# Specification testing in discretized diffusion models: Theory and practice

Gao, Jiti and Casas, Isabel
The University of Adelaide, Universidad Carlos III de Madrid

November 2006

# Specification Testing in Discretized Diffusion Models: Theory and Practice

By Jiti Gao[1]

*The University of Adelaide and The University of Western Australia*

and Isabel Casas[2]

*Universidad Carlos III de Madrid*

**Abstract**: We propose two new tests for the specification of both the drift and the diffusion functions in a discretized version of a semiparametric continuous–time financial econometric model. Theoretically, we establish some asymptotic consistency results for the proposed tests. Practically, a simple selection procedure for the bandwidth parameter involved in each of the proposed tests is established based on the assessment of the power function of the test under study. To the best of our knowledge, this is the first approach of this kind in specification of continuous–time financial econometrics. The proposed theory is supported by good small and medium–sample studies.

KEYWORDS: Continuous–time diffusion process, kernel method, nonparametric testing, power function, size function, time series econometrics.

[1] Jiti Gao is from the School of Economics, The University of Adelaide, Adelaide SA 5005, Australia. Email: jiti.gao@adelaide.edu.au

[2] Department of Statistics and Econometrics, Universidad Carlos III de Madrid, C/Madrid, 126, 28903-Getafe (Madrid) Spain.

## 1. Introduction

Consider a continuous–time diffusion process of the form

$$dr_t = \mu(r_t)dt + \sigma(r_t)dB_t, \tag{1.1}$$

where $\mu(\cdot)$ and $\sigma(\cdot) > 0$ are respectively the univariate drift and volatility functions of the process, and $B_t$ is the standard Brownian motion. During the last decade or so, specification of model (1.1) has attracted a lot of attention in both theoretical studies and practical applications. For example, the practitioner would be interested in knowing which one of the following popular models is more appropriate for a given set of interest rate data:

$$dr = \beta(\alpha - r)dt + \sigma^\delta dB \text{ for } \delta = 0, 0.5, 1, \tag{1.2}$$

$$dr = \beta(\alpha - r)dt + \sigma r^\rho dB \text{ for } 0 < \rho \leq 2, \tag{1.3}$$

$$dr = r\{\kappa - (\sigma^2 - \kappa\alpha)r\}dt + \sigma r^{3/2}dB, \tag{1.4}$$

$$dr = (\alpha_{-1}r^{-1} + \alpha_0 + \alpha_1 r + \alpha_2 r^2)dt + \sigma r^{3/2}dB. \tag{1.5}$$

To make such a choice for a given set of interest rate data, one may specify model (1.1) parametrically to determine whether one of the popular parametric models is appropriate. In the field of continuous–time model specification, some closely related studies include Aït-Sahalia (1996a), who proposes a simple methodology for testing whether the marginal density function of $\{r_t\}$ belongs to a parametric family of density functions; Corradi and White (1999), who establish an asymptotically normal test for the diffusion function; Fan and Zhang (2003), who propose a simultaneous test procedure for the specification of both the drift and diffusion functions; Gao and King (2004), who propose an improved test for a parametric specification of the marginal density function; Kristensen (2004), in which a semiparametric diffusion model is considered and tested; Corradi and Swanson (2005), who propose using a bootstrap specification test; Hong and Li (2005), who establish an asymptotically consistent test for specifying the transitional density function of $\{r_t\}$ parametrically; Arapis and Gao (2006), who consider testing for a parametric specification of the drift function; Chen, Gao and Tang (2008), who develop an empirical likelihood method to establish an adaptive test for the parametric specification of the transitional density function; and Li (2007), who discusses a nonparametric test for the parametric specification of the diffusion function in a diffusion process.

For the implementation of the proposed tests, existing studies use either a single bandwidth based on an optimal estimation procedure (Aït-Sahalia 1996a; Corradi and White 1999; Fan and Zhang 2003; Hong and Li 2005) or a set of suitable bandwidth values (Horowitz and Spokoiny 2001; Gao and King 2004; Arapis and Gao 2006; Chen and Gao 2007; Gao 2007; Chen, Gao and Tang 2008). As is well–known, the first choice is based on an optimal estimation procedure, and therefore may not be optimal for testing purposes. Our own experience and others (Horowitz and Spokoiny 2001) show that the second choice can be arbitrary and problematic in practice. This is probably why in practice Horowitz and Spokoiny (2001) choose an optimal bandwidth based on the assessment of the power function of their test before constructing a suitable set of bandwidth values for the implementation of their test.

This paper mainly considers a semiparametric case where $\mu(\cdot)$ is already pre–specified while the form of $\sigma(\cdot)$ is allowed to be specified nonparametrically. The main motivation for considering such a class of semiparametric diffusion models is as follows: (a) most empirical studies suggest using a simple form for the drift function, such as a polynomial function for interest rate data, while the diffusion function is allowed to be flexible; (b) when the form of the drift function is unknown but sufficiently smooth, it may be well–approximated by a parametric form, such as by a suitable polynomial function; (c) the drift function may be treated as a constant function or even zero when interest is on studying the stochastic volatility of $\{r_t\}$; and (d) the precise form of the diffusion function is very crucial, but it is quite problematic to assume a known form for the diffusion function due to the fact that the instantaneous volatility is normally unobservable.

We first establish a simple kernel test $L(h)$ for the specification of the diffusion function through using a discretized version of such a continuous–time diffusion model, where $h$ is a bandwidth involved in the kernel test. In order to implement the proposed test in practice, we propose a new bootstrap simulation procedure to approximate the $1-\alpha$ quantile, $l_\alpha$, of the distribution of the simple test by a bootstrap simulated critical value $l_\alpha^*$. In theory, we show that the proposed test not only satisfies $P(L(h) > l_\alpha^*) = \alpha + O(\sqrt{h})$ under the null, but also is asymptotically consistent under the alternative.

In practice, we make the best use of the bootstrap to choose a suitable bandwidth such that the power function of the proposed test is maximized at such a bandwidth while the size is controlled by $\alpha$. To the best of our knowledge, the proposed theory and methodology for the specification of a discretized diffusion model is new. In addition, our finite–sample studies show that the proposed test has little size distortion and that it is also quite powerful although the 'distance' between the null and the alternative is made deliberately close.

In summary, the main contribution of this paper is as follows:

(i) It establishes a simple kernel test for specifying the diffusion function parametrically through using a discretized version of the diffusion model. An extension to the parametric specification of the drift function in a semiparametric diffusion model is also discussed.

(ii) The implementation of such a test does not require nonparametrically estimating any higher–order moments of the process. As a result, the main feature of the proposed test is its implementation with ease in practice.

(iii) The resulting theory and methodology for the discretized version is new and potentially useful to provide solutions to such nonparametric and semiparametric testing problems in continuous–time financial models without discretization.

The rest of the paper is organised as follows. Section 2 proposes a simple kernel test and establishes theoretical properties for it. A simulation procedure for implementing the proposed test is given in Section 3. Section 4 concludes the paper with some remarks. Mathematical details are relegated to the appendix. Throughout this paper, we use $a_n = O(b_n)$ to mean that there is some constant $-\infty < c_* \neq 0 < \infty$ such that $\lim_{n\to\infty} \frac{a_n}{b_n} = c_*$, and $a_n = o(b_n)$ to represent $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$.

## 2. New specification tests

### *2.1. Specification of diffusion function*

Throughout the first part of this section, we consider a semiparametric diffusion model of the form

$$dr_t = \mu(r_t, \theta)dt + \sigma(r_t)dB_t, \tag{2.1}$$

where $\mu(r, \theta)$ is a known parametric function indexed by a vector of unknown parameters, $\theta \in \Theta$ (a parameter space), and $\sigma(r)$ is an unknown but sufficiently smooth function. As pointed out in the introductory section, there is sufficient evidence that the assumption of a parametric form for the drift function is not unreasonable. In addition, Arapis and Gao (2006) show that when the drift function is unknown nonparametrically, one may specify the drift function parametrically without knowing the form of $\sigma(r)$.

Similarly to most existing studies, we apply the Euler first–order scheme to derive a discretized version of model (2.1) of the form

$$r_{t\Delta} - r_{(t-1)\Delta} = \mu(r_{(t-1)\Delta}, \theta)\Delta + \sigma(r_{(t-1)\Delta}) \cdot (B_{t\Delta} - B_{(t-1)\Delta}), \ t = 1, 2, \cdots, T, \quad (2.2)$$

where $T$ is the number of observations, $\Delta$ is the time between successive observations. In practice, $\Delta$ is small but fixed, as most continuous-time models in finance are estimated with monthly, weekly, daily, or higher frequency observations.

Model (2.2) implies that the drift function $\mu(\cdot)$ and the diffusion function $\sigma^2(\cdot)$ may be approximated by

$$\begin{aligned} \mu(r_{(t-1)\Delta}, \theta) &\approx E\left[\frac{r_{t\Delta} - r_{(t-1)\Delta}}{\Delta} | r_{(t-1)\Delta}\right] \quad \text{and} \\ \sigma^2(r_{(t-1)\Delta}) &\approx E\left[\frac{\left(r_{t\Delta} - r_{(t-1)\Delta}\right)^2}{\Delta} | r_{(t-1)\Delta}\right] \end{aligned} \quad (2.3)$$

as $\Delta \to 0$. Some existing studies, such as Bandi and Phillips 2003, Nicolau 2003, Arapis and Gao 2006, and Gao 2007, then construct nonparametric estimators of $\mu(\cdot)$ and $\sigma^2(\cdot)$ using (2.3). As a result, such nonparametric estimators of the drift and diffusion functions can only be consistent when $\Delta \to 0$. Naturally, the condition of $\Delta \to 0$ is certainly needed when tests are constructed based on (2.3) (see Li 2007).

Since the construction of the proposed tests $L_{1T}(h)$ and $L_{2T}(h)$ below is based on the discretized version (2.2) and one of the functions is always parametrically specified, the asymptotic biases of the parametric estimators involved in the tests are negligible and also independent of the choice of $\Delta$. Therefore, our theory and methodology remains applicable even when $\Delta$ is fixed.

Let $Y_t = \frac{r_{t\Delta} - r_{(t-1)\Delta}}{\Delta}$, $x_t = r_{(t-1)\Delta}$, $f(x_t, \theta) = \mu(x_t, \theta)$ and $g(x_t) = \Delta^{-1}\sigma^2(x_t)$. Model (2.2) suggests using a discrete semiparametric autoregressive time series model of the form

$$Y_t = f(x_t, \theta) + \epsilon_t \quad \text{with} \quad \epsilon_t = \sqrt{g(x_t)}\ e_t, \tag{2.4}$$

where $\{e_t\}$ is a sequence of independent N(0,1) errors and independent of $\{x_s\}$ for all $s \le t$. So $E[e_t|x_t] = E[e_t] = 0$ and $\text{var}[e_t|x_t] = \text{var}[e_t] = 1$. The main interest of this paper is to test

$$\begin{aligned} \mathcal{H}_{01} &: \quad P\left(g(x_t) = g(x_t, \vartheta_0)\right) = 1 \ \text{ versus} \\ \mathcal{H}_{11} &: \quad P\left(g(x_t) = g(x_t, \vartheta_1) + C_{1T} \cdot D_1(x_t)\right) = 1 \end{aligned} \tag{2.5}$$

for some $\vartheta_0, \vartheta_1 \in \Theta$, where both $\vartheta_0$ and $\vartheta_1$ are chosen such that Assumption A.3(ii) listed in the Appendix A holds, $\Theta$ is a parameter space, $C_{1T}$ is a sequence of real numbers, and $D_1(x_t)$ is a smooth and completely nonparametric function. Note that $\vartheta_0$ may be different from the true value, $\theta_0$, of $\theta$ involved in the drift function.

It should also be pointed out that the probabilities in (2.5) are independent of $t$ since $\{x_t\}$ is assumed to be strictly stationary. In addition, as assumed in Assumption A.4 in the Appendix, the choice of $C_{1T}$ includes both global ($C_{1T} = C_1$ not depending on $T$) and local ($C_{1T}$ tending to zero when $T$ goes to $\infty$) alternatives.

In order to construct our test for $\mathcal{H}_{01}$, we use (2.4) to formulate a regression model of the form

$$\epsilon_t^2 = g(x_t) + \eta_t, \tag{2.6}$$

where the error process $\eta_t = g(x_t)(e_t^2 - 1)$ is of the following properties: under $\mathcal{H}_{01}$

$$E[\eta_t|x_t] = 0 \quad \text{and} \quad E[\eta_t^2|x_t] = 2g^2(x_t, \vartheta_0). \tag{2.7}$$

In general, for any $k \ge 1$ we have under $\mathcal{H}_{01}$

$$E\left[\eta_t^k|x_t\right] = E\left[(e_t^2 - 1)^k\right]\ g^k(x_t, \vartheta_0) \equiv c_k g^k(x_t, \vartheta_0), \tag{2.8}$$

where $c_k = E\left[(e_t^2 - 1)^k\right]$ is a known value for each $k$ using the fact that $e_t \sim N(0,1)$ has all known moments. This implies that all higher–order conditional moments of $\{\eta_t\}$ will be specified if the second conditional moment of $\{\eta_t\}$ is specified.

Since $E[\eta_t|x_t] = 0$ under $\mathcal{H}_{01}$, we have

$$d(\eta) = E\left[\eta_t E\left(\eta_t|x_t\right)\pi(x_t)\right] = E\left[\left(E^2(\eta_t|x_t)\right)\pi(x_t)\right] = 0\ , \tag{2.9}$$

under $\mathcal{H}_{01}$. This would suggest using a kernel–based sample analogue of (2.9) of the form

$$N_T(h) = \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \widehat{\eta}_s \; p_{st} \; \widehat{\eta}_t, \tag{2.10}$$

where $p_{st} = \frac{1}{T\sqrt{h}} K\left(\frac{x_s - x_t}{h}\right)$ and $\widehat{\eta}_t = \left(Y_t - f(x_t, \widehat{\theta})\right)^2 - g(x_t, \widehat{\vartheta}_0)$, in which $\widehat{\theta}$ is a $\sqrt{T}$–consistent estimator of $\theta$ and $\widehat{\vartheta}_0$ is also a $\sqrt{T}$–consistent estimator of $\vartheta_0$ under $\mathcal{H}_{01}$. Similar test statistics for specifying parametric mean functions have been proposed and studied extensively in Fan and Li (1996), Zheng (1996), Li and Wang (1998), Li (1999), Fan and Li (2000), Fan and Linton (2003), Arapis and Gao (2006), Gao (2007) and others.

In view of the definition of $\widehat{\eta}_t$, we may have the following decomposition:

$$\begin{aligned} N_T(h) \;\; &= \;\; \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \widehat{\eta}_s \; p_{st} \; \widehat{\eta}_t = \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \eta_s \; p_{st} \; \eta_t \\ &+ \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \left(f(x_s, \theta) - f(x_s, \widehat{\theta})\right)^2 p_{st} \left(f(x_t, \theta) - f(x_t, \widehat{\theta})\right)^2 \\ &+ \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \left(g(x_s) - g(x_s, \widehat{\vartheta}_0)\right) p_{st} \left(g(x_t) - g(x_t, \widehat{\vartheta}_0)\right) \\ &+ o_P\left(N_T(h)\right), \end{aligned} \tag{2.11}$$

where $\eta_t = g(x_t)(e_t^2 - 1)$.

Also, simple calculations imply that for sufficiently large $T$

$$\text{var}[N_T(h)] = \sigma_g^2 \; (1 + o(1)), \tag{2.12}$$

where $\sigma_g^2 = 2\mu_2^2 \int K^2(u) du$ with $\mu_2 = E[\eta_1^2] = 2E\left[g^2(x_1)\right]$.

For the implementation of $N_T(h)$ in practice, in order to avoid nonparametrically estimating any unknown quantity we estimate $\sigma_g^2$ under $\mathcal{H}_{01}$ by $\widehat{\sigma}_{1T}^2 = 2\widehat{\mu}_2^2 \int K^2(u) du$ with $\widehat{\mu}_2 = \frac{2}{T} \sum_{t=1}^{T} g^2(x_t, \widehat{\vartheta}_0)$.

We then propose using a normalized version of the form

$$
\begin{aligned}
L_{1T}(h) &= \frac{\sum_{s=1}^{T}\sum_{t=1,\neq t}^{T} \widehat{\eta}_s\ p_{st}\ \widehat{\eta}_t}{\widehat{\sigma}_{1T}} = \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \eta_s\ p_{st}\ \eta_t}{\sigma_0} \cdot \frac{\sigma_0}{\widehat{\sigma}_{1T}} \\
&+ \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \left(f(x_s,\theta) - f(x_s,\widehat{\theta})\right)^2 p_{st} \left(f(x_t,\theta) - f(x_t,\widehat{\theta})\right)^2}{\widehat{\sigma}_{1T}} \\
&+ \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \left(g(x_s) - g(x_s,\widehat{\vartheta}_0)\right) p_{st} \left(g(x_t) - g(x_t,\widehat{\vartheta}_0)\right)}{\widehat{\sigma}_{1T}} \\
&+ o_P\left(\frac{N_T(h)}{\widehat{\sigma}_{1T}}\right), \qquad (2.13)
\end{aligned}
$$

where $\sigma_0^2 = 2\mu_0^2 \int K^2(u)du$ with $\mu_0 = 2\ E\left[g^2(x_1,\vartheta_0)\right]$ under $\mathcal{H}_{01}$.

Let

$$L_T(h) = \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \eta_s\ p_{st}\ \eta_t}{\sigma_0}. \qquad (2.14)$$

Lemma A.1 in the Appendix shows that under $\mathcal{H}_{01}$,

$$\lim_{T\to\infty} P\left(L_T(h) \le x\right) = \Phi(x) \qquad (2.15)$$

for $x \in I\!R$, where $\Phi(x)$ denotes the cumulative distribution function of the standard normal random variable.

The following result establishes that $L_{1T}(h)$ is asymptotic normal under $\mathcal{H}_{01}$; its proof is given in the Appendix.

**Theorem 2.1**. *Suppose that Assumptions A.1–A.3(i)(ii)(iv) listed in the Appendix hold. Then under* $\mathcal{H}_{01}$

$$\lim_{T\to\infty} P\left(L_{1T}(h) \le x\right) = \Phi(x). \qquad (2.16)$$

Theorem 2.1 shows that $L_{1T}(h)$ converges in distribution to $N(0,1)$ regardless of the choice of $\Delta$. This is mainly because the marginal density function of $\{x_t : 1 \le t \le T\}$ remains the same when $\{x_t : 1 \le t \le T\}$ is assumed to be strictly stationary. A detailed discussion is similar to that of Arapis and Gao (2006, p.323).

As shown in the Appendix, Assumption A.1(iii) that $\lim_{T\to\infty} h = 0$ and $\lim_{T\to\infty} Th = \infty$ imposes the minimal conditions on $h$ such that the asymptotic normality is the limiting distribution of the proposed test under $\mathcal{H}_{01}$. To the best of our knowledge, such minimal conditions on $h$ have only been assumed by Zheng (1996), and Li and Wang (1998) when the authors consider testing for a parametric specification for the conditional mean of a nonparametric regression model with independent observations.

As pointed out in the literature, such asymptotically normal tests may not be very useful in practice, in particular when the size of the data is not sufficiently large. Thus, the conventional $\alpha$–level asymptotic critical value, $l_{\text{acv}}$, of the standard normality may not be useful in applications. This paper proposes to approximate the $l_{\text{acv}}$ by a Monte Carlo simulated critical value.

**Simulation Procedure**: Let $l_{1\text{cv}}$ be the $1-\alpha$ quantile of the exact finite–sample distribution of $L_{1T}(h)$. Since $l_{1\text{cv}}$ may be unknown in practice, we suggest approximating $l_{1\text{cv}}$ by either a non–random approximate $\alpha$–level critical value, $l_{1\alpha}$, or a stochastic approximate $\alpha$–level critical value, $l_{1\alpha}^*$, using the following simulation procedure:

1. For each $t = 1, 2, \ldots, T$, generate $Y_t^* = f(x_t, \widehat{\theta}) + \sqrt{g(x_t, \widehat{\vartheta}_0)}\ e_t^*$, where the original sample $\mathcal{X}_T = (x_1, \cdots, x_T)$ acts in the resampling as a fixed design, $\{e_t^*\}$ is independent of $\{x_t\}$ and sampled identically distributed from $N(0,1)$. Use the data set $\{(x_t, Y_t^*) : t = 1, 2, \ldots, T\}$ to re-estimate $(\theta, \vartheta_0)$. Let $(\widehat{\theta}^*, \widehat{\vartheta}_0^*)$ denote the pair of the resulting estimates.

2. Define $L_{1T}^*(h)$ to be the version of $L_{1T}(h)$ with $(x_t, Y_t)$ and $(\widehat{\theta}, \widehat{\vartheta}_0)$ being replaced by $(x_t, Y_t^*)$ and $(\widehat{\theta}^*, \widehat{\vartheta}_0^*)$ in the calculation. Let $l_{1\alpha}$ be the $1-\alpha$ quantile of the distribution of $L_{1T}^*(h)$.

3. Repeat the above steps $M$ times and then obtain the empirical distribution of $L_{1T}^*(h)$. The bootstrap distribution of $L_{1T}^*(h)$ given $\mathcal{W}_T = \{(x_t, Y_t) : 1 \leq t \leq T\}$ is defined by $P^*\left(L_{1T}^*(h) \leq x\right) = P\left(L_{1T}^*(h) \leq x | \mathcal{W}_T\right)$. Let $l_{1\alpha}^*$ satisfy $P^*\left(L_{1T}^*(h) \geq l_{1\alpha}^*\right) = \alpha$ and then estimate $l_{1\alpha}$ by $l_{1\alpha}^*$.

It should be pointed out that both $l_{1\alpha}$ and $l_{1\alpha}^*$ may be functions of $h$. We then have the following theorem; its proof is given in the Appendix.

**Theorem 2.2**. (i) *Suppose that Assumptions A.1–A.3 hold. Then under $\mathcal{H}_{01}$ the following equation*

$$\sup_{x \in R^1} |P^*(L_{1T}^*(h) \leq x) - P(L_{1T}(h) \leq x)| = O\left(\sqrt{h}\right) \tag{2.17}$$

*holds in probability with respect to the joint distribution of* $\mathcal{W}_T$.

(ii) *Suppose that Assumptions A.1–A.3 hold. Then under* $\mathcal{H}_{01}$

$$P\left(L_{1T}(h) > l_{1\alpha}^*\right) = \alpha + O\left(\sqrt{h}\right). \tag{2.18}$$

(iii) *Assume that Assumptions A.1–A.4 hold. Then under* $\mathcal{H}_{11}$

$$\lim_{T\to\infty} P\left(L_{1T}(h) > l_{1\alpha}^*\right) = 1. \tag{2.19}$$

For some corresponding test statistics in the time series case (Li and Wang 1998; Fan and Linton 2003), asymptotic results weaker than (2.17)–(2.19) have already been established. In Section 3 below, we will show how to assess the finite–sample properties of (2.18) and (2.19).

For each $h$ we define the following size and power functions

$$S_T(h) = P\left(L_{1T}(h) > l_{1\alpha} | \mathcal{H}_{01} \text{ holds}\right) \text{ and } P_T(h) = P\left(L_{1T}(h) > l_{1\alpha} | \mathcal{H}_{11} \text{ holds}\right). \tag{2.20}$$

Correspondingly, we define $(S_T^*(h), P_T^*(h))$ with $l_{1\alpha}$ replaced by $l_{1\alpha}^*$.

To establish further results, we need to introduce the following notation:

$$\rho(h) = C_\pi \ \Pi(K) \ \sqrt{h}, \tag{2.21}$$

where $C_\pi = \frac{\int \pi^3(x)dx}{\left(\sqrt{\int \pi^2(x)dx}\right)^3}$ and $\Pi(K) = \frac{\sqrt{2}K^{(3)}(0)}{3\left(\sqrt{\int K^2(u)du}\right)^3}$, in which $K^{(3)}(\cdot)$ is the three–time convolution of $K(\cdot)$ with itself.

We now establish the following theoretical results; their proofs are given in the Appendix below.

**Theorem 2.3**. (i) *Suppose that Assumptions A.1–A.4 hold. Then*

$$S_T(h) = 1 - \Phi(l_{1\alpha} - s(h)) - \rho(h)\ (1 - (l_{1\alpha} - s(h))^2)\ \phi(l_{1\alpha} - s(h)) + o\left(\sqrt{h}\right), \tag{2.22}$$

$$S_T^*(h) = 1 - \Phi(l_{1\alpha}^* - s(h)) - \rho(h)\ (1 - (l_{1\alpha}^* - s(h))^2)\ \phi(l_{1\alpha}^* - s(h)) + o\left(\sqrt{h}\right) \tag{2.23}$$

*hold in probability with respect to the joint distribution of* $\mathcal{W}_T$, *where* $\phi(\cdot)$ *is the probability density function of* $N(0,1)$, *and* $s(h) = C_0(g)\sqrt{h}$ *with*

$$C_0(g) = \frac{T \int \left(g(x, \widehat{\vartheta}) - g(x, \vartheta_0)\right)^2 \pi^2(x)dx}{\sigma_0}.$$

(ii) *Suppose that Assumptions A.1–A.4 hold. Then the following equations hold in probability with respect to the joint distribution of* $\mathcal{W}_T$*:*

$$P_T(h) = 1 - \Phi(l_{1\alpha} - r(h)) - \rho(h)\ (1 - (l_{1\alpha} - r(h))^2)\ \phi(l_{1\alpha} - r(h)) + o\left(\sqrt{h}\right), \quad (2.24)$$

$$P_T^*(h) = 1 - \Phi(l_{1\alpha}^* - r(h)) - \rho(h)\ (1 - (l_{1\alpha}^* - r(h))^2)\ \phi(l_{1\alpha}^* - r(h)) + o\left(\sqrt{h}\right) \quad (2.25)$$

*under* $\mathcal{H}_{11}$*, where* $r(h) = \sqrt{h}\left(C_1(g) + D_{1\pi}\ TC_{1T}^2\right)$*, in which*

$$C_1(g) = \frac{T\int \left(g(x,\widehat{\vartheta}) - g(x,\vartheta_1)\right)^2 \pi^2(x)dx}{\sigma_0} \text{ and } D_{1\pi} = \frac{\int D_1^2(x)\pi^2(x)dx}{\sigma_0}. \quad (2.26)$$

As pointed out above, both $l_{1\alpha}$ and $l_{1\alpha}^*$ may be functions of $h$. Theorem 2.4 below gives asymptotically explicit expressions for noth $l_{1\alpha}$ and $l_{1\alpha}^*$. Let $\psi(\alpha) = C_\pi\ \Pi(K)\ (z_\alpha^2 - 1)$ with $z_\alpha$ being the $1-\alpha$ quantile of the standard normal distribution. The proof of Theorem 2.4 is given in the Appendix.

**Theorem 2.4**. *Assume that the conditions of Theorem 2.3(i) hold. Then for T sufficiently large*

$$l_{1\alpha} = l_{1\alpha}(h) \approx z_\alpha + \psi(\alpha)\ \sqrt{h} \quad \text{in probability}, \quad (2.27)$$

$$l_{1\alpha}^* = l_{1\alpha}^*(h) \approx z_\alpha + \psi(\alpha)\ \sqrt{h} \quad \text{in probability}. \quad (2.28)$$

Theorem 2.4 shows that there is an asymptotic correction, $\psi(\alpha)\ \sqrt{h}$, to the normal quantile $z_\alpha$. Section 3 below shows that the size of $L_{1T}(h)$ associated with $l_{1\alpha}$ is more stable than that of $L_{1T}(h)$ based on $z_\alpha$.

We now choose an optimal bandwidth $\widehat{h}_{1\text{test}}$ such that for some small $c_{\min} > 0$

$$\widehat{h}_{1\text{test}} = \arg\max_{h\in H_{1T}} P_T(h) \text{ with } H_{1T} = \{h:\ \alpha - c_{\min} < S_T(h) < \alpha + c_{\min}\}. \quad (2.29)$$

Similarly to Chapter 3 of Gao (2007), it may be shown that the leading term of $\widehat{h}_{1\text{test}}$ may be approximated by

$$\widehat{h}_{1\text{test}} = \left(\widehat{C}_\pi\ \Pi(K)\right)^{-\frac{1}{2}} \left(\widehat{D}_{1\pi}\ T\ C_{1T}^2\right)^{-\frac{3}{2}} (1 + o_P(1)), \quad (2.30)$$

where $\widehat{C}_\pi = \frac{\frac{1}{T}\sum_{t=1}^T \widehat{\pi}^2(x_t)}{\left(\sqrt{\frac{1}{T}\sum_{t=1}^T \widehat{\pi}(x_t)}\right)^3}$ and $\widehat{D}_{1\pi} = \frac{\sum_{t=1}^T \widehat{D}_1^2(x_t)\widehat{\pi}(x_t)}{T\ \widehat{\sigma}_{1T}}$ with $\widehat{D}_1(x_t) = \frac{\sum_{s=1}^T K\left(\frac{x_t - x_s}{\widehat{h}_{\text{cv}}}\right)\left(\widehat{\epsilon}_s^2 - g(x_s,\widehat{\vartheta}_0)\right)}{C_{1T}\ \sum_{u=1}^T K\left(\frac{x_t - x_u}{\widehat{h}_{\text{cv}}}\right)}$ and $\widehat{\pi}(x) = \frac{1}{T\widehat{h}_{\text{cv}}}\sum_{t=1}^T K\left(\frac{x - x_t}{\widehat{h}_{\text{cv}}}\right)$ being a density estimator, in which $\widehat{\epsilon}_t = Y_t - f(x_t,\widehat{\theta})$ and $\widehat{h}_{\text{cv}} = 1.06\ T^{-\frac{1}{5}} \cdot \sqrt{\frac{1}{T-1}\sum_{t=1}^T (x_t - \bar{x})^2}$ with $\bar{x} = \frac{1}{T}\sum_{t=1}^T x_t$.

In Section 3 below, we will show how to practically implement $\widehat{h}_{1\text{test}}$.

*2.2. Specification of drift function*

Throughout the second part of this section, we consider a semiparametric diffusion model of the form

$$dr_t = \mu(r_t)dt + \sigma(r_t, \vartheta)dB_t, \tag{2.31}$$

where $\sigma(r, \vartheta)$ is a positive parametric function indexed by a vector of unknown parameters, $\vartheta \in \Theta$ (a parameter space), and $\mu(r)$ is an unknown but sufficiently smooth function. As pointed out in existing studies, such as Kristensen (2004), there is some evidence that the assumption of a parametric form for the diffusion function is not unreasonable in such cases where the diffusion function is already pre–specified, the main interest is for example to specify whether the drift function should be linear or quadratic. More recently, Arapis and Gao (2006) discuss how to specify the drift function parametrically while the diffusion function is allowed to be unknown nonparametrically.

As for model (2.2), we suggest approximating model (2.31) by a semiparametric autoregressive model of the form

$$Y_t = f(x_t) + \sqrt{g(x_t, \vartheta)}\ e_t, \tag{2.32}$$

where $f(x_t) = \mu(x_t)$, $g(x_t, \vartheta) = \Delta^{-1}\sigma^2(x_t, \vartheta)$, and $\{e_t\}$ is a sequence of independent Normal errors with $E[e_t|x_t] = E[e_t] = 0$ and $\text{var}[e_t|x_t] = \text{var}[e_t] = 1$. Our interest is then to test

$$\begin{aligned} \mathcal{H}_{02} &: P\left\{f(x_t) = f(x_t, \theta_0)\right\} = 1 \text{ versus} \\ \mathcal{H}_{12} &: P\left\{f(x_t) = f(x_t, \theta_1) + C_{2T} \cdot D_2(x_t)\right\} = 1 \end{aligned} \tag{2.33}$$

for some $\theta_0,\ \theta_1 \in \Theta$, where $\Theta$ is a parameter space, $C_{2T}$ is a sequence of real numbers, and $D_2(x)$ is a smooth and completely nonparametric function. Note that $\theta_0$ may be different from the true value, $\vartheta_0$, of $\vartheta$ involved in the diffusion function.

Analogously to the construction of $L_{1T}(h)$, we propose using a normalized version of the form

$$L_{2T}(h) = \frac{\sum_{s=1}^{T} \sum_{t=1,\neq t}^{T} p_{st}\widehat{\epsilon}_s\widehat{\epsilon}_t}{\widehat{\sigma}_{2T}}, \tag{2.34}$$

where $\widehat{\epsilon}_t = Y_t - f(x_t, \widehat{\theta}_0)$ with $\widehat{\theta}_0$ being a $\sqrt{T}$–consistent estimator of $\theta_0$, and $\widehat{\sigma}_{2T}^2 = 2\widehat{\nu}_2^2 \int K^2(u)du$ with $\widehat{\nu}_2 = \frac{1}{T}\sum_{t=1}^T g(x_t, \widehat{\vartheta})$, in which $\widehat{\vartheta}$ is a $\sqrt{T}$–consistent estimator of $\vartheta$. Since the diffusion function is pre–specified parametrically, we need not involve any nonparametric estimator in $\widehat{\sigma}_{2T}^2$.

Similarly to Theorems 2.1–2.4, we may establish the corresponding results for $L_{2T}(h)$. The corresponding $\widehat{h}_{2\text{test}}$ is given as follows:

$$\widehat{h}_{2\text{test}} = \left(\widehat{C}_\pi \ \Pi(K)\right)^{-\frac{1}{2}} \ \left(\widehat{D}_{2\pi} \ T \ C_{2T}^2\right)^{-\frac{3}{2}} \ (1 + o_P(1)), \tag{2.35}$$

where $\widehat{D}_{2\pi} = \frac{\sum_{t=1}^T \widehat{D}_2^2(x_t)\widehat{\pi}(x_t)}{T \ \widehat{\sigma}_{2T}}$, in which $\widehat{D}_2(x_t) = \frac{\sum_{s=1}^T K\left(\frac{x_t - x_s}{\widehat{h}_{\text{cv}}}\right)\left(Y_s - f(x_s, \widehat{\theta}_0)\right)}{C_{2T} \ \sum_{u=1}^T K\left(\frac{x_t - x_u}{\widehat{h}_{\text{cv}}}\right)}$.

As the details are very analogous, we do not wish to repeat them. Instead, we will focus on the implementation of $L_{2T}(h)$ in Section 3.2 below. Since neither $L_{1T}(h)$ nor $L_{2T}(h)$ involve any additional nonparametric estimation, our finite–sample studies in Section 3 show that it is practically easy to implement the proposed tests. In addition, they both have good small and medium–sample properties with respect to the size and power functions.

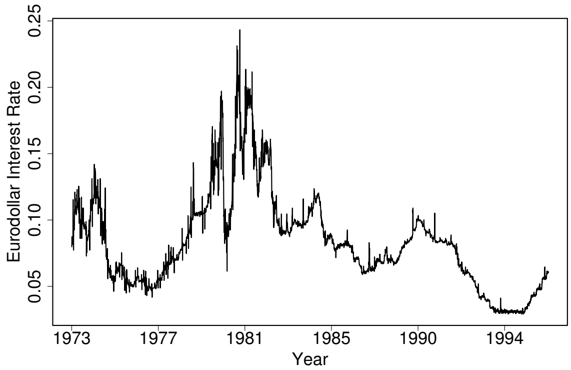
## 3. An example of implementation

Throughout our finite–sample study, we consider testing both the drift and the diffusion functions parametrically for the following model:

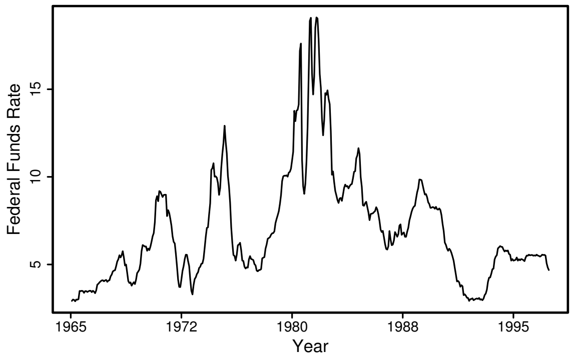
$$dr_t = \beta_0(\alpha_0 - r_t)dt + \sigma_0 r_t^{\rho_0} dB_t, \ t \geq 0, \tag{3.1}$$

where $\alpha_0$, $\beta_0$, $\sigma_0$ and $\rho_0$ are initial parameter values. For the diffusion specification, the initial parameter values estimated from the daily Eurodollar interest rates (June 1, 1973 to February 25, 1995) plotted in Part A of Figure 1, are taken from Hong and Li (2005). For the drift specification, the initial parameter values estimated from the monthly recorded Fed funds (January 1963 to December 1998) plotted in Part B of Figure 1, are taken from Aït–Sahalia (1999). The parameter estimates based on the maximum likelihood method are given in Table 1.

In the first part of our finite–sample study, we approximate the semiparametric continuous–time diffusion model $dr_t = \beta(\alpha - r_t)dt + \sigma(r_t)dB_t$ by a semiparametric




Part A




Part B

Figure 1: Part A: Eurodollar interest rates. Part B: Federal fund rates.

| Parameters | Eurodollar | Fed rate |
|---|---|---|
| $\alpha_0$ | 0.064 | 0.084 |
| $\beta_0$ | 0.62 | 0.087 |
| $\sigma_0$ | 1.48 | 0.779 |
| $\rho_0$ | 1.35 | 1.48 |

Table 1: Initial parameters $\theta_0$ for model (3.1)

time series model of the form

$$Y_t = \beta(\alpha - x_t) + \sqrt{g(x_t)}\; e_t \quad \text{with} \quad g(x_t) = \Delta^{-1}\sigma^2(x_t), \tag{3.2}$$

where $Y_t = \frac{r_{t\Delta} - r_{(t-1)\Delta}}{\Delta}$, $x_t = r_{(t-1)\Delta}$, $\sigma(\cdot) > 0$ is unknown nonparametrically, and $e_t = \frac{B_{t\Delta} - B_{(t-1)\Delta}}{\sqrt{\Delta}} \sim N(0,1)$. Since our finite–sample studies suggest that the resulting size and power values vary little according to the choice of $\Delta$, our finite–sample studies are based on the choice of $\Delta = 1$. We choose $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ throughout this example.

We are interested in testing

$$\mathcal{H}_{01}: \ \sigma(r) = \sigma(r, \vartheta_0) = \sigma_0 r^{\rho_0} \quad \text{versus} \quad \mathcal{H}_{11}: \ \sigma(r) = \sigma(r, \vartheta_0) = \sigma_0 r^{\rho_0} + C_{1T} \tag{3.3}$$

for some $\vartheta_0 = (\sigma_0, \rho_0) \in \Theta$ and $C_{1T} = \sqrt{T^{-1}\text{loglog}(T)}$.

In the second part of our finite–sample study, we approximate the semiparametric continuous–time diffusion model $dr_t = \mu(r_t)dt + \sigma(r_t, \vartheta)dB_t$ by a semiparametric time series model of the form

$$Y_t = \mu(x_t) + \sqrt{g(x_t, \vartheta)}\ e_t \quad \text{with} \quad g(x_t, \vartheta) = \Delta^{-1}\sigma^2(x_t, \vartheta). \tag{3.4}$$

We are also interested in testing

$$\mathcal{H}_{02}: \ \mu(r) = \beta_0(\alpha_0 - r) \quad \text{versus} \quad \mathcal{H}_{12}: \ \mu(r) = \beta_0(\alpha_0 - r) + C_{2T} \tag{3.5}$$

for some $\theta_0 = (\alpha_0, \beta_0) \in \Theta$ and $C_{2T} = \sqrt{T^{-1}\text{loglog}(T)}$.

Because of the choice of $C_{1T}$ and $C_{2T}$, we can easily compute $\widehat{h}_{1\text{test}}$ in (2.30) and $\widehat{h}_{2\text{test}}$ in (2.35). In order to compare the size and power properties of $L_{iT}(h)$ $(i = 1, 2)$ with the most relevant alternatives, we introduce the following simplified notation: for $i = 1, 2$,

$$\begin{aligned}
\alpha_{i0} &= P\left(L_{iT}\left(\widehat{h}_{i\text{test}}\right) > l_{i\alpha}^*\left(\widehat{h}_{i\text{test}}\right) | \mathcal{H}_{0i} \text{ holds}\right), \\
\beta_{i0} &= P\left(L_{iT}\left(\widehat{h}_{i\text{test}}\right) > l_{i\alpha}^*\left(\widehat{h}_{i\text{test}}\right) | \mathcal{H}_{1i} \text{ holds}\right), \\
\alpha_{i1} &= P\left(L_{iT}\left(\widehat{h}_{\text{cv}}\right) > l_{i\alpha}^*\left(\widehat{h}_{\text{cv}}\right) | \mathcal{H}_{0i} \text{ holds}\right), \\
\beta_{i1} &= P\left(L_{iT}\left(\widehat{h}_{\text{cv}}\right) > l_{i\alpha}^*\left(\widehat{h}_{\text{cv}}\right) | \mathcal{H}_{1i} \text{ holds}\right), \\
\alpha_{i2} &= P\left(L_{iT}\left(\widehat{h}_{\text{cv}}\right) > z_\alpha | \mathcal{H}_{0i} \text{ holds}\right), \\
\beta_{i2} &= P\left(L_{iT}\left(\widehat{h}_{\text{cv}}\right) > z_\alpha | \mathcal{H}_{1i} \text{ holds}\right),
\end{aligned} \tag{3.6}$$

where $\widehat{h}_{\text{cv}} = 1.06\ T^{-\frac{1}{5}} \cdot \sqrt{\frac{1}{T-1}\sum_{t=1}^T (x_t - \bar{x})^2}$ with $\bar{x} = \frac{1}{T}\sum_{t=1}^T x_t$.

At the significance level of $\alpha = 1\%$, $5\%$ or $\alpha = 10\%$ with $z_{0.01} = 2.33$ at $\alpha = 1\%$, $z_{0.05} = 1.645$ at $\alpha = 5\%$ and $z_{0.10} = 1.28$ at $\alpha = 10\%$, for each individual case of

$T = 400$, 500 or 600, we apply the Simulation Procedure to obtain the corresponding simulated critical value for each of $l^*_{i\alpha}$ for $i = 1, 2$. We choose $N = 250$ in the Simulation Procedure and use $M = 500$ replications to compute the size and power values for each version. The corresponding results for the size and power are summarized in the following tables. Tables 3.1–3.3 give the results for the diffusion specification while Tables 3.4–3.6 provide the corresponding results for the drift specification.

**Table 3.1. Simulated size and power values at the 1% significance level**

| Sample Size | Null Hypothesis Is True | | | Null Hypothesis Is False | | |
|---|---|---|---|---|---|---|
| $n$ | $\alpha_{10}$ | $\alpha_{11}$ | $\alpha_{12}$ | $\beta_{10}$ | $\beta_{11}$ | $\beta_{12}$ |
| 400 | 0.017 | 0.013 | 0.024 | 0.405 | 0.011 | 0.024 |
| 500 | 0.007 | 0.007 | 0.019 | 0.361 | 0.011 | 0.025 |
| 600 | 0.014 | 0.012 | 0.026 | 0.334 | 0.014 | 0.025 |

**Table 3.2. Simulated size and power values at the 5% significance level**

| Sample Size | Null Hypothesis Is True | | | Null Hypothesis Is False | | |
|---|---|---|---|---|---|---|
| $n$ | $\alpha_{10}$ | $\alpha_{11}$ | $\alpha_{12}$ | $\beta_{10}$ | $\beta_{11}$ | $\beta_{12}$ |
| 400 | 0.039 | 0.053 | 0.056 | 0.515 | 0.060 | 0.064 |
| 500 | 0.034 | 0.044 | 0.046 | 0.491 | 0.057 | 0.059 |
| 600 | 0.047 | 0.042 | 0.047 | 0.492 | 0.047 | 0.052 |

**Table 3.3. Simulated size and power values at the 10% significance level**

| Sample Size | Null Hypothesis Is True | | | Null Hypothesis Is False | | |
|---|---|---|---|---|---|---|
| $n$ | $\alpha_{10}$ | $\alpha_{11}$ | $\alpha_{12}$ | $\beta_{10}$ | $\beta_{11}$ | $\beta_{12}$ |
| 400 | 0.066 | 0.100 | 0.087 | 0.516 | 0.107 | 0.095 |
| 500 | 0.071 | 0.096 | 0.082 | 0.497 | 0.113 | 0.095 |
| 600 | 0.090 | 0.089 | 0.073 | 0.508 | 0.097 | 0.090 |

**Table 3.4. Simulated size and power values at the 1% significance level**

| Sample Size | Null Hypothesis Is True | | | Null Hypothesis Is False | | |
|---|---|---|---|---|---|---|
| $n$ | $\alpha_{20}$ | $\alpha_{21}$ | $\alpha_{22}$ | $\beta_{20}$ | $\beta_{21}$ | $\beta_{22}$ |
| 400 | 0.014 | 0.011 | 0.026 | 0.134 | 0.012 | 0.029 |
| 500 | 0.011 | 0.016 | 0.030 | 0.156 | 0.015 | 0.030 |
| 600 | 0.011 | 0.010 | 0.024 | 0.140 | 0.018 | 0.034 |

**Table 3.5. Simulated size and power values at the 5% significance level**

| Sample Size | Null Hypothesis Is True | | | Null Hypothesis Is False | | |
|---|---|---|---|---|---|---|
| $n$ | $\alpha_{20}$ | $\alpha_{21}$ | $\alpha_{22}$ | $\beta_{20}$ | $\beta_{21}$ | $\beta_{22}$ |
| 400 | 0.053 | 0.055 | 0.071 | 0.230 | 0.053 | 0.066 |
| 500 | 0.063 | 0.056 | 0.068 | 0.246 | 0.054 | 0.066 |
| 600 | 0.044 | 0.048 | 0.062 | 0.231 | 0.057 | 0.075 |

**Table 3.6. Simulated size and power values at the 10% significance level**

| Sample Size | Null Hypothesis Is True | | | Null Hypothesis Is False | | |
|---|---|---|---|---|---|---|
| $n$ | $\alpha_{20}$ | $\alpha_{21}$ | $\alpha_{22}$ | $\beta_{20}$ | $\beta_{21}$ | $\beta_{22}$ |
| 400 | 0.103 | 0.112 | 0.110 | 0.303 | 0.096 | 0.093 |
| 500 | 0.104 | 0.107 | 0.105 | 0.307 | 0.099 | 0.098 |
| 600 | 0.096 | 0.104 | 0.103 | 0.291 | 0.101 | 0.100 |

For the parametric specification of the diffusion function, Tables 3.1–3.3 show that the test $L_{1T}\left(\widehat{h}_{1\text{test}}\right)$ has little size distortion compared with $L_{1T}\left(\widehat{h}_{1\text{cv}}\right)$, as the size values in column 4 of Tables 3.1–3.3 show that the use of an asymptotic critical value may contribute to the size distortion. Moreover, columns 5–7 of Tables 3.1–3.3 show that $L_{1T}\left(\widehat{h}_{1\text{test}}\right)$ has some reasonable power values although the 'distance' between the null and the alternative has been made deliberately close at the rate of $\sqrt{T^{-1}\,\mathrm{loglog}(T)} = 0.0604$ for $T = 500$ or $0.0556$ for $T = 600$. In addition, $L_{1T}(\widehat{h}_{1\mathrm{test}})$ is much more powerful than $L_{1T}(\widehat{h}_{1\mathrm{cv}})$, whose power values are comparable with the corresponding size values. This is mainly because $\lim_{T\to\infty} T\,\sqrt{\widehat{h}_{\mathrm{cv}}}\;C_{1T}^2 = 0$ implies $\lim_{T\to\infty} P_T(h) = \alpha$ when choosing $C_{1T} = \sqrt{T^{-1}\mathrm{loglog}(T)}$ and $\widehat{h}_{\mathrm{cv}}$ proportional to $T^{-\frac{1}{5}}$. For the parametric specification of the drift function, similar conclusions can be made. There are some differences noticed. The main difference is that for each individual case the size is more settled than that for the diffusion case on the one hand, but on the other hand the power is smaller than the corresponding version for each individual case in the diffusion specification. This shows that there is a kind of trade–off between the size and the power of a test.

## 4. Conclusion

In this paper, we establish a new kernel test for the specification of the diffusion function in continuous–time financial models and then propose combining a power–based selection criterion into the implementation of the proposed test in practice. As pointed out briefly in Section 2, the proposed test may also be extended to specify the higher–order moments of the diffusion process. In addition, as can be seen from the discussion in Section 2, we may apply the proposed test for specifying certain continuous–time stochastic volatility models. Such topics are left for future research.


## Acknowledgements

The authors would like to thank Michael McAleer and two referees for their constructive comments on the original version. Acknowledgments also go to Iréne Gijbels for her comments on an earlier version. Thanks also go to the Australian Research Council Discovery Grants Scheme for two Grants under Grant Numbers: DP0558602 and DP0879088.


## Appendix A

This appendix lists the necessary assumptions for the establishment and the proofs of the main results given in Section 2.

A.1. *Assumptions*

Assumption A.1. (i) *Assume that the discrete sequence* $\{r_t : t = 1, 2\cdots\}$ *is strictly stationary and* $\alpha$*-mixing with mixing coefficient* $\alpha(t) \leq C_\alpha \alpha^t$ *defined by*

$$\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Omega_1^s, B \in \Omega_{s+t}^{\infty}\} \tag{A.1}$$

*for all* $s, t \geq 1$*, where* $0 < C_\alpha < \infty$ *and* $0 < \alpha < 1$ *are constants, and* $\{\Omega_i^j\}$ *denotes a sequence of* $\sigma$*–fields generated by* $\{r_t : i \leq t \leq j\}$*. Let* $\Omega_T = \Omega_0^T$ *be the* $\sigma$*–field generated by* $\{r_t : 0 \leq t \leq T\}$*. Let* $\pi_{\tau_1,\tau_2,\cdots,\tau_l}(\cdot)$ *be the joint probability density of* $(x_{1+\tau_1}, \ldots, x_{1+\tau_l})$ *(*$1 \leq l \leq 4$*). Assume that* $\pi_{\tau_1,\tau_2,\cdots,\tau_l}(\cdot)$ *for all* $1 \leq l \leq 4$ *do exist and are continuous.*

(ii) *Assume that the univariate kernel function* $K(\cdot)$ *is a symmetric and bounded probability density function. In addition, we assume the existence of* $K^{(3)}(\cdot)$*, the three–time convolution of* $K(\cdot)$ *with itself. In addition,* $\int K^2(u)du > 0$*.*

(iii) *The bandwidth* $h$ *satisfies both* $\lim_{T\to\infty} h = 0$ *and* $\lim_{T\to\infty} Th = \infty$*.*

Assumption A.2. (i) *The drift and the diffusion functions are three–times differentiable in* $x \in R^+ = (0, \infty)$*. In addition, there exists some constant* $0 < d_0 < \infty$ *such that* $P(\sigma(r_1) > 0) = 1$ *and* $E\left[\sigma^{16+\delta_0}(r_1)\right] \leq d_0$ *for some* $\delta_0 > 0$*. In addition,* $E\left[\sigma^i(x_1, \vartheta_0)\right] > 0$ *for* $i = 2, 4$*.*

(ii) *The integral of* $\overline{\mu}(v, \theta) = \frac{1}{\sigma^2(v,\vartheta_0)} \exp\left(-\int_v^{\bar{v}} 2\frac{\mu(x,\theta)}{\sigma^2(x,\vartheta_0)}dx\right)$ *converges at both boundaries of* $R^+$*, where* $\bar{v}$ *is fixed in* $R^+$*.*

(iii) *The integral of* $s(v, \theta) = \exp\left(\int_v^{\bar{v}} 2\frac{\mu(x,\theta)}{\sigma^2(x,\vartheta_0)}dx\right)$ *diverges at both boundaries of* $R^+$*.*

(iv) *The marginal density* $\pi(\cdot)$ *is strictly positive on* $R^+$*, and the initial random variable* $r_0$ *is distributed as* $\pi(\cdot)$*.*

Assumption A.3. (i) *There exist some absolute constants* $\varepsilon_1 > 0$ *and* $0 < A_{1L} < \infty$ *such that*

$$\lim_{T\to\infty} P\left(\sqrt{T}||\widehat{\theta} - \theta|| > A_{1L}\right) < \varepsilon_1.$$

(ii) *Let* $\mathcal{H}_0$ *be true. Then* $\vartheta_0 \in \Theta$ *and* $\lim_{T\to\infty} P\left(\sqrt{T}||\widehat{\vartheta}_0 - \vartheta_0|| > B_{1L}\right) < \varepsilon_2$ *for any* $\varepsilon_2 > 0$ *and some* $B_{1L} > 0$*.*

*Let* $\mathcal{H}_0$ *be false. Then there is a* $\vartheta_1 \in \Theta$ *such that* $\lim_{T\to\infty} P\left(\sqrt{T}||\widehat{\vartheta}_0 - \vartheta_1|| > B_{2L}\right) < \varepsilon_2$ *for any* $\varepsilon_2 > 0$ *and some* $B_{2L} > 0$*.*

(iii) *There exist some absolute constants* $\varepsilon_3 > 0$*,* $\varepsilon_4 > 0$*, and* $0 < B_{3L}, B_{4L} < \infty$ *such that both*

$$\lim_{T\to\infty} P\left(\sqrt{T}||\widehat{\vartheta}_0^* - \widehat{\vartheta}_0|| > B_{3L}|\Omega_T\right) < \varepsilon_3 \quad \text{and} \quad \lim_{T\to\infty} P\left(\sqrt{T}||\widehat{\theta}^* - \widehat{\theta}|| > B_{4L}|\Omega_T\right) < \varepsilon_4$$

*hold in probability, where* $\widehat{\vartheta}_0^*$ *and* $\widehat{\theta}^*$ *are as defined in the Simulation Procedure above Theorem 2.1.*

(iv) *Let $f(x,\theta)$ and $g(x,\vartheta)$ be twice differentiable with respect to $\theta$ and $\vartheta$, respectively. In addition, the following quantities are assumed to be finite:*

$$C_1(g) = E\left[\left(\left|\left|\frac{\partial g(x_1,\vartheta)}{\partial \vartheta}|_{\vartheta=\vartheta_0}\right|\right|\right)^2\right] \quad \textit{and} \quad \widetilde{C}_1(f) = E\left[\left(\left|\left|\frac{\partial f(x_1,\theta)}{\partial \theta}\right|\right|\right)^4\right],$$

*where $f(x,\theta) = \mu(x,\theta)$, $g(x,\vartheta) = \Delta^{-1}\ \sigma^2(x,\vartheta)$ and $||\cdot||^2$ denotes the Euclidean norm.*

Assumption A.4. *Let $\lim_{T\to\infty} T\sqrt{h}\ C_{1T}^2 = \infty$. Assume that $D_1(x)$ is an unknown and continuous function such that $0 < C_1(D) = E\left[D_1^2(x_1)\right] < \infty$.*

Remark A.1. Assumption A.1(i) is quite natural in this kind of problem. Note that instead of assuming that the continuous–time process $\{r_t : t \geq 0\}$ is strictly stationary as in Li (2007), Assumption A.1(i) assumes only that the discrete sequence $\{r_t : t = 1, 2, \cdots\}$ is strictly stationary. Similar conditions have been used in Aït-Sahalia (1996a) and Hong and Li (2005). This is mainly because we need not require $\Delta \to 0$ as $T \to \infty$ to establish our asymptotic distributions. Assumption A.1(ii) is to ensure the existence of quantities associated with $K(\cdot)$. As pointed out in Section 2, Assumption A.1(iii) imposes the minimal conditions on $h$ such that the asymptotic normality is the limiting distribution of the proposed test.

Assumption A.2 corresponds to Assumptions A1 and A2 of Aït-Sahalia (1996a) to ensure both the existence and uniqueness of a solution of the diffusion process. Assumption A.2(i) requires the existence of the moments of $\sigma(r_1)$. This holds in many cases including the case where the marginal density function $\pi(r)$ of $\{r_t\}$ has compact support. When the marginal density has no compact support, but it satisfies $\lim_{r\to\infty} r^m \pi(r) = 0$ for certain $m > 0$. Obviously, both the Gaussian and $\chi^2$ processes are covered.

Assumption A.3 is for some technical proofs and derivations. Many well–known parametric functions and estimators do satisfy Assumption A.3. In addition, Assumption A.3(i)–(iii) is similar to some existing conditions, such as Assumption 2 of Horowitz and Spokoiny (2001). Assumption A.4 imposes some mild conditions to ensure that both classes of global and local alternatives are included.

A.2. *Technical lemma for the proof of Theorem 2.1*

In order to prove Theorem 2.1, we need to establish the following lemma.

Lemma A.1. *Suppose that Assumptions A.1–A.3(i)(ii) listed in the Appendix hold. Then for $x \in R^1 = (-\infty, \infty)$*

$$\lim_{T\to\infty} P\left(L_T(h) \leq x\right) = \Phi(x). \tag{A.2}$$

PROOF: Recall from (2.14) that under $\mathcal{H}_{01}$

$$\begin{aligned} L_T(h) &= \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \eta_s(\vartheta_0)\ p_{st}\ \eta_t(\vartheta_0)}{\sigma_0} \\ &= \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \sigma^2(x_s,\vartheta_0)(e_s^2-1)\ p_{st}\ (e_t^2-1)\sigma^2(x_t,\vartheta_0)}{\sigma_*}, \end{aligned} \tag{A.3}$$

where $\eta_t(\vartheta_0) = g(x_t,\vartheta_0)(e_t^2-1)$ and $\sigma_* = 2E\left[\sigma^4(x_1,\vartheta_0)\right]\ \sqrt{\int K^2(u)du}$.

Since $L_T(h)$ is a quadratic form of weakly dependent time series independent of $\Delta$, we are able to employ Lemma A.1 of Gao and King (2004) to show that $L_T(h)$ is asymptotically normal. The detail is similar to the proof of Theorem 2.1 of Gao and King (2004).

A.3. *Proof of Theorem 2.1*

In view of the decomposition of $L_{1T}(h)$ in (2.13), using Assumptions A.1–A.3(i)(ii)(iv) and then Lemma A.1, the proof of Theorem 2.1 follows immediately.

A.4. *Technical lemmas for the proof of Theorem 2.2*

Similar to the decomposition of $L_{1T}(h)$ in (2.13), regardless of under $\mathcal{H}_{01}$ or $\mathcal{H}_{11}$, we have

$$\begin{aligned} L_{1T}(h) &= \frac{\sum_{s=1}^T \sum_{t=1,\neq t}^T p_{st}\widehat{\eta}_s\widehat{\eta}_t}{\widehat{\sigma}_{1T}} = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \eta_s\ p_{st}\ \eta_t}{\sigma_0} \cdot \frac{\sigma_0}{\widehat{\sigma}_{1T}} \\ &+ \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(f(x_s,\theta) - f(x_s,\widehat{\theta})\right)^2 p_{st} \left(f(x_t,\theta) - f(x_t,\widehat{\theta})\right)^2}{\sigma_0} \cdot \frac{\sigma_0}{\widehat{\sigma}_{1T}} \\ &+ \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(g(x_s) - g(x_s,\widehat{\vartheta}_0)\right) p_{st} \left(g(x_t) - g(x_t,\widehat{\vartheta}_0)\right)}{\sigma_0} \cdot \frac{\sigma_0}{\widehat{\sigma}_{1T}} + o_P\left(\frac{N_T(h)}{\widehat{\sigma}_{1T}}\right) \\ &\equiv (L_T(h) + F_T(h) + Q_T(h)) \cdot \frac{\sigma_0}{\widehat{\sigma}_{1T}} + R_T(h), \end{aligned} \tag{A.4}$$

where $\sigma_0^2 = 2\mu_0^2 \int K^2(u)du$ is as defined in Section 2,

$$\begin{aligned} L_T(h) &= \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \eta_s\ p_{st}\ \eta_t}{\sigma_0} = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T (e_s^2-1)\ g(x_s)p_{st}g(x_t)\ (e_t^2-1)}{\sigma_0}, \\ F_T(h) &= \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(f(x_s,\theta) - f(x_s,\widehat{\theta})\right)^2 p_{st} \left(f(x_t,\theta) - f(x_t,\widehat{\theta})\right)^2}{\sigma_0}, \\ Q_T(h) &= \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(g(x_s) - g(x_s,\widehat{\vartheta}_0)\right) p_{st} \left(g(x_t) - g(x_t,\widehat{\vartheta}_0)\right)}{\sigma_0}, \end{aligned}$$

and $R_T(h) = L_{1T}(h) - (L_T(h) + F_T(h) + Q_T(h)) \cdot \frac{\sigma_0}{\widehat{\sigma}_{1T}}$ is the remainder term.

In order to prove Theorem 2.2, we need to introduce the following lemmas.

LEMMA A.2. (i) *Suppose that Assumptions A.1–A.3(i)(ii)(iv) hold. Then under* $\mathcal{H}_{01}$

$$R_T(h) = o_P(1) \quad \text{and} \quad F_T(h) = o_P\left(Q_T(h)\right). \tag{A.5}$$

(ii) *Suppose that Assumptions A.1–A.4 hold. Then under* $\mathcal{H}_{11}$

$$R_T(h) = o_P\left(Q_T(h)\right) \quad \text{and} \quad F_T(h) = o_P\left(Q_T(h)\right). \tag{A.6}$$

PROOF: We only prove the first part of (A.5) under $\mathcal{H}_{01}$. The proof of the second part of (A.5) and Its proof of (A.6) under $\mathcal{H}_{11}$ both follow similarly using Assumption A.4(ii).

Observe that under $\mathcal{H}_{01}$ one of the terms involved in $R_T(h)$ is

$$R_{1T}(h) = \frac{2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T}\left(g(x_s,\vartheta_0) - g(x_s,\widehat{\vartheta}_0)\right)p_{st}\eta_t(\vartheta_0)}{\sigma_0}\cdot\frac{\sigma_0}{\widehat{\sigma}_{1T}} = R_{10}(h)\cdot\frac{\sigma_0}{\hat{\sigma}_{1T}}, \tag{A.7}$$

where $R_{10}(h) = \frac{2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T}\left(g(x_s,\vartheta_0)-g(x_s,\widehat{\vartheta}_0)\right)p_{st}\eta_t(\vartheta_0)}{\sigma_0}$.

A Taylor expansion for $g(x_s,\widehat{\vartheta})$ at $\vartheta_0$ implies

$$g(x_s,\widehat{\vartheta}_0) - g(x_s,\vartheta_0) = \frac{\partial g(x_s,\vartheta)}{\partial\vartheta}|_{\vartheta=\vartheta_0}\circ\left(\widehat{\vartheta}_0-\vartheta_0\right) + o_P\left(\widehat{\vartheta}_0-\vartheta_0\right), \tag{A.8}$$

where the symbol "$\circ$" defines the product of two vectors of $a = (a_1,\cdots,a_n)$ and $b = (b_1,\cdots,b_n)$ by $a\circ b = \sum_{i=1}^{n} a_i b_i$.

In view of (A.7) and (A.8), using Assumption A.3, in order to show that (A.5) is true for $R_{10}(h)$, it suffices to show that for any sufficiently small $\psi > 0$

$$E\left[\sum_{s=1}^{T}\sum_{t=1,\neq s}^{T} b_s(\vartheta_0)a_{st}\eta_t(\vartheta_0)\right]^2 < \infty, \tag{A.9}$$

where $\{a_{st}\}$ is as defined before, and $b_s(\vartheta_0) = \frac{\partial g(x_s,\vartheta)}{\partial\vartheta}|_{\vartheta=\vartheta_0}\circ\mathbf{1} = \sum_{i=1}^{d}\frac{\partial g(x_s,\vartheta)}{\partial\vartheta_i}|_{\vartheta_i=\vartheta_{i0}}$, in which $\mathbf{1} = (1,\cdots,1)$ is a $d$–dimensional vector of unit elements, and $\{\vartheta_i\}$ is the $i$–th component of the vector $\vartheta$.

Equation (A.9) follows from

$$\begin{aligned} E\left[\sum_{t=2}^{T}\sum_{s=1}^{t-1} b_s(\vartheta_0)a_{st}\eta_t(\vartheta_0)\right]^2 &= \sum_{t=2}^{T}\sum_{s=1}^{t-1} E\left[b_s(\vartheta_0)a_{st}\eta_t(\vartheta_0)\right]^2 \\ &= \frac{1}{T^2h\sigma_0^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} E\left[K^2\left(\frac{x_s-x_t}{h}\right)b_s^2(\vartheta_0)\right] \\ &= (1+o(1))\ C_0(K)\ E\left[\left|\left|\frac{\partial g(x_1,\vartheta)}{\partial\vartheta}|_{\vartheta=\vartheta_0}\right|\right|^2\right] < \infty \end{aligned} \tag{A.10}$$

using Assumption A.3, where $C_0(K) = \sigma_0^{-2} \int K^2(u)du$.

Similarly, we may show that the first part of equation (A.5) holds for the other terms of $R_T(h)$. Therefore, we complete an outline of the proof of Lemma A.2.

In order to establish a useful lemma, we need to introduce the following notation: Let $g(x_s) = \Delta^{-1}\sigma^2(x_s)$ be as defined before, $b(x_s) = \frac{g(x_s)}{\sqrt{\sigma_0}}$, $z_s = (e_s^2 - 1)$,

$$a_{st} = a(x_s, x_t) = b(x_s)\ \frac{1}{T\sqrt{h}} K\left(\frac{x_s - x_t}{h}\right)\ b(x_t) \ \text{ and } \ L_T(h) \ = \ \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} z_s\ a(x_s, x_t)\ z_t.$$

We now have the following lemma.

Lemma A.3. *Suppose that the conditions of Theorem 2.2(i) hold. Then for any $h$*

$$\sup_{x \in R^1} \left| P\left(L_T(h) \le x\right) - \Phi(x) + \rho(h)\ (x^2 - 1)\phi(x) \right| = O\left(h\right), \tag{A.11}$$

*where $\phi(x)$ denotes the probability density function of $N(0,1)$.*

Proof: In view of the form of $L_T(h)$, we may follow the proof of Lemma A.1 of Gao and Gijbels (2005). Using the fact that $\{x_s\}$ and $\{e_t\}$ are independent for all $s \le t$. we may deal with the conditional probability $P\left(L_T(h) \le x | \mathcal{X}_T\right)$ and then use the dominated convergence theorem to deduce (A.11) unconditionally.

Recall $L_T(h) = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \eta_s\ p_{st}\ \eta_t}{\sigma_g}$ and let $L_T^*(h) = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \eta_s^*\ p_{st}\ \eta_t^*}{\sigma_0}$, where $\eta_s^* = g(x_s)\left(e_s^{*2} - 1\right)$.

Similarly, we define $L_T^*(h)$, $F_T^*(h)$, $Q_T^*(h)$ and $R_T^*(h)$ as the corresponding versions of $L_T(h)$, $F_T(h)$, $Q_T(h)$ and $R_T(h)$ involved in (A.4) with $(x_t, Y_t)$ and $(\widehat{\theta}, \widehat{\vartheta}_0)$ being replaced by $(x_t, Y_t^*)$ and $(\widehat{\theta}^*, \widehat{\vartheta}_0^*)$ respectively.

Lemma A.4. *Suppose that the conditions of Theorem 2.2(i) hold. Then the following*

$$\sup_{x \in R^1} \left| P^*\left(L_T^*(h) \le x\right) - \Phi(x) + \rho(h)\ (x^2 - 1)\phi(x) \right| = O_P\left(h\right) \tag{A.12}$$

*holds in probability.*

Proof: Since the proof follows similarly from that of Lemma A.3 using some conditioning arguments given $\mathcal{W}_T$, we do not wish to repeat the details.

Lemma A.5. (i) *Suppose that the conditions of Theorem 2.2(ii) hold. Then under $\mathcal{H}_{01}$*

$$E\left[Q_T(h)\right] = O\left(\sqrt{h}\right) \quad \text{and} \quad E\left[F_T(h)\right] = o\left(\sqrt{h}\right). \tag{A.13}$$

(ii) *Suppose that the conditions of Theorem 2.2(ii) hold. Then under* $\mathcal{H}_{01}$

$$E^*\left[Q_T^*(h)\right] = O_P\left(\sqrt{h}\right) \quad \textit{and} \quad E^*\left[F_T^*(h)\right] = o_P\left(\sqrt{h}\right) \tag{A.14}$$

*in probability with respect to the joint distribution of* $\mathcal{W}_T$*, where* $E^*[\cdot] = E[\cdot|\mathcal{W}_T]$.

*(iii)* Suppose that the conditions of Theorem 2.2(i) hold. Then under $\mathcal{H}_{01}$

$$E\left[Q_T(h)\right] - E^*\left[Q_T^*(h)\right] = O_P\left(\sqrt{h}\right) \quad \textit{and} \quad E\left[F_T(h)\right] - E^*\left[F_T^*(h)\right] = o_P\left(\sqrt{h}\right) \tag{A.15}$$

*in probability with respect to the joint distribution of* $\mathcal{W}_T$.

PROOF: As the proofs of (i)–(iii) are quite similar, we need only to prove the first part of (iii). In view of (A.4), we have

$$\begin{aligned} Q_T(h) &= \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(g(x_s) - g(x_s, \widehat{\vartheta}_0)\right) p_{st} \left(g(x_t) - g(x_t, \widehat{\vartheta}_0)\right)}{\sigma_0}, \\ Q_T^*(h) &= \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(g(x_s) - g(x_s, \widehat{\vartheta}_0^*)\right) p_{st} \left(g(x_t) - g(x_t, \widehat{\vartheta}_0^*)\right)}{\sigma_0}. \end{aligned} \tag{A.16}$$

Ignoring the higher–order terms, it can be shown that the leading term of $Q_T^*(h) - Q_T(h)$ is represented approximately by

$$Q_T^*(h) - Q_T(h) \approx \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(g(x_s, \widehat{\vartheta}_0) - g(x_s, \widehat{\vartheta}_0^*)\right) p_{st} \left(g(x_t, \widehat{\vartheta}_0) - g(x_t, \widehat{\vartheta}_0^*)\right)}{\sigma_g}. \tag{A.17}$$

Using (A.17), Assumption A.3(iii)(iv) and the fact that

$$E[p_{st}] = \frac{1}{T\sqrt{h}} E\left[K\left(\frac{x_s - x_t}{h}\right)\right] = \frac{\sqrt{h}}{T} \int K(u) du = \frac{\sqrt{h}}{T}, \tag{A.18}$$

we can deduce that

$$E[Q_T(h)] - E^*[Q_T^*(h)] = O_P\left(\sqrt{h}\right), \tag{A.19}$$

which completes an outline of the proof.

LEMMA A.6. *Suppose that the conditions of Theorem 2.2(iii) hold. Then under* $\mathcal{H}_{11}$

$$\lim_{T\to\infty} E\left[Q_T(h)\right] = \infty \quad \text{and} \quad \lim_{T\to\infty} \frac{E\left[F_T(h)\right]}{E\left[Q_T(h)\right]} = 0. \tag{A.20}$$

Proof: In view of the definitions of and $Q_T(h)$ and $F_T(h)$, we need only to show the first part of (A.20). Observe that for $\vartheta_1$ defined in the second part of Assumption A.3(ii),

$$\begin{aligned} Q_T(h) &= \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(g(x_s) - g(x_s, \widehat{\vartheta}_0)\right) p_{st} \left(g(x_t) - g(x_t, \widehat{\vartheta}_0)\right)}{\sigma_0} \\ &= \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(g(x_s) - g(x_s, \vartheta_1)\right) p_{st} \left(g(x_t) - g(x_t, \vartheta_1)\right)}{\sigma_0} \\ &+ \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(g(x_s, \vartheta_1) - g(x_s, \widehat{\vartheta}_0)\right) p_{st} \left(g(x_t, \vartheta_1) - g(x_t, \widehat{\vartheta}_0)\right)}{\sigma_0} \\ &+ o_P\left(Q_T(h)\right). \end{aligned} \tag{A.21}$$

In view of (A.21), using the second part of Assumption A.3(ii), in order to prove (A.20) it suffices to show that for $T \to \infty$ and $h \to 0$,

$$E\left[\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(g(x_s) - g(x_s, \vartheta_1)\right) p_{st} \left(g(x_t) - g(x_t, \vartheta_1)\right)\right] \to \infty. \tag{A.22}$$

Simple calculations imply that as $T \to \infty$ and $h \to 0$

$$E\left[\sum_{t=1}^T \sum_{s=1,\neq t}^T \left(g(x_s) - g(x_s, \vartheta_1)\right) p_{st} \left(g(x_t) - g(x_t, \vartheta_1)\right)\right] = C_{1T}^2 E\left[\sum_{t=1}^T \sum_{s=1,\neq t}^T D_1(x_s) p_{st} D_1(x_t)\right]$$

$$\begin{aligned} &= (1 + o(1))\; C_{1T}^2 \sqrt{h} T \int K(u)du \int D_1^2(v)\pi(v)dv \\ &= (1 + o(1))\; T C_{1T}^2 \sqrt{h} \int D_1^2(v)\pi(v)dv \to \infty \end{aligned} \tag{A.23}$$

using Assumption A.4.

A.5. *Proof of Theorem 2.2*

A.5.1. Proof of Theorem 2.2(i): Recall from (A.4) that

$$L_{1T}(h) = \left(L_T(h) + F_T(h) + Q_T(h)\right) \cdot \frac{\sigma_0}{\widehat{\sigma}_{1T}} + R_T(h), \tag{A.24}$$

$$L_{1T}^*(h) = \left(L_T^*(h) + F_T^*(h) + Q_T^*(h)\right) \cdot \frac{\sigma_0}{\widehat{\sigma}_{1T}^*} + R_T^*(h). \tag{A.25}$$

In view of Assumption A.3, Lemmas A.5 and A.6, we may ignore any terms with orders higher than $\sqrt{h}$ and then consider the following approximations:

$$\begin{aligned} L_{1T}(h) &= L_T(h) + E\left[Q_T(h)\right] + o_P(\sqrt{h}) \quad \text{and} \\ L_{1T}^*(h) &= L_T^*(h) + E^*\left[Q_T^*(h)\right] + o_P(\sqrt{h}). \end{aligned} \tag{A.26}$$

Let $q(h) = E[Q_T(h)]$ and $q^*(h) = E^*[Q_T^*(h)]$. We then apply Lemmas A.3 and A.4 to obtain that

$$
\begin{aligned}
P\left(L_{1T}(h) \le x\right) &= P\left(L_T(h) \le x - q(h) + o_P(\sqrt{h}))\right) \\
&= \Phi(x - q(h)) - \rho(h)\ \left((x - q(h))^2 - 1\right)\ \phi(x - q(h)) \\
&+\ o(\sqrt{h}) \quad \text{and} \qquad\qquad (A.27) \\
P^*\left(L_{1T}^*(h) \le x\right) &= P^*\left(L_T^*(h) \le x - q^*(h) + o_P(\sqrt{h}))\right) \\
&= \Phi(x - q^*(h)) - \rho(h)\ \left((x - q^*(h))^2 - 1\right)\ \phi(x - q^*(h)) \\
&+\ o_P(\sqrt{h}) \qquad\qquad (A.28)
\end{aligned}
$$

hold uniformly over $x \in R^1$.

Theorem 2.2(i) follows consequently from (A.15) and (A.27).

A.5.2. Proof of Theorem 2.2(ii): In view of the definition that $P^*\left(L_{1T}^*(h) \ge l_{1\alpha}^*\right) = \alpha$ and the conclusion from Theorem 2.2(i) that

$$
P\left(L_{1T}(h) \ge l_{1\alpha}^*\right) - P^*\left(L_{1T}^*(h) \ge l_{1\alpha}^*\right) = O_P(\sqrt{h}), \qquad (A.29)
$$

the proof of $P\left(L_{1T}(h) \ge l_{1\alpha}^*\right) = \alpha + O(\sqrt{h})$ follows unconditionally from the dominated convergence theorem.

A.5.2. Proof of Theorem 2.2(iii): Since Theorem 2.2(i) implies that $l_{1\alpha}^* - l_{1\alpha}$ converges to 0 in probability, in order to prove Theorem 2.2(iii), it suffices to show that under $\mathcal{H}_{11}$

$$
\lim_{T\to\infty} P\left(L_{1T}(h) \ge l_{1\alpha}\right) = 1, \qquad (A.30)
$$

which follows from

$$
\begin{aligned}
P\left(L_{1T}(h) \ge l_\alpha\right) &= P\left(L_T(h) \ge \alpha - q(h) + o_P(\sqrt{h}))\right) \\
&= 1 - \Phi(l_{1\alpha} - q(h)) + \rho(h)\ \left((l_{1\alpha} - q(h))^2 - 1\right)\ \phi(l_{1\alpha} - q(h)) \\
&+\ o(\sqrt{h}) \to 1 \qquad\qquad (A.31)
\end{aligned}
$$

using (A.27), the fact that $q(h) \to \infty$ as $T \to \infty$ under $\mathcal{H}_{11}$ implied from Lemma A.6.

Alternatively, the proof of Theorem 2.2(iii) may be completed using Theorem 2.1 and Lemma A.6.

A.6. Proof of Theorem 2.3: Observe that

$$\begin{aligned}
S_T(h) &= P\left(L_{1T}(h) \geq l_{1\alpha} | \mathcal{H}_0\right) = P\left(L_T(h) \geq l_{1\alpha} - Q_T(h) + o_P(Q_T(h)) | \mathcal{H}_{01}\right) \\
&= 1 - P\left(L_T(h) \leq l_{1\alpha} - Q_T(h) + o_P(Q_T(h)) | \mathcal{H}_{01}\right), && \text{(A.32)} \\
S_T^*(h) &= P\left(L_{1T}(h) \geq l_{1\alpha}^* | \mathcal{H}_0\right) = P\left(L_T(h) \geq l_{1\alpha}^* - Q_T(h) + o_P(Q_T(h)) | \mathcal{H}_{01}\right) \\
&= 1 - P\left(L_T(h) \leq l_{1\alpha}^* - Q_T(h) + o_P(Q_T(h)) | \mathcal{H}_{01}\right), && \text{(A.33)} \\
P_T(h) &= P\left(L_{1T}(h) \geq l_{1\alpha} | \mathcal{H}_1\right) = P\left(L_T(h) \geq l_{1\alpha} - Q_T(h) + o_P(Q_T(h)) | \mathcal{H}_{11}\right) \\
&= 1 - P\left(L_T(h) \leq l_{1\alpha} - Q_T(h) + o_P(Q_T(h)) | \mathcal{H}_{11}\right), && \text{(A.34)} \\
P_T^*(h) &= P\left(L_{1T}(h) \geq l_{1\alpha}^* | \mathcal{H}_{11}\right) = P\left(L_T(h) \geq l_{1\alpha}^* - Q_T(h) + o_P(Q_T(h)) | \mathcal{H}_{11}\right) \\
&= 1 - P\left(L_T(h) \leq l_{1\alpha}^* - Q_T(h) + o_P(Q_T(h)) | \mathcal{H}_{11}\right). && \text{(A.35)}
\end{aligned}$$

Using Assumptions A.3(iii)(iv) and A.4, in view of (A.16) and (A.21), a Taylor expansion of $g(\cdot, \vartheta)$ at $\vartheta_0$ implies that for sufficiently large $T$

$$\begin{aligned}
Q_T(h) &= C_0(g)\ \sqrt{h}\ (1 + o_P(1)) \quad \text{under} \quad \mathcal{H}_{01} \quad \text{and} && \text{(A.36)} \\
Q_T(h) &= \sqrt{h}\ \left(C_1(g) + D_{1\pi}\ T\ C_{1T}^2\right)\ (1 + o_P(1)) \ \ \text{under} \ \ \mathcal{H}_{11} && \text{(A.37)}
\end{aligned}$$

hold in probability, where $C_1(g)$ and $D_{1\pi}$ are as defined in Theorem 2.3.

The proof of Theorem 2.3 then follows from Lemmas A.5–A.6 and (A.32)–(A.37).

A.7. Proof of Theorem 2.4: Define $F_{T,h}(x)$ and $F_{T,h}^*(x)$ as the exact finite–sample distributions of $L_{1T}(h)$ and $L_{1T}^*(h)$, respectively. Using existing results (Serfling 1980; Hall 1992) and Theorem 2.3(i) imply

$$\begin{aligned}
l_{1\alpha} - z_\alpha &= \frac{\Phi(z_\alpha) - F_{T,h}(l_\alpha)}{\phi(z_\alpha)} + o_P\left(|l_{1\alpha} - z_\alpha|\right) \\
&= \frac{1}{\phi(z_\alpha)}\ \left((z_\alpha^2 - 1)\ \phi(z_\alpha)\ \psi(\alpha)\sqrt{h}\right) + o_P\left(|l_{1\alpha} - z_\alpha|\right) \\
&= \psi(\alpha)\ \sqrt{h} + o_P\left(|l_{1\alpha} - z_\alpha|\right), \\
l_{1\alpha}^* - z_\alpha &= \frac{\Phi(z_\alpha) - F_{T,h}^*(l_{1\alpha}^*)}{\phi(z_\alpha)} + o_P\left(|l_{1\alpha}^* - z_\alpha|\right) \\
&= \frac{1}{\phi(z_\alpha)}\ \left((z_\alpha^2 - 1)\ \phi(z_\alpha)\ \psi(\alpha)\sqrt{h}\right) + o_P\left(|l_{1\alpha}^* - z_\alpha|\right) \\
&= \psi(\alpha)\ \sqrt{h} + o_P\left(|l_{1\alpha}^* - z_\alpha|\right), && \text{(A.38)}
\end{aligned}$$

where $\psi(\alpha)$ is as defined above Theorem 2.4. The proof is now finished.

## 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., 1999. Transition densities for interest rate and other nonlinear diffusions. Journal of Finance LIV, 1361–1395.

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

Chen, S. X., Gao, J., 2007. An adaptive empirical likelihood test for parametric time series models. Journal of Econometrics 141, 950–972.

Chen, S. X.. Gao, J., Tang, C., 2008. A test for model specification of diffusion processes. Annals of Statistics 36, 167–198.

Corradi, V., Swanson, N., 2005. Bootstrap specification tests for diffusion processes. Journal of Econometrics 124, 117–148.

Corradi, V., White, H., 1999. Specification tests for the variance of a diffusion process. Journal of Time Series Analysis 20, 253–270.

Fan, J., Zhang, C. M., 2003. A re-examination of Stanton's diffusion estimators with applications to financial model validation. Journal of the American Statistical Association 461, 118–134.

Fan, Y., Li, Q., 1996. Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica 64, 865–890.

Fan, Y., Li, Q., 2000. Consistent model specification tests: kernel-based tests versus Bierens' ICM tests. Econometric Theory 16, 1016–1041.

Fan, Y., Linton, O., 2003. Some higher–theory for a consistent nonparametric model specification test. Journal of Statistical Planning and Inference 109, 125–154.

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

Gao, J., King, M. L., 2004. Adaptive testing in continuous–time diffusion models. Econometric Theory 20, 844–882.

Hall, P., 1992. The Bootstrap and Edgeworth Expansions. Springer, New York.

Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous–time models with application to spot interest rates. Review of Financial Studies 18, 37–84.

Horowitz, J., Spokoiny, V., 2001. An adaptive rate–optimal test of a parametric mean–regression model against a nonparametric alternative. Econometrica 69, 599–631.

Kristensen, D., 2004. Estimation in two classes of semiparametric diffusion models. Working paper available from Department of Economics, The University of Wisconsin–Madison.

Li, F., 2007. Testing the parametric specification of the diffusion function in a diffusion process. Econometric Theory 23, 221–250.

Li, Q., 1999. Consistent model specification tests for time series econometric models. Journal of Econometrics 92, 101–147.

Li, Q., Wang, S., 1998. A simple consistent bootstrap tests for a parametric regression functional form. Journal of Econometrics 87, 145–165.

Nicolau, J., 2003. Bias reduction in nonparametric diffusion coefficient estimation. Econometric Theory 19, 754–777.

Serfling, R. J., 1980. Approximation Theorems of Mathematical Statistics. John Wiley & Sons, New York.

Zheng, J. X., 1996. A consistent test of functional form via nonparametric estimation techniques. Journal of Econometrics 75, 263–289.