A Semi-parametric Test for Drift Specification in the Diffusion Model

LIN ZHU*

Indiana University

April 3, 2009

Abstract

In this paper, we propose a misspecification test for the drift coefficient in a semi-parametric diffusion model. Our test is based on the score marked empirical process whose asymptotic behavior will be distorted by the estimation of the drift parameters. We use martingale transformtion to take away the estimation effects which makes our test asymptotic distribution-free. The limit theory relies on both "in-fill" and "long-span" asymptotics. The size and power properties are examined via simulation studies and a empirical work is implemented for testing the mean-reverting spot interest rate model.

Keywords and Phrases: Score marked empirical process; Semi-parametric estimation; Martingale transformation; Asymptotic distribution-free test.

JEL CLASSIFICATION:

1. INTRODUCTION

Since Merton's seminal work around 1970s, continuous-time models have been widely used in economics and finance, for example, asset pricing, derivative valuation and term structure theory. The most commonly used continuous-time models are diffusion processes, which are generally described by means of stochastic differential equations of the form

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad 0 \le t \le T \tag{1}$$

*I am extremely thankful to my advisor Prof. Juan Carlos Escanciano for his numerous guidance and help. All errors are my own. Department of Economics, Indiana University; 100 S. Woodlawn, Wylie Hall; Bloomington, IN 47405-7104; USA, e-mail: lz2@indiana.edu

where W_t is the standard Wiener process, the drift coefficient $\mu(\cdot)$ and the diffusion coefficient $\sigma(\cdot)$ are both assumed to be continuous and may depend on unknown parameters.

In the empirical literature, many parametric models have been proposed in favor of different purposes and most of them are mutually exclusive. For example, in modeling the spot interest rate, different authors specify different diffusion functions $\sigma(\cdot)$ (see Aït-Sahalia (1996b) for a summary), however, almost all these models have specified the same mean-reverting drift function. The mean-reverting behavior can be viewed as a common belief when modeling the dynamics of the instantaneous spot interest rate despite different beliefs on the diffusion function. Hence, in order to reduce the risk of misspecification, we can work on the semi-parametric model where only the drift function is parametrized and the diffusion part is left unspecified.

Another motivation for semi-parametric model comes from the purpose of identifying the source of misspecification. Imagine that, when we jointly test the drift and diffusion functions, even we can reject the null, we still have no idea about which part is misspecified: the drift, the diffusion, or both? Thus, it would be more appropriate to consider the semi-parametric model by only parametrizing the drift function if we are interested in testing the correct specification of the drift. Furthermore, by observing that in (1) the drift is of order dt, and the diffusion is of order $\sqrt{dt}$, therefore, with infinitesimal time change, nonparametric identification becomes much easier for the diffusion term than the drift term, see Jiang and Knight (1997). Hence, with parametric specification for drift, it is more precise to do estimation.

Compared with extensive studies on estimating and testing the parametric diffusion models, there are quite few papers on semi-parametric models, among which Arapis and Gao (2006), Kristensen (2008a,b) and Park (2008) are most relevent. Arapis and Gao (2006) proposed a semi-parametric approach for testing the linearity of the drift function in the spot intestest rate model based on discretized version of the diffusion model. Kristensen (2008a,b) developed a framework for estimating and testing a semi-parametric diffusion model with either drift function or diffusion function specified. Park (2008) applied time change method to estimate the parameters in the drift part.

In this paper, we modify the optimal martingale estimating equation method to make it applicable to the semi-parametric diffusion model and the resulting estimator of the unknown parameters in drift function is $\sqrt{T}$-consistent resorting to both "in-fill" and "long-span" asymptotics framework. Our test is based on the score marked empirical process proposed in Negri and Nishiyama (2008) for diffusion models. However, two proablems make their test not applicable to our testing framework: (1) their test is only for simple hypothesis, that is, under the null, both drift and diffusion are explicitly given including the parameters, but in this paper, we want to test the correct specification of the functional form of drift function with diffusion function unspecified; (2) their test is not feasible in practice in the sense that continuous observations are not obtainable for financial or economic data. Hence, we have to fill up the gaps by taking into account of the estimation effects and discreteness of the data. We utilize the martingale transformation method, proposed in Khmaladze (1981), to take away the estimation effects which makes our test asymptotic distribution-free.

The remainder of this paper is organized as follows. In Section 2 we state the hypothesis of interest and introduce our test statistic. In Section 3 the asymptotic properties of the test are discussed. The size and power performance of the test are examined by Monte Carlo study in Section 4, and we will also test the hypothesis of linear mean-reverting drift in the spot interest rate. Section 5 concludes. The proofs are provided in the Appendix.

2. NULL HYPOTHESIS AND TEST STATISTIC

Throughout this paper, we consider a semiparametric scalar diffusion model of this form

$$dX_t = \mu(X_t, \theta)dt + \sigma(X_t)dW_t \tag{2}$$

where $\mu(\cdot, \theta)$ is a parametric function known upto the finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^d$ for a positive integer d, and $\sigma(\cdot)$ is unknown but sufficiently smooth. Under some regularity conditions, the above stochastic differential equation has a unique strong solution $\{X_t, t > 0\}$ which is stationary and ergodic, see Durrett (1996). The stationary marginal density is given by

$$\pi(x, \theta) = \frac{\xi(\theta)}{\sigma^2(x)} \exp\left\{2\int_{x_0}^{x} \frac{\mu(u, \theta)}{\sigma^2(u)} du\right\} \tag{3}$$

where the process is distributed on $D = (\underline{x}, \overline{x})$ with $-\infty \leq \underline{x} < \overline{x} \leq \infty$ (for example, $\underline{x} = 0, \overline{x} = \infty$ is of particular interest in finance). The lower bound x_0 is irrelevant because $\xi(\theta)$ is chosen to ensure that the density integrates to one. Notice that since $\sigma(\cdot)$ is unknown, we don't have an explicit form for the marginal density. The observations are obtained at discrete-time level, and we assume the observations are equispaced to keep exposition simple although the results can be easily extended to non-equispaced case, however, our method is not applicable to randomly sampled data due to additional unknown structure in conditional moments indued by random sampling (see Duffie and Glynn (2004)). Denote the data as $\{X_{i\Delta} : i = 0, 1, ..., N, \text{ and } N \cdot \Delta = T\}$, where T is the observation horizon. As will be seen later, our asymptotic theory will depend on both "in-fill" and "long-span" asymptotics, that is, $\Delta \to 0$ and $T \to \infty$.

In this paper, as pointed out in the introduction, we are interested in testing the correct specification of the drift coefficient. Formally, our null and alternative hypotheses are

$$\begin{aligned} H_0 &: \exists \text{ some } \theta_0 \in \Theta, \text{such that } \mu(\cdot) = \mu(\cdot, \theta_0) \\ H_1 &: \mu(\cdot) \neq \mu(\cdot, \theta) \text{ for any } \theta \in \Theta \end{aligned} \tag{4}$$

where $\mu(\cdot)$ is the true drift function. Our test statistic is based on the **score marked empirical process** proposed by Negri and Nishiyama (2008) which is tailored for testing diffusion models. In Negri and Nishiyama (2008), the null hypothesis is simple with known parameter in drift function, and diffusion function is also known under both the null and the alternative. However, almost all existing parametric models only specify the functional form and we need to estimate the parameters in the model. As we will show later, it turns out that the estimation effect is not negligible. Moreover, their test statistic is not feasible in the sense that they assume data are continuously observed, thus in this paper, we will overcome this problem by using discretely observed data. Also notice that the underlying assumption for our testing problem is that the observations come from a stationary and ergodic diffusion process.

Before introducing our test statistic, let's first take a look at how to estimate the parameters in the drift coefficient. For the unknown diffusion function $\sigma(\cdot)$, we use Nadaraya-Watson kernel estimator due to its simplicity and good finite-sample performance for small Δ. To estimate the drift parameter θ_0, Kristensen (2008a) provided a $\sqrt{N}$-consistent Pseudo-MLE estimator with small but fixed Δ by imposing stronger conditions on the diffusion process. In this paper, since we are working in the "in-fill" asymptotic framework, we can take another approach, optimal martingale estimating equation method, as in Bibby, Jacobsen and Sørensen (2004):

$$G_N(\theta) = \sum_{i=1}^{N} \frac{\dot{\mu}(X_{(i-1)\Delta},\theta)}{\widehat{\sigma}^2(X_{(i-1)\Delta})} \left[X_{i\Delta} - X_{(i-1)\Delta} - \mu(X_{(i-1)\Delta},\theta) * \Delta\right]$$

where $\dot{\mu}(X_{(i-1)\Delta},\theta)$ is the gradient of $\mu(X_{(i-1)\Delta},\theta)$ w.r.t. θ and $\widehat{\sigma}^2(\cdot)$ is the Nadaraya-Watson kernel estimator. We get our estimator $\widehat{\theta}$ by solving the equation $G_N(\theta) = 0$. As discussed in Bibby, Jacobsen and Sørensen (2004) and Kessler (1997), this estimator is $\sqrt{N\Delta}$(or $\sqrt{T}$)-consistent and asymptotically normal when $\Delta \to 0$ and $N * \Delta^2 \to 0$.

From now on, we assume that a preliminary $\sqrt{T}$-consistent estimator of θ and a consistent estimator of $\sigma(\cdot)$ are available. To make the exposition clear and easy to follow, let's assume at this moment that the data are continuously observed and $\sigma(\cdot)$ is known. We first define two score marked empirical processes with true parameter θ_0 and with estimated parameter $\widehat{\theta}$:

$$V_T(x) = \frac{1}{\sqrt{T}} \int_0^T 1_{(-\infty,x]}(X_t)\sigma^{-1}(X_t)(dX_t - \mu(X_t,\theta_0)dt) \tag{5}$$

$$\begin{aligned} \widetilde{V}_T(x) &= \frac{1}{\sqrt{T}} \int_0^T 1_{(-\infty,x]}(X_t)\sigma^{-1}(X_t)(dX_t - \mu(X_t,\widehat{\theta})dt) \\ &= V_T(x) - \frac{1}{\sqrt{T}} \int_0^T 1_{(-\infty,x]}(X_t)\sigma^{-1}(X_t)(\mu(X_t,\widehat{\theta}) - \mu(X_t,\theta_0))dt \end{aligned} \tag{6}$$

then as in Negri and Nishiyama (2008), $\{V_T(x) : x \in \mathbb{R}\}$ converges weakly to the Gaussian process $\{B \circ F(x) : x \in \mathbb{R}\}$ with mean 0 and covariance function given by $Cov(B \circ F(x), B \circ F(y)) = F(x \wedge y)$

where $B(\cdot)$ is a standard Brownian motion and $F(\cdot)$ is the invariant distribution function of the diffusion process under the null; notice the weak convergence takes place in $(C_B(\mathbb{R}), \mathcal{B})$, the space of continuous and bounded functions on $\mathbb{R}$ with Borel σ-algebra $\mathcal{B}$ induced by the uniform norm, as $T \to \infty$, and we denote the weak convergence as $V_T \Longrightarrow B \circ F$ hereafter. After we replace θ_0 by its estimator $\widehat{\theta}$ under the null, it turns out that the estimation effect is not negligible, that is, the second term in (6) will distort the asymptotic distribution of $\widetilde{V}_T$ such that its asymptotic distribution depends on unknown parameters. Basically, there are two ways to tackle this problem: we can either use bootstrap method (or subsampling) to approximate its finite-sample distribution, or we can take away the estimation effects by applying a proper transformation to $\widetilde{V}_T(x)$ such that the transformed process is asymptotic distribution-free (ADF hereafter). The second approach is known as martingale transformation method proposed by Khmaladze (1981), later Stute, Thies and Zhu (1998) applied this method to model checks for regression model with unknown heterogeneity and Koul and Stute (1999) utilized it for time series models. Briefly, the martingale transformation $\mathcal{T}$ is an isometry in the space orthogonal to the score function determined by the model. It suffices to make sure that the martingale tranformation preserves the distribution of $B \circ F$ and takes away the asymptotic distortion from estimating the unknown parameters.

In our problem, the tranformation $\mathcal{T}$ is defined as

$$(\mathcal{T}\varphi)(x) = \varphi(x) - \int_{\underline{x}}^{x} \sigma^{-1}(y)\dot{\mu}^{T}(y, \theta_0)A^{-1}(y)\int_{y}^{\infty} \sigma^{-1}(u)\dot{\mu}(u, \theta_0)d\varphi(u)\ dF(y) \tag{7}$$

where $A(y) = \int_y^{\infty} \sigma^{-2}(u)\dot{\mu}(u, \theta_0)\dot{\mu}^{T}(u, \theta_0)dF(u)$ is assumed to be positive definite for all $y \in D$. Notice that $\mathcal{T}$ involves a set of unknown things: $\theta_0, \sigma(\cdot)$ and $F(\cdot)$. To get a feasible transformation, we need to replace all unknown factors by their corresponding estimates, and we denote the feasible transformation as $\widehat{\mathcal{T}}$:

$$(\widehat{\mathcal{T}}\varphi)(x) = \varphi(x) - \int_{\underline{x}}^{x} \widehat{\sigma}^{-1}(y)\dot{\mu}^{T}(y, \widehat{\theta})\widehat{A}^{-1}(y)\int_{y}^{\infty} \widehat{\sigma}^{-1}(u)\dot{\mu}(u, \widehat{\theta})d\varphi(u)\ d\widehat{F}(y) \tag{8}$$

where $\widehat{A}(y) = \int_y^{\infty} \widehat{\sigma}^{-2}(u)\dot{\mu}(u, \widehat{\theta})\dot{\mu}^{T}(u, \widehat{\theta})d\widehat{F}(u)$.

Since all financial and economic data are discretely observed, any feasible estimation method or testing procedure for economic models should take the discreteness into consideration. Hence, with discrete observations $\{X_{i\Delta} : i = 0, 1, ..., N, \text{ with } N * \Delta = T\}$, we define two score marked empirical processes here, one with true diffusion term $\sigma(\cdot)$ and another with estimate $\widehat{\sigma}(\cdot)$

$$\widehat{V}^0_{N,\Delta}(x) = \frac{1}{\sqrt{N * \Delta}} \sum_{i=0}^{N-1} 1_{(-\infty,x]}(X_{i\Delta})\sigma^{-1}(X_{i\Delta})(X_{(i+1)\Delta} - X_{i\Delta} - \mu(X_{i\Delta}, \widehat{\theta}) * \triangle). \tag{9}$$

$$\widehat{V}^1_{N,\Delta}(x) = \frac{1}{\sqrt{N * \Delta}} \sum_{i=0}^{N-1} 1_{(-\infty,x]}(X_{i\Delta})\widehat{\sigma}^{-1}(X_{i\Delta})(X_{(i+1)\Delta} - X_{i\Delta} - \mu(X_{i\Delta}, \widehat{\theta}) * \triangle). \tag{10}$$

And our test statistic is of Kolmogorov type based on the transformed process $\widehat{\mathcal{T}}\widehat{V}_{N,\Delta}$, and we have the following weak convergence result:

$$\sup_{x\in\mathbb{R}}\left|\widehat{\mathcal{T}}\widehat{V}_{N,\Delta}(x)\right| \Longrightarrow \sup_{0\le t\le 1}|B(t)|\,. \tag{11}$$

Because of the ADF property of our test statistic and availability of the limit distribution, we can just implement our test by comparing the test statistic with the critical values at any specific significance level. Intuitively, the null hypothesis will be rejected if the test statistic is large enough.

3. ASYMPTOTIC PROPERTY OF THE TEST

3.1 Asymptotic Theory

In this section, we will study the asymptotic properties of our test statistic. To study the behavior of the estimation effects, we need to impose some conditions on the drift function. Following Khmaladze and Koul (2004), we assume for $\theta\in\Theta$:

$$\begin{gathered}\mu(x,\theta)-\mu(x,\theta_0)=\dot{\mu}_\theta^T(x,\theta_0)(\theta-\theta_0)+\rho_\mu(x;\theta,\theta_0),\\ 0<\int\dot{\mu}_\theta^T(x,\theta_0)\dot{\mu}_\theta(x,\theta_0)\;dF(x)<\infty\\ C_\theta:=\int\dot{\mu}_\theta(x,\theta_0)\dot{\mu}_\theta^T(x,\theta_0)\;dF(x)\;\;\text{is positive definite}\\ \int\sup_{\|\theta-\theta_0\|\le\delta}\rho_\mu^2(x;\theta,\theta_0)\;dF(x)=o(\delta^2)\end{gathered} \tag{12}$$

where $F(\cdot)$ is the invariant stationary distribution of the diffusion process and the superscript T stands for transpose. Then under the null hypothesis, we have:

Proposition 1: Under the condition (12), together with $\int_{\underline{x}}^{\overline{x}}\sigma^{-2}(x)dF(x)<+\infty$ and $\sqrt{T}(\widehat{\theta}-\theta_0)=O_P(1)$, it holds that $\widetilde{V}_T(x)=V_T(x)-G^T(x)*\sqrt{T}(\widehat{\theta}-\theta_0)+o_P(1)$ uniformly in $x\in D$ as $T\to\infty$, where $G(x)=\int_{\underline{x}}^{x}\sigma^{-1}(x)\dot{\mu}_\theta(x,\theta_0)\;dF(x)$.

Hence, the estimation of θ_0 distorts the asymptotic distribution along the score functions, which also indicates the way we construct the martingale tranformation. Now, with the martingale transformation $\mathcal{T}$ defined in section 2, it is not diffucult to see $\mathcal{T}(G^T*\sqrt{T}(\widehat{\theta}-\theta_0))=0$ because of its linearity, and our next proposition shows that the tranformation also preserves the distribution of $B\circ F$.

Proposition 2: Let $\mathcal{T}$ be defined by (7) and $\widetilde{V}_T$ defined by (6), then we have $\mathcal{T}\widetilde{V}_T\Longrightarrow B\circ F$ as $T\longrightarrow\infty$ in $(C_B(\mathbb{R}),\mathcal{B})$, the space of continuous and bounded functions endowed with the topology induced by the uniform norm.

Now with discretely observed data, we first examine the difference between $\widetilde{V}_T$ and $\widehat{V}^0_{N,\Delta}$. As we know, the convergence of residual based empirical process depends on the martingale property of the residual, in which we should know $f(X_{i\Delta}) := E(X_{(i+1)\Delta}|X_{i\Delta})$ upto unknown parameters. Since except in a few cases, $E(X_{(i+1)\Delta}|X_{i\Delta})$ generally doesn't have closed form expression for diffusion models, we have to rely on the approximation. Here, $E(X_{(i+1)\Delta}|X_{i\Delta}) = X_{i\Delta}+\mu(X_{i\Delta},\theta)\Delta+O_P(\Delta^2)$ is the first order approximation, see Stanton (1997). Hence, in addition to the "long-span" framework ($T \to \infty$), our test also has to resort to the "in-fill" asymptotics ($\Delta \to 0$). However, our finite-sample simulation result in next section indicates that daily data are good enough to implement our test.

Proposition 3: Under the null hypothesis, with estimators $(\widehat{\theta}, \widehat{\sigma}(\cdot))$ for $(\theta, \sigma(\cdot))$ defined above, then $\widehat{V}^0_{N,\Delta}(x) = \widetilde{V}_T(x) + o_P(1)$ uniformly in $x \in D$, as $\Delta \to 0$.

A consequence of proposition 3 is that $\mathcal{T}\widehat{V}^0_{N,\Delta} = \mathcal{T}\widetilde{V}_T + o_p(1)$, which yields the following theorem by combining with proposition 2:

Theorem 1: Under the null hypothesis, the tranformed score marked empirical process $\mathcal{T}\widehat{V}^0_{N,\Delta}$ converges weakly to $B \circ F$ in the Skorokhod space as $\Delta \to 0$, $T \to \infty$ and $T * \Delta \to 0$.

Theorem 1 is still infeasible for application because both $\mathcal{T}$ and $\widehat{V}^0_{N,\Delta}$ involve unknown quantities like $\sigma(\cdot)$, θ_0 and F. Hence, we need to replace them by corresponding estimates, and the resulting score marked empirical process $\widehat{V}^1_{N,\Delta}$ and the tranformation $\widehat{\mathcal{T}}$ are defined in section 2. (Need some conditions on the estimate of $\sigma(\cdot)$) Then our main theorem follows:

Theorem 2: Under condition (12) and some condition regarding the convergence of $\widehat{\sigma}(\cdot)$, then under H_0, $\widehat{\mathcal{T}}\widehat{V}^1_{N,\Delta} \Longrightarrow B \circ F$ in distribution in the Skorokhod space as $T \to \infty$ and $\Delta \to 0$.

Based on the transformed empirical process, many asymptotic distribution-free test statistics can be formed through some continuous functionals, for example Cramer-von-Mises test and Kolmogorov-Smirnov test. In this paper, we would like to use the Kolmogorov-Smirnov test, then theorem 2 and continuous mapping theorem yield

$$\sup_{x\in\mathbb{R}} \left|\widehat{\mathcal{T}}\widehat{V}_{N,\Delta}(x)\right| \Longrightarrow \sup_{\underline{x}\le x\le\overline{x}} |B \circ F(x)| = \sup_{0\le t\le 1} |B(t)| \text{ in distribution.}$$

It's easy to simulate the distributions of the right-hand side variable. We use the critical values from Bai (2003), namely, 1.94, 2.22 and 2.80 correspondingly at the significance level 10%, 5% and 1%.

3.2 Consistency of the Test

(to be done)

4. SIMULATION STUDIES AND DATA ANALYSIS

4.1 Finite Sample Performance

This section examines the finite-sample performance of our test statistic through some Monte Carlo experiments. We are going to use two spot interest rate models to simulate our data: the CKLS model (also known as CEV model) proposed in Chan, Karolyi, Longstaff,and Sanders (1992) and the nonlinear drift model as in Aït-Sahalia (1996b).

The CKLS model is

$$dX_t = \beta(\alpha - X_t)dt + \sigma X_t^{\gamma} dW_t$$

with parameter values $(\beta, \alpha, \sigma^2, \gamma) = (0.0808, 0.0972, 0.52186, 1.46)$.

The Aït-Sahalia (1996b) nonlinear drift model is

$$dX_t = (\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2)dt + \sigma X_t^{\gamma} dW_t) \tag{13}$$

with parameter values $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \sigma^2, \gamma) = (0.00107, -0.0517, 0.877, -4.604, 0.64754, 1.50)$.

Because there is no closed form for the transition density, we will use Milstein's scheme to simulate data:

$$X_{t+\Delta} = X_t + \mu(X_t)\Delta + \sigma(X_t)\sqrt{\Delta}\varepsilon_t + \frac{1}{2}\sigma^2(X_t)\Delta(\varepsilon_t^2 - 1) \tag{14}$$

where $\{\varepsilon_t : t = 1, 2, ...\}$ are i.i.d. standard normal r.v.'s. The initial value is set to equal the mean interest rate level of the data set in Aït-Sahalia (1996b). To reduce discretization bias, we simulate 100 observations each day and sample the data at daily frequency. Throughout this section, we assume that we only observe daily data but treat the time unit "1" differently. That is, if we treat "1" as one year, then the sampling frequency is $\Delta = \frac{1}{252}$; if we treat "1" as one month, then the sampling frequency becomes $\Delta = \frac{1}{21}$. In principle, for the same data set, we have different ways to specify the sampling frequency and the time span, but different specifications will have different impacts on the size and power performance, and we will see this in the following simulation studies. The intuition is that, the drift property is more reflected in the long time span, thus to gain more power, we should require T to be as large as possible; on the other hand, we also need the sampling frequency to be small to get better estimation precision which in turn can improve size performance.

4.1.1 Size of the Test.—

To examine the size of our test, we simulate data from the CKLS model (13) which has the mean-reverting drift specification $\mu(X_t, \theta) = \beta(\alpha - X_t)$ where α is the long run mean and β is the speed of mean reversion. For the parameters given above and for each sampling frequency Δ, we simulate 1000 data sets of $\{X_{i\Delta} : i = 1, ..., N$, with $N * \Delta = T\}$ for $N = 252, 1260, 2520, 5040, 10080$, respectively. These sample sizes correspond to 1, 5, 10, 20 and 40 years when $\Delta = \frac{1}{252}$, and 12, 60, 120, 240 and 480 months when $\Delta = \frac{1}{21}$.

For each data set, we first nonparametrically estimate the diffusion function via kernel method by using the bandwidth $\widehat{\sigma} * N^{-1/5}$ as suggested in Bosq (1998), where $\widehat{\sigma}$ is the sample standard deviation. Then, we estimate the parameters in the drift term $\theta = (\alpha, \beta)'$ via the optimal martingale estimating equation method. With these estimates and by replacing marginal CDF with empirical CDF, we can compute the test statistic $\sup_{x \in \mathbb{R}} \left| \widehat{\mathcal{T}} \widehat{V}_{N,\Delta}(x) \right|$. However, we cannot take the supreme over the whole real line by observing that we can't estimate $\widehat{A}(y)$ for large y, instead we calculate $\sup_{x \in (-\infty, x_0]} \left| \widehat{\mathcal{T}} \widehat{V}_{N,\Delta}(x) \right|$ where x_0 is the 99% quantile of the dataset. We consider the empirical rejection rates using the asymptotic critical values at the 10%, 5% and 1% levels, repectively.

Table 1 and table 2 report the size perfermance with $\Delta = \frac{1}{252}$ and $\Delta = \frac{1}{21}$ respectively. As we can see, with small sampling frequency Δ, the size performance is very good even with very short time span T; with large Δ, the empirical size is extremely distorted for short time span, and it becomes better as T gets larger, for example, the empirical size agrees with norminal size when $T = 480$ months (40 years).

TABLE 1 Size of the test with $\Delta = \frac{1}{252}$

Daily data with $\Delta = 1/252$	Norminal Szie		
T = number of years	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$T = 1$	0.01	0.042	0.092
$T = 5$	0.009	0.045	0.083
$T = 10$	0.010	0.052	0.106
$T = 20$	0.016	0.059	0.110
$T = 40$	0.017	0.062	0.123

TABLE 2 Size of the test with $\Delta = \frac{1}{21}$

Daily data with $\Delta = 1/21$	Norminal Szie		
T = number of months	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$T = 12$	0.001	0.016	0.038
$T = 60$	0.002	0.035	0.088
$T = 120$	0.003	0.039	0.084
$T = 240$	0.006	0.037	0.086
$T = 480$	0.011	0.046	0.101

4.1.2 Power of the Test.—

To investigate the power of our test, we simulate data from the aforementioned nonlinear drift model (14) and test the null hypothesis that the data is generated from a mean-reverting model (linear drift). For each data set, we estimate the mean-reverting model with unknown diffusion term, then compute our test statistic as in the previous section. And we take the same design as in the size study.

Table 3 and 4 report the power perfermance with $\Delta = \frac{1}{252}$ and $\Delta = \frac{1}{21}$ respectively. As the simulation results indicate, with small time span T (when $\Delta = \frac{1}{252}$), our test almost has no power to detect the misspecification of the drift function; with the same data set, if we increase Δ, hence increase T, then we gain much better power, for example, at the significance level $\alpha = 5\%$, we can reject the null hypothesis 62.9% out of 1000 simulations when $T = 480$ months (40 years). The simulation studies for size and power performance confirms our intuition of the dilemma between selecting sampling frequency and time span.

TABLE 3 Power of the test with $\Delta = \frac{1}{252}$

Daily data with $\Delta = 1/252$	Rejection significance level		
T = number of years	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$T = 1$	0.002	0.020	0.062
$T = 5$	0.002	0.036	0.076
$T = 10$	0.009	0.035	0.078
$T = 20$			
$T = 40$			

TABLE 4 Power of the test with $\Delta = \frac{1}{21}$

Daily data with $\Delta = 1/21$	Rejection significance level		
T = number of months	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$T = 12$	0.003	0.020	0.049
$T = 60$	0.024	0.087	0.163
$T = 120$	0.038	0.159	0.244
$T = 240$	0.112	0.285	0.439
$T = 480$	0.349	0.629	0.754

4.2 Application to Spot Interest Rate Model

(not done yet) As a comparison, we first use our test to reexamine the mean-reverting spot interest rate using the same data set as in Ai-Sahalia (1996b). The daily data are from June 1973 to February 1995 with a total 5505 observations.

5. CONCLUSIONS

(to be filled)

APPENDIX

PROOF OF PROPOSITION 1: Under (12), we can write (6) as

$$\begin{aligned}
\widetilde{V}_T(x) &= \frac{1}{\sqrt{T}} \int_0^T 1_{(-\infty,x]}(X_t)\sigma^{-1}(X_t)(dX_t - \mu(X_t, \widehat{\theta})dt) \\
&= V_T(x) - \frac{1}{\sqrt{T}} \int_0^T 1_{(-\infty,x]}(X_t)\sigma^{-1}(X_t)(\mu(X_t, \widehat{\theta}) - \mu(X_t, \theta_0))dt \\
&= V_T(x) - \frac{1}{T} \int_0^T 1_{(-\infty,x]}(X_t)\sigma^{-1}(X_t)\dot{\mu}_\theta^T(X_t, \theta_0)dt \cdot \sqrt{T}(\widehat{\theta} - \theta_0) - \frac{1}{\sqrt{T}} \int_0^T 1_{(-\infty,x]}(X_t)\sigma^{-1}(X_t)\rho_\mu(X_t; \widehat{\theta}, \theta_0 \\
&= V_T(x) - I_1(x) - I_2(x)
\end{aligned}$$

$I_1(x) = (G^T(x) + o_P(1)) * \sqrt{T}(\widehat{\theta} - \theta_0) = G^T(x) * \sqrt{T}(\widehat{\theta} - \theta_0) + o_P(1)$ follows from ergodicity and stationarity of the diffusion process; we also need to show $I_2 = o_P(1)$, as $T \to \infty$.

$$\begin{aligned}
I_2 &= \int_0^T \frac{1}{\sqrt{T}} 1_{(-\infty,x]}(X_t)\sigma^{-1}(X_t)\rho_\mu(X_t; \widehat{\theta}, \theta_0)dt \\
&\leq \left(\int_0^T \frac{1}{T} 1_{(-\infty,x]}(X_t)\sigma^{-2}(X_t)dt\right)^{1/2} * \left(\int_0^T \rho_\mu^2(X_t; \widehat{\theta}, \theta_0)dt\right)^2
\end{aligned}$$

The condition (12) implies that for every $0 < k < \infty$, for any $\delta > 0$, there exists T_δ such that with probability at least $1 - \delta$ the following holds for all $T > T_\delta$:

$$E\left[\sup_{\|\theta - \theta_0\| \leq T^{-1/2}k} \int_0^T \rho_\mu^2(X_t; \theta, \theta_0)dt\right] \leq \int_0^T E\left[\sup_{\|\theta - \theta_0\| \leq T^{-1/2}k} \rho_\mu^2(X_t; \theta, \theta_0)\right] dt = T * o_P(T^{-1}k^2) = o_P(k^2)$$

Thus, with $\left\|\widehat{\theta}-\theta_0\right\| = O_P(T^{-1/2})$, we have $I_2 = o_P(1)$ uniformly in $x \in D$. □

Proof of Proposition 2: From Lemma 3.1 in Stute, Thies and Zhu (1998), $Cov(\mathcal{T}(B\circ F)(x), \mathcal{T}(B\circ F)(y)) = F(x\wedge y)$. And we also have

$$\begin{aligned}
& \mathcal{T}\widetilde{V}_T(x) - \mathcal{T}V_T(x) \\
= \;& (\widetilde{V}_T(x) - V_T(x)) - \int_{\underline{x}}^{x} \sigma^{-1}(y)\dot{\mu}^T(y,\theta_0)A^{-1}(y)\int_y^{\infty}\sigma^{-1}(u)\dot{\mu}(u,\theta_0)(\widetilde{V}_T(du) - V_T(du))\; dF(y) \\
= \;& (\widetilde{V}_T(x) - V_T(x)) \\
& + \int_{\underline{x}}^{x} \sigma^{-1}(y)\dot{\mu}^T(y,\theta_0)A^{-1}(y) * \frac{1}{\sqrt{T}}\int_0^T 1_{(y,\infty)}(X_t)\sigma^{-2}(X_t)\dot{\mu}(X_t,\theta_0)(\mu(X_t,\widehat{\theta}) - \mu(X_t,\theta_0))dt\; dF(y) \\
= \;& (-G^T(x)T^{1/2}(\widehat{\theta}-\theta_0) + o_{P(1)}) \\
& +(\int_{\underline{x}}^{x} \sigma^{-1}(y)\dot{\mu}^T(y,\theta_0)A^{-1}(y)\int_y^{\infty}\sigma^{-2}(u)\dot{\mu}(u,\theta_0)\dot{\mu}^T(u,\theta_0)dF(u)\; dF(y) * T^{1/2}(\widehat{\theta}-\theta_0) + o_P(1)) \\
= \;& o_P(1)
\end{aligned}$$

together with $\mathcal{T}V_T \Longrightarrow \mathcal{T}(B\circ F) = B\circ F$, the conclusion follows.□

Proof of Proposition 3: Let $Y_t(x) = 1_{(-\infty,x]}(X_t)\sigma^{-1}(X_t)$, $Y_i(x) = 1_{(-\infty,x]}(X_{i\Delta})\sigma^{-1}(X_{i\Delta})$ and

$$\begin{aligned}
\widehat{V}(x,\theta) \;&:=\; \frac{1}{\sqrt{T}}\sum_{i=0}^{N-1} 1_{(-\infty,x]}(X_{i\Delta})\sigma^{-1}(X_{i\Delta})(X_{(i+1)\Delta} - X_{i\Delta} - \mu(X_{i\Delta},\theta) * \triangle), \\
\widetilde{V}(x,\theta) \;&=\; \frac{1}{\sqrt{T}}\int_0^T 1_{(-\infty,x]}(X_t)\sigma^{-1}(X_t)(dX_t - \mu(X_t,\theta)dt)
\end{aligned}$$

then for each $x, Y_t(x)$ is a caglad process, and we have for each (x,θ), $\widehat{V}(x,\theta) \xrightarrow{P} \widetilde{V}(x,\theta)$ as $\Delta \to 0$. By observing that $E[\left|1_{(-\infty,x_1)}(X_t) - 1_{(-\infty,x_2)}(X_t)\right|] = F(|x_1 - x_2|)$ and together with the fact that $\theta \in \Theta$ is compact and $\widehat{V}(x,\theta)$ is continuous in θ, then the uniform convergence (in x and θ) follows, thus Prop. 3 holds.

Theorem 1 just follows from proposition 2 and 3.

Proof of Theorem 2:

REFERENCES

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

Aït-Sahalia, Y. (2007): "Estimating Continuous-Time Models Using Discretely Sampled Data", *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress, 2007*

Arapis, M.and J. Gao, 2006: "Empirical comparisons in short-term interest rate models using nonparametric methods", *Journal of Financial Econometrics* 4, 310 345.

Bibby, B. M., M. Jacobsen and M. Sørensen, 2004: "Estimating Functions for Discretely Sampled Diffusion-Type Models", working paper.

Bosq, D. (1998) "Nonparametric Statistics for Stochastic Processes", *Springer*

Cai, Z. and Hong, Y., 2008: "Some Recent Developments in Nonparametric Finance", working paper.

Duffie, D. and P. Glynn, 2004: "Estimation of continuous-time Markov processes sampled at random time intervals", *Econometrica* 72, 1773-1808.

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.

Kanaya, S. (2008) "Non-parametric specification testing for continuous-time markov processes: do the processes follow diffusions?". working paper.

Khmaladze, E. V. (1981): "Martingale approach in the theory of goodness-of-fit tests", *Theory Probab. Appl.* 26 240–257.

Khmaladze, E. V. and Koul, H. (2004): "Martingale transforms goodness-of-fit tests in regression models", *Ann. Statist.* 32 995–1034

Koul, H. L. and Stute, W. (1999): "Nonparametric model checks for time series", *Ann. Statist.* 27 204–237.

Kristensen, D., 2008: "Pseudo-Maximum Likelihood Estimation in Two Classes of Semiparametric Diffusion Models", working paper.

Kristensen, D., 2008: "Nonparametric Estimation and Misspecification Testing of Diffusion Models", working paper.

Negri, I. and Y. Nishiyama (2008): "Goodness of fit test for ergodic diffusion processes"

Park, J. Y., 2008: "Martingale Regression and Time Change", working paper.

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

Stute, W., Thies, S. and Zhu, L.-X. (1998): "Model checks for regression: An innovation process approach", *Ann. Statist.* 26 1916–1934.

Sundaresan, S. (2000). "Continuous-Time Methods in Finance: A Review and an Assessment." *Journal of Finance* 55, 1569–1622.