ is likely to be caused by the neglected nonlinear drift component which is contained in $\{\varepsilon_{t,\Delta}\}$. The significance of the ARCH-in-mean effects ($M_{2,\Delta}(p)$) suggests that $\varepsilon_{t,\Delta}$ contains a volatility-type component in mean. Because we have used a daily data set, the discretization bias is expected to be small or negligible. This finding has important implications on the drift dynamics of interest rate. The significance of skewness-in-mean effects ($M_{3,\Delta}p)$) implies that the interest rate drift dynamics is not symmetric, i.e., the mean-reverting behavior to the long-run interest rate level from below is different from the mean-reverting behavior from large interest rate values. Interestingly, the test statistic with the choice of $l = 4$ is insignificant, suggesting that extreme interest rate changes may have little impact on the drift dynamics of the interest rate. The evidence of separate inference based on derivative tests $M_{l,\Delta}(p)$ for $l = 1,2,3,4$ thus reveals information on possible patterns of the drift dynamics of the interest rate process.

To examine whether some nonlinear drift models in the literature can adequately capture the drift dynamics of the interest rate, we apply the test $M_{0,\Delta}(p)$ to the following two popular nonlinear drift models in the literature:

- Ahn and Gao's (1999) Inverse-Feller drift model:

$$\mu(Y_t,\theta) = Y_t[\beta - (\sigma^2 - \beta\alpha)Y_t].$$

- Aït-Sahalia's (1996) nonlinear drift model:

$$\mu(Y_t, \theta) = \alpha_{-1} Y_t^{-1} + \alpha_0 + \alpha_1 Y_t + \alpha_2 Y_t^2.$$

Both Ahn and Gao (1999) and Aït-Sahalia (1996) specify a level-effect diffusion function in a diffusion model. We do not need to specify the diffusion function here because we allow for a more general error structure. For these nonlinear drift models, the OLS estimator cannot be used, because it will suffer from discretization bias and is not consistent. However, the conditional GMM estimator in (3.12) remains consistent.

We apply the generalized spectral derivative test $M_{0,\Delta}(p)$ to the estimated discrete-time residuals of Ahn and Gao's (1999) model and Aït-sahalia's (1996) model and respectively. The results are reported in the table 5. We first consider results on Ahn and Gao's (1999) nonlinear drift model, which specifies a quadratic drift without intercept (it does not nest the linear drift model due to the zero intercept restriction). Compared to the linear drift model, the $M_{1,\Delta}(p)$ values decrease dramatically, suggesting that Ahn and Gao's model can capture some nonlinear drift dynamics. In particular, the quadratic term in the drift function provides an additional flexibility (beyond the linear drift) in capturing the movements of the interest rate at the extreme levels. On the other hand, the values of derivative test statistics $M_{1,\Delta}(p), M_{2,\Delta}(p)$ and $M_{3,\Delta}(p)$ remain as large as for those for the linear drift model ($M_{2,\Delta}(p)$ is even larger). This indicates that Ahn and Gao's model cannot capture much of the nonlinear drift dynamics. The significance of $M_{3,\Delta}(p)$ is apparently due to the fact that Ahn and Gao's model specifies a symmetric drift dynamics and therefore cannot capture asymmetric drift dynamics.

Next, we consider the results on Aït-Sahalia's (1996) nonlinear drift model, which has an asymmetric drift dynamics and nests both the linear drift model and Ahn and Gao's (1999) quadratic drift model as special cases. The test statistic $M_{0,\Delta}(p)$ for Aït-Sahalia's model has much smaller values than those for the linear drift model and Ahn and Gao's model. This suggests that Aït-Sahalia's (1996) nonlinear drift best captures the drift dynamics of the interest rate process among the three drift models under study, although it is still firmly rejected by $M_{0,\Delta}(p)$. Interestingly, the correlation-based test $M_{1,\Delta}(p)$ becomes insignificant for Ait-Sahalia's nonlinear drift model, which suggests that the misspecified component contained in the discrete-time residual is serially uncorrelated, although it is not a *m.d.s.* The ARCH-in-Mean type test statistic $M_{2,\Delta}(p)$, although still significant, becomes smaller than those for the linear drift model and Ahn and Gao's quadratic drift model. The symmetry test statistic $M_{3,\Delta}(p)$ also becomes much smaller than before, indicating that Aït-Sahalia's nonlinear drift model can capture part of asymmetric drift dynamics as well as the interest rate movements at the extreme levels. This is consistent with the fact that Aït-Sahalia's nonlinear drift model allows for asymmetric drift dynamics while the linear drift and Ahn and Gao's quadratic drift models exclude it. Unlike for the linear drift and Ahn and Gao's models, the derivative test $M_{4,\Delta}(p)$ now becomes significant.

In summary, we have the following findings:

- There exists rather strong evidence of nonlinear drift for the 7-day Eurodollar rate and such evidence is robust to different estimation methods, the presence of level effect, stochastic volatility and jumps. The evidence of nonlinear drift is not likely to be spurious given the reasonable finite sample performance of the generalized spectral derivative test as demonstrated in the empirically realistic simulation study.

- Some popular nonlinear drift models substantially improve the goodness of fit over the linear drift model. In particular, Aït-Sahalia's (1996) nonlinear drift model outperforms both the linear drift model and Ahn and Gao's (1999) quadratic drift model, and it can capture some important features of the drift dynamics, including the asymmetric mean-reverting drift dynamics and interest rate movements at the extreme levels.

- However, the nonlinear drift models are still severely misspecified. The nonlinear drift dynamics of the 7-day Eurodollar rate appears subtle and complicated. There exists room for further improving the modelling of the drift function for the short term interest rate.

5. Conclusion

In this paper, we have addressed an unresolved important issue in the interest rate modelling literature, namely whether the drift of the interest rate process is linear or nonlinear. We use a new approach that avoids the undesired features of smoothed nonparametric estimation and is robust to a very general error structure that allows for stochastic volatility and jumps. The new approach is made possible by first considering a continuous-time regression for the interest rate and then testing it via a generalized spectral derivative approach and discretely sampled interest rate data. In an empirically realistic simulation study, the generalized spectral derivative test is shown to provide a reliable inference in finite samples for highly persistent interest rate data in a continuous-time modelling setup. With this test, we document very strong evidence that the drift of the interest rates is highly nonlinear, with significant volatility-in-mean type effects and asymmetric mean-reverting dynamics. These findings are robust to various estimation methods of drift parameters, to a general error structure that allows for level effect, stochastic volatility and jumps. We also document that such popular nonlinear drift models such as Aït-Sahalia's (1996a) nonlinear drift model and Ahn and Gao's (1991) Inverse-Feller drift model can capture some nonlinear drift dynamics of the short rate, and Aït-Sahalia's nonlinear drift model outperforms Ahn and Gao's quadratic drift model due to its flexibility to capture asymmetric mean-reverting feature. However, these nonlinear drift models are still firmly rejected. Our results indicate that there exists room for better modelling the dynamics of the interest rate process by considering nonlinear drift, no matter in one-factor or multi-factor continuous-time modelling.

References

Aït-Sahalia, Y. (1996a), "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies* 9, 385-426.

Aït-Sahalia, Y. (1996b), "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica* 64, 527-560.

Aït-Sahalia, Y., Fan and H. Peng (2008), "Nonparametric Transition-Based Tests for Diffusions," forthcoming in *Journal of American Statistical Association.*

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.

Andersen, T.G., L. Benzoni and J. Lund (2004), "Stochastic Volatility, Mean Drift, and Jumps in the Short-Term Interest Rate," Working Paper, Econometric Society 2004 North American Winter Meetings 432, Econometric Society.

Arapis M. and Gao J. (2006), "Empirical Comparisons in Short-Term Interest Rate Models Using Nonparametric Methods," *Journal of Financial Econometrics* 4(2), 310-345.

Bandi F. and Phillips P. (2007), "A Simple Approach to the Parametric Estimation of Potentially Nonstationary Diffusions" *Journal of Econometrics* 137(2), 354-395.

Black F. and Karasinski P. (1991), "Bond and Option Pricing When Short Rates are Lognormal,".*Financial Analysts Journal*, 52–59.

Brennan, M. J., and E. S. Schwartz (1979), "A Continuous Time Approach to the Pricing of Bonds," *Journal of Banking and Finance* 3, 133-155.

Brennan, M. J., and E. S. Schwartz (1982), "An Equilibrium Model of Bond Pricing and a Test of Market Efficiency," *Journal of Financial and Quantitative Analysis* 17, 210-239.

Carrasco M., Chernov M., Florens J. and Ghysels E. (2002), "Efficient Estimation of Jump Diffusions and General Dynamic Models with a Continuum of Moment Conditions," CIRANO Working Papers.

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.

Chapman, D. and N. Pearson (2000), "Is the Short Rate Drift Actually Nonlinear?," *Journal of Finance* , 355-388.

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.

Corradi, V and H. White (1999), "Specification Tests for the Variance of a Diffusion," *Journal of Time Series Analysis* 20, 253-270.

Courtadon G. (1982), "The Pricing of Options on Default-Free Bonds, *Journal of Financial and Quantitative Analysis* 17, 75–100.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica* 53, 321-346.

Das S.R, (2002), "The surprise element: jumps in interest rates," *Journal of Econometrics* 106, 27–65.

Duffie D., Pan J. and Singleton K. (2000), "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica* 68, 1343–1376.

Durham, G. B. (2003), "Likelihood-Based Specification Analysis of Continuous-Time Models of the Short-Term Interest Rate," *Journal of Financial Economics* 70, 463–487.

Engle R., D. Lilien and R. Robins(1987), "Estimation of Time Varying Risk Premia in the Term Structure: the ARCH-M Model," *Econometrica* 55, 391-407.

French K., and R. Roll (1986), "Stock Return variances: the arrival of information and the reaction of traders," *Journal of Financial Economics* 17, 5–26.

Gao, J. (2007), *Nonlinear Time Series: Semiparametric and Nonparametric Methods,* Chapman & Hall/CRC, London.

Gao, J. and Casas, I. (2008), "Specification Testing in Discretized Diffusion models," *Journal of Econometrics* 147, 131-140.

Gao, J. and King, M (2004), "Adaptive Testing in Continuous-Time Diffusion models," *Econometric Theory* 20, 844-883

Hardle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.

Hong, Y. and Y. Lee (2005), "Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroskedasticity of Unknown Form," *Review of Economic Studies* 72, 499-541.

Hong, Y. and H. Li (2005), "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates," *Review of Financial Studies* 18, 38-84.

Kristensen, D. (2008a), "Pseudo-Maximum Likelihood Estimation in Two Classes of Semiparametric Diffusion Models," Working Paper, Department of Economics, Columbia University.

Kristensen, D. (2008b), "Nonparametric Estimation and Misspecification Testing of Diffusion Models," Working Paper, Department of Economics, Columbia University.

Jiang, G. J. and J. 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. (1998), "Nonparametric Modeling of US Interest Rate Term Structure Dynamics and Implications on the Prices of Derivative Securities," *Journal of Financial and Quantitative Analysis* 33, 465-497.

Johannes, M. (2004), "The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models," *Journal of Finance* 59, 227-260.

Jones, C. S. (1999), "Bayesian Estimation of Continuous-Time Finance Models," working article, University of Rochester.

Jones, C. S. (2003), "Nonlinear Mean Reversion in the Short-Term Interest Rate," *Review of Financial Studies* 16, 793-843.

Lo, A.W., and J. Wang (1995), "Implementing option pricing models when asset returns are predictable," Journal of Finance, L,1, 87–129.

Li, F.(2007), "Testing the Parametric Specification of the Diffusion Function in a Diffusion Process," Econometric Theory, 23, 221-250.

Li, M., N. D. Pearson, and A. M. Poteshman (2004), "Conditional Estimation of Diffusion Processes," *Journal of Financial Economics* 74, 31–66.

Neftci, S.N. (2000), *Introduction to the Mathematics of Financial Derivatives*, 2nd ed., Academic Press.

Newey, W. and K. West (1987) "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

Park, J.Y. (2008), "Martingale Regression," Working Paper, Department of Economics, Texas A&M.

Pearson N.D., and Sun T. S (1994), "Exploiting the Conditional Density in Estimating the Term Structure: An application to the Cox, Ingersoll, and Ross Model," *Journal of Finance* 49, 1279-1304.

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.

Phillips, P.C.B. and J. Yu (2008). "A Two-stage Realized Volatility Approach to Estimation of Diffusion Processes with Discrete Data", forthcoming *Journal of Econometrics.*

Pritsker, M. (1998), "Nonparametric Density Estimation of Tests of Continuous Time Interest Rate Models," *Review of Financial Studies* 11, 449-487.

Protter, P. (2005), Stochastic Integration and Differential Equations, 2nd ed., Springer-Verlag, Heidelberg.

Revuz, D. and M, Yor (2005), *Continuous Martingales and Brownian Motion*, 3rd ed, Springer.

Sam, A. G. and G. J. Jiang (2007), "Nonparametric Estimation of the Short Rate Diffusion Process from a Panel of Yields," Forthcoming in *Journal of Financial and Quantitative Analysis.*

Song, Z. (2009), "Generating Conditional Moment Restrictions with Complete Identification: A GMM Estimator for Diffusion Models," Working Paper, Cornell University.

Stanton, R. (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance* 52, 1973-2002.

Takamiazawa, H. (2008), "Is Nonlinear Drift Implied by the Short End of the Term Structure?," *Review of Financial Studies* 21, 311-346.

Vasicek, O. (1977), "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics* 5, 177-188.

TABLE 1. Empirical Sizes Under the CIR Model

T	$T=250$		$T=500$		$T=1000$		$T=250$		$T=500$		$T=1000$	
	$\hat{M}_{0,\Delta}(\hat{p}_0)$						$\hat{M}_{0,\Delta}(\hat{p}_0)$					
α	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
	Low Persistence $(\alpha,\sigma^2)=(0.85837{,}0.002185)$						High Persistence $(\alpha,\sigma^2)=(0.214592{,}0.000546)$					
$\bar{p}$	OLS						OLS					
2	9.1	6.0	9.8	6.1	9.7	6.6	8.2	5.0	10.7	7.1	9.9	5.6
4	9.1	6.0	9.8	6.1	9.7	6.6	8.2	5.0	10.7	7.1	9.9	5.6
6	9.1	6.0	9.8	6.1	9.7	6.6	8.2	5.0	10.7	7.1	9.9	5.6
8	9.2	6.0	9.8	6.1	9.7	6.6	8.5	5.1	10.7	7.1	9.9	5.6
10	10.0	6.1	10.0	6.3	9.7	6.6	8.4	5.5	10.8	7.5	10.2	5.6
	Moment-Based Estimator						Moment-Based Estimator					
2	9.5	6.8	9.6	6.5	10.3	6.0	9.1	5.7	9.4	6.6	10.0	6.5
4	9.5	6.8	9.6	6.5	10.3	6.0	9.1	5.7	9.4	6.6	10.0	6.5
6	9.5	6.8	9.6	6.5	10.3	6.0	9.1	5.7	9.4	6.6	10.0	6.5
8	9.5	7.0	9.7	6.5	10.3	6.0	9.3	5.9	9.4	6.6	10.0	6.5
10	10.0	7.1	10.9	6.5	10.3	6.1	9.9	6.3	9.9	7.0	10.0	6.5

Notes : (i) 1000 iterations;
(ii) DGP [CIR model]: $dY_t = \alpha(\beta - Y_t)dt + \sigma\sqrt{Y_t}dW_t$, with $\beta = 0.085711$,
$(\alpha,\sigma^2) = (0.214592{,}0.000546)$ and $(\alpha,\sigma^2) = (0.85837{,}0.002185)$, respectively;
(iii) $\hat{M}_{0,\Delta}(\hat{p}_0)$, generalized spectral test in (3.9);
(iv) $\hat{p}_0$ is a plug-in data driven lag order based on the preliminary bandwidth $\bar{p} = 2{,}4{,}6{,}8{,}10$.

TABLE 2. Empirical Sizes Under Stochastic Volatility

T	$T=250$		$T=500$		$T=1000$		$T=250$		$T=500$		$T=1000$	
	$\hat{M}_{0,\Delta}(\hat{p}_0)$						$\hat{M}_{0,\Delta}(\hat{p}_0)$					
$\bar{p}$	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
	Moment-Based Estimator						OLS					
2	9.3	6.1	11.7	8.5	10.5	6.6	8.8	5.4	9.1	5.9	10.3	6.7
4	9.3	6.1	11.7	8.5	10.5	6.6	8.8	5.4	9.1	5.9	10.3	6.7
6	9.3	6.1	11.7	8.5	10.5	6.6	8.8	5.4	9.1	5.9	10.3	6.7
8	9.5	6.2	11.7	8.5	10.5	6.6	8.9	5.5	9.1	5.9	10.3	6.7
10	10.0	6.6	12.3	9.0	10.7	6.8	9.3	5.4	10.3	6.2	10.4	6.8

Notes : (i) 1000 iterations;
(ii) DGP [SV with linear drift]: $dY_t = \alpha_1(\beta_1 - X_t)dt + \sigma_t Y_t^{\gamma} dW_t, d\log\sigma_t^2 = \alpha_2(\beta_2 - \log\sigma_t^2)dt + \xi dB_t$ with $(\alpha_1, \beta_1, \gamma, \alpha_2, \xi) =$ (0.1633,0.0595,0.5443,1.0397,-6.3599);
(iii) $\hat{M}_{0,\Delta}(\hat{p}_0)$, generalized spectral test in (3.9);
(iv) $\hat{p}_0$ is a plug-in data driven lag order based on the preliminary bandwidth $\bar{p} =$ 2,4,6,8,10.

TABLE 3. Empirical Sizes Under Affine Jump Diffusion

T	$T=250$		$T=500$		$T=1000$		$T=250$		$T=500$		$T=1000$	
	$\hat{M}_{0,\Delta}(\hat{p}_0)$						$\hat{M}_{0,\Delta}(\hat{p}_0)$					
$\bar{p}$	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
	Moment-Based Estimator						OLS					
2	6.1	3.9	5.0	3.5	6.4	4.3	2.4	1.7	2.7	1.6	4.0	2.1
4	6.1	3.9	5.0	3.5	6.4	4.3	2.4	1.7	2.7	1.6	4.0	2.1
6	6.1	3.9	5.0	3.5	6.4	4.3	2.4	1.7	2.7	1.6	4.0	2.1
8	6.2	4.0	4.9	3.4	6.4	4.3	2.4	1.7	2.7	1.6	4.0	2.1
10	5.9	3.9	5.1	3.2	6.4	4.4	2.4	1.6	2.7	1.6	4.0	2.1

Notes : (i) 1000 iterations;
(ii) DGPs: CIR model with jumps: $dX_t = \phi(\beta - X_t)dt + \sigma\sqrt{X_t}dW_t + JdN_t$, $J \sim N\left(\mu_J, \sigma_J^2\right)$, $N_t \sim Po(\lambda)$ with $(\phi, \beta, \sigma, \mu_J, \sigma_J) = (0.8542, 0.0330, 0.017, 0.0004, 0.0058)$ and $\lambda=5$;
(iii) $\hat{M}_{0,\Delta}(\hat{p}_0)$, generalized spectral test in (3.9);
(iv) $\hat{p}_0$ is a plug-in data driven lag order based on the preliminary bandwidth $\bar{p} = 2,4,6,8,10$.

TABLE 4. Linearity Testing for 7-Day Eurodollar Rates

	Moment-Based Estimator									
	$\hat{\alpha} = 0.10035$, $\hat{\beta} = 0.227$									
$\bar{p}$	$\hat{M}_{0,\Delta}$	P-value	$\hat{M}_{1,\Delta}$	P-value	$\hat{M}_{2,\Delta}$	P-value	$\hat{M}_{3,\Delta}$	P-value	$\hat{M}_{4,\Delta}$	P-value
2	53.84	0.000	95.09	0.000	1.741	0.041	12.79	0.000	0.169	0.433
5	65.40	0.000	121.97	0.000	1.741	0.042	15.90	0.000	0.208	0.418
10	123.7	0.000	123.77	0.000	2.513	0.006	17.98	0.000	0.769	0.221
	OLS									
	$\hat{\alpha} = 1.335$, $\hat{\beta} = 0.0083$									
$\bar{p}$	$\hat{M}_{0,\Delta}$	P-value	$\hat{M}_{1,\Delta}$	P-value	$\hat{M}_{2,\Delta}$	P-value	$\hat{M}_{3,\Delta}$	P-value	$\hat{M}_{4,\Delta}$	P-value
2	90.06	0.000	93.44	0.000	3.416	0.0003	12.37	0.000	0.277	0.391
5	119.5	0.000	121.01	0.000	3.412	0.0003	15.51	0.000	0.277	0.391
10	134.8	0.000	123.41	0.000	5.048	0.000	17.53	0.000	0.882	0.189

Notes: (1) The sample period for daily 7-day Eurodollar rate is from 6/01/1973 to 2/25/1995;
(2) The test statistic $\hat{M}_{0,\Delta}(\hat{p}_0)$ and derivative tests $\hat{M}_{1,\Delta}(\hat{p}_0)$, $\hat{M}_{2,\Delta}(\hat{p}_0)$, $\hat{M}_{3,\Delta}(\hat{p}_0)$, and $\hat{M}_{4,\Delta}(\hat{p}_0)$ are applied to the estimated discrete-time model residuals by the moment-based estimation and OLS estimation;
(3) $\hat{p}_0$ is a plug-in data driven lag order based on the preliminary bandwidth $\bar{p}$ = 2,5,10.

TABLE 5. Diagnostic Testing for Nonlinear Drift Models of 7-Day Eurodollar Rates

	Ahn and Gao: $\mu(Y_t,\theta) = Y_t[\beta - (\sigma^2 - \beta\alpha)Y_t]$									
	$\hat\alpha = 0.551$, $\hat\beta = 4.11$									
$\bar p$	$\hat M_{0,\Delta}$	P-value	$\hat M_{1,\Delta}$	P-value	$\hat M_{2,\Delta}$	P-value	$\hat M_{3,\Delta}$	P-value	$\hat M_{4,\Delta}$	P-value
2	52.44	0.000	93.11	0.000	2.70	0.000	12.94	0.000	0.297	0.383
5	65.29	0.000	120.9	0.000	3.06	0.000	16.03	0.000	0.297	0.383
10	66.63	0.000	123.1	0.000	4.41	0.000	17.97	0.000	1.120	0.452

	Aït-Sahalia: $\mu(Y_t,\theta) = \alpha_{-1}Y_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2$									
	$\hat\alpha_{-1} = -0.0018$, $\hat\alpha_0 = 37.13$, $\hat\alpha_1 = 1.24$ $\hat\alpha_2 = -9.32$									
$\bar p$	$\hat M_{0,\Delta}$	P-value	$\hat M_{1,\Delta}$	P-value	$\hat M_{2,\Delta}$	P-value	$\hat M_{3,\Delta}$	P-value	$\hat M_{4,\Delta}$	P-value
2	41.31	0.000	0.761	0.223	2.55	0.005	5.25	0.000	8.82	0.000
5	48.27	0.000	0.569	0.285	2.12	0.017	4.60	0.000	8.02	0.000
10	47.56	0.000	0.425	0.335	1.66	0.048	3.68	0.0001	6.49	0.000

Notes: (1) The sample period for daily 7-day Eurodollar rate is from 6/01/1973 to 2/25/1995;
(2) The test statistic $\hat M_{0,\Delta}(\hat p_0)$ and derivative tests $\hat M_{1,\Delta}(\hat p_0)$, $\hat M_{2,\Delta}(\hat p_0)$, $\hat M_{3,\Delta}(\hat p_0)$, and $\hat M_{4,\Delta}(\hat p_0)$ are applied to the estimated discrete-time model residuals of the nonlinear drift models of Ahn and Gao (1999) and Aït-Sahalia (1996) by the moment-based estimation;
(3) $\hat p_0$ is a plug-in data driven lag order based on the preliminary bandwidth $\bar p = 2,5,10$.