

ADAPTIVE TESTING IN CONTINUOUS-TIME DIFFUSION MODELS

JITI GAO

The University of Western Australia

MAXWELL KING

Monash University

We propose an optimal test procedure for testing the marginal density functions of a class of nonlinear diffusion processes. The proposed test is not only an optimal one but also avoids undersmoothing. An adaptive test is constructed, and its asymptotic properties are investigated. To show the asymptotic properties, we establish some general results for moment inequalities and asymptotic distributions for strictly stationary processes under the α -mixing condition. These results are applicable to some other estimation and testing of strictly stationary processes with the α -mixing condition. An example of implementation is given to demonstrate that the proposed model specification procedure is applicable to economic and financial model specification and can be implemented in practice. To ensure the applicability and implementation, we propose a computer-intensive simulation scheme for the choice of a suitable bandwidth involved in the kernel estimation and also a simulated critical value for the proposed adaptive test. Our finite sample studies support both the proposed theory and the simulation procedure.

1. INTRODUCTION AND MOTIVATION

Continuous-time diffusion processes arise in many applications in econometrics, but perhaps nowhere do they play as large a role as in finance. Following the pathbreaking work of Black and Scholes (1973), the use of continuous-time diffusion processes has become a common feature of many applications, especially asset pricing models. This is probably due to the following two reasons.

The authors thank the co-editor and three anonymous referees for their constructive comments and suggestions. The first author also thanks Song Xi Chen for some constructive suggestions, in particular the suggestion on using the local linear form instead of the Nadaraya–Watson kernel form in equation (2.6), and Yongmiao Hong for sending a working paper. The authors acknowledge comments from seminar participants at the International Chinese Statistical Association Meeting in Hong Kong in July 2001, the Western Australian Branch Meeting of the Statistical Society of Australia in September 2001, the University of Western Australia, and Monash University. Thanks also go to the Australian Research Council for its financial support. Address correspondence to Jiti Gao, School of Mathematics and Statistics, the University of Western Australia, Crawley, WA 6009, Australia; e-mail: jiti@maths.uwa.edu.au; or to Maxwell King, Faculty of Business and Economics, Monash University, Clayton, VIC 3800, Australia; e-mail: max.king@buseco.monash.edu.au.

The first one is that continuous-time diffusion processes are able to mimic some important macroeconomic and financial phenomena (see Sundaresan, 2001). The second reason is that various parametric diffusion processes have already been used nicely to model financial data. In both theory and practice, however, one needs to specify whether a parametric diffusion process is appropriate for a given set of financial data. In other words, one needs to determine whether it is appropriate to use a diffusion process with both the drift and the volatility assumed to be parametric for a given set of financial data. To justify whether the use of parametric diffusion processes is appropriate or not for a given set of financial data, empirical researchers have recently shown a preference for nonparametric alternatives. Aït-Sahalia (1996a, 1996b) was among the first to pioneer the nonparametric approach. Other related studies include Jiang and Knight (1997), Stanton (1997), Chapman and Pearson (2000), Gao and King (2001), Hong and Li (2004), and Fan and Zhang (2003). Aït-Sahalia (1996a) considers testing the marginal density functions of a class of diffusion processes under the β -mixing condition. Pritsker (1998) conducts a finite sample simulation of a nonparametric kernel test proposed in Aït-Sahalia (1996a). The principal result of Pritsker (1998) is that the test rejects true models much too often when asymptotic critical values are used. This suggests that the use of an asymptotic critical value may not be suitable in the finite sample analysis of a test power. In addition, the use of an estimation-based bandwidth in the nonparametric kernel test may also contribute to the poor performance of the test in finite sample studies, because an estimation-based optimal bandwidth may not necessarily imply that the corresponding test is optimal. We have been motivated by these two aspects to establish a simulation procedure for the choice of both an appropriate critical value and a test optimum bandwidth to improve the test proposed in Aït-Sahalia (1996b).

Recently, Horowitz and Spokoiny (2001) have developed a new test of a parametric model of a conditional mean function against a nonparametric alternative. The test adapts to the unknown smoothness of the alternative model and is uniformly consistent against alternatives whose distance from the parametric model converges to zero at the fastest possible rate. This rate is slower than $T^{-1/2}$, where T is the number of observations. To the best of our knowledge, the problem of extending the approach of Horowitz and Spokoiny (2001) to construct an adaptive and optimal test for marginal density functions has not been considered. This paper then proposes an adaptive test for testing marginal density functions. The proposed test has an optimal-rate property. In theory, the proposed test is consistent against some local alternatives with an optimal rate as stated in Section 3. In practice, we demonstrate how to apply the test in Section 4 through using a simulated example. Our studies show that the proposed test has some advantages over the test proposed in Aït-Sahalia (1996a).

The rest of the paper is organized as follows. Section 2 discusses the testing of the marginal density. An adaptive test procedure is proposed in Section 3. Section 4 provides an example of implementation. Section 5 concludes the paper

with some remarks on extensions. Mathematical assumptions and proofs are relegated to Appendixes A–C.

2. TESTING MARGINAL DENSITY FUNCTIONS

Consider a continuous-time diffusion process of the form

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

where $\mu(\cdot)$ and $\sigma(\cdot) > 0$ are, respectively, the univariate drift and volatility functions of the process indexed by θ and B_t is standard Brownian motion.

Let $\{r_t\}$ satisfy model (2.1) and $f(\cdot, \theta)$ be a parametric form of the marginal density function of $\{r_t\}$. Within the diffusion process, $f(\cdot, \theta)$ is completely determined by the corresponding drift $\mu(\cdot, \theta)$ and the diffusion $\sigma(\cdot, \theta)$ (see Aït-Sahalia, 1996a, expression (6)) given by

$$f(x, \theta) = \frac{\xi(\theta)}{\sigma^2(x, \theta)} \exp \left\{ \int_{x_0}^x \frac{2\mu(u, \theta)}{\sigma^2(u, \theta)} du \right\}, \tag{2.2}$$

where $\{r_t\}$ is distributed on $D = (x_{\min}, x_{\max})$ with $-\infty \leq x_{\min} < x_{\max} \leq \infty$, both the lower bound x_0 and $\xi(\theta)$ can be chosen to ensure that $f(x, \theta)$ is a probability density, and θ is an unknown parameter vector. Let Θ denote a parameter space in R^q and $\theta_0 \in \Theta$ denote the true value of θ .

Let $f(x)$ be a nonparametric form of the density function. The null and alternative hypotheses are

$$\mathcal{H}_0 : f(x) = f(x, \theta_0) \quad \text{versus} \quad \mathcal{H}_1 : f(x) = f(x, \theta_1) + C_T \Delta_T(x), \quad \theta_1 \in \Theta, \tag{2.3}$$

where $0 \leq C_T \leq 1$, $\lim_{T \rightarrow \infty} C_T = 0$, and $\Delta_T(x)$ is a continuous function satisfying $\int \Delta_T(x) dx = 0$ and $f(x) \geq 0$ under \mathcal{H}_1 . Theoretically, this requires that under \mathcal{H}_1 , the alternative function is still a probability density. In practice, the form of $\Delta_T(x)$ needs to be constructed. The simple and natural choice of $\Delta_T(x)$ is $\Delta_T(x) = f_1(x, \theta_1) - f(x, \theta_1)$, where $f_1(x, \theta)$ is another specified density function and $\theta_1 \in \Theta$. For example, $f(x, \theta)$ is the marginal density of $\{r_t\}$ satisfying the CIR model proposed in Cox, Ingersoll, and Ross (1985), and $f_1(x, \theta)$ is the marginal density of $\{r_t\}$ satisfying the AG model proposed in Ahn and Gao (1999).

For this case, the hypothesis structure (2.3) can be written as

$$\mathcal{H}_0 : f(x) = f(x, \theta_0) \quad \text{versus} \quad \mathcal{H}_1 : f(x) = f(x, \theta_1) + C_T (f_1(x, \theta_1) - f(x, \theta_1)).$$

This is equivalent to

$$\mathcal{H}_0 : f(x) = f(x, \theta_0) \quad \text{versus} \quad \mathcal{H}_1 : f(x) = (1 - C_T)f(x, \theta_1) + C_T f_1(x, \theta_1).$$

This basically requires us to test whether $\{r_t\}$ is sampled from $f(x, \theta_0)$ or from $f(x, \theta_1)$ with probability $1 - C_T$ and from $f_1(x, \theta_1)$ with probability C_T . Obviously, such a structure of the null hypothesis versus a sequence of local alternatives naturally extends the usual structure of the null hypothesis against a global alternative of the form

$$\mathcal{H}'_0 : f(x) = f(x, \theta_0) \quad \text{versus} \quad \mathcal{H}'_1 : f(x) = f_1(x, \theta_1), \quad \theta_1 \in \Theta.$$

For the diffusion process, we observe the process at dates $\{t\Delta | t = 0, 1, \dots\}$, where $\Delta > 0$ is generally small but fixed. Let $X_t = r_{(t-1)\Delta}$ for $t \geq 1$ throughout this section. Let $k(\cdot)$ be a kernel function, $k_h(\cdot) = h^{-1}k(\cdot/h)$, and $\hat{f}(x) = (1/T) \sum_{t=1}^T k_h(x - X_t)$ be the standard kernel density estimator of $f(x)$. Intuitively, it is natural to compare $\hat{f}(x)$ and $f(x, \theta)$ directly.

In a seminal paper, Aït-Sahalia (1996a) uses a test statistic of the form

$$\hat{M}_T(h) = \frac{Th}{T} \sum_{t=1}^T (\hat{f}(X_t) - f(X_t, \hat{\theta}_M))^2 = h \sum_{t=1}^T (\hat{f}(X_t) - f(X_t, \hat{\theta}_M))^2,$$

where $\hat{\theta}_M = \arg \min_{\theta \in \Theta} (1/T) \sum_{t=1}^T (\hat{f}(X_t) - f(X_t, \theta))^2$.

It then follows from (13) of Aït-Sahalia (1996a) that as $T \rightarrow \infty$

$$\hat{L}_{0T}(h) = \frac{\hat{M}_T(h) - \hat{\mu}_T(h)}{\sqrt{h} \hat{\sigma}_T(h)} \rightarrow_D N(0, 1) \tag{2.4}$$

under the β -mixing and some other conditions, where

$$\hat{\mu}_T(h) = R(k) \cdot \left(\frac{1}{T} \sum_{t=1}^T \hat{f}(X_t) \right) \quad \text{and} \quad \hat{\sigma}_T^2(h) = 2k^{(4)}(0) \cdot \left(\frac{1}{T} \sum_{t=1}^T \hat{f}^3(X_t) \right),$$

in which $R(k) = \int k^2(u) du < \infty$ and $k^{(j)}(0)$ denotes the j -times convolution product of $k(\cdot)$ given by

$$k^{(4)}(0) = \int_{-\infty}^{\infty} L^2(x) dx \quad \text{with} \quad L(x) = \int_{-\infty}^{\infty} k(y)k(x+y) dy. \tag{2.5}$$

The preceding test statistic is based on $\hat{f}(x) - f(x, \hat{\theta}_M)$, which measures directly the difference between $\hat{f}(x)$ and $f(x, \hat{\theta}_M)$. It can be shown that under H_0 ,

$$E[\hat{f}(x) - f(x, \hat{\theta}_M)]^2 = O(h^4).$$

This implies that it has the same order as the mean square error of $\hat{f}(x)$ if h is chosen to be $O(T^{-1/5})$. Thus, to obtain an asymptotically normal distribution with zero mean, h has to satisfy $\lim_{T \rightarrow \infty} Th^{4.5} = 0$ as required in Assumption A5 of Aït-Sahalia (1996a). This implies undersmoothing.

To reduce the bias and avoid undersmoothing, we propose a nonparametric estimator, $\tilde{f}(x, \tilde{\theta})$, of $f(x, \theta)$ of the form

$$\tilde{f}(x, \tilde{\theta}) = \sum_{t=1}^T w_t(x) f(X_t, \tilde{\theta}), \tag{2.6}$$

where $\tilde{\theta}$ is a consistent estimator of θ , $w_t(x) = w_t(x, h) = (1/T)k_h(x - X_t) \times [(s_2(x) - s_1(x)(x - X_t))/(s_2(x)s_0(x) - s_1^2(x))]$, and $s_r(x) = (1/T)\sum_{s=1}^T k_h(x - X_s)(x - X_s)^r$ for $r = 0, 1, 2$.

We also define

$$\tilde{f}(x) = \sum_{t=1}^T w_t(x) f(X_t).$$

If $\tilde{\theta}$ is a \sqrt{T} -consistent estimator of θ , then we have

$$E\{\tilde{f}(x, \tilde{\theta}) - \tilde{f}(x, \theta)\}^2 = O\left(\frac{1}{T}\right).$$

It follows from Fan and Gijbels (1996) that

$$E[\hat{f}(x) - f(x)] = \frac{1}{2} h^2 \sigma_k^2 f^{(2)}(x) + c_k f^{(3)}(\xi_1) h^3$$

and

$$E[\tilde{f}(x) - f(x)] = \frac{1}{2} h^2 \sigma_k^2 f^{(2)}(x) + d_k f^{(3)}(\xi_2) h^3,$$

provided that the first three derivatives of $f(x)$ exist, where ξ_1 and ξ_2 lie between x and h and x , c_k , and d_k are constants depending on functionals of $k(\cdot)$, and $\sigma_k^2 = \int x^2 k(x) dx$.

This implies that as $T \rightarrow \infty$

$$E[\hat{f}(x) - \tilde{f}(x)] = (c_k f^{(3)}(\xi_1) - d_k f^{(3)}(\xi_2)) h^3. \tag{2.7}$$

As can be seen from (2.7), the use of the difference $\hat{f}(x) - \tilde{f}(x, \tilde{\theta})$ can avoid undersmoothing. In other words, we can still assume $\limsup_{T \rightarrow \infty} Th^5 < \infty$.

Let us now establish our test statistic. We first have a look at the following distance function:

$$D(f, \theta) = \int (f(x) - f(x, \theta))^2 f(x) dx.$$

This naturally suggests estimating $D(f, \theta)$ by

$$D(\hat{f}, \tilde{\theta}) = \int (\hat{f}(x) - \tilde{f}(x, \tilde{\theta}))^2 \hat{f}(x) dx.$$

We then propose using a test statistic of the form

$$\hat{N}_T = \hat{N}_T(h) = Th \int (\hat{f}(x) - \tilde{f}(x, \tilde{\theta}))^2 \hat{f}(x) dx. \tag{2.8}$$

We now state the main results of this section. Their proofs are relegated to Appendix A.

THEOREM 2.1. (i) *Suppose that Assumptions A.1–A.5 in Appendix A hold. Then under \mathcal{H}_0 in (2.3) we have*

$$L_T(h) = \frac{\hat{N}_T(h) - \mu_0}{\sqrt{h}\sigma_0} \rightarrow_D N(0,1) \quad \text{as } T \rightarrow \infty,$$

where $\mu_0 = R(k) \int_{-\infty}^{\infty} f^2(x) dx$ and $\sigma_0^2 = 2k^{(4)}(0) \int f^4(x) dx$.

(ii) *Assume that the conditions of (i) hold. In addition, assume that there is a random data-driven \hat{h} such that $(\hat{h}/h) - 1 \rightarrow_p 0$ as $T \rightarrow \infty$. Then under \mathcal{H}_0 in (2.3) we have*

$$L_T(\hat{h}) = \frac{\hat{N}_T(\hat{h}) - \mu_0}{\sqrt{\hat{h}}\sigma_0} \rightarrow_D N(0,1) \quad \text{as } T \rightarrow \infty.$$

THEOREM 2.2. (i) *Suppose that Assumptions A.1–A.5 in Appendix A hold. Then under \mathcal{H}_0 in (2.3) we have*

$$\hat{L}_T(h) = \frac{\hat{N}_T(h) - \hat{\mu}_T(h)}{\sqrt{h}\hat{\sigma}_T(h)} \rightarrow_D N(0,1) \quad \text{as } T \rightarrow \infty,$$

where $\hat{\mu}_T(h)$ and $\hat{\sigma}_T(h)$ are as defined in (2.4).

(ii) *Assume that the conditions of (i) hold. In addition, assume that there is a random data-driven \hat{h} such that $(\hat{h}/h) - 1 \rightarrow_p 0$ as $T \rightarrow \infty$. Then under \mathcal{H}_0 in (2.3) we have*

$$\hat{L}_T(\hat{h}) = \frac{\hat{N}_T(\hat{h}) - \hat{\mu}_T(\hat{h})}{\sqrt{\hat{h}}\hat{\sigma}_T(\hat{h})} \rightarrow_D N(0,1) \quad \text{as } T \rightarrow \infty.$$

Remark 2.1. (i) Similar to \hat{L}_{0T} of (2.4), one may replace $\hat{N}_T(h)$ by

$$\tilde{N}_T(h) = \frac{Th}{T} \sum_{t=1}^T (\hat{f}(X_t) - \tilde{f}(X_t, \tilde{\theta}))^2 = h \sum_{t=1}^T (\hat{f}(X_t) - \tilde{f}(X_t, \tilde{\theta}))^2.$$

(ii) As can be seen from Theorem 2.2(i), we need to estimate both the asymptotic mean and variance of $\hat{N}_T(h)$ involved in practice. It is possible to avoid estimating this kind of unknown quantity by introducing a weight function into $\hat{N}_T(h)$. In both theory and practice, however, the asymptotic power of the test may depend on the choice of such a weight function. We therefore follow a

suggestion made by two of the referees and use the natural form $\hat{N}_T(h)$ to construct an adaptive test in Section 3.

(iii) Theorem 2.2(i) establishes an asymptotic normality test statistic. Theorem 2.2(ii) shows that the asymptotic normality remains unchanged when h is replaced with the random data-driven \hat{h} , which is known as the plug-in method. Fan and Gijbels (1996, pp. 152–154) have shown that the plug-in method has some advantages in applications. Whether the proposed test statistic $\hat{L}_T(h)$ is optimal has not been discussed. A modified form of the test statistic is shown to be optimal, and the detailed discussion is given in Section 3.

3. AN ADAPTIVE TEST PROCEDURE

Section 2 establishes the asymptotic normality of the test statistic for testing the marginal densities. The test statistic has nontrivial power only if C_T converges more slowly than $T^{-1/2}$. To improve the asymptotic power properties of the test, we consider extending the approach of Horowitz and Spokoiny (2001) for testing nonparametric regression functions. It is assumed that a marginal density function g belongs to a class of s -times ($s \geq 2$) differentiable density functions on R^1 , such as a Hölder, Sobolev, or Besov class, \mathcal{G} , which is separated from the null hypothesis by some distance C_T that converges to zero as $T \rightarrow \infty$. The objective of this section is to find the fastest rate at which C_T can approach zero while permitting consistent testing uniformly over \mathcal{G} . This rate is called the optimal rate of testing. A test is consistent uniformly over \mathcal{G} if

$$\lim_{T \rightarrow \infty} \inf_{g \in \mathcal{G}} P(\mathcal{H}_0 \text{ is rejected against } g) = 1. \tag{3.1}$$

Thus, the optimal rate of testing is the fastest rate at which C_T can approach zero while maintaining (3.1).

3.1. Asymptotic Behavior of the Test Statistic under the Null Hypothesis

As can be seen in Section 2, the proposed test statistic depends on the bandwidth. This section then suggests using

$$L^* = \max_{h \in H_T} \hat{L}_T(h) = \max_{h \in H_T} \frac{\hat{N}_T(h) - \hat{\mu}_T(h)}{\sqrt{h} \hat{\sigma}_T(h)}, \tag{3.2}$$

where $H_T = \{h = h_{\max} a^k : h \geq h_{\min}, k = 0, 1, 2, \dots\}$, in which $0 < h_{\min} < h_{\max}$ and $0 < a < 1$. Let J_T denote the number of elements of H_T . In this case, $J_T \leq \log_{1/a}(h_{\max}/h_{\min})$. Detailed conditions on h_{\min} and h_{\max} will be given in Assumption B.3 in Appendix B.

Simulation Scheme. We discuss how to obtain a critical value for L^* . The exact α -level critical value, l_α^* ($0 < \alpha < 1$), is the $1 - \alpha$ quantile of the exact finite sample distribution of L^* . Because θ_0 is unknown, l_α^* cannot be evaluated in practice. We therefore suggest choosing a simulated α -level critical value, l_α , by using the following simulation procedure.

1. For the simulation, we either use resamples of the sampled data X_t or generate the data X_t from the marginal density $f(x, \theta_0)$ or the corresponding transition density with an initial value of θ_0 under \mathcal{H}_0 .
2. The true value θ_0 is estimated based on the simulated $\{X_t\}$, and the resulting estimate is denoted by $\hat{\theta}$.
3. We choose H_T as specified following (3.2) with h_{\min} and h_{\max} satisfying Assumption B.3 in Appendix B and then compute L^* of (3.2) using the simulated $\{X_t\}$ and $\hat{\theta}$.
4. Repeat the preceding steps M times and produce M versions of L^* , L_m^* for $m = 1, 2, \dots, M$. The simulated critical value l_α is then the $(1 - \alpha)\%$ percentile of the M values, L_m^* for $m = 1, 2, \dots, M$, of L^* .

We now state the following result, and its proof is relegated to Appendix B.

THEOREM 3.1. *Assume that Assumptions A.1, A.3, and A.4 in Appendix A and B.1–B.3 listed in Appendix B hold. Then under \mathcal{H}_0*

$$\lim_{T \rightarrow \infty} P(L^* > l_\alpha) = \alpha.$$

The main result on the behavior of the test statistic L^* under \mathcal{H}_0 is that l_α is an asymptotically correct α -level critical value under any model in \mathcal{H}_0 .

3.2. Consistency against a Fixed Alternative

We now show that L^* is consistent against a fixed alternative model. Assume that model (1.1) holds. Let the parameter set Θ be an open subset of R^q . Let $\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta\}$ satisfy Assumption B.1 in Appendix B. For convenience, let

$$F(\theta) = (f(X_1, \theta), \dots, f(X_T, \theta))^\tau \quad \text{and} \quad \bar{f} = (f(X_1), \dots, f(X_T))^\tau.$$

Measure the distance between f and \mathcal{F} by the normalized l_2 distance

$$\rho(f, \mathcal{F}) = \left[\inf_{\theta \in \Theta} \left(\frac{1}{T} \|\bar{f} - F(\theta)\|^2 \right) \right]^{1/2},$$

where $\|\cdot\|$ denotes the euclidean norm. If \mathcal{H}_0 is false, then $\rho(f, \mathcal{F}) \geq C_\rho$ for all sufficiently large T and some $C_\rho > 0$. A consistent test will reject a false \mathcal{H}_0 with probability approaching one as $T \rightarrow \infty$.

The following theorem establishes a consistency result, and its proof is relegated to Appendix B.

THEOREM 3.2. *Assume that the conditions of Theorem 3.1 hold. In addition, if there is a $C_\rho > 0$ such that $\lim_{T \rightarrow \infty} P(\rho(f, \mathcal{F}) \geq C_\rho) = 1$ holds then*

$$\lim_{T \rightarrow \infty} P(L^* > l_\alpha) = 1.$$

3.3. Consistency against a Sequence of Local Alternatives

In this section, we consider the consistency of L^* under local alternatives of the form

$$f_T(x) = f(x, \theta_1) + C_T \Delta_T(x) \tag{3.3}$$

with $C_T \geq C_0 T^{-1/2} \sqrt{\log \log T}$ for some constant $C_0 > 0$ and $\theta_1 \in \Theta$.

Let

$$\bar{f}_T = (f_T(X_1), \dots, f_T(X_T))^T \quad \text{and} \quad \bar{\Delta}_T = (\Delta_T(X_1), \dots, \Delta_T(X_T))^T.$$

We now have that

$$\frac{1}{T} \|\bar{f}_T - F(\theta_1)\|^2 = \frac{C_T^2}{T} \|\bar{\Delta}_T\|^2 = \frac{C_T^2}{T} \sum_{i=1}^T |\Delta_T(X_i)|^2. \tag{3.4}$$

To ensure that the rate of convergence of \bar{f}_T to the parametric model $F(\theta_1)$ is the same as the rate of convergence of C_T to zero, in view of (3.4), we need to assume that $\Delta_T(x)$ is a continuous function that is normalized so that

$$\lim_{T \rightarrow \infty} P\left(\frac{1}{T} \sum_{i=1}^T |\Delta_T(X_i)|^2 \geq \delta\right) = 1 \tag{3.5}$$

for some $\delta > 0$. When $\Delta_T(\cdot)$ does not depend on T , condition (3.5) can be replaced by $E[\Delta^2(X_1)] > 0$, which holds automatically when $\Delta(\cdot) \neq 0$.

We now state the following consistency result, and its proof is relegated to Appendix B.

THEOREM 3.3. *Assume that Assumptions A.1, A.3, and A.4 in Appendix A and B.1–B.3 with $h_{\max} = c_{\max}(\log \log T)^{-1}$ for some constant $c_{\max} > 0$ in Appendix B hold. Let $\tilde{\theta}$ be a \sqrt{T} -consistent estimator of θ . Let f_T satisfy (3.3) with $C_T \geq CT^{-1/2} \sqrt{\log \log T}$ for some constant $C > 0$. In addition, let condition (3.5) hold. Then*

$$\lim_{T \rightarrow \infty} P(L^* > l_\alpha) = 1.$$

The result shows that the power of the adaptive, rate-optimal test approaches one as $T \rightarrow \infty$ for any function $\Delta_T(\cdot)$ and sequence $\{C_T\}$ that satisfy the conditions of Theorem 3.3.

3.4. Consistency against a Sequence of Smooth Alternatives

This section establishes that L^* is consistent uniformly over alternatives in a Hölder smoothness class whose distance from the parametric model approaches zero at the fastest possible rate. It can be shown that we can extend the results to Sobolev and Besov classes under more technical conditions.

Before specifying our smoothness classes, we introduce the following notation. Define the Hölder norm

$$\|f\|_{H,s} = \sup_{x \in S_f} \sum_{1 \leq j \leq s} \left| \frac{d^j f(x)}{dx^j} \right|, \tag{3.6}$$

where $S_f = \{x \in R^1 : f(x) > 0\}$.

The smoothness classes that we consider consist of functions $f \in S(H, s) \equiv \{f : \|f\|_{H,s} \leq c_H\}$ for some (unknown) $s \geq 2$ and $c_H < \infty$.

For some $s \geq 2$ and all sufficiently large $C_f < \infty$, define

$$B_{H,T} = \left\{ f \in S(H, s) : \lim_{T \rightarrow \infty} P(\rho(f, \mathcal{F}) \geq C_f (T^{-1} \sqrt{\log \log T})^{2s/(4s+1)}) = 1 \right\}, \tag{3.7}$$

where $\rho(f, \mathcal{F})$ is as defined in Section 3.2.

We now state the following consistency result, and its proof is relegated to Appendix B.

THEOREM 3.4. *Assume that Assumptions A.1, A.3, and A.4 in Appendix A and B.1–B.3 in Appendix B hold. Then for $0 < \alpha < 1$ and $B_{H,T}$ as defined in (3.7)*

$$\lim_{T \rightarrow \infty} \inf_{f \in B_{H,T}} P(L^* > l_\alpha) = 1.$$

Remark 3.1. Theorems 3.1–3.4 show that we have established some consistency results for the proposed test given in (3.2). Such consistency results correspond to Theorems 1–4 of Horowitz and Spokoiny (2001) for a fixed design regression case. In our case, we deal with the case where the observations are stationary and α -mixing time series. In addition, the optimum version L^* is asymptotically consistent as established in Theorem 3.2. This is one of the advantages of our test over existing ones, such as the natural competitor proposed in Ait-Sahalia (1996a). In Section 4, we show that our test also outperforms the natural competitor in the finite sample case.

4. EXAMPLE OF IMPLEMENTATION AND APPLICATION IN DIFFUSION MODELS

This section illustrates the proposed adaptive test by the following example. As the bootstrap simulation procedure for selecting both the bandwidth and simulated critical values is extremely computationally demanding, especially for large numbers of data, we only consider using the CIR model proposed by Cox et al. (1985) and show how to implement the adaptive test statistic L^* of (3.2) in practice through using a simulated example. The main reason for choosing the model is not only because both the marginal and transition density functions have closed forms but also because the model has been studied extensively in the literature. See, for example, Ait-Sahalia (1999) and Hong and Li (2004).

Example 4.1

We consider using the CIR model given by

$$dr_t = \kappa(\beta - r_t) dt + \sigma\sqrt{r_t} dB_t, \tag{4.1}$$

where $\kappa > 0$, $\beta > 0$, and $\sigma > 0$ are unknown parameters and B_t is standard Brownian motion. It can be shown that $\{r_t\}$ is distributed on $R^+ = (0, \infty)$ if $2\kappa\beta/\sigma^2 \geq 1$. Furthermore, it follows from Lemma 3.1 of Masry and Tjøstheim (1995) that the process $\{r_t\}$ satisfies Assumption A.1(i). Alternatively, one may apply Assumption A.3' of Ait-Sahalia (1996b, p. 552) to verify that $\{r_t\}$ is strictly stationary and α -mixing.

As a result of (2.2), the marginal density function of $\{r_t\}$ satisfying model (4.1) is

$$f(x, \theta) = \frac{(2\kappa)^{\nu+1}}{\Gamma(\nu+1)\sigma^{2(\nu+1)}} \cdot x^\nu \cdot \exp\left(-\frac{2\kappa}{\sigma^2}x\right), \quad x \in R^+ = (0, \infty), \tag{4.2}$$

where $\theta = (\beta, \kappa, \sigma)$, $\nu = 2\kappa\beta/\sigma^2 - 1$, and $\Gamma(\cdot)$ is the usual gamma function. Let θ_0 be the true value of θ .

To construct a sequence of local alternatives, we also consider using a marginal density of the form

$$f_1(x, \theta) = \frac{1}{\Gamma(2 + \nu_1)} (\nu_1\beta)^{2+\nu_1} \cdot x^{-(3+\nu_1)} \cdot \exp(-\nu_1\beta x^{-1}), \quad x \in R^+, \tag{4.3}$$

where $\nu_1 = 2\kappa/\sigma^2$. It is known that $f_1(x, \theta)$ is the marginal density of $\{r_t\}$ satisfying the AG model proposed in Ahn and Gao (1999)

$$dr_t = \kappa(\beta - r_t)r_t dt + \sigma r_t^{1.5} dB_t, \quad t = 1, 2, \dots, \tag{4.4}$$

with parameter values $\kappa > 0$, $\beta > 0$, and $\sigma > 0$. The necessary and sufficient conditions for stationarity and unattainability of 0 and ∞ in finite expected time are the pairs $\kappa > 0$ and $\beta > 0$ (see Ahn and Gao, 1999). To show that $\{r_t\}$ is

strictly stationary and α -mixing, as explained in Appendix A of Ahn and Gao (1999, pp. 755–756), one needs only to verify Assumption A.3' of Ait-Sahalia (1996b, p. 552). It is easy to see that such an assumption holds for the marginal density, drift, and diffusion functions given in (4.3) and (4.4).

The corresponding structure of the test problem (2.3) for this example can be constructed as

$$\mathcal{H}_0: f(x) = f(x, \theta_0) \quad \text{versus} \quad \mathcal{H}_1: f_T(x) = f(x, \theta_1) + C_T \Delta_T(x), \tag{4.5}$$

where

$$C_T = (T^{-1} \sqrt{\log \log T})^{4/9} \quad \text{and} \quad \Delta_T(x) = f_1(x, \theta_1) - f(x, \theta_1), \tag{4.6}$$

in which $\theta_1 \in \Theta$. The reason for choosing such $\Delta_T(\cdot)$ as the local shift function is to ensure that the models under \mathcal{H}_1 fluctuate closely around those under \mathcal{H}_0 . The choice of (4.5) and (4.6) ensures that (3.7) holds with $s = 2$. This implies that the adaptive test is consistent against the sequence with an optimal rate. Note that Assumptions B.1 and B.2 hold.

In the following simulation, we consider using a class of alternatives of the form

$$f_\psi(x, \theta_1) = f(x, \theta_1) + \psi \cdot (f_1(x, \theta_1) - f(x, \theta_1)) = (1 - \psi)f(x, \theta_1) + \psi f_1(x, \theta_1), \tag{4.7}$$

where $\theta_1 \in \Theta$ and $0 < \psi < 1$ is defined as the truncation parameter to be chosen.

To compute the nonparametric estimators involved, we choose the normal kernel function given by

$$k(x) = \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} \tag{4.8}$$

throughout the simulation. Observe that Assumptions A.1–A.4 hold. For the CIR and AG models, we simulate the data from their marginal density and transitional functions, which all have closed forms.

In the detailed simulation, we simulate the data from (4.2) for the CIR model, (4.3) for the AG model, and then (4.7) under \mathcal{H}_1 . Using the simulated data, we compute

$$L^* = \max_{h \in H_T} \hat{L}_T(h) = \sup_{h \in H_T} \frac{\hat{N}_T(h) - \hat{\mu}_T(h)}{\sqrt{h \hat{\sigma}_T(h)}}, \tag{4.9}$$

in which $R(k) = 1/2\sqrt{\pi}$ and $k^{(4)}(0) = 1/2\sqrt{2\pi}$ are used after the choice of (4.8) and H_T is as defined following (3.2) with $h_{\min} = T^{-(11/36)}$, $h_{\max} = 2(\log \log T)^{-1}$, and $a = \frac{35}{36}$. Note that Assumption B.3 holds.

To compare L^* with $\hat{L}_{0T}(h)$ in (2.4), we construct a test statistic of the form

$$L_0^* = \hat{L}_{0T}(h_*) = \frac{\hat{M}_T(h_*) - \hat{\mu}_T(h_*)}{\sqrt{h_* \hat{\sigma}_T(h_*)}}, \tag{4.10}$$

where h_* is chosen by using the following procedure.

- We simulate X_t with probability $1 - \psi$ from the CIR model and with probability ψ from the AG model with an initial value of θ_1 under \mathcal{H}_1 .
- Use the simulated data $\{X_t : t = 1, 2, \dots, T\}$ to estimate θ_1 .
- Compute the resulting function of h given by

$$\hat{L}_{0T}(h) = \frac{\hat{M}_T(h) - \hat{\mu}_T(h)}{\sqrt{h \hat{\sigma}_T(h)}}.$$

- Repeat the preceding steps $Q = 1,000$ times and produce Q versions of $\hat{L}_{0T}(h)$ denoted by $\hat{L}_{0T,m}(h)$ for $m = 1, 2, \dots, Q$. Use the Q functions of h , $\hat{L}_{0T,m}(h)$ for $m = 1, 2, \dots, Q$, to construct their empirical bootstrap distribution function, that is,

$$F_h(u) = \frac{1}{Q} \sum_{m=1}^Q I(\hat{L}_{0T,m}(h) \leq u),$$

where $I(U \leq u)$ is the usual indicator function.

- For a given asymptotic critical value ecv_α at the level α (e.g., $ecv_{0.005} = 1.645$ at the 5% level), we then calculate the following power function:

$$\psi(h) = 1 - F_h(ecv_\alpha).$$

- Find approximately at which h value the power function $\psi(h)$ is maximized. Denote the maximizer by h_* .

We then consider using the same choice of the parameter values as in (17) of Pritsker (1998). This means that the baseline model is model (4.1) with $\kappa = 0.89218$, $\beta = 0.090495$, and $\sigma = \sqrt{0.032742}$. In this example, the same parameter values were also used as θ_1 in computing the power of the tests L^* and L_0^* . The truncation parameter was chosen as $\psi = 0$ under \mathcal{H}_0 , whereas the truncation parameter was chosen as $\psi = \frac{1}{33}$ under \mathcal{H}_1 . Three different sizes of sample $T = 1,000, 2,755$, or $5,500$ were then considered. The corresponding simulated critical values, l_α and $l_{0\alpha}$, of L^* and L_0^* at the α level are then found by using the simulation scheme proposed in Section 3.1. The sizes of the tests were then computed based on the data simulated under \mathcal{H}_0 , and the power values of the tests were calculated based on the data generated under \mathcal{H}_1 . In implementing the simulation procedure, we used $M = 1,000$ involved in the simulation

scheme proposed in Section 3.1. The number of simulations in producing Table 1 was also 1,000. Both the size and the power of L^* and L_0^* are given in Table 1.

Remark 4.1. (i) As can be seen from Table 1, the power values of both L^* and L_0^* look reasonable when $\psi = \frac{1}{33}$, or about 3%. This may show that both L^* and L_0^* are practically applicable to the medium sample case, because the difference between the null hypothesis and its alternative was made deliberately close. We also computed the power of the tests for the case where $\psi = \frac{1}{20}$, or 5%. Our small sample results showed that the power of L^* was already 100% even when $T = 1,000$. In general, it is true that the power increases as ψ increases for each case. Observe that L^* is slightly more powerful than L_0^* , although h_* involved in $L_0^* = \hat{L}_{0T}(h_*)$ has been chosen based on the assessment of its power. We observe that the sizes of the two tests are also close to either 5% in the first half of Table 1 or 1% in the second half of Table 1.

(ii) We also examined the dependence of the power on the choice of the initial parameter values. Our experience suggests that the power of the tests mainly depends on the choice of the truncation parameter ψ . This is both understandable and expected, because the test statistics finally depend only on the estimation and reestimation procedure of the vector of the initial parameters rather than the initial parameter values themselves. This is probably why artificial values or parameter values estimated from a set of real data are used as initial values for starting a simulation procedure. For example, Hong and Li (2004) use the parameter values estimated from the U.S. interest rate series for their simulation procedure.

(iii) Compared with existing results (see Pritsker, 1998), both the size and power of L_0^* have been significantly improved. This is probably because (a) the choice of h involved in $\hat{L}_{0T}(h)$ is based on the assessment of the power of $\hat{L}_{0T}(h)$ rather than using an estimation-based optimal value and (b) to avoid using the

TABLE 1. Rejection rates for the marginal density tests

Observation	Null hypothesis is true		Null hypothesis is false	
	L^*	L_0^*	L^*	L_0^*
<i>The 5% level</i>				
1,000	0.044	0.039	0.721	0.618
2,755	0.055	0.058	0.874	0.712
5,500	0.052	0.042	0.992	0.887
<i>The 1% level</i>				
1,000	0.007	0.017	0.526	0.497
2,755	0.014	0.019	0.673	0.532
5,500	0.009	0.006	0.869	0.795

asymptotic distribution of $\hat{L}_{0T}(h)$ and then an asymptotic critical value of 1.645 at the 5% or 2.33 at the 1% level, we have used the bootstrap-based simulated critical value, $l_{0\alpha}$, at the level α . We also computed both the power and size values for the case where h was chosen by using a cross-validation criterion, and the resulting sizes and power values were similar to those obtained by Pritsker (1998), although L^* always performed better than L_0^* . This further demonstrates that the asymptotic distribution of either $\hat{L}_T(h)$ or $\hat{L}_{0T}(h)$ can only provide some kind of idea about the asymptotic behavior. In practice, we strongly suggest using the proposed bootstrap simulation procedure for choosing a simulated critical value rather than an asymptotic critical value.

5. CONCLUSION

In this paper, we have considered testing the general continuous-time diffusion model (1.1) under the α -mixing condition. The results for continuous-time models under the α -mixing condition complement some existing results under the β -mixing condition. See, for example, Ait-Sahalia (1996a). Moreover, an adaptive and optimal test procedure has been established. This extension corresponds to Horowitz and Spokoiny (2001) for the fixed design nonparametric regression and then to Chen, Gao, and Li (2001) for a nonparametric time series regression model. To deal with the α -mixing condition, we have established some novel results for moment inequalities (see Lemma C.2) and limit theorems (see Lemma A.1) for degenerate U -statistics of strongly dependent processes. Both Lemmas A.1 and C.2 are applicable to some other estimation and testing of diffusion processes with the α -mixing condition (for more about various mixing conditions, see Doukhan, 1995). In addition, we have demonstrated how to implement the proposed test procedure in practice through using a simulated example.

The results given in this paper can be extended in a number of directions. First, it is possible to consider testing for both the marginal and transition density functions simultaneously, because the transition density can capture the full dynamics of a diffusion process and, in particular, can distinguish the diffusion processes that have the same marginal density but different transition densities. Second, the results of this paper for the short-range dependent continuous-time case can be extended to the long-range dependent continuous-time case. Third, one probably can relax the strict stationarity and the mixing condition, as the recent work by Ait-Sahalia (1999) and Karlsen and Tjøstheim (2001) indicates that it is possible to do such work without the stationarity and the mixing condition. This part is particularly important for two reasons: (i) for the long-range dependent case one needs to avoid assuming both the long-range dependence and the mixing condition, as they contradict each other; and (ii) some important models are nonstationary. These are some issues left for future research.

REFERENCES

- Ahn, D.H. & B. Gao (1999) A parametric nonlinear model of term structure dynamics. *Review of Financial Studies* 12, 721–762.
- Ait-Sahalia, Y. (1996a) Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9, 385–426.
- Ait-Sahalia, Y. (1996b) Nonparametric pricing of interest rate derivative securities. *Econometrica* 64, 527–560.
- Ait-Sahalia, Y. (1999) Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance* 54, 1361–1395.
- Black, F. & M. Scholes (1973) The pricing of options and corporate liabilities. *Journal of Political Economy* 3, 637–654.
- Chapman, D. & N. Pearson (2000) Is the short rate drift actually nonlinear? *Journal of Finance* 54, 355–388.
- Chen, S., J. Gao, & M. Li (2001) Simultaneous Specification Tests for Nonparametric Regression with Application to Diffusion Model Testing. Working paper, Department of Statistics and Applied Probability, National University of Singapore. Available at www.stat.nus.edu.sg/~stacsx.
- Cox, J., E. Ingersoll, & S. Ross (1985) An intertemporal general equilibrium model of asset prices. *Econometrica* 53, 363–384.
- Doukhan, P. (1995) *Mixing-Properties and Examples*. Lecture Notes in Statistics. Springer-Verlag.
- Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. Chapman and Hall.
- Fan, J. & C. Zhang (2003) A re-examination of Stanton's diffusion estimation with applications to financial model validation. *Journal of the American Statistical Association* 457, 118–134.
- Fan, Y. & Q. Li (1998) Central limit theorem for degenerate U -statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics* 10, 245–271.
- Gao, J. & M. King (2001) Estimation and Model Specification Testing in Nonparametric and Semiparametric Regression. Working paper, School of Mathematics and Statistics, the University of Western Australia, Australia. Available at www.maths.uwa.edu.au/~jiti/jems.pdf.
- Gao, J., H. Tong, & R. Wolff (2002) Model specification tests in nonparametric stochastic regression models. *Journal of Multivariate Analysis* 83, 324–359.
- Hjellvik, V., Q. Yao, & D. Tjøstheim (1996) Linearity testing using local polynomial approximation. Discussion paper 60, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin, Spandauerstr. 1, 10178, Berlin.
- Hong, Y. & H. Li (2004) Nonparametric specification testing for continuous-time models with application to spot interest rates. *Review of Financial Studies* 17, forthcoming.
- Horowitz, J. & V. Spokoiny (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599–632.
- Jiang, G. & J. Knight (1997) A nonparametric approach to the estimation of diffusion processes with an application to a short-term interest rate model. *Econometric Theory* 13, 615–645.
- Karlsen, H. & D. Tjøstheim (2001) Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372–416.
- Masry, E. & D. Tjøstheim (1995) Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* 11, 258–289.
- Pritsker, M. (1998) Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies* 11, 449–487.
- Roussas, G. & D. Ioannides (1987) Moment inequalities for mixing sequences of random variables. *Stochastic Analysis and Applications* 5, 61–120.
- Stanton, R. (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52, 1973–2002.
- Sundaresan, S. (2001) Continuous-time methods in finance: A review and an assessment. *Journal of Finance* 55, 1569–1622.

APPENDIX A

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

A.1. Assumptions. Let the parameter set Θ be an open subset of R^q . Let $\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta\}$. Define $\nabla_{\theta} f(x, \theta) = \partial f(x, \theta) / \partial \theta$, $\nabla_{\theta}^2 f(x, \theta) = \partial^2 f(x, \theta) / \partial \theta \partial \theta'$, and $\nabla_{\theta}^3 f(x, \theta) = \partial^3 f(x, \theta) / \partial \theta \partial \theta' \partial \theta''$ whenever these derivatives exist. For any $q \times q$ matrix D , define

$$\|D\|_{\infty} = \sup_{v \in R^q} \frac{\|Dv\|}{\|v\|},$$

where $\|v\|^2 = \sum_{i=1}^q v_i^2$ for $v = (v_1, \dots, v_q)'$.

Assumption A.1. (i) Assume that the process $\{r_t\}$ is strictly stationary and α -mixing with the mixing coefficient $\alpha(t) = C_{\alpha} \alpha^t$ defined by

$$\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Omega_s^{\sigma}, B \in \Omega_{s+t}^{\sigma}\}$$

for all $s, t \geq 1$, where $0 < C_{\alpha} < \infty$ and $0 < \alpha < 1$ are constants and Ω_t^{σ} denotes the σ -field generated by $\{r_i : i \leq t\}$.

(ii) Assume that the univariate kernel function $k(\cdot)$ is nonnegative, symmetric, and four-times differentiable on $R^1 = (-\infty, \infty)$. In addition, $\int_{-\infty}^{\infty} x^2 k(x) dx < \infty$ and $\int_{-\infty}^{\infty} k^2(x) dx < \infty$.

Assumption A.2. (i) The parameter space $\Theta \subset R^q$ is compact. In a neighborhood of the true parameter θ_0 , $f(x, \theta)$ is twice continuously differentiable in θ ; $E[(\partial f(x, \theta) / \partial \theta)(\partial f(x, \theta) / \partial \theta)']$ is of full rank. In addition, assume that $G(x)$ is a positive and integrable function with $E[G(X_t)] < \infty$ uniformly in $t \geq 1$ such that $\sup_{\theta \in \Theta} |f(X_t, \theta)|^2 \leq G(X_t)$ and $\sup_{\theta \in \Theta} \|\nabla_{\theta}^j f(X_t, \theta)\|^2 \leq G(X_t)$ for $j = 1, 2, 3$, where for $B = \{b_{ij} : 1 \leq i, j \leq q\}$, $\|B\|^2 = \sum_{i=1}^q \sum_{j=1}^q b_{ij}^2$.

(ii) Assume that $\tilde{\theta}$ is a \sqrt{T} -consistent estimator of θ_0 .

Assumption A.3. For every $\theta \in \Theta$:

(i) The drift and the diffusion functions are three times continuously differentiable in $x \in R^+ = (0, \infty)$, and $\sigma > 0$ on R^+ .

(ii) The integral of $\bar{\mu}(v, \theta) = [1/\sigma^2(v, \theta)] \exp(-\int_v^{\bar{v}} 2[\mu(x, \theta)/\sigma^2(x, \theta)] dx)$ converges at both boundaries of D , where \bar{v} is fixed in D .

(iii) The integral of $s(v, \theta) = \exp(\int_v^{\bar{v}} 2[\mu(x, \theta)/\sigma^2(x, \theta)] dx)$ diverges at both boundaries of D .

Assumption A.4. (i) Assume that the first three derivatives of $f(x)$ are continuous on D and that $f(x) > c_f > 0$ on the interior of D for some $c_f > 0$. In addition, both $f(x)$ and $f^2(x)$ are integrable on D .

(ii) The initial random variable r_0 is distributed as $f(x)$.

(iii) The true drift and diffusion functions satisfy Assumption A.3.

Assumption A.5. The bandwidth parameter h satisfies that

$$\lim_{T \rightarrow \infty} h = 0, \quad \lim_{T \rightarrow \infty} Th^2 = \infty, \quad \text{and} \quad \limsup_{T \rightarrow \infty} Th^5 < \infty.$$

Remark A.1. Assumptions A.1–A.4 are quite natural in this kind of problem. Assumptions A.2–A.4 correspond to Assumptions A0, A1, and A3 of Ait-Sahalia (1996a). Assumption A.1 is the exception. Assumption A.1(i) assumes the α -mixing condition, which is weaker than the β -mixing condition. Assumption A.1(ii) is quite general, allowing the use of the standard normal kernel. Assumption A.5 ensures that the theoretically optimum value of $h_{\text{optimal}} = CT^{-1/5}$ can be included. This is important, because there may be cases in which h_{optimal} is also optimal for testing purposes.

A.2. Technical Lemmas. The following lemmas are necessary for the proof of the main results stated in Section 2.

LEMMA A.1. Let ξ_t be an r -dimensional strictly stationary and strong mixing (α -mixing) stochastic process. Let $\phi(\cdot, \cdot)$ be a symmetric Borel function defined on $R^r \times R^r$. Assume that for any fixed $x \in R^r$, $E[\phi(\xi_1, x)] = 0$ and $E[\phi(\xi_i, \xi_j) | \Omega_0^{i-1}] = 0$ for any $i < j$, where Ω_i^j denotes the σ -field generated by $\{\xi_s : i \leq s \leq j\}$. Let $\phi_{st} = \phi(\xi_s, \xi_t)$, $\sigma_{st}^2 = \text{var}(\phi_{st})$, and $\sigma_T^2 = \sum_{1 \leq s < t \leq T} \sigma_{st}^2$. For some small constant $0 < \delta < 1$, let

$$M_{T1} = \max_{1 \leq i < j < k \leq T} \max \left\{ E|\phi_{ik}\phi_{jk}|^{1+\delta}, \int |\phi_{ik}\phi_{jk}|^{1+\delta} dP(\xi_i) dP(\xi_j, \xi_k) \right\},$$

$$M_{T21} = \max_{1 \leq i < j < k \leq T} \max \left\{ E|\phi_{ik}\phi_{jk}|^{2(1+\delta)}, \int |\phi_{ik}\phi_{jk}|^{2(1+\delta)} dP(\xi_i) dP(\xi_j, \xi_k) \right\},$$

$$M_{T22} = \max_{1 \leq i < j < k \leq T} \max \left\{ \int |\phi_{ik}\phi_{jk}|^{2(1+\delta)} dP(\xi_i, \xi_j) dP(\xi_k), \right. \\ \left. \int |\phi_{ik}\phi_{jk}|^{2(1+\delta)} dP(\xi_i) dP(\xi_j) dP(\xi_k) \right\},$$

$$M_{T3} = \max_{1 \leq i < j < k \leq T} E|\phi_{ik}\phi_{jk}|^2, \quad M_{T4} = \max_{\substack{1 < i, j, k \leq 2T \\ i, j, k \text{ different}}} \left\{ \max_P \int |\phi_{i_1}\phi_{j_k}|^{2(1+\delta)} dP \right\},$$

where the maximization over P in the equation for M_{T4} is taken over the four probability measures $P(\xi_1, \xi_i, \xi_j, \xi_k)$, $P(\xi_1)P(\xi_i, \xi_j, \xi_k)$, $P(\xi_1)P(\xi_{i_1})P(\xi_{i_2}, \xi_{i_3})$, and $P(\xi_1)P(\xi_i)P(\xi_j)P(\xi_k)$, where (i_1, i_2, i_3) is the permutation of (i, j, k) in ascending order;

$$M_{T51} = \max_{1 \leq i < j < k \leq T} \max \left\{ E \left| \int \phi_{ik}\phi_{jk}\phi_{ik}\phi_{jk} dP(\xi_i) \right|^{2(1+\delta)} \right\},$$

$$M_{T52} = \max_{1 \leq i < j < k \leq T} \max \left\{ \int \left| \int \phi_{ik}\phi_{jk}\phi_{ik}\phi_{jk} dP(\xi_i) \right|^{2(1+\delta)} dP(\xi_j) dP(\xi_k) \right\},$$

$$M_{T6} = \max_{1 \leq i < j < k \leq T} E \left| \int \phi_{ik}\phi_{jk} dP(\xi_i) \right|^2.$$

Assume that all the M'_T 's are finite. Let

$$M_T = \max\{T^2 M_{T1}^{1/(1+\delta)}, T^2 M_{T51}^{1/[2(1+\delta)]}, T^2 M_{T52}^{1/[2(1+\delta)]}, T^2 M_{T6}^{1/2}\},$$

$$N_T = \max\{T^{3/2} M_{T21}^{1/[2(1+\delta)]}, T^{3/2} M_{T22}^{1/[2(1+\delta)]}, T^{3/2} M_{T3}^{1/2}, T^{3/2} M_{T4}^{1/[2(1+\delta)]}\},$$

If $\lim_{T \rightarrow \infty} (\max\{M_T, N_T\}/\sigma_T^2) = 0$, then

$$\frac{1}{\sigma_T} \sum_{1 \leq s < t \leq T} \phi(\xi_s, \xi_t) \rightarrow_D N(0, 1) \quad \text{as } T \rightarrow \infty.$$

Remark A.2. Lemma A.1 establishes central limit theorems for degenerate U -statistics of strongly dependent processes. It should be pointed out that the conclusion of Lemma A.1 remains true when the martingale assumption that $E[\phi(\xi_i, \xi_j)|\Omega_0^{j-1}] = 0$ for any $i < j$ is removed. Such a martingale assumption is used only for a direct application of an existing central limit theorem (CLT) for martingales. Without such a condition, one needs only to decompose

$$\begin{aligned} \phi(\xi_i, \xi_j) &= \phi(\xi_i, \xi_j) - E[\phi(\xi_i, \xi_j)|\Omega_0^{j-1}] + E[\phi(\xi_i, \xi_j)|\Omega_0^{j-k}] \\ &\equiv \tilde{\phi}(\xi_i, \xi_j) + E[\phi(\xi_i, \xi_j)|\Omega_0^{j-1}] \end{aligned}$$

and then apply the martingale CLT to $\tilde{\phi}(\xi, \xi_j)$. Using the condition that $E[\phi(\xi_1, x)] = 0$ for each given x , one can show that the terms involving $E[\phi(\xi_i, \xi_j)|\Omega_0^{j-1}]$ are negligible (see Roussas and Ioannides, 1987, Theorem 5.5). Thus, as assumed in Lemma 3.2 of Hjellvik, Yao, and Tjøstheim (1996) and Theorem 2.1 of Fan and Li (1998), the condition that $E[\phi(\xi_1, x)] = 0$ for each given x is the key assumption.

Proof. Let

$$U_T = \sum_{1 \leq s < t \leq T} \phi(\xi_s, \xi_t), \quad V_t = \sum_{s=1}^{t-1} \phi(\xi_s, \xi_t),$$

$$\text{var}(U_T) = \sum_{t=2}^T \text{var}(V_t) = \sum_{t=2}^T E[V_t^2] = \sum_{1 \leq s < t \leq T} \sigma_{st}^2 + 2 \sum_{1 \leq i < j < k \leq T} E[\phi_{ik} \phi_{jk}].$$

To prove Lemma A.1, it suffices to show that as $T \rightarrow \infty$

$$\sigma_T^{-2} \sum_{t=2}^T V_t^2 \rightarrow_p 1 \tag{A.1}$$

and

$$\sigma_T^{-4} \sum_{t=2}^T E[V_t^4] \rightarrow 0. \tag{A.2}$$

By Lemma C.1 (with $\eta_1 = \phi_{ik}$, $\eta_2 = \phi_{jk}$, $l = 2$, $p_i = 2(1 + \delta)$, and $Q = 1/(1 + \delta)$),

$$E|\phi_{ik} \phi_{jk}| \leq 10M_{T1}^{1/(1+\delta)} \alpha^{\delta/(1+\delta)} (j - i).$$

Therefore,

$$\sum_{1 \leq i < j < k \leq T} E|\phi_{ik}\phi_{jk}| \leq 10T^2M_{T1}^{1/(1+\delta)} \sum_{i=1}^T \left(1 - \frac{i}{T}\right) \alpha^{\delta/(1+\delta)}(i) \leq CT^2M_{T1}^{1/(1+\delta)},$$

because $\sum_{i=1}^{\infty} \alpha^{\delta/(1+\delta)}(i) < \infty$.

Observe that

$$\begin{aligned} E\left[\sum_{t=2}^T V_t^2 - \sigma_T^2\right]^2 &= E\left\{\sum_{1 \leq s < t \leq T} [\phi_{st}^2 - \sigma_{st}^2] + 2 \sum_{1 \leq i < j < k \leq T} \phi_{ik}\phi_{jk}\right\}^2 \\ &\leq 2E\left\{\sum_{1 \leq s < t \leq T} [\phi_{st}^2 - \sigma_{st}^2]\right\}^2 + 8E\left\{\sum_{1 \leq i < j < k \leq T} \phi_{ik}\phi_{jk}\right\}^2. \end{aligned} \tag{A.3}$$

Let $\eta_{ijk} = \frac{1}{3}(\phi_{ik}\phi_{jk} + \phi_{ij}\phi_{kj} + \phi_{ji}\phi_{ki})$ and $\eta_{ij} = \frac{1}{3} \int \phi_{ik}\phi_{jk} dP(\xi_k)$.

Then by Lemma C.2(i) in Appendix C,

$$\begin{aligned} E\left\{\sum_{1 \leq i < j < k \leq T} \phi_{ik}\phi_{jk}\right\}^2 &= E\left\{\sum_{1 \leq i < j < k \leq T} \eta_{ijk}\right\}^2 \\ &\leq 2E\left\{\sum_{1 \leq i < j < k \leq T} [\eta_{ijk} - \eta_{ij} - \eta_{ik} - \eta_{jk}]\right\}^2 \\ &\quad + 8T^2E\left\{\sum_{1 \leq s < t \leq T} \eta_{st}\right\}^2 \\ &\leq C\{T^3M_{T21}^{1/(1+\delta)} + T^3M_{T21}^{1/(1+\delta)} + T^3M_{T3} + T^4M_{T51}^{1/(1+\delta)} \\ &\quad + T^4M_{T52}^{1/(1+\delta)} + T^4M_{T6}\} = o(\sigma_T^4). \end{aligned} \tag{A.4}$$

Let $C_\phi = \int \phi_{12}^2 \phi_{34}^2 dP(\xi_1) dP(\xi_2) dP(\xi_3) dP(\xi_4)$, where $P(\xi)$ denotes the probability measure of ξ .

Using Lemma C.1 repeatedly, we have that for different i, j, k, l

$$\begin{aligned} |E[\phi_{ij}^2 \phi_{kl}^2] - C_\phi| &\leq 10\{\alpha(\Delta(i, j, k, l))\}^{1-[1/(1+\delta)]} M_{T4}^{1/(1+\delta)} \\ &= 10M_{T4}^{1/(1+\delta)} \{\alpha(\Delta(i, j, k, l))\}^{\delta/(1+\delta)}, \end{aligned} \tag{A.5}$$

where $\Delta(i, j, k, l)$ is the minimum increment in the sequence that is the permutation of i, j, k, l in ascending order.

Similar to (A.5), one can have for all different i, j, k, l

$$|\sigma_{ij}^2 \sigma_{kl}^2 - C_\phi| \leq 10M_{T4}^{1/(1+\delta)} \{\alpha(\Delta(i, j, k, l))\}^{\delta/(1+\delta)}.$$

Therefore,

$$\begin{aligned} E\left\{\sum_{1 \leq i < j \leq T} (\phi_{ij}^2 - \sigma_{ij}^2)\right\}^2 &= \sum_{i < j, k < l, i, j, k, l \text{ different}} |E(\phi_{ij}^2 \phi_{kl}^2) - \sigma_{ij}^2 \sigma_{kl}^2| + O(T^3M_{T3}^{1/(1+\delta)}) \\ &\leq 2\left\{\sum_{i < j < k < l} + \sum_{i < k < l < j} + \sum_{i < k < j < l}\right\} \\ &\quad \times [|E[\phi_{ij}^2 \phi_{kl}^2] - C_\phi| + |\sigma_{ij}^2 \sigma_{kl}^2 - C_\phi|] + O(T^3M_{T3}) \\ &\leq C_1 T^3 M_{T4}^{1/(1+\delta)} + O(T^3 M_{T3}) = o(\sigma_T^4). \end{aligned} \tag{A.6}$$

It now follows from (A.3)–(A.5) that for any $\epsilon > 0$

$$P \left\{ \left| \frac{1}{\sigma_T^2} \sum_{k=2}^T V_k^2 - 1 \right| \geq \epsilon \right\} \leq \frac{1}{\sigma_T^4 \epsilon^2} E \left[\sum_{k=2}^T V_k^2 - \sigma_T^2 \right] \rightarrow 0.$$

Thus, (A.1) holds.

Note that for $2 \leq k \leq T$,

$$\begin{aligned} E[V_k^4] &= E \left\{ \sum_{i=1}^{k-1} \phi_{ik}^2 + 2 \sum_{1 \leq i < j < k} \phi_{ik} \phi_{jk} \right\}^2 \\ &= E \left\{ \sum_{i=1}^k \phi_{ik}^4 + 6 \sum_{1 \leq i < j < k} \phi_{ik}^2 \phi_{jk}^2 + 4 \sum_{l=1}^{k-1} \sum_{1 \leq i < j < k} \phi_{ik}^2 \phi_{ik} \phi_{jk} \right\} \\ &\quad + 4E \left\{ \sum_{1 \leq i < j < k, 1 \leq s < t < k, (i,j) \neq (s,t)} \phi_{ik} \phi_{jk} \phi_{sk} \phi_{tk} \right\} \\ &= 4 \sum_{l=1}^{k-1} \sum_{1 \leq i < j < k} E[\phi_{ik}^2 \phi_{ik} \phi_{jk}] + 4 \sum_{1 \leq i < j < k, 1 \leq s < t < k, (i,j) \neq (s,t)} E[\phi_{ik} \phi_{jk} \phi_{sk} \phi_{tk}] \\ &\quad + O(T^2 M_{T3}). \end{aligned} \tag{A.7}$$

It is easy to see that

$$\int |\phi_{ik} \phi_{jk} \phi_{sk} \phi_{tk}|^{1+\delta} dP \leq \left\{ \int |\phi_{ik} \phi_{jk}|^{2(1+\delta)} dP \int |\phi_{sk} \phi_{tk}|^{2(1+\delta)} dP \right\}^{1/2} \leq M_{T4}.$$

Similar to (A.5), one can have for any $(i, j) \neq (s, t)$,

$$|E[\phi_{ik} \phi_{jk} \phi_{sk} \phi_{tk}]| \leq 10M_{T4}^{1/(1+\delta)} \{\alpha(\Delta(i, j, s, t))\}^{\delta/(1+\delta)},$$

where $\Delta(\cdot)$ is as defined in (A.5).

Consequently, the first two terms on the right-hand side of (A.7) are of order $O(T^3 M_{T4}^{1/(1+\delta)})$, because $\sum_{k=1}^\infty k^2 \{\alpha(k)\}^{\delta/(1+\delta)} < \infty$.

Thus, (A.2) follows from

$$\sum_{k=2}^T E[V_k^4] = O(T^3 M_{T4}^{1/(1+\delta)}) = o(\sigma_T^4). \tag{A.8}$$

This finishes the proof. ■

Before stating the following lemmas, we define the following notation.

$$\epsilon_t(x) = k((x - X_t)/h) - E[k((x - X_t)/h)],$$

$$\phi_{st} = (Th)^{-1} \int \epsilon_s(x) \epsilon_t(x) f(x) dx,$$

$$w_t(x) = \frac{1}{T} k_h(x - X_t) \frac{s_2(x) - s_1(x)(x - X_t)}{s_2(x)s_0(x) - s_1^2(x)},$$

$$s_r(x) = \frac{1}{T} \sum_{s=1}^T k_h(x - X_s)(x - X_s)^r, \quad r = 0, 1, 2,$$

$$\lambda_t(\theta) = \lambda(X_t, \theta) = f(X_t) - f(X_t, \theta) = f(X_t, \theta_0) - f(X_t, \theta),$$

$$\lambda(\theta) = (\lambda_1(\theta), \dots, \lambda_T(\theta))^T,$$

$$a_{st} = Th \int w_s(x) w_t(x) f(x) dx, \quad b_{st} = \int \epsilon_s(x) w_t(x) f(x) dx,$$

$$N_{0T} = N_{0T}(h) = \sum_{s=1}^T \sum_{t=1}^T \phi_{st},$$

$$\begin{aligned} N_T(h) &= N_T(h, \theta) = (Th) \int \{\hat{f}(x) - \tilde{f}(x, \theta)\}^2 f(x) dx \\ &= N_{0T}(h) + Q_T(\theta) + \Pi_T(\theta) + P_{1T} + P_{2T} + R_T(\theta), \end{aligned}$$

using

$$\begin{aligned} \hat{f}(x) - \tilde{f}(x, \theta) &= \hat{f}(x) - E[\hat{f}(x)] + E[\hat{f}(x)] - E[\tilde{f}(x)] \\ &\quad + E[\tilde{f}(x)] - \tilde{f}(x) + \tilde{f}(x) - \tilde{f}(x, \theta), \end{aligned}$$

where

$$Q_T(\theta) = \lambda(\theta)^T A \lambda(\theta) = \sum_{s=1}^T \sum_{t=1}^T a_{st} \lambda_s(\theta) \lambda_t(\theta),$$

$$\Pi_T(\theta) = \sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda_t(\theta),$$

$$P_{1T} = \int (E[\hat{f}(x) - \tilde{f}(x)])^2 f(x) dx,$$

$$P_{2T} = \int (\tilde{f}(x) - E[\tilde{f}(x)])^2 f(x) dx,$$

in which A is the $T \times T$ matrix with $\{a_{st}\}$ as its (s, t) element.

We assume without loss of generality throughout the rest of this paper that

$$\int k(x) dx = \int k^2(x) dx = R(k) \equiv 1.$$

LEMMA A.2. *Under the conditions of Theorem 2.1, we have as $T \rightarrow \infty$*

$$E[N_{0T}(h)] = \int f^2(x) dx \quad \text{and} \quad \text{var}[N_{0T}(h)] = 2h k^{(4)}(0) \int f^4(x) dx(1 + o(1)). \quad (\mathbf{A.9})$$

Proof. We now prove (A.9). It follows from Assumptions A.2 and A.3 that as $T \rightarrow \infty$

$$\begin{aligned} E[\phi_{it}] &= (Th)^{-1} \int E[\epsilon_t^2(x)]f(x) dx \\ &= \frac{1}{Th} \int \left\{ E\left[k^2\left(\frac{x - X_t}{h}\right) \right] - E^2\left[k\left(\frac{x - X_t}{h}\right) \right] \right\} f(x) dx \\ &= \frac{1}{T} (1 + o(1)) \int k^2(u) du \cdot \int f^2(x) dx. \end{aligned}$$

This completes the proof of the first part of (A.9). For the proof of the second part of (A.9), let

$$\sigma_{st}^2 = E[\phi_{st}^2] \quad \text{and} \quad \sigma_T^2 = 2 \sum_{1 \leq s, t \leq T} \sigma_{st}^2.$$

Then

$$\sigma_T^2 = 2 \sum_{1 \leq s, t \leq T} \sigma_{st}^2 = 2 \sum_{t=1}^T \sum_{s=1}^T E[\phi_{st}^2].$$

We first look at the main component of σ_T^2 . We now have

$$\phi_{1st}^2 = \iint \frac{1}{(Th)^2} k\left(\frac{x - X_s}{h}\right) k\left(\frac{y - X_s}{h}\right) k\left(\frac{x - X_t}{h}\right) k\left(\frac{y - X_t}{h}\right) f(x)f(y) dx dy.$$

Using Assumptions A.1–A.4, we have as $T \rightarrow \infty$

$$\begin{aligned} a_{st}(x, y) &\equiv E\left[k\left(\frac{x - X_s}{h}\right) k\left(\frac{y - X_s}{h}\right) k\left(\frac{x - X_t}{h}\right) k\left(\frac{y - X_t}{h}\right) \right] \\ &= \iint k\left(\frac{x - u}{h}\right) k\left(\frac{y - u}{h}\right) k\left(\frac{x - v}{h}\right) k\left(\frac{y - v}{h}\right) f(u, v) dudv \\ &= \iint k\left(\frac{x - y}{h} + \frac{y - u}{h}\right) k\left(\frac{y - u}{h}\right) k\left(\frac{x - v}{h}\right) k\left(\frac{x - v}{h} - \frac{x - y}{h}\right) f(u, v) dudv \\ &= h^2 L\left(\frac{x - y}{h}\right) L\left(\frac{y - x}{h}\right) f(x, y)(1 + o(1)), \end{aligned}$$

where $L(x) = \int k(x + y)k(y) dy$ is as defined in (2.5).

Therefore, as $T \rightarrow \infty$

$$\begin{aligned} \sum_{1 \leq s, t \leq T} E[\phi_{st}^2] &= \frac{1}{(Th)^2} \sum_{1 \leq s, t \leq T} \iint f(x)f(y)a_{st}(x, y) dx dy \\ &= hk^{(4)}(0) \left(\int f^4(x) dx \right) (1 + o(1)), \end{aligned} \tag{A.10}$$

where $k^{(4)}(\cdot)$ is as defined in (2.5).

Similarly, one can show that as $T \rightarrow \infty$

$$\sum_{1 \leq s, t \leq T} E[\phi_{st}^2 - \phi_{1st}^2] = o(h). \tag{A.11}$$

We now deal with the remainder term of $\text{var}[N_{0T}(h)]$. By Lemma C.1 (with $\eta_1 = \phi_{ik}$, $\eta_2 = \phi_{jk}$, $l = 2$, $p_i = 2(1 + \delta)$, and $Q = 1/(1 + \delta)$),

$$E|\phi_{ik} \phi_{jk}| \leq 10M_{T1}^{1/(1+\delta)} \alpha^{\delta/(1+\delta)} (j - i),$$

where M_{T1} is as defined in Lemma A.1.

Therefore, using the fact that $\sum_{i=1}^{\infty} \alpha^{\delta/(1+\delta)}(i) < \infty$,

$$\sum_{1 \leq i < j < k \leq T} E|\phi_{ik} \phi_{jk}| \leq 10T^2 M_{T1}^{1/(1+\delta)} \sum_{i=1}^T \left(1 - \frac{i}{T}\right) \alpha^{\delta/(1+\delta)}(i) \leq CT^2 M_{T1}^{1/(1+\delta)} = o(\sigma_T^2), \tag{A.12}$$

whose proof is similar to that of (A.17), which follows.

Equations (A.10)–(A.12) imply

$$\begin{aligned} \text{var}[L_{0T}(h)] &= 2 \left[\sum_{1 \leq s, t \leq T} \text{var}(\phi_{st}) + 2 \sum_{1 \leq i < j < k \leq T} E(\phi_{ik} \phi_{jk}) \right] \\ &= 2hk^{(4)}(0) \left(\int f^4(x) dx \right) (1 + o(1)). \end{aligned}$$

This finishes the proof of the second part of (A.9). ■

LEMMA A.3. *Under the conditions of Theorem 2.1, we have as $T \rightarrow \infty$*

$$\begin{aligned} E[\hat{f}(x) - f(x)] &= \frac{1}{2} h^2 \sigma_k^2 f^{(2)}(x) + c_k f^{(3)}(x) h^3 (1 + o(1)), \\ E[\tilde{f}(x) - f(x)] &= \frac{1}{2} h^2 \sigma_k^2 f^{(2)}(x) + d_k f^{(3)}(x) h^3 (1 + o(1)), \end{aligned} \tag{A.13}$$

and

$$\text{var}[\tilde{f}(x)] = E[(\tilde{f}(x) - E[\tilde{f}(x)])^2] = C_1 \frac{h^2}{Th} (1 + o(1)), \tag{A.14}$$

where C_1 is a constant and $\hat{f}(x)$, $\tilde{f}(x)$, c_k , and d_k are as defined in Section 2.

Proof. We now give only the proof of (A.14) in some detail, as the proofs of (A.13) and (A.14) are similar and quite standard and the details follow similarly from some existing results. See, for example, Fan and Gijbels (1996).

In view of the definition of $w_t(x)$ and the second equation of (A.13), to prove (A.14), it suffices to show that as $T \rightarrow \infty$

$$\begin{aligned} E \left\{ \frac{1}{Th} \sum_{t=1}^T k \left(\frac{x-u}{h} \right) (s_2(x) - s_1(x)(x - X_t)) [f(x) - f(X_t)] \right\}^2 \\ = \frac{T}{(Th)^2} \int k^2 \left(\frac{x-u}{h} \right) (s_2(x) - s_1(x)(x-u))^2 [f(x) - f(u)]^2 f(u) du \\ = \frac{Th}{(Th)^2} \int k^2(v) (s_2(x) - s_1(x)(vh))^2 [f(x) - f(x-vh)]^2 f(x-vh) dv \\ = C \frac{h^2}{Th} (1 + o(1)) \end{aligned}$$

using a Taylor expansion to $f(x) - f(x - vh)$. This finishes the proof of (A.14). ■

A.3. Proof of Theorem 2.1.

Proof of Theorem 2.1(i). To prove Theorem 2.1(i), in view of Remark A.2 and Lemma A.3, it suffices to show that

$$\frac{N_{0T}(h) - \mu_0}{\sqrt{h}\sigma_0} \rightarrow N(0,1) \quad \text{as } T \rightarrow \infty.$$

To apply Lemma A.1, let $\xi_t = X_t$ and $\phi(\xi_s, \xi_t) = \phi_{st}$ defined previously. Let M_T and N_T be defined as in Lemma A.1. We now verify only the following condition listed in Lemma A.1:

$$\frac{\max\{M_T, N_T\}}{\sigma_h^2} \rightarrow 0 \quad \text{as } T \rightarrow \infty \tag{A.15}$$

for $M_{T1}, M_{T21}, M_{T3}, M_{T51}, M_{T52}$, and M_{T6} , where $\sigma_h^2 = h\sigma_0^2$. The others follow similarly.

For the M_T part, one justifies only

$$\frac{T^2 M_{T1}^{1/(1+\delta)}}{\sigma_h^2} \rightarrow 0 \quad \text{as } T \rightarrow \infty.$$

The others follow similarly.

Let $\psi_{st} = (1/Th) \int k((x - X_s)/h) k((x - X_t)/h) p(x) dx$. We now have

$$\begin{aligned} \psi_{ik} \psi_{jk} &= (Th)^{-2} \iint k \left(\frac{x - X_i}{h} \right) k \left(\frac{x - X_k}{h} \right) k \left(\frac{y - X_j}{h} \right) k \left(\frac{y - X_k}{h} \right) f(x) f(y) dx dy \\ &= T^{-2} \iint k(u) k \left(u + \frac{X_i - X_k}{h} \right) k(v) k \left(v + \frac{X_j - X_k}{h} \right) f(X_i + uh) f(X_j + vh) dudv \\ &= T^{-2} f(X_i) f(X_j) L \left(\frac{X_i - X_k}{h} \right) L \left(\frac{X_j - X_k}{h} \right) (1 + o(1)), \end{aligned}$$

where $L(\cdot)$ is as defined previously.

For any given $1 < \zeta < 2$ and T sufficiently large, we obtain

$$\begin{aligned}
 M_{T11} &= E|\psi_{ik}\psi_{jk}|^\zeta \\
 &= T^{-2\zeta} \iiint |f(u)f(v)|^\zeta \left| L\left(\frac{u-w}{h}\right) \right|^\zeta \left| L\left(\frac{v-w}{h}\right) \right|^\zeta f(u,v,w) dudvdw \\
 &= T^{-2\zeta} h^2 \iiint |f(z+xh)f(z+yh)|^{\zeta_2} |L(x)L(y)|^\zeta f(z+xh, z+yh, z) dx dy dz \\
 &= c_p T^{-2\zeta} h^2,
 \end{aligned} \tag{A.16}$$

using Assumption A.1(ii), where $f(x, y, z)$ denotes the joint density of (X_i, X_j, X_k) and C_p is a constant.

Thus, as $T \rightarrow \infty$

$$\frac{T^2 M_{T11}^{1/(1+\delta)}}{\sigma_h^2} = C \frac{T^2 (T^{-2\zeta} h^2)^{1/\zeta}}{h} = h^{(2-\zeta)/\zeta} \rightarrow 0. \tag{A.17}$$

Hence, (A.17) shows that (A.15) holds for the first part of M_{T1} . The proof for the second part of M_{T1} follows in a similar way. Similarly, we have that as $T \rightarrow \infty$

$$\begin{aligned}
 M_{T3} &= E|\psi_{ik}\psi_{jk}|^2 = (Th)^{-4} h^4 E \left[f^2(X_i) f^2(X_j) L^2\left(\frac{X_i - X_k}{h}\right) L^2\left(\frac{X_j - X_k}{h}\right) \right] \\
 &= T^{-4} \iiint f^2(x) f^2(y) L^2\left(\frac{x-z}{h}\right) L^2\left(\frac{y-z}{h}\right) f(x, y, z) dx dy dz \\
 &= T^{-4} h^2 \iiint f^2(uh+w) f^2(vh+w) L^2(u) L^2(v) dudvdw = CT^{-4} h^2,
 \end{aligned}$$

using Assumption A.1(ii).

This implies that as $T \rightarrow \infty$

$$\frac{T^{3/2} M_{T3}^{1/2}}{\sigma_h^2} = C \frac{T^{3/2} T^{-2} h}{h} = CT^{-1/2} \rightarrow 0. \tag{A.18}$$

Thus, (A.18) now shows that (A.15) holds for M_{T3} . It follows from the structure of $\{\psi_{ij}\}$ that (A.15) holds automatically for M_{T51}, M_{T52} , and M_{T6} , because $E[\phi_{st}] = 0$ for $s \neq t$.

We now prove that (A.15) holds for M_{T21} . For some $0 < \delta < 1$ and $1 \leq i < j < k \leq T$, let $M_{T21} = E[|\psi_{ik}\psi_{jk}|^{2(1+\delta)}]$. Similar to (A.16) and (A.17), we obtain that as $T \rightarrow \infty$

$$\frac{T^{3/2} M_{T21}^{1/2(1+\delta)}}{\sigma_h^2} \rightarrow 0$$

using the fact that $\lim_{T \rightarrow \infty} Th = \infty$.

This completes the proof of (A.15) for M_{T21} , and thus (A.15) holds for the first part of $\{\phi_{st}\}$. Similarly, one can show that (A.15) holds for the other parts of $\{\phi_{st}\}$. Thus, we have shown that under \mathcal{H}_0

$$\frac{N_{0T} - \mu_0}{\sigma_h} \rightarrow N(0,1) \text{ as } T \rightarrow \infty.$$

The proof of Theorem 2.1(i) is therefore finished. ■

Proof of Theorem 2.1(ii). Note that as $T \rightarrow \infty$

$$\begin{aligned} \hat{L}_T(\hat{h}) &= \frac{\hat{N}_T(\hat{h}) - \mu_0}{\sqrt{\hat{h}\sigma_0}} = \frac{\frac{\hat{N}_T(\hat{h})}{\hat{N}_T(h)} [\hat{N}_T(h) - \mu_0] + \left[\frac{\hat{N}_T(\hat{h})}{\hat{N}_T(h)} - \mu_0 \right]}{\left[\frac{\sqrt{\hat{h}}}{\sqrt{h}} - 1 \right] \sigma_h + \sigma_h} \\ &= \frac{\hat{N}_T(h) - \mu_0}{\sigma_h} (1 + o_p(1)) \end{aligned}$$

using the continuity of $\hat{N}_T(h)$ in h . This completes the proof of Theorem 2.1(ii). ■

Proof of Theorem 2.2. The proof follows from Theorem 2.1 and the following standard result:

$$\frac{1}{T} \sum_{i=1}^T \hat{f}^i(X_i) = E[f^i(X_i)] + O_p\left(\frac{1}{\sqrt{Th}}\right) + O(h^2) \text{ for } i = 1, 3. \quad \blacksquare$$

APPENDIX B

This Appendix lists the necessary assumptions for the establishment and the proof of the main results given in Section 3.

B.1. Assumptions.

Assumption B.1. The parameter set Θ is an open subset of R^q for some $q \geq 1$. The parametric family $\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta\}$ satisfies the following conditions.

- (i) Assumption A.2(i) holds.
- (ii) For each $\theta \in \Theta$, $f(x, \theta)$ is continuous with respect to $x \in D$.
- (iii) Assume that there is a finite $C_I > 0$ such that for every $\varepsilon > 0$

$$\inf_{\theta, \theta' \in \Theta : \|\theta - \theta'\| \geq \varepsilon} [f(X_1, \theta) - f(X_1, \theta')]^2 \geq C_I \varepsilon^2$$

holds with probability one (almost surely).

Assumption B.2. (i) Let \mathcal{H}_0 be true. Then $\theta_0 \in \Theta$ and

$$\lim_{T \rightarrow \infty} P(\sqrt{T} \|\tilde{\theta} - \theta_0\| > C_L) < \varepsilon$$

for any $\varepsilon > 0$ and all sufficiently large C_L .

(ii) Let \mathcal{H}_0 be false. Then there is a $\theta^* \in \Theta$ such that

$$\lim_{T \rightarrow \infty} P(\sqrt{T} \|\tilde{\theta} - \theta^*\| > C_L) < \varepsilon$$

for any $\varepsilon > 0$ and all sufficiently large C_L .

Assumption B.3. (i) Assume that the set H_T has the structure of (3.2) with $c_{\min} T^{-\gamma} = h_{\min} < h_{\max} = c_{\max} (\log \log T)^{-1}$, where γ , c_{\min} , and c_{\max} are some constants satisfying $0 < \gamma < 1$ and $0 < c_{\min}, c_{\max} < \infty$.

(ii) Assume that $\Delta_T(x)$ is continuous in $x \in D$ and satisfies $\int_{-\infty}^{\infty} \Delta_T(x) dx = 0$ for all $T \geq 1$.

Remark B.1. Assumptions B.1(i) and B.1(ii) are quite standard in this kind of problem. See Assumptions 1(i) and (ii) of Horowitz and Spokoiny (2001). Assumption B.1(iii) is required to ensure that the marginal density function is identifiable. A similar condition is used in Assumption 1(iii) of Horowitz and Spokoiny (2001). It can be shown that Assumption B.1(iii) holds when $f(x, \theta)$ belongs to classes of simple linear and certain nonlinear functions in θ . The identifiability assumption is imposed to exclude the case where $f(x, \theta)$ is flat as a function of θ over certain range of θ and some value of x , because such a function may be neither identifiable nor a probability density. Assumption B.2 is needed to ensure that the true version of θ under \mathcal{H}_0 or \mathcal{H}_1 can be estimated by a \sqrt{T} -consistent estimator. Assumption B.3(i) imposes some conditions on both h_{\min} and h_{\max} . The theoretical condition on h_{\min} is quite general. In practice, we would suggest using $\gamma = \frac{1}{5}$ to include the estimation-based optimal bandwidth $h_{\text{optimal}} = Cn^{-[1/(2s+1)]}$, because the estimation-based optimal value may also be optimal for testing purposes in some cases. The restriction on h_{\max} is required only for the proof of Theorem 3.3. It should be noted that h_{\max} is not necessarily the optimal bandwidth such that the power of the resulting test is maximized. As explained at the beginning of Section 2, both the existence and reasonableness of Assumption B.3(ii) can be justified. Unlike the regression setting discussed in Horowitz and Spokoiny (2001), we need to assume $\int_{x_{\min}}^{x_{\max}} \Delta_T(x) dx = 0$ to ensure that the alternative is also a probability density. As the main results in Section 2 are only concerned with the null hypothesis, we do not need to assume such a rigorous condition for the main results.

This paper considers using only a set of discrete bandwidths for constructing the adaptive test. It is believed that some corresponding results of Theorems 3.1–3.4 can be established for the case where H_T is an interval of continuous bandwidth values. As H_T is always chosen as a set of discrete bandwidths in practice, we therefore think that such an extension from a set of discrete bandwidths to an interval of continuous bandwidth values may just be for theoretical and technical consideration. As such an extension also involves much more tedious and technical details, we do not discuss this issue in detail in this paper.

B.2. Technical Lemmas. Before stating the necessary lemmas for the proof of the results given in Section 3, we introduce the following notation.

LEMMA B.1. *Suppose that the conditions of Theorem 2.1 hold.*

(i) *For every $\delta > 0$*

$$\max_{h \in H_T} h^{-1} \sup_{\|\theta - \theta_0\| \leq \delta} Q_T(\theta) \leq CT\delta^2$$

in probability, where $C > 0$ is a constant.

(ii) *For each $\theta \in \Theta$ and sufficiently large T*

$$Q_T(\theta) \geq Ch \cdot \lambda(\theta)^\tau \lambda(\theta) \quad \text{in probability.}$$

Proof. (i) It follows from the definition of $Q_T(\theta)$ that

$$Q_T(\theta) \leq \|A\|_\infty \|\lambda(\theta)\|^2. \tag{B.1}$$

To prove Lemma B.1(i), one first needs to show that

$$\|A\|_\infty \leq Ch \tag{B.2}$$

in probability for some constant $C > 0$.

Using the conditions of Lemma B.1, we now have

$$\begin{aligned} \|A\|_\infty &\leq \max_{1 \leq t \leq T} \sum_{s=1}^T a_{st} = C(1 + o_p(1)) \max_{1 \leq t \leq T} \int k\left(\frac{x - X_t}{h}\right) [s_2(x) - s_1(x)(x - X_t)] dx \\ &= C(1 + o_p(1))h \max_{1 \leq t \leq T} \int k(u) [s_2(X_t + uh) - s_1(X_t + uh)uh] du \\ &= C(1 + o_p(1))h \left[s_2(X_t) \int k(u) du - s_1(X_t)h \int uk(u) du \right] \leq Ch \end{aligned}$$

in probability.

In view of (B.2), to prove Lemma B.1(i), it suffices to show that

$$\sup_{\|\theta - \theta_0\| \leq \delta} \|\lambda(\theta)\|^2 \leq CT\delta^2 \tag{B.3}$$

in probability.

A Taylor series expansion to $f(X_t, \theta) - f(X_t, \theta_0)$ and an application of Assumption B.1(i) imply (B.3). This finishes the proof of Lemma B.1(i).

(ii) Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues of A , respectively. In view of

$$\lambda_{\min}(A) \cdot \|\lambda(\theta)\|^2 \leq Q_T(\theta) \leq \lambda_{\max}(A) \cdot \|\lambda(\theta)\|^2, \tag{B.4}$$

to prove Lemma B.1(ii), it suffices to show that for n large enough

$$\lambda_{\min}(A) \geq Ch(1 + o_p(1)) \quad \text{in probability} \tag{B.5}$$

for some $C > 0$. Similar to the proof of Lemma A.2 of Gao, Tong, and Wolff (2002), one can easily finish the proof of (B.5). ■

Without loss of generality, we consider the case of $q = 1$ in the following lemmas and their proofs. Define

$$\psi_i(X_t, \theta) = f^{(i)}(X_t, \theta) = \frac{d^i f(X_t, \theta)}{d\theta^i} \quad \text{for } i = 1, 2, 3.$$

LEMMA B.2. *Under the conditions of Theorem 3.1, we have for any given $\theta \in \Theta$ and $i = 1, 2$*

$$J_T^{-1/2} \max_{h \in H_T} h^{-1/2} \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \psi_i(X_t, \theta) \right| = O_p(1). \tag{B.6}$$

Proof. It suffices to show that for any large constant $C_0 > 0$

$$\begin{aligned} & P \left[J_T^{-1/2} \max_{h \in H_T} h^{-1/2} \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \psi_i(X_t, \theta) \right| > C_0 \right] \\ & \leq \sum_{h \in H_T} P \left[\left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \psi_i(X_t, \theta) \right| > C_0 J_T^{1/2} h^{1/2} \right] \\ & \leq \sum_{h \in H_T} \frac{1}{C_0^2 J_T h} E \left[\sum_{s=1}^T \sum_{t=1}^T b_{st} \psi_i(X_t, \theta) \right]^2 \\ & \leq \sum_{h \in H_T} \frac{1}{C_0^2 J_T h} \left\{ \sum_{s=1}^T \sum_{t=1}^T E [b_{st} \psi_i(X_t, \theta)]^2 + \Lambda_{iT}(\theta) \right\}, \end{aligned} \tag{B.7}$$

where

$$\Lambda_{iT}(\theta) = E \left[\sum_{s=1}^T \sum_{t=1}^T b_{st} \psi_i(X_t, \theta) \right]^2 - \sum_{s=1}^T \sum_{t=1}^T E [b_{st} \psi_i(X_t, \theta)]^2.$$

Similar to the proof of (A.1), one can show that as $T \rightarrow \infty$

$$\sum_{s=1}^T \sum_{t=1}^T E [b_{st} \psi_i(X_t, \theta)]^2 = C(\theta) h (1 + o(1)) \tag{B.8}$$

for some function $C(\theta)$.

Using Lemmas C.1 and C.2 in Appendix C and the fact that $E[\epsilon_t(x)] = 0$ for $x \in D$, one can show that as $T \rightarrow \infty$

$$\Lambda_{iT}(\theta) = o(h) \quad \text{for } i = 1, 2, 3. \tag{B.9}$$

Thus, equations (B.7)–(B.9) complete the proof. ■

LEMMA B.3. *Under the conditions of Theorem 3.1, we have as $T \rightarrow \infty$*

$$J_T^{-1/2} \max_{h \in H_T} h^{-1/2} \max_{1 \leq s \leq T} \left| \sum_{t=1}^T b_{st} \right| = O_p(1). \tag{B.10}$$

Proof. Similar to (B.7), we have for large constant $C_0 > 0$

$$\begin{aligned}
 & P \left[J_T^{-1/2} \max_{h \in H_T} h^{-1/2} \max_{1 \leq s \leq T} \left| \sum_{t=1}^T b_{st} \right| > C_0 \right] \\
 & \leq \sum_{h \in H_T} \sum_{s=1}^T P \left[J_T^{-1/2} h^{-1/2} \left| \sum_{t=1}^T b_{st} \right| > C_0 \right] \leq \frac{1}{C_0^2 J_T} \sum_{h \in H_T} h^{-1} \sum_{s=1}^T E \left[\left| \sum_{t=1}^T b_{st} \right|^2 \right] \\
 & = \frac{1}{C_0^2 J_T} \sum_{h \in H_T} h^{-1} \left\{ \sum_{s=1}^T \sum_{t=1}^T E[b_{st}^2] + \sum_{s=1}^T \sum_{t_1 \neq t_2}^T E[b_{st_1} b_{st_2}] \right\}. \tag{B.11}
 \end{aligned}$$

Similar to (B.8), we can have as $T \rightarrow \infty$

$$\sum_{s=1}^T \sum_{t=1}^T E[b_{st}^2] = Ch(1 + o(1)). \tag{B.12}$$

Analogous to (B.9), one can show that as $T \rightarrow \infty$

$$\sum_{s=1}^T \sum_{t_1 \neq t_2}^T E[b_{st_1} b_{st_2} \xi_{t_1} \xi_{t_2}] = o(h). \tag{B.13}$$

Thus, equations (B.11)–(B.13) complete the proof of (B.10). ■

LEMMA B.4. *Under the conditions of Theorem 3.1, we have for each $u > 0$,*

$$\max_{h \in H_T} \sup_{|\theta - \theta_0| \leq T^{-1/2}u} h^{-1/2} \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda_t(\theta) \right| = O_p(J_T^{1/2} T^{-1/2}) \tag{B.14}$$

under \mathcal{H}_0 .

Proof. We now prove (B.14). Using a Taylor series expansion to $f(X_t, \theta) - f(X_t, \theta_0)$ and Assumption B.1, we have for θ' between θ and θ_0

$$\begin{aligned}
 \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda_t(\theta) \right| &= \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} [f(X_t, \theta) - f(X_t, \theta_0)] \right| \\
 &\leq \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \psi_1(X_t, \theta_0) \right| |\theta - \theta_0| + \frac{1}{2} \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \psi_2(X_t, \theta_0) \right| \\
 &\quad \times |\theta - \theta_0|^2 + \frac{1}{6} \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \psi_3(X_t, \theta') \right| |\theta - \theta_0|^3 \\
 &\leq \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \psi_1(X_t, \theta_0) \right| |\theta - \theta_0| + \frac{1}{2} T |\theta - \theta_0|^2 \left| \sum_{s=1}^T b_{st} \psi_2(X_t, \theta_0) \right| \\
 &\quad + \frac{1}{6} T |\theta - \theta_0|^3 \max_{1 \leq t \leq T} \left| \sum_{s=1}^T b_{st} \right| \cdot \max_{1 \leq t \leq T} |\psi_3(X_t, \theta')|. \tag{B.15}
 \end{aligned}$$

Hence, (B.4), (B.10), (B.15), and Assumption B.1(i) imply

$$\max_{h \in H_T} \sup_{|\theta - \theta_0| \leq T^{-1/2}u} h^{-1/2} \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda_t(\theta) \right| \leq O_p(J_T^{1/2} T^{-1/2}). \tag{B.16}$$

The proof of (B.14) follows from (B.15) and (B.16). ■

LEMMA B.5. *Suppose that the conditions of Theorem 3.1 hold. Then for every $u > 0$, some $h \in H_T$, and as $T \rightarrow \infty$*

$$\sup_{|\theta - \theta^*| \leq T^{-1/2}u} \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda(X_t, \theta) \right| = o_p(Q_T(\theta^*)) \tag{B.17}$$

under \mathcal{H}_1 .

Proof. In view of the definition of $Q_n(\theta)$, to prove (B.17), it suffices to show that as $T \rightarrow \infty$

$$\sup_{|\theta - \theta^*| \leq T^{-1/2}u} \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda(X_t, \theta) \right| = o_p(q_T),$$

where $q_T = E[Q_T(\theta^*)]$.

Note that

$$\begin{aligned} \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda(X_t, \theta) \right| &\leq \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda(X_t, \theta^*) \right| \\ &\quad + \left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \psi_1(X_t, \theta^*) \right| |\theta - \theta^*| \\ &\quad + \frac{1}{2} T |\theta - \theta^*|^2 \left| \sum_{s=1}^T b_{st} \psi_2(X_t, \theta^*) \right| \\ &\quad + \frac{1}{6} T |\theta - \theta^*|^2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^T b_{st} \right| \cdot \max_{1 \leq t \leq T} |\psi_3(X_t, \theta')|, \end{aligned} \tag{B.18}$$

where θ' lies between θ and θ^* .

In view of (B.6), (B.10), (B.18), and Assumptions B.1(i) and B.2(ii), to prove (B.17), it suffices to show that for any $\delta > 0$,

$$P \left[\left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda(X_t, \theta^*) \right| > \delta T q_T \right] \rightarrow 0 \tag{B.19}$$

as $T \rightarrow \infty$.

Similar to (B.8) and (B.9), one can show that as $T \rightarrow \infty$

$$E \left[\sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda(X_t, \theta^*) \right]^2 = Ch(1 + o(1)). \tag{B.20}$$

Thus, equations (B.19) and (B.20) imply that as $T \rightarrow \infty$

$$P \left[\left| \sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda(X_t, \theta^*) \right| > \delta q_T \right] \leq \frac{1}{\delta^2 q_T^2} E \left[\sum_{s=1}^T \sum_{t=1}^T b_{st} \lambda(X_t, \theta^*) \right]^2 = \frac{Ch}{q_T^2} \rightarrow 0 \tag{B.21}$$

using $q_T = CTh(1 + o(1))$ given in the proof of Lemma B.1(ii), where C is a constant independent of T . Lemma B.5 therefore follows from (B.21). ■

Recall the notation introduced in (A.9). We assume without loss of generality that $k^{(4)}(0) = 1$ in Lemma A.2. Define

$$L_0(h) = \frac{N_{0T}(h) - \mu_0}{\sqrt{2h}} \quad \text{and} \quad L_T(h) = \frac{N_T(h) - \mu_0}{\sqrt{2h}}. \tag{B.22}$$

LEMMA B.6. *Suppose that the conditions of Theorem 3.1 hold. Then as $T \rightarrow \infty$*

$$\hat{L}_T(h) = L_T(h) + o_p(1) \tag{B.23}$$

uniformly over $h \in H_T$.

Proof. The proof of (B.23) follows from (2.7) and (2.8) immediately. ■

LEMMA B.7. *Suppose that the conditions of Theorem 3.1 hold. Then $\max_{h \in H_T} L_0(h)$ and $\max_{h \in H_T} L_T(h)$ have identical asymptotic distributions under \mathcal{H}_0 .*

Proof. Note that $Q_T(\theta_0) = 0$ under \mathcal{H}_0 and that Lemmas A.3 and B.1–B.5 imply as $T \rightarrow \infty$

$$\max_{h \in H_T} \frac{P_T(\theta_0)}{\sqrt{2h}} = o_p(1), \quad \max_{h \in H_T} \frac{P_{1T} + P_{2T}}{\sqrt{2h}} = o_p(1), \quad \max_{h \in H_T} \frac{R_T(\theta_0)}{\sqrt{2h}} = o_p(1). \tag{B.24}$$

Therefore, equations (B.21), (B.22), and (B.24) complete the proof of Lemma B.7. ■

LEMMA B.8. *Suppose that the conditions of Theorem 3.1 hold. Then for any $x \geq 0$, $h \in H_T$, and all sufficiently large T*

$$P(L_0(h) > x) \leq \exp\left(-\frac{x^2}{4}\right).$$

Proof. It follows from the beginning of the proof of Theorem 2.1(i) that for any small $\delta > 0$ there exists a large integer $T_0 \geq 1$ such that for $T \geq T_0$

$$|P(L_0(h) \leq x) - \Phi(x)| < \delta,$$

where $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^x e^{-(u^2/2)} du$.

This implies for any $T \geq T_0$ and $x \geq 0$

$$\begin{aligned} P(L_0(h) > x) &\leq 1 - \Phi(x) + \delta \\ &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-(u^2/2)} du + \delta = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-(u^2/4)} e^{-(u^2/4)} du + \delta \\ &\leq e^{-(x^2/4)} \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-(u^2/4)} du + \delta \leq e^{-(x^2/4)} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-(u^2/4)} du + \delta \\ &= e^{-(x^2/4)} \frac{\sqrt{2}}{\sqrt{2\pi}} \int_0^\infty e^{-(v^2/2)} dv + \delta = \frac{\sqrt{2}}{2} e^{-(x^2/4)} + \delta \end{aligned}$$

using $(1/\sqrt{2\pi}) \int_0^\infty e^{-(v^2/2)} dv = \frac{1}{2}$.

The proof follows by letting $0 < \delta \leq (1 - \sqrt{2}/2)e^{-(x^2/4)}$ for any $x \geq 0$. ■

For $0 < \alpha < 1$, define \tilde{l}_α to be the $1 - \alpha$ quantile of $\max_{h \in H_T} L_0(h)$.

LEMMA B.9. *Suppose that the conditions of Theorem 3.1 hold. Then for large enough T*

$$\tilde{l}_\alpha \leq 2\sqrt{\log(J_T) - \log(\alpha)}.$$

Proof. The proof is trivial.

LEMMA B.10. *Suppose that the conditions of Theorem 3.1 hold. Suppose that*

$$\lim_{T \rightarrow \infty} P\left(\frac{Q_T(\theta^*)}{\sqrt{2h}} \geq 2\tilde{l}_\alpha^*\right) = 1 \tag{B.25}$$

for some $h \in H_T$, where

$$\tilde{l}_\alpha^* = \max\left(\tilde{l}_\alpha, \sqrt{2 \log(J_T) + \sqrt{2 \log(J_T)}}\right).$$

Then

$$\lim_{T \rightarrow \infty} P(L^* > l_\alpha) = 1.$$

Proof. To prove Lemma B.10, in view of Lemmas B.6 and B.7, it suffices to show that

$$\lim_{T \rightarrow \infty} P\left(\max_{h \in H_T} L_T(h) > \tilde{l}_\alpha\right) = 1,$$

which holds if

$$\lim_{T \rightarrow \infty} P(L_T(h) > \tilde{l}_\alpha) = 1$$

for some $h \in H_T$. For any $h \in H_T$, using (B.21) and then (B.17) we have

$$\begin{aligned} L_T(h) &= L_0(h) + \frac{Q_T(\theta^*) + R_T(\theta^*) + P_{1T} + P_{2T}}{\sqrt{2h}} \\ &= L_0(h) + \frac{Q_T(\theta^*)(1 + o_p(1))}{\sqrt{2h}}. \end{aligned} \tag{B.26}$$

On the other hand, condition (B.25) implies that as $T \rightarrow \infty$

$$P\left(\frac{Q_T(\theta^*)}{\sqrt{2h}} < 2\tilde{l}_\alpha^*\right) \rightarrow 0. \tag{B.27}$$

Observe that

$$\begin{aligned} P(L_T(h) > \tilde{l}_\alpha) &= P\left(L_T(h) > \tilde{l}_\alpha, \frac{Q_T(\theta^*)}{\sqrt{2h}} \geq 2\tilde{l}_\alpha^*\right) + P\left(L_T(h) > \tilde{l}_\alpha, \frac{Q_T(\theta^*)}{\sqrt{2h}} < 2\tilde{l}_\alpha^*\right) \\ &\equiv I_{1T} + I_{2T}. \end{aligned}$$

Thus, it follows from (B.26) that as $T \rightarrow \infty$

$$\begin{aligned} I_{1T} &= P\left(L_0(h) + \frac{Q_T(\theta^*) + \Pi_T(\theta^*) + R_T(\theta^*) + \sum_{i=1}^2 P_{iT}}{\sqrt{2h}} > \tilde{l}_\alpha \mid \frac{Q_T(\theta^*)}{\sqrt{2h}} \geq 2\tilde{l}_\alpha^*\right) \\ &\quad \times P\left(\frac{Q_T(\theta^*)}{\sqrt{2h}} \geq 2\tilde{l}_\alpha^*\right) \\ &\geq P\left(L_0(h) > \tilde{l}_\alpha - 2\tilde{l}_\alpha^* \mid \frac{Q_T(\theta^*)}{\sqrt{2h}} \geq 2\tilde{l}_\alpha^*\right) P\left(\frac{Q_T(\theta^*)}{2h^{d/2}} \geq 2\tilde{l}_\alpha^*\right) \rightarrow 1 \end{aligned}$$

because $L_0(h)$ is asymptotically normal and therefore bounded in probability and $\tilde{l}_\alpha - 2\tilde{l}_\alpha^* \rightarrow -\infty$ as $T \rightarrow \infty$.

Because of (B.27), as $T \rightarrow \infty$

$$I_{2T} \leq P\left(\frac{Q_T(\theta^*)}{\sqrt{2h}} < 2\tilde{l}_\alpha^*\right) \rightarrow 0.$$

This finishes the proof. ■

B.3. Proofs of Theorems 3.1–3.4.

Proof of Theorem 3.1. The proof follows from Lemmas B.6 and B.7.

Proof of Theorem 3.2. This proof is similar to that of Theorem 3.3, which follows, using Lemma B.1(ii). Alternatively, one can follow the corresponding proof of Theorem 2 of Horowitz and Spokoiny (2001) by using Lemma B.1(ii) and the condition that

$$\lim_{T \rightarrow \infty} P(\rho(f, \mathcal{F}) \geq C_\rho) = 1.$$

to verify (B.25). ■

Proof of Theorem 3.3. Condition (3.5) ensures that the rate of convergence of f_T to the parametric model $F(\theta_1)$ is the same as the rate of convergence of C_T to zero. In particular, when (3.5) holds,

$$\lim_{T \rightarrow \infty} P\left(\inf_{\theta \in \Theta} (\|\bar{f}_T - F(\theta)\|^2) \geq T\delta C_T^2\right) = 1. \tag{B.28}$$

In view of Lemma B.10, to complete the proof of Theorem 3.3, it suffices to verify (B.25). This verification follows from Lemma B.1(ii) and (B.28). ■

Proof of Theorem 3.4. For the proof of Theorem 3.4, one needs to use the conditions of Theorem 3.4 to finish the proof. In our proof, we mainly use Lemma B.1(ii) and the condition of Theorem 3.4 that

$$\lim_{T \rightarrow \infty} P(\rho(f, \mathcal{F}) \geq C_m(T^{-1} \sqrt{\log \log T})^{2s/(4s+1)}) = 1$$

to verify (B.25). ■

APPENDIX C

The following two technical lemmas have already been used in the proofs of Lemma A.1 and Theorem 2.1. The two lemmas are of general interest in themselves and can be used for other nonparametric estimation and testing problems associated with the α -mixing condition.

LEMMA C.1. *Suppose that M_m^n are the σ -fields generated by a stationary α -mixing process ξ_i with the mixing coefficient $\alpha(i)$. For some positive integers m let $\eta_i \in M_{s_i}^{t_i}$ where $s_1 < t_1 < s_2 < t_2 < \dots < t_m$ and suppose $t_i - s_i > \tau$ for all i . Assume further that*

$$\|\eta_i\|_{p_i}^{p_i} = E|\eta_i|^{p_i} < \infty$$

for some $p_i > 1$ for which

$$Q = \sum_{i=1}^l \frac{1}{p_i} < 1.$$

Then

$$\left| E\left[\prod_{i=1}^l \eta_i\right] - \prod_{i=1}^l E[\eta_i]\right| \leq 10(l-1)\alpha(\tau)^{(1-Q)} \prod_{i=1}^l \|\eta_i\|_{p_i}.$$

Proof. See Roussas and Ioannides (1987).

LEMMA C.2. (i) *Let $\psi(\cdot, \cdot, \cdot)$ be a symmetric Borel function defined on $R^r \times R^r \times R^r$. Let the process ξ_i be defined as in Lemma A.1. Assume that for any fixed $x, y \in R^r$, $E[\psi(\xi_1, x, y)] = 0$. Then*

$$E\left\{\sum_{1 \leq i < j < k \leq T} \psi(\xi_i, \xi_j, \xi_k)\right\}^2 \leq CT^3 M^{1/(1+\delta)},$$

where $0 < \delta < 1$ is a small constant, $C > 0$ is a constant independent of T and the function ψ , $M = \max\{M_1, M_2, M_3\}$, and

$$\begin{aligned}
 M_1 &= \max_{1 < i < j \leq T} \max \left\{ E|\psi(\xi_1, \xi_i, \xi_j)|^{2(1+\delta)}, \int |\psi(\xi_1, \xi_i, \xi_j)|^{2(1+\delta)} dP(\xi_1) dP(\xi_i, \xi_j) \right\}, \\
 M_2 &= \max_{1 < i < j \leq T} \max \left\{ \int |\psi(\xi_1, \xi_i, \xi_j)|^{2(1+\delta)} dP(\xi_j) dP(\xi_1, \xi_i) \right\}, \\
 M_3 &= \max_{1 < i < j \leq T} \max \left\{ \int |\psi(\xi_1, \xi_i, \xi_j)|^{2(1+\delta)} dP(\xi_1) dP(\xi_i) dP(\xi_j) \right\}. \tag{C.1}
 \end{aligned}$$

(ii) Let $\phi(\cdot, \cdot)$ be a symmetric Borel function defined on $R^r \times R^r$. Let the process ξ_i be defined as in Lemma A.1. Assume that for any fixed $x \in R^r$, $E[\phi(\xi_1, x)] = 0$. Then

$$E \left\{ \sum_{1 \leq i < j \leq T} \phi(\xi_i, \xi_j) \right\}^2 \leq CT^2 M_4^{1/(1+\delta)},$$

where $\delta > 0$ is a constant, $C > 0$ is a constant independent of T and the function ϕ , and

$$M_4 = \max_{1 < i < j \leq T} \max \left\{ E|\phi(\xi_1, \xi_i)|^{2(1+\delta)}, \int |\phi(\xi_1, \xi_i)|^{2(1+\delta)} dP(\xi_1) dP(\xi_i) \right\}. \tag{C.2}$$

Proof. As the proof of (ii) is similar to that of (i), one proves only (i). Let i_1, \dots, i_6 be distinct integers and $1 \leq i_j \leq T$, let $1 \leq k_1 < \dots < k_6 \leq T$ be the permutation of i_1, \dots, i_6 in ascending order, and let d_c be the c th largest difference among $k_{j+1} - k_j$, $j = 1, \dots, 5$. Let

$$H(k_1, \dots, k_6) = \psi(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) \psi(\xi_{i_4}, \xi_{i_5}, \xi_{i_6}).$$

By Lemma C.1 (with $\eta_1 = \psi(\xi_{i_1}, \xi_{i_2}, \xi_{i_3})$, $\eta_2 = \psi(\xi_{i_4}, \xi_{i_5}, \xi_{i_6})$, $l = 2$, $p_i = 2(1 + \delta)$ and $Q = 1/(1 + \delta)$),

$$|E[H(k_1, \dots, k_6)]| \leq \begin{cases} 10M^{1/(1+\delta)} \alpha^{\delta/(1+\delta)} (k_6 - k_5) & \text{if } k_6 - k_5 = d_1 \\ 10M^{1/(1+\delta)} \alpha^{\delta/(1+\delta)} (k_2 - k_1) & \text{if } k_2 - k_1 = d_1. \end{cases}$$

Thus,

$$\begin{aligned}
 &\sum_{\substack{1 \leq k_1 < \dots < k_6 \leq T \\ k_2 - k_1 = d_1}} |E[H(k_1, \dots, k_6)]| \\
 &\leq \sum_{k_1=1}^{T-5} \sum_{k_2=k_1+\max_{j \geq 3}\{k_j-k_{j-1}\}} \sum_{k_3=k_2+1}^{T-3} \dots \sum_{k_6=k_5+1}^T \{10M^{1/(1+\delta)} \alpha^{\delta/(1+\delta)} (k_2 - k_1)\} \\
 &\leq 10M^{1/(1+\delta)} \sum_{k_1=1}^{T-5} \sum_{k_2=k_1+1}^{T-4} (k_2 - k_1)^2 \alpha^{\delta/(1+\delta)} (k_2 - k_1) \\
 &\leq 10TM^{1/(1+\delta)} \sum_{k=1}^T k^4 \alpha^{\delta/(1+\delta)} (k) \leq CTM^{1/(1+\delta)}. \tag{C.3}
 \end{aligned}$$

Similarly,

$$\sum_{\substack{1 \leq k_1 < \dots < k_6 \leq T \\ k_2 - k_1 = d_1}} |E[H(k_1, \dots, k_6)]| \leq CTM^{1/(1+\delta)}. \tag{C.4}$$

Analogously, it can be shown in a similar way that

$$\sum_{\substack{1 \leq k_1 < \dots < k_6 \leq T \\ k_6 - k_5 = d_2 \text{ or } k_2 - k_1 = d_2}} |E[H(k_1, \dots, k_6)]| \leq CT^2M^{1/(1+\delta)}, \tag{C.5}$$

$$\sum_{\substack{1 \leq k_1 < \dots < k_6 \leq T \\ k_6 - k_5 = d_3 \text{ or } k_2 - k_1 = d_3}} |E[H(k_1, \dots, k_6)]| \leq CTM^{1/(1+\delta)}. \tag{C.6}$$

On the other hand, if $\{k_6 - k_5, k_2 - k_1\} = \{d_4, d_5\}$, by using Lemma C.1 three times we have the inequality

$$|E[H(k_1, \dots, k_6)]| \leq 10M^{1/(1+\delta)} \sum_{i=1}^3 \alpha^{\delta/(1+\delta)}(d_i).$$

Hence,

$$\begin{aligned} & \sum_{\substack{1 \leq k_1 < \dots < k_6 \leq T \\ \{k_6 - k_5, k_2 - k_1\} = \{d_4, d_5\}}} |E[H(k_1, \dots, k_6)]| \\ & \leq \sum_{\substack{1 \leq k_1 < \dots < k_6 \leq T \\ \max\{k_6 - k_5, k_2 - k_1\} \\ \leq \min_{2 \leq j \leq 4} \{k_{j+1} - k_j\}}} \{10M^{1/(1+\delta)} [\alpha^{\delta/(1+\delta)}(k_3 - k_2) \\ & \quad + \alpha^{\delta/(1+\delta)}(k_4 - k_3) + \alpha^{\delta/(1+\delta)}(k_5 - k_4)]\} \\ & \leq 30M^{1/(1+\delta)} \sum_{\substack{1 \leq k_1 < \dots < k_6 \leq T \\ \max\{k_6 - k_5, k_2 - k_1\} \leq d_3}} \alpha^{1/(1+\delta)}(d_3) \leq 30CT^3M^{1/(1+\delta)}. \end{aligned} \tag{C.7}$$

It follows from (C.3)–(C.7) that

$$\sum_{\substack{1 \leq i, j, k, r, s, t \leq T \\ i, j, k, r, s, t \text{ different}}} |E[\psi(\xi_i, \xi_j, \xi_k)\psi(\xi_r, \xi_s, \xi_t)]| \leq CT^3M^{1/(1+\delta)}. \tag{C.8}$$

Similar to (C.8), one can show that

$$\sum_{\substack{1 \leq i, j, k, r, s, t \leq T \\ i, j, k, s, t \text{ different}}} |E[\psi(\xi_i, \xi_j, \xi_k)\psi(\xi_i, \xi_s, \xi_t)]| \leq CT^3M^{1/(1+\delta)}, \tag{C.9}$$

$$\sum_{\substack{1 \leq i, j, k, l \leq T \\ i, j, k, l \text{ different}}} |E[\psi(\xi_i, \xi_j, \xi_k)\psi(\xi_i, \xi_j, \xi_l)]| \leq CT^3M^{1/(1+\delta)}. \tag{C.10}$$

Finally, it is easy to see that

$$\sum_{1 \leq i < j < k \leq T} E[\psi(\xi_i, \xi_j, \xi_k)^2] \leq T^3 \max_{1 < i < j} E[\psi(\xi_1, \xi_i, \xi_j)^2]. \tag{C.11}$$

The conclusion of Lemma C.2(i) follows immediately from (C.8)–(C.11). ■