Econometric Theory, **23**, 2007, 221–250. Printed in the United States of America.
DOI: 10.1017/S0266466607070107

TESTING THE PARAMETRIC SPECIFICATION OF THE DIFFUSION FUNCTION IN A DIFFUSION PROCESS

FUCHUN LI
Bank of Canada

A new consistent test is proposed for the parametric specification of the diffusion function in a diffusion process without any restrictions on the functional form of the drift function. The data are assumed to be sampled discretely in a time interval that can be fixed or lengthened to infinity. The test statistic is shown to follow an asymptotic normal distribution under the null hypothesis that the parametric diffusion function is correctly specified. Monte Carlo simulations are conducted to examine the finite-sample performance of the test, revealing that the test has good size and power.

1. INTRODUCTION

In economics and finance, continuous-time models have been widely used to study the dynamics of underlying state variables, such as asset prices, exchange rates, or spot interest rates. The basic modeling approach in this literature is to assume that the underlying state variables follow a stochastic differential equation.

In the parametric specification of a stochastic differential equation, it is assumed that the functional forms of the drift and diffusion functions are known, apart from a finite number of unknown parameters. Given the parametric specification of a stochastic differential equation, researchers have proposed many different methods to estimate the unknown parameters and to derive the statistical inferences from the discrete observations.

The validity of these estimation and inference procedures, however, is conditional on the hypothesis that the continuous-time model described by a stochastic differential equation is correctly specified. Unfortunately, economic theory typically does not suggest functional forms for the continuous-time model.

The author is grateful to Yacine Aït-Sahalia, John Knight, Oliver Linton (the co-editor), Greg Tkacz, Jun Yang, and three anonymous referees for helpful comments and suggestions. He also thanks seminar participants at the Bank of Canada, the 2004 Semiparametrics Conference in Rio de Janeiro, and the 2005 Econometric Study Group in Bristol. The views expressed in this paper are those of the author. No responsibility for them should be attributed to the Bank of Canada. Address correspondence to Fuchun Li, Department of Monetary and Financial Analysis, Bank of Canada, 234 Wellington, Ottawa, Canada K1A 0G9; e-mail: fuchunli@bank-banque-canada.ca.

© 2007 Cambridge University Press 0266-4666/07 $12.00

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

Model misspecification may lead to misleading conclusions in inference and hypothesis testing. This motivates the development of model specification tests for continuous-time models.

Gallant and Tauchen (1996) proposed a minimum chi-square specification test for continuous-time models using the efficient method of moments. Aït-Sahalia (1996b) proposed a specification test[1] by comparing the model-implied parametric density to the same density estimated nonparametrically. Diebold, Gunther, and Tay (1998), Thompson (2001), and Hong and Li (2005) proposed transition-density-based specification tests based on the fact that the probability integral transform of the model-implied transition density would be distributed as an independent and identical uniform distribution under the correct model specification. Li and Tkacz (2006) proposed a parametric bootstrap procedure to approximate the finite-sample distribution of a goodness-of-fit test statistic of a parametric transition density. Corradi and Swanson (2006) proposed a Kolmogorov-type conditional distribution test. As the limiting distribution of their test statistic is nuisance parameters free, Corradi and Swanson (2006) used a nonparametric bootstrap procedure to construct the critical values.

It is noted that the null hypothesis of all the previously mentioned tests is that both the drift and diffusion functions are specified correctly. Under such a null hypothesis, although these tests can detect a wide range of model misspecifications, they cannot reveal possible sources of model misspecifications. However, for the specification analysis of the continuous-time model, when a misspecified model is rejected, one would like to explore the possible reasons for the rejection. Specifically, is the rejection due to the misspecification from the drift function or the diffusion function? When economic theory provides little guidance about the specification of the drift and diffusion functions, it is advantageous to be able to develop a reliable test that can detect whether the model misspecification comes from the drift function or the diffusion function. It should be mentioned that transition-density-based tests can be used to test the specification of the diffusion function (drift function) only through presupposing both the correct specification of the drift function (diffusion function) and the availability of the closed-form expression of the model-implied transition density. Unfortunately, even if the drift function (diffusion function) is specified correctly, the closed-form expression of a transition density still cannot be available for most continuous-time models. In fact, Wong (1964) showed that a stationary diffusion process has a closed-form expression of its transition density only if it has a linear specification for the drift function and a quadratic specification for the diffusion function. As a consequence, these transition-density-based tests are only applicable to very rare continuous-time models.

The limitations of the previously mentioned tests and the recent developments in nonparametric estimation techniques of a continuous-time model prompt us to make use of nonparametric estimation techniques in developing tests for a parametric form of a continuous-time model by directly testing the specifica-

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

tions of its drift and diffusion functions, without relying on the model-implied density function or model-implied moment condition. Corradi and White (1999) provided a first step in this direction. With knowledge of the functional form of the drift function not required, Corradi and White (1999) proposed a specification test for the diffusion function based on discrete sampling observations.

As pointed out by Corradi and White (1999), however, their test can only be used to test a parametric diffusion function at a given point, and the time span of observations is fixed. Thus, their test cannot be used to detect diffusion function misspecifications over a continuous range of the state variable.

Based on discrete observations, we propose a new test of the functional specification of the diffusion function without any restriction on the functional form for the drift function. The test can be used to test the parametric specification of the diffusion function over a time interval that can be fixed or lengthened to infinity. Using theories of degenerate *U*-statistics, the test statistic is shown to be asymptotically distributed standard normal under the null hypothesis, although diverging to infinity if the parametric specification is misspecified over a significant range. 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. Section 5 concludes. The proofs are provided in the Appendix.

2. THE HYPOTHESIS AND TEST STATISTIC

The model we consider is the following autonomous stochastic differential equation:

$$dx_t = \mu(x_t)\,dt + \sigma(x_t)\,dw_t, \tag{1}$$

with initial condition x_{t_0}, where x_t is the state variable and w_t is a standard Brownian motion process. The functions $\mu(.)$ and $\sigma^2(.)$ are, respectively, the drift function (the instantaneous conditional mean) and the diffusion function (the instantaneous conditional variance) of the process x_t.

We assume that the process x_t is observed at $\{t = t_1, t_2, \ldots, t_n\}$ in the time interval $[t_0, T]$. Although all our results could be derived for nonequispaced data, the only complication being notational, for ease of presentation we will discuss only the equispaced data case. Consequently, we let $\{x_t = x_{t_0+\triangle_n}, x_{t_0+2\triangle_n}, \ldots, x_{t_0+n\triangle_n}\}$ be n equispaced observations on the process x_t at dates $\{t_1 = t_0 + \triangle_n, t_2 = t_0 + 2\triangle_n, \ldots, t_n = t_0 + n\triangle_n\}$, where $\triangle_n = (T - t_0)/n$ is the sampling interval.

We use the notation $x_{n,j}$ to express the observation on the process x_t at $t = t_0 + j\triangle_n$, i.e., $x_{n,j} \equiv x_{t_0+j\triangle_n}$, where $j = 1,2,\ldots,n$.

A parametric family of the specification of the diffusion function $\sigma^2(.)$ is $\{\sigma_0^2(x,\theta) : \theta \in \Theta\}$, with Θ being a subset of R^{d_1}. We want to justify the use of

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

a parametric specification of the diffusion function without knowledge of the functional form of the drift function. Thus, the null hypothesis to be tested is that the parametric specification of $\sigma^2(.)$ is correct:

$$\mathbf{H_0}: \sigma^2(x) = \sigma_0^2(x, \theta_0) \quad \text{almost everywhere} \quad \text{for some } \theta_0 \in \Theta. \tag{2}$$

The alternative hypothesis is $\sigma^2(x) \neq \sigma_0^2(x, \theta)$ for all $\theta \in \Theta$ over a significant range, i.e.,

$$\mathbf{H_1}: \sigma^2(x) \neq \sigma_0^2(x, \theta) \quad \text{on a subset } S \text{ with positive measure} \quad \text{for any } \theta \in \Theta. \tag{3}$$

Our testing approach is based on the squared error goodness-of-fit function between $\sigma^2(.)$ and $\sigma_0^2(x, \theta)$,

$$\begin{aligned} I(\theta, \sigma_0^2) &\equiv E\{[(\sigma^2(x_t) - \sigma_0^2(x_t, \theta))\pi(x_t)]^2 a(x_t)\} \\ &= \int [(\sigma^2(x) - \sigma_0^2(x, \theta))\pi(x)]^2 a(x)\, dF(x), \end{aligned} \tag{4}$$

where $\pi(x)$ and $F(x)$ are, respectively, the unknown density function and cumulative distribution function of x_t. Distance measures similar to (4) were used as a basis for testing the model specifications of either a parametric density function or a general regression function by, e.g., Bickel and Rosenblatt (1973), Hall (1984), Fan (1994), Aït-Sahalia, Bickel, and Stoker (2001), and Li and Tkacz (2006).

The introduction of the density function $\pi(x)$ in (4) is to avoid the problem of trimming the small values of the random denominator in nonparametric estimation of the diffusion function. The inclusion of the weighting function $a(x)$ in (4) allows us to focus goodness-of-fit testing on particular ranges of the state variable. Specifically, we will assume $a(x)$ to be bounded with compact support $S \subset R$ (Assumption 8 in Section 3). This assumption will help us to avoid technical problems in proving uniform convergence of the nonparametric estimations of the marginal density and diffusion function on S. By choosing an appropriate $a(x)$, the specification test can be tailored to the empirical question of interest. In practice, $a(x)$ can typically be chosen as the indicator function of a compact set related to the empirical question of interest. For example, to infer the behavior of the short-term interest rate within a range of levels, say, [0.05, 0.10], only the paths of the state variable that cross the interval [0.05, 0.10] are used in the specification analysis.

Under the null hypothesis H_0, we have $I(\theta_0, \sigma_0^2) = 0$, and under the alternative H_1, we have $I(\theta, \sigma_0^2) > 0$ for any $\theta \in \Theta$. Hence, the measure $I(\theta, \sigma_0^2) > 0$ can be used as an indicator for the misspecification of the diffusion function $\sigma^2(.)$.

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

If $\sigma^2(x), \theta_0$, and $\pi(x)$ were available, then we could estimate $I(\theta_0, \sigma_0^2)$ by its sample analogue $(1/n)\sum_{i=1}^{n}[(\sigma^2(x_{n,j}) - \sigma_0^2(x_{n,j}, \theta_0))\pi(x_{n,j})]^2 a(x_{n,j})$. To get a feasible test statistic, we need to estimate $\sigma^2(x), \theta_0$, and $\pi(x)$.

Under both H_0 and H_1, the true, unknown $\sigma^2(x)$ can be estimated by the kernel method,[2] which is proposed by Jiang and Knight (1997) and Bandi and Phillips (2003),

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^{n-1} K\left(\frac{x_{n,i} - x}{h_n}\right)[x_{n,i+1} - x_{n,i}]^2}{\triangle_n \sum_{i=1}^{n-1} K\left(\frac{x_{n,i} - x}{h_n}\right)}, \tag{5}$$

where $K(.)$ is a kernel function, and h_n is a sequence of bandwidth parameters.

As can be seen from (5), the nonparametric estimator $\hat{\sigma}^2(x)$ is built without imposing any restrictions on the functional form of the drift function. The derivation of the asymptotic distribution of the nonparametric estimator $\hat{\sigma}^2(x)$ depends crucially on the assumption $\triangle_n \to 0$ as $n \to 0$ (Jiang and Knight, 1997; Bandi and Phillips, 2003). In fact, Nicolau (2003) showed that without the assumption $\triangle_n \to 0$ as $n \to 0$, the nonparametric estimator (5) is not consistent. In contrast, in a semiparametric model with the drift function specified parametrically, the semiparametric diffusion function estimator proposed by Aït-Sahalia (1996a) and Kristensen (2004) required that the sampling interval $\triangle_n$ be fixed to obtain asymptotic results. Because the objective of this paper is to construct a test for the specification of the diffusion function without any functional form specification of the drift function, the nonparametric estimation procedure (5) based on $\triangle_n \to 0$ as $n \to 0$ is used to construct our test statistic. The estimator of $\theta_0, \hat{\theta}_n$, is defined as follows:

$$\hat{\theta}_n = \text{argmin}_{\theta \in \Theta} \sum_{i=1}^{n-1} [\log \sigma_0^2(x_{n,i}, \theta) + (\sigma_0^2(x_{n,i}, \theta)\triangle_n)^{-1}(x_{n,i+1} - x_{n,i})^2]. \tag{6}$$

Corradi and White (1999) provided regularity conditions under which $\hat{\theta}_n$ is a quasi-maximum likelihood estimator. These regularity conditions will be given in the assumptions listed in Section 3.

The parametric function $\sigma_0^2(x_{n,i}, \theta_0)$ is estimated by $\sigma_0^2(x_{n,i}, \hat{\theta}_n)$. The unknown density function of $x_t, \pi(x)$, can be consistently estimated by the kernel estimator,[3]

$$\hat{\pi}(x) = \frac{1}{nh_n} \sum_{i=1}^{n-1} K\left(\frac{x_{n,i} - x}{h_n}\right). \tag{7}$$

Let $\hat{F}(x)$ be the empirical cumulative distribution estimator of $F(x)$. Inserting the preceding estimates into the definition of $I(\theta_0, \sigma_0^2)$, given by (4), yields the following estimator of $I(\theta_0, \sigma_0^2)$:

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

$$I_n = \int [(\hat{\sigma}_n^2(x) - \sigma_0^2(x, \hat{\theta}_n))\hat{\pi}(x)]^2 a(x)\, d\hat{F}(x)$$

$$= \frac{1}{n}\sum_{i=1}^{n} [(\hat{\sigma}_n^2(x_{n,i}) - \sigma_0^2(x_{n,i}, \hat{\theta}_n))\hat{\pi}(x_{n,i})]^2 a(x_{n,i}). \quad \textbf{(8)}$$

Our test statistic is a properly centered and scaled version of I_n,

$$J_n \equiv nh_n^{1/2}\left[I_n - \frac{2}{n^2 h_n}\sum_{i=1}^{n} (\hat{\sigma}_n^2(x_{n,i}))^2 \hat{\pi}(x_{n,i}) a(x_{n,i}) \int K^2(u)\, du\right] \Big/ \hat{v}_n, \quad \textbf{(9)}$$

where

$$\hat{v}_n^2 = \frac{8}{n}\sum_{i=1}^{n} (\hat{\sigma}_n^2(x_{n,i}))^4 \hat{\pi}^3(x_{n,i}) a^2(x_{n,i}) \int \left[\int K(u)K(w+u)\, du\right]^2 dw. \quad \textbf{(10)}$$

3. ASSUMPTIONS AND THE LIMITING DISTRIBUTION OF THE TEST STATISTIC

We specify assumptions for the functions $\mu(.), \sigma(.)$, and the parametric family of $\{\sigma_0^2(x,\theta) : \theta \in \Theta\}$, under which the asymptotic validity of this test statistic, J_n, can be established.

Assumption 1. Let $D = (l, r)$ be an open interval with $-\infty \le l < r \le \infty$. The functions $\mu(.)$ and $\sigma(.)$ are at least twice continuously differentiable on D.

Assumption 2. $\sigma^2(x) > 0$ for any $x \in D$.

Assumption 3. The global growth condition is satisfied, i.e., there exists a positive constant, C_D, such that for every $x, y \in D$,

$$\mu^2(x) + \sigma^2(x) \le C_D(1 + x^2). \quad \textbf{(11)}$$

Assumption 4. We define $S(\alpha)$ as

$$S(\alpha) = \int_c^{\alpha} \exp\left\{\int_c^{y}\left[-\frac{2\mu(x)}{\sigma^2(x)}\right]dx\right\}dy, \quad \textbf{(12)}$$

where c is a generic fixed number belonging to D. We require $S(\alpha)$ to satisfy $\lim_{\alpha\to l} S(\alpha) = -\infty$, and $\lim_{\alpha\to r} S(\alpha) = \infty$.

Assumption 5. $\lim_{|x|\to\infty} |\sigma(x)/[2\mu(x) - \sigma(x)\sigma'(x)]| < \infty$, $\lim_{|x|\to\infty} \sigma(x)\pi(x) = 0$.

Assumption 6. The parametric space, Θ, is compact. For any $\theta \in \Theta$ the given function $\sigma_0(x,\theta)$ satisfies Assumptions 1–5. The functions $\partial\sigma_0^2(x,\theta)/\partial\theta$, $\partial^2\sigma_0^2(x,\theta)/\partial\theta\partial\theta'$, $\partial\sigma_0^2(x,\theta)/\partial x$, and $\partial^2\sigma_0^2(x,\theta)/\partial x\partial\theta$ exist and are continuous on $D \times \Theta$.

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

Assumption 7. For almost all $(x,\theta) \in D \times \Theta, \sigma_0^2(x,\theta_1) \neq \sigma_0^2(x,\theta_1)$ if $\theta_1 \neq \theta_2$. For at least finitely many x, there exists a constant, C_1, such that $0 < C_1 \leq \sigma_0(x,\theta_0) \leq C_1^{-1}$. The function $Pr_{x_{t_0}}(\sigma_0(x,\theta) > 0) = 1$ for $(t,\theta) \in [t_0,T] \times \Theta$, where $Pr_{x_{t_0}}$ denotes the probability measure generated by the initial value x_{t_0}.

Assumption 8. $a(x)$ is a given Borel measurable function and bounded with compact support, $S \subset D$. The function $\pi(x)$ and its derivative are continuous and bounded on D, and $\pi(x)$ is bounded away from zero on the compact support, S, of $a(x)$. There exists $\alpha > 0$, such that $\int \exp(\alpha x^2)\pi(x)\,dx < \infty$.

Assumption 9. $K(.)$ is a bounded and symmetric function about 0, with $\int K(u)\,du = 1$ and $\int |u| K(u)\,du < \infty$, and $\int uK(u)\,du = 0$.

Assumptions 1 and 2 ensure the existence and uniqueness of a strong solution to the stochastic differential equation (1). Assumption 3 is used to establish some important moment inequalities (e.g., Friedman, 1975, Thm. 2.2, p. 127) that are needed to derive asymptotic results. Assumption 4 is used to guarantee that the exit time from D is infinite (Ikeda and Watanabe, 1981, Thm. 3.1, p. 362). Assumption 5 is from Hansen and Scheinkman (1995, p. 801). Aït-Sahalia (1996a, p. 552) proved that under Assumption 5 the various classical mixing properties of the discrete observations from the stochastic differential equation (1) are satisfied. In particular, the observation process is absolutely regular with geometric decay rate. Without this assumption, the central limit theorem (CLT) for second-order degenerate U-statistics of absolutely regular processes can fail. Assumptions 6 and 7 were used by Corradi and White (1999, Thm. 3.2, p. 261) to ensure that $\hat{\theta}_n - \theta^* = O_p(1/\sqrt{n})$, where $\theta^* \in \Theta$ and $\theta_* = \theta_0$ under the null hypothesis. Assumption 8 requires $a(x)$ to be bounded with compact support. As we mentioned earlier, without this assumption, we can not prove uniform convergence of the nonparametric estimations of the marginal density and diffusion function on S. In practice, S can typically be taken as the indicator function of a compact set related to the empirical question of interest. Assumption 9 is a standard regularity condition imposed on a kernel function.

The asymptotic null distribution and consistency of J_n are provided in the following theorem.

THEOREM 1. *Suppose Assumptions 1–9 hold and* $h_n = O(n^{-1/\gamma})$, *where* $1.5 < \gamma < 4.5$. *If* T *is either fixed or* $T \to \infty, Th_n^{1/2} \to 0$, *and* $\triangle_n \to 0$ *as* $n \to 0$, *then*

(i) under $H_0, J_n \to N(0,1)$ *in distribution as* $n \to \infty$, *and* $\hat{v}_n^2$ *is a consistent estimator of* v^2, *where* $v^2 = 8\int \sigma^8(x)\pi^4(x)a^2(x)\,dx \int [\int K(u)K(w+u)\,du]^2\,dw$;

(ii) under $H_1, Pr(J_n \geq B_n) \to 1$, *for any nonstochastic sequence* $B_n = o(nh_n^{1/2})$.

Proof. See the Appendix.

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

4. MONTE CARLO SIMULATION

In this section, we examine the finite-sample performance of the test using Monte Carlo simulations. As stated in Section 1, assuming that the drift function is specified correctly and that the closed-form expression of the model-implied transition density is available, transition-density-based tests can also be used for the specification of the diffusion function in a diffusion process. Hong and Li (2005) conducted a simulation study to examine the size and power of their tests. For comparison, we adopt simulation designs that are similar to those of Hong and Li (2005). To examine the test's size performance, we simulate data from the Vasicek (1977), Cox, Ingersoll, and Ross (1985) (CIR hereafter), and drift-misspecified CIR (DMCIR hereafter) models, respectively.

The Vasicek model is

$$dx_t = \beta(\alpha - x_t)\,dt + \sigma dw_t. \tag{13}$$

The CIR model is

$$dx_t = \beta(\alpha - x_t)\,dt + \sigma\sqrt{x_t}\,dw_t. \tag{14}$$

The DMCIR model is

$$dx_t = (\alpha_{-1}x_t^{-1} + \alpha_0 + \alpha_1 x_t + \alpha_2 x_t^2)\,dt + \sigma\sqrt{x_t}\,dw_t. \tag{15}$$

For Vasicek's model, the null hypothesis is that the diffusion function is a constant; i.e., $H_0: \sigma^2(x) = \text{constant}$. Under the null hypothesis, the estimator of $\theta = \sigma^2$ is given by $\hat{\sigma}_n^2 = \sum_{t=1}^{n-1}(x_{n,t+1} - x_{n,t})^2/T$. Under the assumption that the drift function is correctly specified as $\beta(\alpha - x)$, the Hong and Li (2005) test can be used to test the null hypothesis by testing whether the data are generated from a normal transition density (Hong and Li, 2005, p. 48). In the Vasicek model, the parameter β determines the persistence of the process. The smaller β is, the higher the level of persistence in the process, and, consequently, the slower the convergence to the long-run mean, α.

As with Hong and Li (2005) and Pritsker (1998), to examine the impact of the level of persistence on the size performance of our test, we consider both low and high levels of persistent dependence and adopt their parameter values. The parameter values for low and high levels of persistent dependence are, respectively, $(\beta, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185)$ and $(\beta, \alpha, \sigma^2) = (0.214592, 0.089102, 0.000546)$.

To examine the test's size performance when the drift function is misspecified, we consider two cases. For case 1, the data are assumed to be from the CIR model, but they are generated from the DMCIR model. For case 2, the data are assumed to be from the DMCIR model, but they are generated from the CIR model. For both cases, we test the null hypothesis that the diffusion function is $\sigma^2 x$; i.e., $H_0: \sigma^2(x) = \sigma^2 x$. Obviously, in both the CIR model and the DMCIR model, with the drift functions being misspecified, the diffusion

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

functions are correctly specified. The parameter values of the CIR model are taken as $(\beta, \alpha, \sigma^2) = (0.89218, 0.090495, 0.032742)$, which are from Hong and Li (2005), whereas the parameter values of the DMCIR model are taken as $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \sigma^2) = (0.00107, -0.0517, 0.877, -4.604, 0.032742)$. Under the null hypothesis, the estimator of the parameter, $\theta = \sigma^2$, is given by $\hat{\sigma}_n^2 = \sum_{t=1}^{n-1} (x_{n,t+1} - x_{n,t})^2 / T x_{n,t}$.

Because the Vasicek and CIR models have closed-form transition density and marginal density functions (Pritsker, 1998, p. 456; Hong and Li, 2005, p. 49), the simulated sample path can be constructed by their transition densities. The initial values are drawn from their marginal densities. The discrete observations of sample size n are generated over a time period $[0, T]$ with a sampling interval of $\triangle_n = T/n$. For the DMCIR model (15), because its transition density has no closed form, data are simulated using Milstein's scheme (see (19) in this section). Throughout the experiment, we generate 500 realizations of a random sample $\{x_{n,j}\}_{j=1}^{n}$ for sample sizes $n = 250$, 500, 1,000, 2,500, respectively. We discard the first 500 observations to eliminate any start-up effects. T is set to 1 and 5 to consider the impact of the sample interval on the test performance.

To study the test's power performance, we consider two cases. For case 1, the null hypothesis stipulates that the data are generated by a model with a constant diffusion function; i.e., $H_0: \sigma^2(x) =$ constant. However, we simulate data from three different models: the CIR model, the Chan, Karolyi, Longstaff, and Sanders (1992) (CKLS hereafter) model, and the Aït-Sahalia (1996b) nonlinear drift model. The same parameter values in Hong and Li (2005) are again adopted. If we impose the assumption that the drift function is correctly specified as $\beta(\alpha - x)$, the Hong and Li (2005) test can be used to test the null hypothesis by testing whether the data are from a normal transition density.

The CKLS model is

$$dx_t = \beta(\alpha - x_t)\, dt + \sigma x_t^{\rho}\, dw_t, \tag{16}$$

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

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

$$dx_t = (\alpha_{-1} x_t^{-1} + \alpha_0 + \alpha_1 x_t + \alpha_2 x_t^2)\, dt + \sigma \sqrt{x_t}\, dw_t \tag{17}$$

with parameter values $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \sigma^2, \rho) = (0.00107, -0517, 0.877, -4.604, 0.64754, 1.50)$.

For case 2, the null hypothesis stipulates that the data are generated by a model with the diffusion function $\sigma^2 x$. However, the data are simulated from three different models: the CKLS model (16), the nonlinear drift model (17), and a modified CKLS (MCKLS) model.

The MCKLS model is

$$dx_t = \beta(\alpha - x_t)\, dt + [\sigma/(3^{3/2}\alpha)](x_t + 2)^{1.5}\, dw_t, \tag{18}$$

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

with the parameter values $(\beta, \alpha, \sigma^2) = (0.89218, 0.090495, 0.032742)$ used in the CIR model. Note that the process (18) has a nonlinear diffusion function and the same drift function as in the CIR model. Particularly, the linear diffusion function in the CIR model is tangential to the diffusion function in the MCKLS model at point 0.090495. This design helps in evaluating the test's power for testing curvature. For the CKLS model (16), the Aït-Sahalia (1996b) nonlinear drift model (17), and the MCKLS model (18), because their transition densities have no closed forms, we simulate data using Milstein's scheme,

$$x_{t+\triangle_n} = x_t + \mu(x_t)\triangle_n + \sigma(x_t)\sqrt{\triangle_n}\varepsilon_t + \frac{1}{2}\sigma^2(x_t)\triangle_n(\varepsilon_t^2 - 1), \tag{19}$$

where ε_t is a standard normal distribution. The initial value is set to equal the average interest rate level of the data set in Aït-Sahalia (1996b).

Throughout this experiment, we use the standard normal kernel. The bandwidth parameter h_n is chosen according to $h_n = c\sigma_x n^{-1/\gamma}$, where σ_x is the standard deviation of observations. We choose $\gamma = 2.1, 3.5$. The choice of h_n satisfies the conditions of Theorem 1. To check the sensitivity of the test with respect to the choice of bandwidth h_n, we change h_n through different values of c: $c = 0.5, 1, 1.5$. The function $a(x)$ is the indicator function of the interval $S = \{x | x \in [0.002, 2]\}$. The critical value z_α is from the standard normal distribution; i.e., $z_{0.01} = 2.33$, $z_{0.05} = 1.645$, and $z_{0.1} = 1.28$.

Table 1 reports the estimated size of the test. Four general conclusions can be drawn from the table. First, the test has satisfactory size performance at all three levels for sample sizes as small as $n = 250$. In contrast, it is clear that, under the same simulation setting, the Hong and Li tests show strong overrejections under the 1% level (Hong and Li, 2005, Figs. 1 and 2), and the size of their tests is about 2.1% on average, even if the sample size increases to 5,500. Second, the impact of the level of the persistent dependence on the size of the test is minimal, which suggests that the test achieves robustness to the persistent dependence. This result can be explained by the fact that the test statistic is independent of the specification of the drift function, which determines the level of the persistent dependence. Third, the test still exhibits a satisfactory size performance even if the drift function is misspecified. In contrast, the Monte Carlo simulation shows that, under the null hypothesis that the data are generated from the CIR model, the power of the Hong and Li (2005) tests for rejecting the DMCIR model is about 59%, even if the sample size is increased to 2,500 across lag orders from 1 to 20. In other words, the Hong and Li (2005) tests strongly reject the correct null hypothesis of $\sigma^2(x) = \sigma^2 x$. It is obvious that the rejection arises because of the misspecification of the transition density. Fourth, note that the estimated size of the test is quite stable over different choices of bandwidth, particularly for large samples.

Table 2 reports the estimated power of the test when the null hypothesis is that the diffusion function is a constant but in fact the data are generated from

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

TABLE 1. Percentage of rejections of the true H_0

	$c = 0.5$			$c = 1$			$c = 1.5$		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%
Vasicek model with low level of persistent dependence									
250	0.006	0.026	0.040	0.010	0.032	0.044	0.012	0.028	0.054
500	0.018	0.046	0.062	0.010	0.036	0.076	0.020	0.046	0.072
1,000	0.020	0.058	0.086	0.014	0.062	0.090	0.012	0.054	0.082
2,500	0.010	0.054	0.088	0.012	0.056	0.096	0.008	0.052	0.090
Vasicek model with high level of persistent dependence									
250	0.010	0.022	0.044	0.008	0.028	0.046	0.022	0.048	0.072
500	0.012	0.046	0.064	0.020	0.046	0.084	0.020	0.052	0.074
1,000	0.008	0.044	0.068	0.020	0.038	0.090	0.020	0.048	0.076
2,500	0.008	0.046	0.076	0.010	0.048	0.092	0.012	0.050	0.086
Case 1: Data are assumed to be from CIR model but in fact are generated from DMCIR model									
250	0.020	0.030	0.050	0.022	0.046	0.074	0.004	0.024	0.044
500	0.016	0.056	0.108	0.014	0.044	0.072	0.022	0.060	0.090
1,000	0.014	0.052	0.066	0.012	0.054	0.104	0.020	0.058	0.094
2,500	0.012	0.052	0.077	0.010	0.048	0.102	0.014	0.054	0.098
Case 2: Data are assumed to be from DMCIR model but in fact are generated from CIR model									
250	0.014	0.040	0.064	0.014	0.042	0.068	0.020	0.042	0.068
500	0.016	0.050	0.076	0.020	0.050	0.074	0.028	0.056	0.082
1,000	0.016	0.054	0.082	0.020	0.048	0.072	0.020	0.052	0.084
2,500	0.014	0.052	0.090	0.012	0.052	0.078	0.012	0.050	0.096

the CIR model, the CKLS model, and the Aït-Sahalia (1996b) nonlinear drift model, respectively. Table 3 reports the estimated power of the test when the null hypothesis is that the diffusion function is $\sigma^2 x$ but in fact the data are generated from the CKLS model, the Aït-Sahalia (1996b) nonlinear drift model, and the MCKLS model, respectively.

The simulation results of the test's power performance lead to three conclusions. First, Tables 2 and 3 indicate that the test detects the misspecifications of the diffusion functions quite well in both the Vasicek and CIR models against their respective alternatives. For a given alternative, the test's power always increases rapidly with respect to the sample size, in line with the test's consistency property. By comparison, the power of the Hong and Li (2005) tests in detecting the Vasicek model against the CIR model is about 50% when n increases to 2,500, which is noticeably worse than against the CKLS model and the Aït-Sahalia (1996b) nonlinear drift model (Hong and Li, 2005, Fig. 3). However, under the same simulation setting, the power of the test is above 90%.

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

TABLE 2. Percentage of rejections of the false H_0

	$c = 0.5$			$c = 1$			$c = 1.5$		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%
Cox, Ingersoll, and Ross model (CIR)									
250	0.114	0.166	0.202	0.146	0.216	0.280	0.196	0.282	0.344
500	0.230	0.324	0.400	0.308	0.416	0.474	0.362	0.458	0.578
1,000	0.534	0.632	0.696	0.596	0.692	0.748	0.668	0.738	0.792
2,500	0.904	0.922	0.950	0.942	0.968	0.980	0.944	0.960	0.974
Chan et al. model (CKLS)									
250	0.262	0.360	0.426	0.270	0.370	0.452	0.328	0.424	0.496
500	0.752	0.860	0.908	0.822	0.892	0.926	0.900	0.946	0.960
1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Aït-Sahalia nonlinear drift model									
250	0.276	0.362	0.438	0.342	0.432	0.482	0.424	0.518	0.582
500	0.660	0.714	0.752	0.718	0.798	0.832	0.736	0.802	0.836
1,000	0.904	0.938	0.952	0.936	0.956	0.968	0.952	0.974	0.978
2,500	0.994	0.998	0.998	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 3. Percentage of rejections of the false H_0

	$c = 0.5$			$c = 1$			$c = 1.5$		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%
Chan et al. model (CKLS)									
250	0.176	0.260	0.314	0.250	0.360	0.434	0.316	0.374	0.430
500	0.394	0.490	0.0.550	0.446	0.552	0.608	0.528	0.624	0.668
1,000	0.604	0.680	0.718	0.676	0.728	0.782	0.696	0.776	0.818
2,500	0.842	0.892	0.912	0.878	0.904	0.922	0.904	0.938	0.956
Aït-Sahalia nonlinear drift model									
250	0.058	0.110	0.148	0.060	0.102	0.156	0.064	0.124	0.162
500	0.228	0.330	0.402	0.252	0.350	0.420	0.302	0.396	0.458
1,000	0.700	0.824	0.880	0.784	0.858	0.890	0.824	0.875	0.908
2,500	0.974	0.990	0.996	0.992	0.992	0.996	1.000	1.000	1.000
Modified CKLS (MCKLS) model									
250	0.010	0.010	0.012	0.014	0.016	0.018	0.026	0.044	0.104
500	0.052	0.064	0.114	0.062	0.090	0.126	0.092	0.098	0.142
1,000	0.154	0.245	0.330	0.256	0.356	0.420	0.328	0.386	0.472
2,500	0.468	0.630	0.662	0.618	0.706	0.740	0.670	0.734	0.802
5,500	0.818	0.854	0.886	0.834	0.862	0.890	0.850	0.884	0.904

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

Second, the test has good power in detecting the CIR model against the MCKLS model when the sample size increases to 2,500. This test, however, has a lower power in detecting the CIR model against the MCKLS model than the CKLS model and the Aït-Sahalia (1996b) nonlinear drift model. This result can be explained by the fact that the diffusion function in the MCKLS model is closer to the diffusion function in the CIR model than in the two other models. Third, even though the test's power is already quite stable over different choices of h_n for large samples ($n = 1{,}000, n = 2{,}500$), the higher the value of the bandwidth, h_n (i.e., the higher the value of c), the higher the test's power. This result can be explained by the fact that the test statistic diverges to $+\infty$ at the rate of $nh_n^{1/2}$ under the alternative. Hence, a higher h_n (in a certain range) will lead to a more powerful test against some fixed alternatives (in finite samples). This result does not mean that one should use a very large value of h_n in practice, because it would oversmooth the data and hence obliterate any deviation of the data from the null data-generating process. Of course, one should not use a very small value of h_n, because it could result in an inaccurate kernel estimation. Specifically, an h_n that is too small tends to make the test less powerful. Because the test is based on high-frequency data, the large sample sizes available should make the choice of h_n less crucial than the moderate sample size. How to choose the bandwidth optimally, so that the test's power is maximized and the size is kept under control, is left for future research.

Simulation results for $T = 5$ are not presented but are available from the author. They are qualitatively similar to those for $T = 1$.

5. CONCLUSION

In this paper, we propose a consistent test for the parametric specification of the diffusion function in a diffusion process without any restrictions on the functional form of the drift function. The test is based on a comparison of the kernel estimate of the true unknown diffusion function with the parametric specification of the diffusion function. It is shown to have the standard normal distribution under the null hypothesis. The Monte Carlo simulation results suggest that the overall performance of the test is satisfactory.

Extensions to multidimensional diffusion processes (including unobservable state variables) and applications to evaluate the performance of a variety of specifications for the diffusion function in the spot interest rate process (Durham, 2003) will be considered for future work. It would also be useful to develop a test for the parametric specification of the drift function in a diffusion process.

NOTES

1. Pritsker (1998) uses a Monte Carlo simulation method to examine the specification test of Aït-Sahalia (1996b) and finds that it has a poor finite-sample performance. The main reasons are the highly persistent dependence of the interest rate and the slow convergence of the nonparametric estimator for the density function (Pritsker, 1998, p. 451).

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

2. Downing (1999) and Chapman and Pearson (2000) provide Monte Carlo simulation evidence that contrary to the simulation results of the kernel density estimator by Pritsker (1998), the finite-sample performance of the nonparametric estimator of the diffusion function is very good and notably successful.

3. The drift parameters do not enter our test statistic, and as a result, our test does not suffer the quantitatively significant bias from the estimator of the level of the persistent dependence that the Aït-Sahalia test does (Pritsker, 1998). Indeed, as pointed out by Pritsker (1998), a nonparametric estimator of the density function suffers from the small sample bias. However, the large sample size available (e.g., the high-frequency data) makes it possible to estimate the density function precisely (Hong and Li, 2005; Lavergne and Vuong, 2000; Li, 1999). Even though the nonparametric method has some cost associated with it (i.e., very large samples and an accurate measurement of the function; Pagan and Ullah, 1999), it remains a reliable and powerful tool for analyzing financial data.

REFERENCES

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

Aït-Sahalia, Y. (1996b) Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 2, 385–426.

Aït-Sahalia, Y., P.J. Bickel, & T.M. Stoker (2001) Goodness-of-fit tests for regression using kernel methods. *Journal of Econometrics* 105, 363–412.

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

Bickel, P.J. & M. Rosenblatt (1973) On some global measures of the deviations of density function estimates. *Annals of Statistics* 1, 1071–1095.

Chan, K.C., G.A. Karolyi, F.A. Longstaff, & A.B. Sanders (1992) An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* 47, 1209–1227.

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

Corradi, V. & N.R. Swanson (2006) Bootstrap conditional distribution tests under dynamic misspecification. *Journal of Econometrics* 133, 779–806.

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

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

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

Downing, C.T. (1999) Nonparametric Estimation of Multifactor Continuous Time Interest-Rate Models. Working paper 1999-62, Board of Governors of the Federal Reserve System.

Durham, G.B. (2003) Likelihood-based specification analysis of continuous-time models of the short-term interest rate, *Journal of Financial Economics* 70, 436–487.

Fan, Y. (1994) Testing the goodness-of-fit of a parametric density function by kernel method. *Econometric Theory* 10, 316–356.

Fan, Y. & Q. Li (1999) Central limit theorem for degenerate U-statistic of absolute regular processes with application to model specification testing. *Journal of Nonparametric Statistics* 10, 245–271.

Friedman, A. (1975) *Stochastic Differential Equations and Applications*, vol. 1. Academic Press.

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

Györfi, L., W. Härdle, P. Sarda, & P. Vieu (1989) *Nonparametric Curve Estimation from Time Series*. Lecture Notes in Statistics 60. Springer-Verlag.

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14, 1–16.
Hansen, L.P. & J.A. Scheinkman (1995) Back to the future: Generating moment implications for continuous time Markov processes. *Econometrica* 63, 767–804.
Hong, Y. & H. Li (2005) Nonparametric specification testing for continuous-time models with application to spot interest rates. *Review of Financial Studies* 18, 37–84.
Ikeda, N. & S. Watanabe (1981) *Stochastic Differential Equations and Diffusion Processes*. North-Holland.
Jiang, G.J. & J.L. Knight (1997) A nonparametric approach to the estimation of diffusion process, with an application to a short-term interest rate model. *Econometric Theory* 13, 615–645.
Kristensen, D. (2004) Estimation in Two Classes of Semiparametric Diffusion Models. Mimeo, Department of Economics, University of Wisconsin-Madison.
Lavergne, P. & Q. Vuong (2000) Nonparametric significance testing. *Econometric Theory* 16, 576–601.
Li, F. (2005) Testing the Parametric Specification of the Diffusion Function in a Diffusion Process. Working paper, Bank of Canada.
Li, Q. (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101–147.
Li, F. & G. Tkacz (2006) A consistent bootstrap test for conditional density functions with time-series data. *Journal of Econometrics* 133, 841–862.
Nicolau, J. (2003) Bias reduction in nonparametric diffusion coefficient estimation. *Econometric Theory* 19, 754–777.
Pagan, A. & A. Ullah (1999) *Nonparametric Econometrics*. Cambridge University Press.
Pritsker, M. (1998) Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies* 11, 449–487.
Thompson, T. (2001) Specification Tests for Continuous-Time Models. Working paper, Department of Economics, Harvard University.
Vasicek, O. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.
Wong, E. (1964) The construction of a class of stationary Markov processes. In R. Bellman (ed.), *Proceedings of Symposia in Applied Mathematics on Stochastic Processes in Mathematical Physics and Engineering*, vol. 16, pp. 264–276. American Mathematical Society.
Yoshihara, K. (1976) Limiting behavior of U-statistic for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie Verwandte. Gebiete* 35, 237–252.

APPENDIX: Proofs

Let $\{X_{n,l}\} \equiv \{X_{n,l}, : l \leq n; n \geq 1\}$ be a triangle array of random variables and $M_{s,t}^{n}$ denote the sigma algebra generated by $(X_{n,s}, \ldots, X_{n,t})$ for $s \leq t$.

DEFINITION.

(i) *$\{X_{n,l}\}$ is said to be a strictly stationary triangular array of random variables if for positive integers s_1, s_2, and k, $\{X_{n,s_1}, X_{n,s_1+1}, \ldots, X_{n,s_1+k}\}$ and $\{X_{n,s_2}, X_{n,s_2+1}, \ldots, X_{n,s_2+k}\}$ have the same joint distribution, where $s_1 + k \leq n, s_2 + k \leq n$.*

(ii) *Let $\{X_{n,l}\}$ be a strictly stationary triangular array of random variables and $\beta_{n,\tau} = \sup_{s+\tau \leq n} E[\sup_{A \in M_{s+\tau,n}} \{P(A|M_{1,s}^{n}(n)) - P(A)\}]$. Then, $\{X_{n,l}\}$ is said to satisfy*

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

an absolute regularity condition with the mixing coefficient $\beta_{n,\tau}$ *if* $\beta_{n,\tau} \to 0$ *as* $\tau \to \infty$.

LEMMA 1. *Let* $\xi_{n,1}, \ldots, \xi_{n,n}$ *be random vectors taking values in* R^{d_2} *satisfying an absolute regularity condition with the mixing coefficient* $\beta_{n,\tau}$. *Let* $a(y_1, \ldots, y_k)$ *be a Borel measurable function such that, for some* $\delta > 0$,

$$\overline{M}_n = \max\{E[|a(\xi_{n,i_1}, \ldots, \xi_{n,i_k})|^{1+\delta}], E[E_{\xi_{n,i_1}, \ldots, \xi_{n,i_j}}[|a(\xi_{n,i_1}, \ldots, \xi_{n,i_k})|^{1+\delta}]]\}$$

exists. Then,

$$|E[a(\xi_{n,i_1}, \ldots, \xi_{n,i_k})] - E[E_{\xi_{n,i_1}, \ldots, \xi_{n,i_j}}[a(\xi_{n,i_1}, \ldots, \xi_{n,i_k})]]| \le 4\overline{M}_n^{1/(1+\delta)} \beta_{n,m}^{\delta/(1+\delta)},$$

where

$$E_{\xi_{n,i_1}, \ldots, \xi_{n,i_j}}[a(\xi_{n,i_1}, \ldots, \xi_{n,i_k})] = \int a(z_{i_1}, \ldots, z_{i_j}, \xi_{n,i_{j+1}}, \ldots, \xi_{n,i_k})\, dF_n^{i_1, \ldots, i_j}(z_{i_1}, \ldots, z_{i_j})$$

and $m = i_{j+1} - i_j$, $F_n^{i_1, \ldots, i_j}(z_{i_1}, \ldots, z_{i_j})$ *is the distribution function of random vectors* $\xi_{n,i_1}, \ldots, \xi_{n,i_j}$ *and* $i_1 < i_2 < \cdots < i_j$.

Proof. Lemma 1 follows straightforwardly from Lemma 1 in Yoshihara (1976). ■

For notational simplicity, h_n and $K((x_{n,j} - x)/h_n)$ are expressed by h and $K_j(x)$, respectively. We use E^j to express the conditional expectation with respect to the σ-field generated by $\{x_u : u \le t_0 + j\triangle_n\}$. The symbol C denotes a generic big enough positive constant. The notation $A \sim B$ means that A has an order no larger than that of B. Also let $B(.,.,.)$ be a Borel measurable function and $F_n(x, \bar{x})$ be the joint distribution function for $(x_{n,i}, x_{n,j})$, where $i \ne j$. We denote $E_i[B(x_{n,i}, x_{n,j}, x_{n,k})] \equiv \int B(x, x_{n,j}, x_{n,k})\, dF(x)$, $E_k[B(x_{n,i}, x_{n,j}, x_{n,k})] \equiv \int B(x_{n,i}, x_{n,j}, x)\, dF(x)$, and $E_{i,j}[B(x_{n,i}, x_{n,j}, x_{n,k})] \equiv \int B(x, \bar{x}, x_{n,k})\, dF_n(x, \bar{x})$.

LEMMA 2. *Suppose that Assumptions 1–4 and 6 hold,* $E(x^{2d}) < \infty$ *for some positive integer d, and* $t' \in [t_0, T)$. *Then,*

(i) for $t' > t_0 + j\triangle_n$:

$$E^j(x_{t'} - x_{n,j})^{2d} \le D_n(1 + x_{n,j}^{2d})(t' - t_0 - j\triangle_n)^d, \tag{A.1}$$

where $D_n = 2^{2(2d-1)} C_D^{2d}\{(t' - t_0 - j\triangle_n)^d + [d(2d-1)]^d\}$.

(ii) Let $r_n(z_{n,j}, x, \hat{\theta}_n) = (1/nh\triangle_n) K_j(x)[(x_{n,j+1} - x_{n,j})^2 - \sigma_0^2(x, \hat{\theta}_n)\triangle_n]$, *where* $z_{n,j} = (x_{n,j}, x_{n,j+1})$. *Then, under the null hypothesis, for* $x \in S$:

$$E[r_n(z_{n,j}, x, \hat{\theta}_n)] = O(n^{-1}\triangle_n^{1/2}) + O(n^{-1}h^2) + O(n^{-3/2}h^{1/\xi} - 1), \tag{A.2}$$

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

and

$$\int Er_n^2(z_{n,j}, x, \hat{\theta}_n)a(x)\,dF(x) = \frac{2}{n^2h}\int \sigma^4(x)\pi(x)a(x)\,dF(x)\int K^2(x)\,dx + O(\triangle_n^{1/2}(n^2h)^{-1} + n^{-2}), \tag{A.3}$$

where $\xi > 1$.

Proof of (i) of Lemma 2. (A.1) directly follows Theorem 2.2 of Friedman (1975), simply by replacing the unconditional expectation by conditional expectation. ■

Proof of (ii) of Lemma 2. For notational simplicity, we denote $\int_{t_0+j\triangle_n}^{t_0+(j+1)\triangle_n} g(u)\,du$ by $\int_{\triangle_n} g(u)\,du$, where $g(u)$ is any integrable function. An application of Itô's formula yields

$$\begin{aligned} &Er_n(z_{n,j}, x, \hat{\theta}_n) \\ &= \frac{1}{n\triangle_n h}E\left\{K_j(x)E^j\left[\int_{\triangle_n}(x_u - x_{n,j})\mu(x_u)\,du + \int_{\triangle_n}(\sigma^2(x_u) - \sigma^2(x_{n,j}))\,du\right]\right\} \\ &\quad + \frac{1}{n\triangle_n h}E\left\{K_j(x)\int_{\triangle_n}(\sigma^2(x_{n,j}) - \sigma^2(x))\,du\right\} \\ &\quad - \frac{1}{nh}E[K_j(x)(\sigma_0^2(x,\hat{\theta}_n) - \sigma^2(x))] \\ &\equiv A_{n1} + A_{n2} + A_{n3}. \end{aligned}$$

We will prove (ii) of Lemma 2 by showing $A_{n1} = O(\triangle_n^{1/2}n^{-1}), A_{n2} = O(n^{-1}h)$, and $A_{n3} = O(n^{-3/2}h^{1/\xi-1})$. From Assumptions 3 and 4, (A.1), and Schwarz's inequality, it is straightforward to have that $A_{n1} = O(\triangle_n^{1/2}n^{-1})$. For A_{n2}, by Assumption 4 we have

$$\begin{aligned} A_{n2} &= \frac{1}{nh}\int K\left(\frac{u-x}{h}\right)(\sigma^2(u) - \sigma^2(x))\pi(u)\,du \\ &= n^{-1}\int K(v)(\sigma^2(x+hv) - \sigma^2(x))\pi(x+hv)\,dv \\ &= O(n^{-1}h^2). \end{aligned}$$

For A_{n3}, under H_0, from Assumption 6 and Hölder's inequality, there exists $\xi > 1$ such that

$$|A_{n3}| \le C(nh)^{-1}[E(K_j(x))^{\xi}]^{1/\xi}[E|\hat{\theta}_n - \theta_0|^{\eta}]^{1/\eta} = O(n^{-3/2}h^{1/\xi-1}),$$

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

where $\eta = (1 - \xi^{-1})^{-1}$. Hence (A.2) holds. We show (A.3).

$$\begin{aligned}
&\int Er_n^2(z_{n,j}, x, \hat{\theta}_n)a(x)\,dF(x) \\
&\quad = (nh\triangle_n)^{-1}E\left\{\int K_j^2(x)[(x_{n,j+1} - x_{n,j})^2 - \sigma^2(x,\theta_0)\triangle_n]^2 a(x)\,dF(x)\right\} \\
&\qquad + (nh)^{-2}E\left\{\int K_j^2(x)v^2(x,\theta_0,\hat{\theta}_n)a(x)\,dF(x)\right\} \\
&\qquad + (nh)^{-2}\triangle_n^{-1}E\left\{\int K_j^2(x)[(x_{n,j+1} - x_{n,j})^2 - \sigma^2(x,\theta_0)\triangle_n]v(x,\theta_0,\hat{\theta}_n)a(x)\,dF(x)\right\} \\
&\quad \equiv A_{n3}^1 + A_{n3}^2 + A_{n3}^3,
\end{aligned}$$

where $v(x,\theta_0,\hat{\theta}_n) = \sigma_0^2(x,\hat{\theta}_n) - \sigma_0^2(x,\theta_0)$.

Using (i) of Lemma 2, the mean value theorem (to $\sigma_0^2(x,\hat{\theta}_n) - \sigma_0^2(x,\theta_0)$), $\hat{\theta}_n - \theta_0 = O_p(n^{-1/2})$, and changing the variable by $(x_{n,j} - x)/h = u$, it is straightforward to have $A_{n3}^2 = O((n^3h)^{-1})$ and $A_{n3}^3 = O(\triangle_n^{1/2}(n^{5/2}h)^{-1}) + O(n^{-5/2})$. We next consider A_{n3}^1.

Applying Itô's formula to $(x_{n,j+1} - x_{n,j})^4$ and $(x_{n,j+1} - x_{n,j})^2$, we have

$$\begin{aligned}
A_{n3}^1 = (nh\triangle_n)^{-2}\Bigg\{&\int E\left[K_j^2(x)\int_{\triangle_n} 4E^j(x_u - x_{n,j})^3\mu(x_u)\,du\right]a(x)\,dF(x) \\
&- 4\triangle_n\int E\left[K_j^2(x)\int_{\triangle_n} E^j(x_u - x_{n,j})\mu(x_u)\,du\right]\sigma^2(x)a(x)\,dF(x) \\
&+ 6\int E\left[K_j^2(x)\int_{\triangle_n} E^j(x_u - x_{n,j})^2(\sigma^2(u) - \sigma^2(x))\,du\right]a(x)\,dF(x) \\
&+ 12\int E\left[K_j(x)\int_{\triangle_n}\int_{t_0+\triangle_n}^{u} E^j(x_s - x_{n,j})\mu(x_s)\,ds\,du\right]\sigma^2(x)a(x)\,dF(x) \\
&+ 6\int E\left[K_j^2(x)\int_{\triangle_n}\int_{t_0+\triangle_n}^{u} E^j(\sigma^2(x_s) - \sigma^2(x))\,ds\,du\right]\sigma^2(x)a(x)\,dF(x) \\
&- 2\int E\left[K_j^2(x)\triangle_n\int_{t_0+j\triangle_n}^{u} E^j(\sigma^2(x_u) - \sigma^2(x))\,du\right]\sigma^2(x)a(x)\,dF(x) \\
&+ 2\triangle_n^2\int E[K_j^2(x)]\sigma^4(x)a(x)\,dF(x)\Bigg\} \\
\equiv\ & A_{n3}^{11} + A_{n3}^{12} + A_{n3}^{13} + A_{n3}^{14} + A_{n3}^{15} + A_{n3}^{16} + A_{n3}^{17}.
\end{aligned}$$

By using (i) of Lemma 2, it can be shown that $A_{n3}^{11} = O(\triangle_n^{1/2}(n^2h)^{-1}), A_{n3}^{12} = O(\triangle_n^{1/2}(n^2h)^{-1})$, $A_{n3}^{13} = O(\triangle_n^{1/2}(n^2h)^{-1} + n^{-2}), A_{n3}^{14} = O(\triangle_n^{1/2}(n^2h)^{-1}), A_{n3}^{15} = O(\triangle_n^{1/2}(n^2h)^{-1} + n^{-2}), A_{n3}^{16} = O(\triangle_n^{1/2}(n^2h)^{-1} + n^{-2})$, and

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

$$A_{n3}^{17} = 2(n^2h)^{-1}\int \pi(x+hu)\sigma^4(x)a(x)\,dF(x)\int K^2(u)\,du$$

$$= 2(n^2h)^{-1}\int \pi(x)\sigma^4(x)a(x)\,dF(x)\int K^2(u)\,du + O(n^{-2}).$$

Summarizing the preceding discussion, we have shown that

$$\max\{A_{n3}^2, A_{n3}^3, A_{n3}^{11}, A_{n3}^{12}, A_{n3}^{14}\} = O(\triangle_n^{1/2}(n^2h)^{-1}),$$

$$\max\{A_{n3}^{13}, A_{n3}^{15}, A_{n3}^{16}\} = O(\triangle_n^{1/2}(n^2h)^{-1} + n^{-2}),$$

$$A_{n3}^{17} = 2(n^2h)^{-1}\int \pi(x)\sigma^4(x)a(x)\,dF(x)\int K^2(u)\,du + O(n^{-2}).$$

These results imply that (A.3) holds. ■

LEMMA 3. *Under Assumptions 1–9 and the null hypothesis, I_n can be written as*

$$I_n = \bar{I}_n + o_p((nh^{1/2})^{-1}), \tag{A.4}$$

where $\bar{I}_n = \int[(\hat{\sigma}_n^2(x) - \sigma_0^2(x,\hat{\theta}_n))\hat{\pi}(x)]^2 a(x)\,dF(x)$.

Lemma 3 indicates that the only difference between $\bar{I}_n$ and I_n is that the latter is average over the empirical conditional distribution function instead of F. Lemma 3 shows that this difference is inconsequential for the asymptotic distribution of the test statistic.

Proof of Lemma 3. We need to prove that

$$\varepsilon_n \equiv \int[(\hat{\sigma}_n^2(x) - \sigma_0^2(x,\hat{\theta}_n))\hat{\pi}(x)]^2 a(x)\,d(\hat{F}(x) - F(x))$$

$$= \sum_{i,j=1}^{n}\int (nh\triangle_n)^{-2}K_i(x)K_j(x)[(x_{n,i+1} - x_{n,i})^2 - \sigma_0^2(x,\hat{\theta}_n)\triangle_n]$$

$$\times[(x_{n,j+1} - x_{n,j})^2 - \sigma_0^2(x,\hat{\theta}_n)\triangle_n]a(x)\,d(\hat{F}(x) - F(x))$$

$$= o_p((nh^{1/2})^{-1}). \tag{A.5}$$

First, under Assumption 5 the observed data sequence $\{x_{n,t}\}$ is absolutely regular with geometric decay rate; i.e., $\{x_{n,t}\}$ satisfies the absolutely regular condition with mixing coefficient $\beta_{n,m} = O(\lambda^m)$ (Aït-Sahalia, 1996a, p. 552), where λ is a positive constant determined by the integral operator of the diffusion process (1). Let $m = [b\log n]$ and $\kappa = \log\lambda$ where b is a sufficiently large positive constant. Then we have $\beta_{n,m} = O(\lambda^m) = O(\lambda^{-b\kappa log_\lambda n}) = O(n^{-b\kappa})$.

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

Recalling $\sigma_0^2(x,\hat{\theta}_n) - \sigma_0^2(x,\theta_0) = v(x,\theta_0,\hat{\theta}_n)$, ε_n can be written as follows:

$$\begin{aligned}\varepsilon_n = n^{-3}(\triangle_n h)^{-2} \sum_{i,j,k=1}^{n} \Big\{ & K_i(x_{n,k})K_j(x_{n,k})[(x_{n,i+1}-x_{n,i})^2 - \sigma_0^2(x_{n,k},\hat{\theta}_n)\triangle_n] \\ & \times [(x_{n,j+1}-x_{n,j})^2 - \sigma_0^2(x_{n,k},\hat{\theta}_n)\triangle_n]a(x_{n,k}) \\ & - \int K_i(x)K_j(x)[(x_{n,i+1}-x_{n,i})^2 - \sigma_0^2(x,\theta_0)\triangle_n] \\ & \times [(x_{n,j+1}-x_{n,j})^2 - \sigma_0^2(x,\theta_0)\triangle_n]a(x)\,dF(x)\Big\} \\ - 2n^{-3}h^{-2}\triangle_n^{-1} \sum_{i,j,k=1}^{n} \Big\{ & K_i(x_{n,k})K_j(x_{n,k})[(x_{n,i+1}-x_{n,i})^2 - \sigma_0^2(x_{n,k},\theta_0)\triangle_n] \\ & \times v(x_{n,k},\theta_0,\hat{\theta}_n)a(x_{n,k}) \\ & - \int K_i(x)K_j(x)[(x_{n,i+1}-x_{n,i})^2 - \sigma_0^2(x,\theta_0)\triangle_n] \\ & \times v(x,\theta_0,\hat{\theta}_n)a(x)\,dF(x)\Big\} \\ + n^{-1}(nh)^{-2} \sum_{i,j,k=1}^{n} \Big\{ & K_i(x_{n,k})K_j(x_{n,k})v(x_{n,k},\theta_0,\hat{\theta}_n) \\ & - \int K_i(x)K_j(x)v^2(x,\theta_0,\hat{\theta}_n)a(x)\,dF(x)\Big\} \\ \equiv \varepsilon_{n1} - 2\varepsilon_{n2} + \varepsilon_{n3}. & \end{aligned} \tag{A.6}$$

We shall show that $\varepsilon_{ni} = o_p((nh^{1/2})^{-1})$ for $i = 1,2,3$. We will first show that $\varepsilon_{n1} = o_p((nh^{1/2})^{-1})$. We define $W_{i,j,k}(x_{n,k})$ as follows:

$$\begin{aligned} W_{i,j,k}(x_{n,k}) = {} & K_i(x_{n,k})K_j(x_{n,k})[(x_{n,i+1}-x_{n,i})^2 - \sigma_0^2(x_{n,k},\theta_0)\triangle_n] \\ & \times [(x_{n,j+1}-x_{n,j})^2 - \sigma_0^2(x_{n,k},\theta_0)\triangle_n]a(x_{n,k}) \\ & - \int K_i(x)K_j(x)[(x_{n,i+1}-x_{n,i})^2 - \sigma_0^2(x,\theta_0)\triangle_n] \\ & \times [(x_{n,j+1}-x_{n,j})^2 - \sigma_0^2(x,\theta_0)\triangle_n]a(x)\,dF(x); \end{aligned}$$

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

then,

$$\begin{aligned}\varepsilon_{n1} &= n^{-3}(h\triangle_n)^{-2}\sum_{i,j,k=1}^{n} W_{i,j,k}(x_{n,k}) \\ &= n^{-3}(h\triangle_n)^{-2}\sum_{i=1}^{n} W_{i,i,i}(x_{n,i}) + n^{-3}(h\triangle_n)^{-2}\sum_{i\neq k} W_{i,i,k}(x_{n,k}) \\ &\quad + n^{-3}(h\triangle_n)^{-2}\sum_{i\neq j,k} W_{i,j,k}(x_{n,k}) \\ &\equiv \varepsilon_{n1}^1 + \varepsilon_{n1}^2 + \varepsilon_{n1}^3. \end{aligned} \tag{A.7}$$

To prove $\varepsilon_{n1} = o_p((nh^{1/2})^{-1})$, we need to prove that $\varepsilon_{n1}^i = o_p((nh^{1/2})^{-1})$ for $i = 1,2,3$. For ε_{n1}^1, because of $(x_{n,i+1} - x_{n,i})^{2d} = O_p(\triangle_n^d)$ for $d = 1,2$ by (A.1) in Lemma 2, we immediately obtain

$$\varepsilon_{n1}^1 = O_p((nh\triangle_n)^{-2}\triangle_n^2) = o_p((nh^{1/2})^{-1}). \tag{A.8}$$

For ε_{n1}^2, we denote $(\varepsilon_{n1}^2)^2$ as

$$\begin{aligned}(\varepsilon_{n1}^2)^2 &= n^{-2}(nh\triangle_n)^{-4}\sum_{i\neq k} W_{i,i,k}\sum_{i'\neq k'} W_{i',i',k'} \\ &= (\varepsilon_{n11}^2)^2 + (\varepsilon_{n12}^2)^2 + (\varepsilon_{n13}^2)^2,\end{aligned}$$

where $(\varepsilon_{n11}^2)^2$,$(\varepsilon_{n12}^2)^2$, and $(\varepsilon_{n13}^2)^2$, respectively, denote the cases where the summation indices satisfy $\min\{|k-k'|,|k-i|,|k-i'|\} > m+1$, either $i = i', k = k'$ or $i = k'$, $k = i'$ and all remaining cases. For $(\varepsilon_{n11}^2)^2$, using $E_k(W_{i,i,k}(x_{n,k})) = 0$ and Lemma 1, we have

$$E(n^2h(\varepsilon_{n11}^2)^2) \leq 0 + Ch^{-3}(n\triangle_n)^{-4}(h\triangle_n^4)n^4\beta_{n,m}^{1/2} = O(h^{-2}\beta_{n,m}^{1/2}) = o(1). \tag{A.9}$$

For $(\varepsilon_{n12}^2)^2$,

$$E(n^2h(\varepsilon_{n12}^2)^2) \leq Ch^{-3}(n\triangle_n)^{-4}(n^2\triangle_n^4) = O((nh^{3/2})^{-2}) = o(1). \tag{A.10}$$

For $(\varepsilon_{n13}^2)^2$, using the Hölder inequality, we have

$$E(n^2h(\varepsilon_{n13}^2)^2) \leq C(nh^3)^{-1}h^{4/3}m = O(m(nh^{1.5})^{-1}) = o(1). \tag{A.11}$$

From (A.9)–(A.11), we have

$$\varepsilon_{n1}^2 = o_p((nh^{1/2})^{-1}). \tag{A.12}$$

To evaluate ε_{n1}^3, we consider the second moment of $(nh^{1/2}\varepsilon_{n1}^3)^2$,

$$E(nh^{1/2}\varepsilon_{n1}^3)^2 = h^{-3}(n\triangle_n)^{-4}\sum_{i\neq j,k}\sum_{i'\neq j',k'} E[W_{i,j,k}W_{i',j',k'}]. \tag{A.13}$$

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

We consider four different cases: (a) for any two summation indices l and L from $k, i, j, k', i', j', |l - L| > m + 1$ for all $L \neq l$; (b) there exist exactly four different summation indices such that for any index l from these four indices, $|l - L| > m + 1$ for all $L \neq l$; (c) there exist exactly three different summation indices such that for any index l from these three indices, $|l - L| > m + 1$ for all $L \neq l$; and (d) all the other remaining cases. We will use EA_s to denote these cases (s = a, b, c, d). For case (a), using $E_k W_{i,j,k} = 0$ or $E_{k'} W_{i',j',k'} = 0$ and Lemma 1, we have

$$EA_a \leq 0 + C(h^{-3})(n\triangle_n)^{-4} n^6 (h\triangle_n^4)\beta_{n,m}^{1/2} = Cn^2 h^{-2}\beta_{n,m}^{1/2} = o(1). \tag{A.14}$$

For case (b), we only need to consider the case $|k - k'| \leq m + 1$, because otherwise we will have k or k' is at least m periods away from any other indices and by Lemma 1, we know it is bounded by $Cn^6\beta_{n,m}^{1/2} = o(1)$. With $|k - k'| \leq m + 1$, we must have at least $m + 1$ periods away from any other indices for $l = i, j, i', j'$. Hence, repeating the application of Lemma 1 yields

$$\begin{aligned} EA_b &\leq h^{-3}(n\triangle_n)^{-4} \sum_{i \neq j,k} \sum_{i' \neq j',k'} E[E_j E_i (W_{i,j,k}) E_{j'} E_{i'} (W_{i',j',k'})] + Cn^6\beta_{n,m}^{1/2} \\ &\leq Ch^{-3}(n\triangle_n)^{-4}(\triangle_n^4 h^8 + \triangle_n^6 h^4) mn^5 + Cn^6\beta_{n,m}^{1/2} = o(1). \end{aligned} \tag{A.15}$$

For case (c), we only need to consider $|k - k'| \leq m + 1$, $|k - l| \leq m + 1$ for exactly one $l \in \{i, j, i', j'\}$, because otherwise it will be bounded by $Cn^6\beta_{n,m}^{1/2}$ by Lemma 1. By symmetry we only need to consider $l = i$. Repeating the application of Lemma 1 yields

$$\begin{aligned} EA_c &\leq h^{-3}(n\triangle_n)^{-4} \sum_{i \neq j,k} \sum_{i' \neq j',k'} E[E_j (W_{i,j,k}) E_{j'} E_{i'} (W_{i',j',k'})] + Cn^6\beta_{n,m}^{1/2} \\ &\leq Ch^{-3}(n\triangle_n)^{-4}(\triangle_n^2 h^4 + \triangle_n^3 h^2)(\triangle_n^2 h^2 + \triangle_n^{5/2} h) n^4 m^2 + Cn^6\beta_{n,m}^{1/2} \\ &= C(h^3 + h^2\triangle_n^{1/2} + \triangle_n h + \triangle_n^{3/2}) m^2 + Cn^6\beta_{n,m}^{1/2} = o(1). \end{aligned} \tag{A.16}$$

For case (d), for any three different l's, $|k - l| \leq m + 1$ for all $k \neq l$; case (d) has at most $m^3 n^3$ terms. Hence, using Lemma 1, we have

$$EA_d \leq Ch^{-3}(n\triangle_n)^{-4} m^3 n^3 h^2 \triangle_n^4 + Cn^6\beta_{n,m}^{1/2} = Cm^3 (nh)^{-1} = o(1). \tag{A.17}$$

From (A.13)–(A.17), we have

$$\varepsilon_{n1}^3 = o_p((nh^{1/2})^{-1}). \tag{A.18}$$

Finally, from (A.8), (A.12), and (A.18), $\varepsilon_{n1} = o_p((nh^{1/2})^{-1})$.

To prove $\varepsilon_{n2} = o_p((nh^{1/2})^{-1})$, expanding $\sigma_0^2(x, \hat{\theta}_n)$ around θ_0 yields

$$\sigma_0^2(x, \hat{\theta}_n) = \sigma_0^2(x, \theta_0) + \nabla_\theta' \sigma_0^2(x, \theta_0)(\hat{\theta}_n - \theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)' \nabla_\theta^2 \sigma_0^2(x, \bar{\theta}_n)(\hat{\theta}_n - \theta_0). \tag{A.19}$$

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

Using (A.19), ε_{n2} can be expressed as

$$\begin{aligned}\varepsilon_{n2} = {} & -(n\triangle_n)^{-2}(nh)^{-1}\sum_{i,j,k=1}^{n}\Bigg\{K_i(x_{n,k})K_j(x_{n,k}) \\ & \times[(x_{n,i+1}-x_{n,i})^2-\sigma_0^2(x,\theta_0)\triangle_n]\nabla'_\theta\sigma_0^2(x_{n,j},\theta_0)a(x_{n,k}) \\ & -\int K_i(x)K_j(x)[(x_{n,i+1}-x_{n,i})^2-\sigma_0^2(x,\theta_0)\triangle_n] \\ & \times\nabla'_\theta\sigma_0^2(x,\theta_0)a(x)\,dF(x)\Bigg\}(\hat\theta_n-\theta_0) \\ & +(\hat\theta_n-\theta_0)'(n\triangle_n)^{-1}(nh)^{-2}\sum_{i,j,k=1}^{n}\Bigg\{K_i(x_{n,k})K_j(x_{n,k}) \\ & \times[(x_{n,i+1}-x_{n,i})^2-\sigma_0^2(x_{n,k},\theta_0)\triangle_n] \\ & \times\nabla_\theta^2\sigma_0^2(x_{n,j},\bar\theta_n)a(x_{n,k}) \\ & -\int K_i(x)K_j(x) \\ & \times[(x_{n,i+1}-x_{n,i})^2-\sigma_0^2(x,\theta_0)\triangle_n] \\ & \times\nabla_\theta^2\sigma_0^2(x,\bar\theta_n)a(x)\,dF(x)\Bigg\}(\hat\theta_n-\theta_0) \\ \equiv {} & \varepsilon_{n2}^1(\hat\theta_n-\theta_0)+(\hat\theta_n-\theta_0)'\varepsilon_{n2}^2(\hat\theta_n-\theta_0). \end{aligned} \tag{A.20}$$

Using the same approach as we used to prove that $\varepsilon_{n1}=o_p((nh^{1/2})^{-1})$, we can prove that $\varepsilon_{n2}^1=o_p((nh^{1/2})^{-1})$. By $(x_{n,i+1}-x_{n,i})^2=O_p(\triangle_n)$ and Assumption 5, it follows that $\varepsilon_{n2}^2=O_p(1)$. Hence, $\varepsilon_{n2}=o_p((nh^{1/2})^{-1})$.

Finally, for ε_{n3}, we have

$$\begin{aligned}\varepsilon_{n3} = {} & n^{-1}(nh)^{-2}\sum_{i,k=1}^{n}\Bigg\{K_i(x_{n,k})K_i(x_{n,k})v^2(x_{n,k},\theta_0,\hat\theta_n) \\ & -\int K_i(x)K_j(x)v^2(x_{n,k},\theta_0,\hat\theta_n)a(x)\,dF(x)\Bigg\} \\ & +n^{-1}(nh)^{-2}\sum_{i\neq j}^{n}\Bigg\{K_i(x_{n,k})K_j(x_{n,k})v^2(x_{n,k},\theta_0,\hat\theta_n) \\ & -\int K_i(x)K_j(x)v^2(x_{n,k},\theta_0,\hat\theta_n)a(x)\,dF(x)\Bigg\} \\ \equiv {} & \varepsilon_{n3}^1+\varepsilon_{n3}^2. \end{aligned} \tag{A.21}$$

It is easy to show that $\varepsilon_{n3}^1=O_p(n^{-2}h^{-1})=o_p(1)$ and $\varepsilon_{n3}^2=O_p(n^{-1}h^{-1})=o_p(1)$. Hence, $\varepsilon_{n3}=o_p(nh^{1/2})$. ■

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

We will need a CLT for degenerate U-statistics of the triangular arrays of random variables $\{X_{n,l}, l \le n; n \ge 1\}$, which is used in deriving the asymptotic distribution of the test statistic proposed in this paper. We write

$$U_n = \sum \sum_{1 \le s \le t \le n} H_n(X_{n,t}, X_{n,s}),$$

where $H_n(.,.)$ depends on n and satisfies $\int H_n(x,y)\, dF_n(x) = 0$ for all y and $F_n(.)$ is the marginal distribution function of $X_{n,t}$. For every n, let $\{\widetilde{X}_{n,l}\}_{l=1}^n$ be an independent identically distributed (i.i.d.) sequence having the same marginal distribution as $\{x_{n,l}\}$. Following the same approach as in Fan and Li (1999), we define $\sigma_n^2 \equiv E[H_n^2(\widetilde{X}_{n,1}, \widetilde{X}_{n,2})]$. In Lemma 4, which follows, Assumptions (A1)–(A3) in Fan and Li (1999) are said to be satisfied by $\{X_{n,t}\}$ if the conditions in Assumptions (A1)–(A3) in Fan and Li (1999) calculated by every row of $\{X_{n,t}\}$ are satisfied.

LEMMA 4. *Let $\{X_{n,t}\}$ be strictly stationary and satisfy the absolutely regular condition with mixing coefficient $\beta_{n,m}$. If Assumptions (A1)–(A3) in Fan and Li (1999) are satisfied by $\{X_{n,t}\}$, then we have $(\sqrt{2}U_n)/n\sigma_n \to N[0,1]$ in distribution as $n \to 0$.*

Proof of Lemma 4. The proof follows in the same way as the proof of Theorem 2.1 in Fan and Li (1999) and thus is omitted here.

Proof of part (i) of Theorem 1. From Lemma 3, we will complete the proof of part (i) of Theorem 1 by showing

(a) $\bar{J}_n \equiv (nh^{1/2})[\bar{I}_n - (2/n^2h)\sum_{t=1}^n (\hat{\sigma}_n^2(x_{n,t}))^2 a(x_{n,t})\hat{\pi}(x_{n,t}) \int K^2(u)\, du] \to N[0, v^2]$

in distribution as $n \to \infty$.

(b) $\hat{v}_n^2 \to v^2$ in probability as $n \to \infty$.

Proof of (a). $\bar{I}_n$ can be rewritten as

$$\bar{I}_n = \int \left\{ \sum_{j=1}^{n-1} \frac{1}{nh\triangle_n} K_j(x)[(x_{n,j+1} - x_{n,j}) - \sigma_0^2(x, \hat{\theta}_n)\triangle_n] \right\}^2 a(x)\, dF(x).$$

Let $\bar{r}_n(z_{n,j}, x, \theta) = r_n(z_{n,j}, x, \theta) - Er_n(z_{n,j}, x, \theta)$, where $z_{n,j} = (x_{n,j}, x_{n,j+1})$. We decompose $\bar{I}_n$ according to

$$\begin{aligned}
\bar{I}_n &= 2 \sum_{1 \le j < k \le n} \int \bar{r}_n(z_{n,j}, x, \hat{\theta}_n)\bar{r}_n(z_{n,k}, x, \hat{\theta}_n) a(x)\, dF(x) \\
&\quad + \sum_{j=1}^n \int \bar{r}_n^2(z_{n,j}, x, \hat{\theta}_n) a(x)\, dF(x) \\
&\quad + 2(n-1) \sum_{j=1}^n \int \bar{r}_n(z_{n,j}, x, \hat{\theta}_n) E(\bar{r}_n(z_{n,j}, x, \hat{\theta}_n)) a(x)\, dF(x) \\
&\quad + n(n-1) \int [E\bar{r}_n(z_{n,j}, x, \hat{\theta}_n)]^2 a(x)\, dF(x) \\
&\equiv \bar{I}_{n1} + \bar{I}_{n2} + \bar{I}_{n3} + \bar{I}_{n4}. \qquad \textbf{(A.22)}
\end{aligned}$$

We will show under our assumptions that $\bar{I}_{n1}$ is asymptotically normal in distribution and $\bar{I}_{n3}$ and $\bar{I}_{n4}$ are asymptotically negligible in probability, whereas $\bar{I}_{n2}$ gives a bias

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

term. First, we prove that $nh^{1/2}(\bar{I}_{n2} - E\bar{I}_{n2}) \to 0$ in probability under the null hypothesis.

$$\begin{aligned}\operatorname{var}(\bar{I}_{n2}) &= \sum_{j=1}^{n} \operatorname{var}\left(\int r_n^2(z_{n,j}, x, \hat{\theta}_n) a(x)\, dF(x) \right) \\ &\quad + 2 \sum_{1 \le j < k \le n} \left\{ E\left[\int r_n^2(z_{n,j}, x, \hat{\theta}_n) a(x)\, dF(x) \int r_n^2(z_{n,k}, x, \hat{\theta}_n) a(x)\, dF(x) \right] \right. \\ &\quad \left. - \int E r_n^2(z_{n,j}, x, \hat{\theta}_n) a(x)\, dF(x) \int E r_n^2(z_{n,k}, x, \hat{\theta}_n) a(x)\, dF(x) \right\} \\ &\equiv \bar{I}_{n2}^1 + \bar{I}_{n2}^2. \end{aligned} \tag{A.23}$$

Changing variable by $(x_{n,j} - x)/h = u$ and using (i) of Lemma 2, we have

$$\begin{aligned} n^{-1}\bar{I}_{n2}^1 &\le E\left[\int r_n^2(z_{n,j}, x, \hat{\theta}_n) a(x)\, dF(x) \right]^2 \\ &= E\left\{ \int (nh\triangle_n)^2 K_j^2(x) [(x_{n,j+1} - x_{n,j})^4 - 2\sigma_0^2(x, \hat{\theta}_n)\triangle_n (x_{n,j+1} - x_{n,j})^2 \right. \\ &\quad \left. + \sigma_0^4(x, \hat{\theta}_n)\triangle_n^2] a(x)\, dF(x) \right\}^2 \\ &\le C((n\triangle_n)^4 h^2)^{-1} E\left\{ \left[\int K^2(u) a(x_{n,j} + hu) \pi(x_{n,j} + hu)\, du \right]^2 (x_{n,j+1} - x_{n,j})^8 \right\} \\ &\quad + C(n^4h^2)^{-1} E\left[\int K^2(u) \sigma_0^8(x_{n,j} + hu, \hat{\theta}_n) a(x_{n,j} + hu) \pi(x_{n,j} + hu)\, du \right]^2 \\ &\quad + C(n^4h^2)^{-1} E\left\{ \left[\int K^2(u) \sigma_0^4(x_{n,j} + hu, \hat{\theta}_n) a(x_{n,j} + hu) \pi(x_{n,j} + hu)\, du \right]^2 \right. \\ &\quad \left. \times (x_{n,j+1} - x_{n,j})^4 \right\} \\ &= O((n^4h^2)^{-1}). \end{aligned} \tag{A.24}$$

For $\bar{I}_{n2}^2$, we consider two different cases: (a) $\min\{|j - k| > m + 1\}$ and (b) $\min\{|j - k| \le m + 1\}$. We will use EB_a and EB_b to denote cases (a) and (b), respectively. By Schwarz's inequality and (A.24), we have

$$E\left[\int r_n^2(z_{n,j}, x, \hat{\theta}_n) a(x)\, dF(x) \int r_n^2(z_{n,k}, y, \hat{\theta}_n) a(y)\, dF(y) \right]^2 = O((n^4h^2)^{-1}). \tag{A.25}$$

Hence, by Lemma 1, we have

$$EB_a \le 0 + C(nh)^{-2}\beta_{n,m}^{1/2}, \tag{A.26}$$

$$EB_b \le Cmn^{-3}h^{-2}. \tag{A.27}$$

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

By (A.3), (A.25)–(A.27), and Chebyshev's inequality, it follows that

$$nh^{1/2}\left[\bar{I}_{n2} - \frac{2}{nh}\int \sigma^4(x)\pi(x)a(x)\,dF(x)\int K^2(x)\,dx\right] = o_p(1), \tag{A.28}$$

which characterizes the asymptotic bias term in our test statistic. Using Theorem 3.3.2 and Remarks 3.3.4 and 3.3.5 in Györfi, Härdle, Sarda, and Vieu (1989, Sect. III.3), one can establish the following results:

$$\sup_{x\in S}|\hat{\pi}(x) - \pi(x)| = O_p(h^2 + (nh)^{-1/2}ln^{1/2}(n)), \tag{A.29}$$

$$\sup_{x\in S}|\hat{\sigma}^2(x) - \sigma^2(x)| = O_p(h^2 + (nh)^{-1/2}ln^{1/2}(n)). \tag{A.30}$$

The proofs of (A.29) and (A.30) are available from the author upon request. By the CLT it is easy to verify that

$$\frac{2}{nh}\int \sigma^4(x)\pi(x)a(x)\,dF(x) = \frac{2}{n^2h}\sum_{i=1}^{n}\sigma^4(x_{n,i})\pi(x_{n,i})a(x_{n,i}) + O_p(n^{-3/2}h^{-1}). \tag{A.31}$$

From (A.28)–(A.31), we obtain the sample analogue of (A.28),

$$nh^{1/2}\left[\bar{I}_{n2} - \frac{2}{n^2h}\sum_{i=1}^{n}(\hat{\sigma}^2(x_{n,i}))^2a(x_{n,i})\hat{\pi}(x_{n,i})\int K^2(u)\,du\right] = o_p(1). \tag{A.32}$$

To prove $nh^{1/2}\bar{I}_{n3} = o_p(1)$, we evaluate $E(nh^{1/2}\bar{I}_{n3})^2$. Using Lemma 2 and choosing $1 < \xi < \frac{4}{3}$ in (A.2), we have

$$\begin{aligned} E(nh^{1/2}\bar{I}_{n3})^2 &= n^2(n-1)^2h\sum_{j,k=1}^{n}\iint E[r_n(z_{n,j},x,\hat{\theta}_n)r_n(z_{n,k},y,\hat{\theta}_n)] \\ &\quad\times E[r_n(z_{n,1},x,\hat{\theta}_n)]E[r_n(z_{n,1},y,\hat{\theta}_n)]a(x)a(y)\,dF(x)\,dF(y) \\ &\quad - n^4(n-1)^2h\left\{\int [Er_n(z_{n,1},x,\hat{\theta}_n)]^2a(x)\,dF(x)\right\}^2 = o(1). \end{aligned} \tag{A.33}$$

From $E(\bar{I}_{n3}) = 0$, (A.33), and Chebyshev's inequality, it follows that $nh^{1/2}\bar{I}_{n3} = o_p(1)$. For $\bar{I}_{n4}$, choosing $1 < \xi < \frac{4}{3}$ from (A.2) we have

$$\begin{aligned} nh^{1/2}\bar{I}_{n4} &= n^2(n-1)h^{1/2}\int [Er_n(z_{n,1},x,\hat{\theta}_n)]^2a(x)\,dF(x) \\ &= O(Th^{1/2}) + O(nh^{4.5}) + O(h^{2/\xi-3/2}) = o(1). \end{aligned} \tag{A.34}$$

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

Next, we express $\bar{I}_{n1}$ in U-statistic form as

$$\begin{aligned}
\bar{I}_{n1} &= 2\sum_{1\le j<k\le n}\int \bar{r}_n(z_{n,j},x,\theta_0)\bar{r}_n(z_{n,k},x,\theta_0)a(x)\,dF(x)\\
&\quad + (nh)^{-2}\sum_{j\neq k}\int\{K_j(x)[\sigma_0^2(x,\hat{\theta}_n)-\sigma_0^2(x,\theta_0)]-E[K_j(x)(\sigma_0^2(x,\hat{\theta}_n)-\sigma_0^2(x,\theta_0))]\}\\
&\quad \times\{K_k(x)[\sigma_0^2(x,\hat{\theta}_n)-\sigma_0^2(x,\theta_0)]-E[K_k(x)(\sigma_0^2(x,\hat{\theta}_n)-\sigma_0^2(x,\theta_0))]\}a(x)\,dF(x)\\
&\quad - 2\sum_{j\neq k}\int \bar{r}_n(z_{n,j},x,\theta_0)\{K_k(x)(\sigma_0^2(x,\hat{\theta}_n)-\sigma_0^2(x,\theta_0))\\
&\qquad\qquad - E[K_k(x)(\sigma_0^2(x,\hat{\theta}_n)-\sigma_0^2(x,\theta_0))]\}a(x)\,dF(x)\\
&\equiv \bar{I}_{n11}+\bar{I}_{n12}-2\bar{I}_{n13}. \qquad \textbf{(A.35)}
\end{aligned}$$

Using the Hölder inequality, we have $\bar{I}_{n12}=o_p((nh^{1/2})^{-1})$ and $\bar{I}_{n13}=o_p((nh^{1/2})^{-1})$. We will use Lemma 4 to prove that $\bar{I}_{n11}$ is asymptotically normal in distribution. We now verify that Assumptions (A1)–(A3) in Fan and Li (1999) are satisfied under Assumptions 1–9.

Let $H(z_{n,j},z_{n,k})\equiv\int\bar{r}_n(z_{n,j},x,\theta_0)\bar{r}_n(z_{n,k},x,\theta_0)a(x)\,dF(x)$ and $\{\tilde{z}_{n,j}\}_{j=1}^n$ be an i.i.d. sequence having the same marginal distribution as $\{z_{n,j}\}_{j=1}^n$. Then we have

$$\begin{aligned}
\sigma_n^2 &\equiv E[H^2(\tilde{z}_{n,1},\tilde{z}_{n,2})]=\int E[\bar{r}_n(\tilde{z}_{n,1},x,\theta_0)\bar{r}_n(\tilde{z}_{n,1},y,\theta_0)]\\
&\quad\times E[\bar{r}_n(\tilde{z}_{n,2},x,\theta_0)\bar{r}_n(\tilde{z}_{n,2},y,\theta_0)]a(x)a(y)\,dF(x)\,dF(y)\\
&= \iint E[r_n(\tilde{z}_{n,1},x,\theta_0)r_n(\tilde{z}_{n,1},y,\theta_0)]\\
&\quad\times E[r_n(\tilde{z}_{n,2},x,\theta_0)r_n(\tilde{z}_{n,2},y,\theta_0)]a(x)a(y)\,dF(x)\,dF(y)\\
&\quad - 2\iint E[r_n(\tilde{z}_{n,1},x,\theta_0)r_n(\tilde{z}_{n,1},y,\theta_0)]\\
&\quad\times E[r_n(\tilde{z}_{n,2},x,\theta_0)]E[r_n(\tilde{z}_{n,2},y,\theta_0)]a(x)a(y)\,dF(x)\,dF(y)\\
&\quad + \iint E[r_n(\tilde{z}_{n,1},x,\theta_0)]E[r_n(\tilde{z}_{n,1},y,\theta_0)]\\
&\quad\times E[r_n(\tilde{z}_{n,2},x,\theta_0)]E[r_n(\tilde{z}_{n,2},y,\theta_0)]a(x)a(y)\,dF(x)\,dF(y).
\end{aligned}$$

To evaluate σ_n^2, we first evaluate $E[r_n(\tilde{z}_{n,1},x,\theta_0)r_n(\tilde{z}_{n,1},y,\theta_0)]$.

$$\begin{aligned}
&E[r_n(\tilde{z}_{n,1},x,\theta_0)r_n(\tilde{z}_{n,1},y,\theta_0)]\\
&\quad=(nh\triangle_n)^{-2}E\{K_1(x)K_1(y)E_1[(x_{n,2}-x_{n,1})^2-\sigma^2(x)\triangle_n]^2\}\\
&\qquad+\sigma^2(x)\triangle_n(nh\triangle_n)^{-2}E\{K_1(x)K_1(y)E_1[(x_{n,2}-x_{n,1})^2-\sigma^2(x)\triangle_n]\}\\
&\qquad-\sigma^2(y)\triangle_n(nh\triangle_n)^{-2}E\{K_1(x)K_1(y)E_1[(x_{n,2}-x_{n,1})^2-\sigma^2(x)\triangle_n]\}.
\end{aligned}$$

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

Applying Itô's formula to $(x_{n,2} - x_{n,1})^2$ and $(x_{n,2} - x_{n,1})^4$, and using similar arguments as in the proof of Lemma 2, we obtain

$$
\begin{aligned}
&E[r_n(\tilde{z}_{n,1}, x, \theta_0) r_n(\tilde{z}_{n,1}, y, \theta_0)] \\
&\quad = 2\sigma^4(x)(n^2h)^{-1} \int K(u) K\left(\frac{x-y}{h} + u\right) \pi(x + hu)\, du \\
&\qquad + \int K(u) K\left(\frac{x-y}{h} + u\right) \pi(x + hu)\, du \times O(n^{-2} + \triangle_n^{1/2}(n^2h)^{-1}) \\
&\quad = 2\sigma^4(x)(n^2h)^{-1} \int K(u) K\left(\frac{x-y}{h} + u\right) \pi(x + hu)\, du \\
&\qquad + \int K(u) K\left(\frac{x-y}{h} + u\right) \pi(x + hu)\, du \times (n^2h)^{-1} \qquad \textbf{(A.36)}
\end{aligned}
$$

and

$$
E[r_n(\tilde{z}_{n,1}, x, \theta_0)] E[r_n(\tilde{z}_{n,1}, y, \theta_0)] = O(n^{-2}\triangle_n) + O(n^{-2}h^2) + O(n^{-2}h\triangle_n^{1/2}). \qquad \textbf{(A.37)}
$$

From (A.36) and (A.37), it follows that $E[r_n(\tilde{z}_{n,1}, x, \theta_0)] E[r_n(\tilde{z}_{n,1}, y, \theta_0)] = o(E[r_n(\tilde{z}_{n,1}, x, \theta_0) r_n(\tilde{z}_{n,1}, y, \theta_0)])$. Hence, we get

$$
\begin{aligned}
\sigma_n^2 &= \iint E[r_n(\tilde{z}_{n,1}, x, \theta_0) r_n(\tilde{z}_{n,1}, y, \theta_0)] \\
&\qquad \times E[r_n(\tilde{z}_{n,2}, x, \theta_0) r_n(\tilde{z}_{n,2}, y, \theta_0)] a(x) a(y)\, dF(x)\, dF(y) \\
&\qquad + o\left(\iint E[r_n(\tilde{z}_{n,1}, x, \theta_0) r_n(\tilde{z}_{n,1}, y, \theta_0)] \right. \\
&\qquad\qquad \left. \times E[r_n(\tilde{z}_{n,2}, x, \theta_0) r_n(\tilde{z}_{n,2}, y, \theta_0)] a(x) a(y)\, dF(x)\, dF(y) \right) \\
&= \frac{4}{n^4h^2} \iint \sigma^4(x)\sigma^4(y) \iint K(u) K(v) K\left(\frac{x-y}{h} + u\right) K\left(\frac{x-y}{h} + v\right) \\
&\qquad \times \pi(x + hu)\pi(x + hv)\, du\, dv a(x) a(y) \pi(x)\pi(y)\, dx\, dy + o(n^{-4}h^{-1}) \\
&= 4(n^4h)^{-1} \int \sigma^8(y)\pi^4(y)a^2(y)\, dy \iiint K(u)K(v)K(w+u)K(w+v)\, du\, dv\, dw \\
&\qquad + o((n^4h)^{-1}) \\
&= 4(n^4h)^{-1} \int \sigma^8(y)\pi^4(y)a^2(y)\, dy \int \left[\int K(u)K(w+u)\, du \right]^2 dw + o((n^4h)^{-1}).
\end{aligned}
$$

(A.38)

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

To conserve space, the detailed proof of the rest of the conditions of the CLT in Lemma 4 is not incorporated here, but it is available in Li (2005) or from the author.

$$\mu_{n4} \equiv E[H^4(\tilde{z}_{n,1}, \tilde{z}_{n,2})] = O((n^8 h^3)^{-1}) \tag{A.39}$$

$$\gamma_{n11} \equiv \max_{t\neq s, t'\neq s'} E[H(z_{n,t}, z_{n,s})H(z_{n,t'}, z_{n,s'})] = O(n^{-4}h^{4/\eta-4}), \tag{A.40}$$

where $1 < \eta = (1 - \zeta^{-1})^{-1} < \frac{4}{3}$ and ζ is slightly larger than 2.

$$\gamma_{n22} \equiv \max_{t\neq s, t'\neq s'} E[H^2(z_{n,t}, z_{n,s})H^2(z_{n,t'}, z_{n,s'})] = O(n^{-8}h^{4/\eta-6}), \tag{A.41}$$

$$\gamma_{n13} \equiv \max_{t\neq s, t'\neq s'} E[H(z_{n,t}, z_{n,s})H^3(z_{n,t'}, z_{n,s'})] = O(n^{-8}h^{4/\eta-6}), \tag{A.42}$$

where $1 < \eta \leq \frac{4}{3}$.

$$\tilde{\gamma}_{n14} \equiv \max_{t\neq s} \int \{E[H(z, z_{n,t})H(z, z_{n,s})]\}^2 \, dF_n(z) = O(n^{-8}h^{5/\eta-5}), \tag{A.43}$$

where $1 < \eta \leq \frac{5}{3}$.

$$\tilde{\gamma}_{n22} \equiv E[H^2(\tilde{z}_{n,1}, \tilde{z}_{n,2})H^2(\tilde{z}_{n,1}, \tilde{z}_{n,3})] = O(n^{-8}h^{2/\eta-4}), \tag{A.44}$$

where $1 < \eta < 2$. Summarizing the preceding discussion, we have shown that $\sigma_n^2 = O(n^{-4}h^{-1})$, $\mu_{n4} = O(n^{-8}h^{-3})$, $\gamma_n \equiv \max\{\gamma_{n11}, \bar{\gamma}_{n22}, \bar{\gamma}_{n14}\} = O(n^{-4}h^{4/\eta-4})$, and $v_n \equiv \max\{\gamma_{n22}, \gamma_{n13}\} = O(n^{-8}h^{4/\eta-6})$. These results, together with the assumption $h = O(n^{-1/\gamma})$ and $1.5 < \gamma < 4.5$, imply (A1)(i)–(iii) in Fan and Li (1999).

Define $G_n(z_t, z_s) \equiv E[H(z_{n,1}, z_t)H(z_{n,1}, z_s)]$. Then,

$$\sigma_{nG}^2 \equiv E[G_n^2(z_t, z_s)] = O(n^{-8}h^{3/\eta-4}), \tag{A.45}$$

where $1 < \eta \leq \frac{3}{2}$.

$$\mu_{nG2} \equiv \max_{t\neq s} \int G^2(z_{n,t}, z_{n,s}) \, dQ(z_{n,t}, z_{n,s}) = O(n^{-8}h^{3/\eta-4}), \tag{A.46}$$

where $1 < \eta < \frac{3}{2}$.

$$\begin{aligned}\gamma_{nG11} \equiv \max\{&\max_{s\neq s'\neq s''} |E[G(z_{n,s}, z_{n,s})G(z_{n,s'}, z_{n,s''})]|,\\ &\max_{s\neq s'\neq s''} |E[G(z_{n,s}, z_{n,s'})G(z_{n,s}, z_{n,s''})]|,\\ &\max_{s\neq s'\neq s''\neq \bar{s}} |E[G(z_{n,s}, z_{n,s'})G(z_{n,s''}, z_{n,\bar{s}})]|\}\\ = O(n^{-8}h^{3/\eta-4}),\end{aligned} \tag{A.47}$$

where $1 < \eta < \frac{3}{2}$. Thus, (A2)(i)–(iii) in Fan and Li (1999) are satisfied.

Finally it is easy to show that M_n in Fan and Li (1999) is bounded by some positive constant. From $\beta_{n,m} = O(n^{-b\kappa})$, we have $m^2 n^2 \beta_{n,m}^{1/2}/\sigma_n^4 = o(1)$, provided we choose b sufficiently large. Thus (A3)(i) and (ii) in Fan and Li (1999) are satisfied. ■

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press

Proof of (b). From Assumption 8, (A.29), and (A.30), we have

$$\inf_{x\in S} \pi(x) \equiv C_0 > 0, \tag{A.48}$$

$$\sup_{x\in S} |(\hat{\sigma}^2(x)\hat{\pi}(x))^4 - (\sigma^2(x)\pi(x))^4| = o_p(1). \tag{A.49}$$

From (A.48) and (A.49) it follows that

$$\begin{aligned}
&\left| \frac{1}{n}\sum_{i=1}^{n} \left[\frac{(\hat{\sigma}^2(x_{n,i})\hat{\pi}(x_{n,i}))^4}{\hat{\pi}(x_{n,i})} a^2(x_{n,i}) - \frac{(\sigma^2(x_{n,i})\pi(x_{n,i}))^4}{\pi(x_{n,i})} a^2(x_{n,i}) \right] \right| \\
&\quad \le \sup_{x\in S} \left| \frac{(\hat{\sigma}^2(x)\hat{\pi}(x))^4 - (\sigma^2(x)\pi(x))^4}{\pi(x)} a^2(x) \right| \\
&\qquad + \sup_{x\in S} \left| \frac{(\hat{\sigma}^2(x)\hat{\pi}(x))^4 (\hat{\pi}(x) - \pi(x))}{\pi(x)\hat{\pi}(x)} a^2(x) \right| \\
&\quad = o_p(1),
\end{aligned}$$

which implies

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \frac{(\hat{\sigma}^2(x_{n,i})\hat{\pi}(x_{n,i}))^4}{\hat{\pi}(x_{n,i})} a^2(x_{n,i}) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n} \frac{(\sigma^2(x_{n,i})\pi(x_{n,i}))^4}{\pi(x_{n,i})} a^2(x_{n,i}) + o_p(1) \\
&\quad = \int \sigma^8(x)\pi^4(x)a^2(x)\,dx + o_p(1)
\end{aligned}$$

by the law of large numbers. Thus, $v_n^2 = v^2 + o_p(1)$. ■■

Proof of part (ii) of Theorem 1. Given Assumptions 1–9, from Theorem 3.2 in Corradi and White (1999), $\hat{\theta}_n$ is an $\sqrt{n}$-consistent estimator of some $\theta^* \in \Theta$. We can show by using similar arguments as those in the proof of part (i) of Theorem 1 that, under H_1,

$$I_n = \int [(\sigma^2(x) - \sigma_0^2(x,\theta^*))\pi(x)]^2 a(x)\,dF(x) + o_p(1),$$

where $\int [(\sigma^2(x) - \sigma_0^2(x,\theta^*))\pi(x)]^2 a(x)\,dF(x) > 0$ and $v_n^2 \to v^2$ in probability. Hence, we have $J_n = O_p(nh^{1/2})$. ■

https://doi.org/10.1017/S0266466607070107 Published online by Cambridge University Press