Bandwidth Selection in Nonparametric Kernel Testing

Jiti GAO and Irène GIJBELS

We propose a sound approach to bandwidth selection in nonparametric kernel testing. The main idea is to find an Edgeworth expansion of the asymptotic distribution of the test concerned. Due to the involvement of a kernel bandwidth in the leading term of the Edgeworth expansion, we are able to establish closed-form expressions to explicitly represent the leading terms of both the size and power functions and then determine how the bandwidth should be chosen according to certain requirements for both the size and power functions. For example, when a significance level is given, we can choose the bandwidth such that the power function is maximized while the size function is controlled by the significance level. Both asymptotic theory and methodology are established. In addition, we develop an easy implementation procedure for the practical realization of the established methodology and illustrate this on two simulated examples and a real data example.

KEY WORDS: Choice of bandwidth parameter; Edgeworth expansion; Nonparametric kernel testing; Power function; Size function.

1. INTRODUCTION

Consider a nonparametric regression model of the form

$$Y_i = m(X_i) + e_i, \qquad i = 1, 2, \ldots, n, \tag{1}$$

where $\{X_i\}$ is a sequence of strictly stationary time series variables, $\{e_i\}$ is a sequence of independent and identically distributed (iid) errors with $E[e_1] = 0$ and $0 < E[e_1^2] = \sigma^2 < \infty$, $m(\cdot)$ is an unknown function defined over $\mathbb{R}^d$ for $d \geq 1$, and n is the number of observations. We assume that $\{X_i\}$ and $\{e_j\}$ are independent for all $1 \leq i \leq j \leq n$.

To avoid the so-called "curse of dimensionality" problem, we consider mainly the case of $1 \leq d \leq 3$ in this article. For the case of $d \geq 4$, various dimension-reduction estimation and specification methods have been discussed extensively in several monographs (e.g., Fan and Gijbels 1996; Hart 1997; Fan and Yao 2003; Li and Racine 2007).

There is a vast literature on testing a parametric regression model (null hypothesis) versus a nonparametric model, especially for the case of iid X_i's (random or fixed design case). Many goodness-of-fit testing procedures are based on evaluating a distance between a parametric estimate of the regression function m (assuming that the null hypothesis is true) and a nonparametric estimate of that function. Among the popular choices for a nonparametric kernel estimator for m are the Nadaraya–Watson estimator, the Gasser–Müller estimator, and a local linear (polynomial) estimator. Earlier works following this approach of evaluating such a distance include those of Härdle and Mammen (1993), Weihrather (1993), and González-Manteiga and Cao (1993), among others. Härdle and Mammen (1993) considered a weighted L_2-distance between a parametric estimator and a nonparametric Nadaraya–Watson estimator of the regression function. The asymptotic distribution of their test statistic under the null hypothesis depends on the unknown error variance (the conditional error variance function).

Jiti Gao is Professor of Economics, School of Economics, The University of Adelaide, Adelaide SA 5005, Australia (E-mail: *jiti.gao@adelaide.edu.au*). Irene Gijbels is Professor of Statistics, Department of Mathematics, University of Leuven, Celestijnenlaan 200B, B-3001, Leuven, Belgium (E-mail: *irene.gijbels@wis.kuleuven.be*). The authors thank the editor, an associate editor, and three reviewers for their very valuable comments, which led to a considerably improved version of the original manuscript. The first author thanks Jiying Yin for his excellent computing assistance and the Australian Research Council Discovery (grants DP0558602 and DP0879088) for financial support. The second author thanks the School of Mathematics and Statistics at the University of Western Australia for its kind hospitality and support. The second author also gratefully acknowledges the K.U. Leuven Research Fund (GOA/2007/4) and the Flemish Science Foundation, FWO Belgium (project G.0328.08) for financial support.

Weihrather (1993) instead used a Gasser–Müller nonparametric estimator in the fixed-design regression case, divided by an estimator of the error variance, and considered a discretized version of the L_2-distance. González-Manteiga and Cao (1993) also considered the fixed design regression case but relied on minimum distance estimation of the parametric model, seeking to minimize a weighted L_2-type distance between the parametric model and a pilot nonparametric estimator.

Another approach to the same testing problem was introduced by Dette (1999), who focused on the integrated conditional variance function and used as a test statistic the difference of a parametric estimator and a nonparametric (Nadaraya–Watson–based) estimator of this integrated variance. This estimator (asymptotically) has been shown to correspond to test statistics based on a weighted L_2-distance between a parametric and nonparametric estimator of the regression function, as in the aforementioned articles, using an appropriate weight function in defining the L_2-distance. Dette (1999) studied the asymptotic distribution of the test statistic under fixed alternatives. Such alternatives are to be distinguished from the so-called "sequences of local alternatives," where the difference between the regression function under the alternative and that under the null hypothesis depends on the sample size n and decreases with n. The latter setup is the one that we consider here.

The aforementioned works and several more recent goodness-of-fit tests (see, e.g., Zhang and Dette 2004 and references therein) all rely on nonparametric kernel-type regression estimators, and the resulting test statistics are of a similar form (at least in first-order asymptotics) and depend on a bandwidth parameter. The choice of the bandwidth parameter in such goodness-of-fit testing procedures is the main focus in this article. Roughly speaking, the literature distinguishes two approaches to deal with this bandwidth parameter choice in nonparametric and semiparametric kernel methods used for constructing model specification tests for the mean function of model (1). One of these approaches is to use an estimation-based optimal bandwidth value, such as a cross-validation bandwidth. The other approach is to consider a set of suitable values for the bandwidth and proceed from there.

Studies based on the first approach include those of Härdle and Mammen (1993) for testing nonparametric regression with iid designs and errors; Hjellvik and Tjøstheim (1995) and Hjellvik, Yao, and Tjøstheim (1998) for testing linearity in dependent time series cases; Li (1999) for specification testing

Journal of the American Statistical Association
December 2008, Vol. 103, No. 484, Theory and Methods
DOI 10.1198/016214508000000968

in econometric time series cases; Chen, Härdle, and Li (2003) for using empirical likelihood–based tests; and Juhl and Xiao (2005) for testing structural change in nonparametric time series regression. As has been pointed out in the literature, such choices cannot be justified in both theory and practice, because estimation-based optimal values may not be optimal for testing purposes.

Nonparametric tests involving the second approach of choosing either a set of suitable bandwidth values for the kernel case or a sequence of positive integers for the smoothing spline case include those of Fan (1996), Fan, Zhang, and Zhang (2001), and Horowitz and Spokoiny (2001). The practical implementation of choosing such sets or sequences is problematic, however. This is probably why Horowitz and Spokoiny (2001) developed their theoretical results based on a set of suitable bandwidths on the one hand, but chose their practical bandwidth values based on the assessment of the power function of their test on the other hand. Apart from such test statistics based on nonparametric kernel, nonparametric series, spline smoothing, and wavelet methods, there exist test statistics that are constructed and studied based on empirical distributions. Such studies have recently been summarized by Zhu (2005).

To the best of our knowledge, the idea of choosing the appropriate smoothing parameter such that the size of the test under consideration is preserved while maximizing the power against a given alternative was first explored analytically by Kulasekera and Wang (1997), who proposed using a nonparametric kernel test to check whether the mean functions of two data sets can be identical in a nonparametric fixed-design setting. Other, closely related studies have compared power values of the same test at different bandwidths or different tests at the same bandwidth; such studies include those of Hart (1997), Hjellvik et al. (1998), Hunsberger and Follmann (2001), and Zhang and Dette (2004). The latter authors compared three main types of nonparametric kernel tests proposed by Härdle and Mammen (1993), Zheng (1996), and Fan et al. (2001).

On the issue of size correction, some recent studies have been published. For example, Fan and Linton (2003) developed an Edgeworth expansion for the size function of their test and then proposed using corrected asymptotic critical values to improve the small to medium sample size properties of their test. Other related studies include those of Nishiyama and Robinson (2000, 2005) and Horowitz (2003), who developed some useful Edgeworth expansions for bootstrap distributions of partial-sum type tests for improving the size performance.

The present work is motivated by such existing studies, especially those of Kulasekera and Wang (1997), Fan and Linton (2003), Dette and Spreckelsen (2004), and Zhang and Dette (2004), to develop a solid theory to support a power function–based bandwidth selection procedure such that the power of the proposed test is maximized while the size remains under control when using nonparametric kernel testing in the parametric specification of a nonparametric regression model of the form (1) associated with the hypothesis form of (2).

Before stating the main results of this article, we introduce some notational details. Our main interest is to test a parametric null hypothesis of the form

$$\begin{aligned}&\mathcal{H}_0: m(x) = m_{\theta_0}(x) \text{ versus a sequence of alternatives}\\ &\text{of the form}\\ &\mathcal{H}_1: m(x) = m_{\theta_1}(x) + \Delta_n(x) \text{ for all } x \in \mathbb{R}^d,\end{aligned} \tag{2}$$

where both θ_0 and $\theta_1 \in \Theta$ are unknown parameters, Θ is a parameter space of $\mathbb{R}^p$, and $\Delta_n(x)$ is a sequence of nonparametrically unknown functions over $\mathbb{R}^d$. With $\Delta_n(x)$ not being equal to 0, the function $m_{\theta_1}(x)$ in $\mathcal{H}_1$ is the projection of the true function on the null model.

Note that $m(x)$ under $\mathcal{H}_1$ in (2) is semiparametric when $\{\Delta_n(x)\}$ is unknown nonparametrically. Also note that instead of requiring (2) for all $x \in \mathbb{R}^d$, it may be assumed that (2) holds with probability 1 for $x = X_i$. Some first-order asymptotic properties for both the size and power functions of a nonparametric kernel test for the case where $\Delta_n(\cdot) \equiv \Delta(\cdot)$, corresponding to a class of fixed alternatives (not depending on n), have been discussed in the literature (e.g., Dette and Spreckelsen 2004). In this article we focus on studying higher-order asymptotic properties of such kernel tests for the case where $\{\Delta_n(\cdot)\}$ is a sequence of local alternatives in the sense that $\lim_{n\to\infty} \Delta_n(x) = 0$ for all $x \in \mathbb{R}^d$.

Let $K(\cdot)$ be the probability kernel density function, and let h be the bandwidth involved in the construction of a nonparametric kernel test statistic denoted by $\widehat{T}_n(h)$. To implement the kernel test in practice, we propose a new bootstrap simulation procedure to approximate the $1-\alpha$ quantile of the distribution of the kernel test by a bootstrap-simulated critical value, l_α. Let $\alpha_n(h) = P(\widehat{T}_n(h) > l_\alpha | \mathcal{H}_0)$ and $\beta_n(h) = P(\widehat{T}_n(h) > l_\alpha | \mathcal{H}_1)$ be the size and power functions. In Theorem 2 we show that

$$\begin{aligned}\alpha_n(h) &= 1 - \Phi(l_\alpha - s_n)\\ &\quad - \kappa_n(1 - (l_\alpha - s_n)^2)\phi(l_\alpha - s_n) + o(\sqrt{h^d})\end{aligned} \tag{3}$$

and

$$\begin{aligned}\beta_n(h) &= 1 - \Phi(l_\alpha - r_n)\\ &\quad - \kappa_n(1 - (l_\alpha - r_n)^2)\phi(l_\alpha - r_n) + o(\sqrt{h^d}),\end{aligned} \tag{4}$$

where $s_n = p_1\sqrt{h^d}$, $r_n = p_2 n\delta_n^2\sqrt{h^d}$, $\kappa_n = p_3\sqrt{h^d}$, and $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution and density function of the standard Normal random variable, in which all p_i's are positive constants and $\delta_n^2 = \int \Delta_n^2(x)\pi^2(x)\,dx$, with $\pi(\cdot)$ the marginal density function of $\{X_i\}$.

Our aim is to choose a bandwidth h_{ew} such that $\beta_n(h_{\mathrm{ew}}) = \max_{h\in H_n(\alpha)} \beta_n(h)$ with $H_n(\alpha) = \{h : \alpha - c_{\min} < \alpha_n(h) < \alpha + c_{\min}\}$ for some small $0 < c_{\min} < \alpha$. Our detailed study reported in Section 3 shows that h_{ew} is proportional to $(n\delta_n^2)^{-3/(2d)}$. Such an established relationship between δ_n and h_{ew} demonstrates that the choice of an optimal rate of h_{ew} depends on that of an order of δ_n.

If δ_n is chosen proportional to $n^{-(d+12)/(6(d+4))}$ for a sequence of local alternatives under $\mathcal{H}_1$, then the optimal rate of h_{ew} is proportional to $n^{-1/(d+4)}$, which is the order of a nonparametric cross-validation estimation-based bandwidth often used for testing purposes. When considering a sequence of local alternatives with $\delta_n = O(n^{-1/2}\sqrt{\log\log n})$ chosen as the optimal rate for testing in this kind of kernel testing (Horowitz and Spokoiny 2001), the optimal rate of h_{ew} is proportional to $(\log\log n)^{-3/(2d)}$.

The rest of the article is organized as follows. Section 2 points out that existing nonparametric kernel tests can be decomposed with quadratic forms of $\{e_i\}$ as leading terms in the decomposition. This motivates the discussion on establishing Edgeworth expansions for such quadratic forms. Section 3 applies the Edgeworth expansions to study both the size and power functions of a representative kernel test. Section 4 presents several examples of implementation. Section 5 provides some concluding remarks. Mathematical assumptions and proofs are provided in the Appendix.

2. NONPARAMETRIC KERNEL TESTING

As mentioned in Section 1, various authors have discussed and studied nonparametric kernel test statistics based on a (weighted) L_2-distance function between a nonparametric kernel estimator and a parametric counterpart of the mean function. It can be shown that the leading term of each of these nonparametric kernel test statistics is of a quadratic form (see, e.g., Chen et al. 2003),

$$P_n(h) = \sum_{i=1}^{n}\sum_{j=1}^{n} e_i w(X_i) L_h(X_i - X_j) w(X_j) e_j, \qquad (5)$$

where $L_h(\cdot) = \frac{1}{n\sqrt{h^d}} L(\frac{\cdot}{h})$, $L(x) = \int K(y)K(x+y)\,dy$, and $w(\cdot)$ is a suitable weight function probably depending on either $\pi(\cdot)$ or $\sigma^2(\cdot)$ or both, in which $K(\cdot)$ is a probability kernel function, h is a bandwidth parameter, and both are involved in a nonparametric kernel estimation of $m(\cdot)$.

In this article we concentrate on a second group of nonparametric kernel test statistics using a different distance function. Rewrite model (1) into a notational version of the form under $\mathcal{H}_0$,

$$Y = m_{\theta_0}(X) + e, \qquad (6)$$

where X is assumed to be random and θ_0 is the true value of θ under $\mathcal{H}_0$. Obviously, $E[e|X] = 0$ under $\mathcal{H}_0$. Previous studies (Zheng 1996; Li and Wang 1998; Li 1999; Fan and Linton 2003; Dette and Spreckelsen 2004; Juhl and Xiao 2005) have proposed using a distance function of the form

$$E[eE(e|X)\pi(X)] = E[(E^2(e|X))\pi(X)], \qquad (7)$$

where $\pi(\cdot)$ is the marginal density function of X.

This suggests using a normalized kernel-based sample analog of (7) of the form

$$T_n(h) = \frac{1}{n\sqrt{h^d}\sigma_n} \sum_{i=1}^{n} \sum_{j=1,\neq i}^{n} e_i K\left(\frac{X_i - X_j}{h}\right) e_j, \qquad (8)$$

where $\sigma_n^2 = 2\mu_2^2 \nu_2 \int K^2(u)\,du$ with $\mu_k = E[e_1^k]$ for $k \ge 1$ and $\nu_l = E[\pi^l(X_1)]$ for $l \ge 1$.

It can be easily seen that $T_n(h)$ is the leading term of the following quadratic form

$$Q_n(h) = \frac{1}{n\sqrt{h^d}\sigma_n} \sum_{i=1}^{n}\sum_{j=1}^{n} e_i K\left(\frac{X_i - X_j}{h}\right) e_j. \qquad (9)$$

In summary, both equations (5) and (9) can be generally written as

$$R_n(h) = \sum_{i=1}^{n}\sum_{j=1}^{n} e_i \phi_n(X_i, X_j) e_j, \qquad (10)$$

where $\phi_n(\cdot,\cdot)$ may depend on n, the bandwidth h, and the kernel function K.

Thus it is of general interest to study asymptotic distributions and their Edgeworth expansions for quadratic forms of type (10). To present the main ideas of establishing Edgeworth expansions for such quadratic forms, we focus on $T_n(h)$ in the rest of this article. This is because the main approach to establishing an Edgeworth expansion for the asymptotic distribution of each of such tests is the same as that for $T_n(h)$.

Because $T_n(h)$ involves some unknown quantities, we estimate it through a stochastically normalized version of the form

$$\widehat{T}_n(h) = \frac{\sum_{i=1}^{n} \sum_{j=1,\neq i}^{n} \widehat{e}_i K(\frac{X_i - X_j}{h}) \widehat{e}_j}{n\sqrt{h^d}\widehat{\sigma}_n}, \qquad (11)$$

where $\widehat{e}_i = Y_i - m_{\widehat{\theta}}(X_i)$ and $\widehat{\sigma}_n^2 = 2\widehat{\mu}_2^2 \widehat{\nu}_2 \int K^2(u)\,du$ with $\widehat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{e}_i^2$ and $\widehat{\nu}_2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{\pi}^2(X_i)$, where $\widehat{\theta}$ is a $\sqrt{n}$-consistent estimator of θ_0 under $\mathcal{H}_0$ and $\widehat{\pi}(x) = \frac{1}{n b_{\mathrm{cv}}^d} \times \sum_{i=1}^{n} K(\frac{x - X_i}{b_{\mathrm{cv}}})$ is the conventional nonparametric kernel density estimator, with $\widehat{b}_{\mathrm{cv}}$ a bandwidth parameter chosen by cross-validation (see, e.g., Silverman 1986).

Similar to existing results (Li 1999), it may be shown that for each given h,

$$\widehat{T}_n(h) = T_n(h) + o_P(\sqrt{h^d}). \qquad (12)$$

Thus we may use the distribution of $\widehat{T}_n(h)$ to approximate that of $T_n(h)$. Let l_α^e $(0 < \alpha < 1)$ be the $1 - \alpha$ quantile of the exact finite-sample distribution of $\widehat{T}_n(h)$. Because l_α^e may not be evaluated in practice, we suggest choosing either a nonrandom approximate α-level critical value, l_α, or a stochastic approximate α-level critical value, l_α^*, using the following simulation procedure:

- Generate $Y_i^* = m_{\widehat{\theta}}(X_i) + \sqrt{\widehat{\mu}_2} e_i^*$ for $1 \le i \le n$, where $\{e_i^*\}$ is a sequence of iid random samples drawn from a prespecified distribution, such as N(0, 1). Use the data set $\{(X_i, Y_i^*) : i = 1, 2, \ldots, n\}$ to estimate $\widehat{\theta}$ by $\widehat{\theta}^*$, and compute $\widehat{T}_n(h)$. Let l_α be the $1 - \alpha$ quantile of the distribution of

$$\widehat{T}_n^*(h) = \frac{\sum_{i=1}^{n} \sum_{j=1,\neq i}^{n} \widehat{e}_i^* K(\frac{X_i - X_j}{h}) \widehat{e}_j^*}{n\sqrt{h^d}\widehat{\sigma}_n^*}, \qquad (13)$$

where $\widehat{e}_i^* = Y_i^* - m_{\widehat{\theta}^*}(X_i)$ and $\widehat{\sigma}_n^{*2} = 2\widehat{\mu}_2^{*2}\widehat{\nu}_2 \int K^2(u)\,du$ with $\widehat{\mu}_2^* = \frac{1}{n}\sum_{i=1}^{n} \widehat{e}_i^{*2}$.

In the simulation process, the original sample $\mathcal{X}_n = (X_1, \ldots, X_n)$ acts in the resampling as a fixed design even when $\{X_i\}$ is a sequence of random regressors.

- Repeat the foregoing step M times and produce M versions of $\widehat{T}_n^*(h)$, denoted by $\widehat{T}_{n,m}^*(h)$ for $m = 1, 2, \ldots, M$. Use the M values of $\widehat{T}_{n,m}^*(h)$ to construct their empirical distribution function. The bootstrap distribution of $\widehat{T}_n^*(h)$, given $\mathcal{W}_n = \{(X_i, Y_i) : 1 \le i \le n\}$, is defined by $P^*(\widehat{T}_n^*(h) \le x) = P(\widehat{T}_n^*(h) \le x | \mathcal{W}_n)$. Let l_α^* $(0 < \alpha < 1)$ satisfy $P^*(\widehat{T}_n^*(h) \ge l_\alpha^*) = \alpha$, and then estimate l_α by l_α^*.

Note that both $l_\alpha = l_\alpha(h)$ and $l_\alpha^* = l_\alpha^*(h)$ depend on h. It should be pointed out that the choice of a prespecified distribution does not have much impact on either the theoretical or practical results. In addition, we also may use a wild bootstrap procedure to generate a sequence of resamples for $\{e_i^*\}$.

Also note that the foregoing simulation is based on the so-called "regression bootstrap" simulation procedure discussed in the literature (e.g., Li and Wang 1998; Franke, Kreiss, and Mammen 2002; Li and Racine 2007). When $X_i = Y_{i-1}$, we also may use a recursive simulation procedure, commonly used in the literature (see, e.g., Hjellvik and Tjøstheim 1995; Franke et al. 2002).

Because the choice of a simulation procedure does not affect the establishment of our theory, we establish our main results based on the proposed simulation procedure. We now have the following results in Theorems 1 and 2, the proofs of which are provided in the Appendix.

Theorem 1. Suppose that Assumptions A.1 and A.2 in the Appendix hold. Then, under $\mathcal{H}_0$,

$$\sup_{x \in R^1} |P^*(\widehat{T}_n^*(h) \le x) - P(\widehat{T}_n(h) \le x)| = O(\sqrt{h^d}) \tag{14}$$

holds in probability with respect to the joint distribution of $\mathcal{W}_n$, and

$$P(\widehat{T}_n(h) > l_\alpha^*) = \alpha + O(\sqrt{h^d}). \tag{15}$$

For an equivalent test, Li and Wang (1998) established some results weaker than (14). Fan and Linton (2003) considered some higher-order approximations to the size function of the test discussed by Li and Wang (1998).

For each h, we define the following size and power functions:

$$\alpha_n(h) = P(\widehat{T}_n(h) > l_\alpha | \mathcal{H}_0) \quad \text{and} \quad \beta_n(h) = P(\widehat{T}_n(h) > l_\alpha | \mathcal{H}_1). \tag{16}$$

Correspondingly, we define $(\alpha_n^*(h), \beta_n^*(h))$, with l_α replaced by l_α^*.

Before we discuss how to choose an optimal bandwidth in Section 3, we give Edgeworth expansions of both the size and power functions in Theorem 2. To express the Edgeworth expansions, we need to introduce the following notation. Let

$$\kappa_n = \frac{\sqrt{h^d}\left(\frac{\mu_3^2 K^2(0)}{nh^d} + \frac{4\mu_2^3 \nu_3}{3} K^{(3)}(0)\right)}{\sigma_n^3}, \tag{17}$$

where $\nu_l = E[\pi^l(X_1)] = \int \pi^{l+1}(x)\,dx$ and $K^{(3)}(\cdot)$ is the three-time convolution of $K(\cdot)$ with itself.

Theorem 2. a. Suppose that Assumptions A.1 and A.2 in the Appendix hold. Then

$$\alpha_n(h) = 1 - \Phi(l_\alpha - s_n) - \kappa_n(1 - (l_\alpha - s_n)^2)\phi(l_\alpha - s_n) + o(\sqrt{h^d}) \tag{18}$$

and

$$\alpha_n^*(h) = 1 - \Phi(l_\alpha^* - s_n) - \kappa_n(1 - (l_\alpha^* - s_n)^2)\phi(l_\alpha^* - s_n) + o(\sqrt{h^d}) \tag{19}$$

hold in probability with respect to the joint distribution of $\mathcal{W}_n$, where $\Phi(\cdot)$ and $\phi(\cdot)$ are the probability distribution and density functions of N(0, 1), and $s_n = C_0(m)\sqrt{h^d}$ with

$$C_0(m) = \frac{\int (\frac{\partial m_{\theta_0}(x)}{\partial \theta})^\tau (E[(\frac{m_{\theta_0}(X_1)}{\partial \theta})(\frac{m_{\theta_0}(X_1)}{\partial \theta})^\tau])^{-1} (\frac{m_{\theta_0}(x)}{\partial \theta}) \pi^2(x)\,dx}{\sqrt{2\nu_2 \int K^2(v)\,dv}}.$$

b. Suppose that Assumptions A.1–A.3 in the Appendix hold. Then the following equations hold in probability with respect to the joint distribution of $\mathcal{W}_n$:

$$\beta_n(h) = 1 - \Phi(l_\alpha - r_n) - \kappa_n(1 - (l_\alpha - r_n)^2)\phi(l_\alpha - r_n) + o(\sqrt{h^d}) \tag{20}$$

and

$$\beta_n^*(h) = 1 - \Phi(l_\alpha^* - r_n) - \kappa_n(1 - (l_\alpha^* - r_n)^2)\phi(l_\alpha^* - r_n) + o(\sqrt{h^d}), \tag{21}$$

where $r_n = nC_n^2\sqrt{h^d}$, where

$$C_n^2 = \frac{\int \Delta_n^2(x)\pi^2(x)\,dx}{\sigma^2\sqrt{2\nu_2 \int K^2(v)\,dv}}. \tag{22}$$

As expected, the rate of r_n depends on the form of $\Delta_n(\cdot)$.

To simplify the following expressions, let z_α be the $1-\alpha$ quantile of the standard normal distribution and $d_j = (z_\alpha^2 - 1)c_j$ for $j = 1, 2$, where

$$c_1 = \frac{4K^{(3)}(0)\mu_2^3\nu_3}{3\sigma_n^3} \quad \text{and} \quad c_2 = \frac{\mu_3^2 K^2(0)}{\sigma_n^3}. \tag{23}$$

Let $d_0 = d_1 - C_0(m)$. A corollary of Theorem 2 is given in Theorem 3.

Theorem 3. Suppose that the conditions of Theorem 2(a) hold. Then, under $\mathcal{H}_0$,

$$l_\alpha \approx z_\alpha + d_0\sqrt{h^d} + d_2\frac{1}{n\sqrt{h^d}} \quad \text{in probability} \tag{24}$$

and

$$l_\alpha^* \approx z_\alpha + d_0\sqrt{h^d} + d_2\frac{1}{n\sqrt{h^d}} \quad \text{in probability.} \tag{25}$$

Theorem 3 shows that the size distortion of the proposed test is $d_0\sqrt{h^d} + d_2\frac{1}{n\sqrt{h^d}}$ when using the standard asymptotic normality in practice. A similar result was obtained by Fan and Linton (2003). We also show that the bootstrap-simulated critical value is approximated explicitly by $z_\alpha + d_0\sqrt{h^d} + d_2\frac{1}{n\sqrt{h^d}}$.

As the main objective of this work, in Section 3 we propose a suitable selection criterion for the choice of h such that while the size function is appropriately controlled, the power function is maximized at this h. A closed-form expression of the power function–based optimal bandwidth is given.

3. POWER FUNCTION–BASED BANDWIDTH CHOICE

We use the Edgeworth expansions established in Section 2 to choose a suitable bandwidth such that the power function $\beta_n(h)$ is maximized while the size function $\alpha_n(h)$ is controlled by a significance level. We thus define

$$h_{\text{ew}} = \arg\max_{h \in H_n(\alpha)} \beta_n(h),$$
$$\text{with } H_n(\alpha) = \{h : \alpha - c_{\min} < \alpha_n(h) < \alpha + c_{\min}\} \quad (26)$$

for some arbitrarily small $c_{\min} > 0$.

We now discuss how to solve the optimization problem (26). It follows from (17) and (23) that

$$\kappa_n = \frac{\sqrt{h^d}\left(\frac{\mu_3^2 K^2(0)}{nh^d} + \frac{4\mu_2^3\nu_3}{3} K^{(3)}(0)\right)}{\sigma_n^3} = c_1\sqrt{h^d} + c_2\frac{1}{n\sqrt{h^d}}. \quad (27)$$

Let $x = \sqrt{h^d}$. Rewrite κ_n as $\kappa_n = c_1 x + c_2 n^{-1}x^{-1}$. Let $\gamma_n = (z_\alpha^2 - 1)\kappa_n$,

$$l_\alpha - r_n \approx z_\alpha + \gamma_n - r_n = z_\alpha + (d_1 - nC_n^2)x + d_4 x^{-1} \equiv z_\alpha + d_3 x + d_4 x^{-1} \quad (28)$$

and

$$l_\alpha - s_n \approx z_\alpha + \gamma_n - s_n \approx z_\alpha + (d_1 - C_0(m))x + d_4 x^{-1} = z_\alpha + d_0 x + d_4 x^{-1}, \quad (29)$$

where $d_0 = d_1 - C_0(m)$, $d_1 = (z_\alpha^2 - 1)c_1$, $d_3 = d_1 - nC_n^2$, and $d_4 = c_2(z_\alpha^2 - 1)n^{-1}$. Note that $\lim_{n\to\infty} d_4 = 0$. Because Assumption A.3 implies that $\lim_{n\to\infty} nC_n^2 = +\infty$, we thus have

$$\lim_{n\to\infty} d_3 = -\infty \quad \text{when} \quad \lim_{n\to\infty} nC_n^2 = +\infty. \quad (30)$$

Because of this, we treat d_3 as a sufficiently large negative value when nC_n^2 is viewed as a sufficiently large positive value in the finite-sample analysis of this section.

Ignoring the higher-order terms [i.e., terms of order $o(x + n^{-1}x^{-1})$ or smaller], we now rewrite the power and size functions $\beta_n(h)$ and $\alpha_n(h)$ simply as functions of $x = \sqrt{h^d}$ as follows:

$$\begin{aligned}\beta_n(h) &\approx 1 - \Phi(l_\alpha - r_n) - \kappa_n(1 - (l_\alpha - r_n)^2)\phi(l_\alpha - r_n)\\ &\approx 1 - \Phi(z_\alpha + d_3 x + d_4 x^{-1}) - (c_1 x + c_2 n^{-1}x^{-1})\\ &\quad \times (1 - (z_\alpha + d_3 x + d_4 x^{-1})^2)\phi(z_\alpha + d_3 x + d_4 x^{-1})\\ &\equiv \beta(x), \end{aligned} \quad (31)$$

$$\begin{aligned}\alpha_n(h) &\approx 1 - \Phi(l_\alpha - s_n) - \kappa_n(1 - (l_\alpha - s_n)^2)\phi(l_\alpha - s_n)\\ &\approx 1 - \Phi(z_\alpha + d_0 x + d_4 x^{-1}) - (c_1 x + c_2 n^{-1}x^{-1})\\ &\quad \times (1 - (z_\alpha + d_0 x + d_4 x^{-1})^2)\phi(z_\alpha + d_0 x + d_4 x^{-1})\\ &\equiv \alpha(x). \end{aligned} \quad (32)$$

Our objective is then to find $x_{\text{ew}} = \sqrt{h_{\text{ew}}^d}$ such that

$$x_{\text{ew}} = \arg\max_{x \in H_n(\alpha)} \beta(x)$$
$$\text{with } H_n(\alpha) = \{x : \alpha - c_{\min} < \alpha(x) < \alpha + c_{\min}\}, \quad (33)$$

where $c_{\min}$ is chosen as $c_{\min} = \frac{\alpha}{10}$, for example. Finding roots of $\beta'(x) = 0$ implies that the leading order of the unique real root of the equation is given approximately by

$$h_{\text{ew}} = x_{\text{ew}}^{2/d} = a_1^{-1/(2d)} t_n^{-3/(2d)}, \quad (34)$$

where $t_n = nC_n^2$ and $a_1 = \frac{\sqrt{2}K^{(3)}(0)}{3(\sqrt{\int K^2(u)\,du})^3} \cdot c(\pi)$ with $c(\pi) = \frac{\int \pi^3(x)\,dx}{(\sqrt{\int \pi^2(x)\,dx})^3}$, in which C_n^2 is as defined in Theorem 2(b).

It also can be shown that h_{ew} is the maximizer of the power function $\beta_n(h)$ at $h = h_{\text{ew}}$ such that

$$\beta_n''(x)|_{x=\sqrt{h_{\text{ew}}^d}} < 0, \quad (35)$$

at least for sufficiently large n. Detailed derivations of (34) and (35) are given in the supplemental material (available at *http://www.amstat.org/publications/jasa/supplemental_materials*).

Furthermore, the choice of h_{ew} satisfies both Assumptions A.1(e) and A.3, that

$$\lim_{n\to\infty} nh_{\text{ew}}^d = +\infty \quad \text{and} \quad \lim_{n\to\infty} n\sqrt{h_{\text{ew}}^d} C_n^2 = +\infty.$$

This implies that the choice of h_{ew} is valid to ensure that $\lim_{n\to\infty} \beta_n(h_{\text{ew}}) = 1$.

When both $\sigma^2 = \mu_2 = E[e_1^2]$ and the marginal density function $\pi(\cdot)$ of $\{X_i\}$ are unknown in practice, we propose using an estimated version of h_{ew},

$$\widehat{h}_{\text{ew}} = \widehat{a}_1^{-1/(2d)} \widehat{t}_n^{-3/(2d)}, \quad (36)$$

where

$$\widehat{t}_n = n\widehat{C}_n^2, \qquad \text{with } \widehat{C}_n^2 = \frac{\frac{1}{n}\sum_{i=1}^n \widehat{\Delta}_n^2(X_i)\widehat{\pi}(X_i)}{\widehat{\mu}_2\sqrt{2\widehat{\nu}_2 \int K^2(v)\,dv}}$$

and

$$\widehat{a}_1 = \frac{\sqrt{2}K^{(3)}(0)}{3(\sqrt{\int K^2(u)\,du})^3}\widehat{c}(\pi),$$
$$\text{with } \widehat{c}(\pi) = \frac{\frac{1}{n}\sum_{i=1}^n \widehat{\pi}^2(X_i)}{(\sqrt{\frac{1}{n}\sum_{i=1}^n \widehat{\pi}(X_i)})^3},$$

in which $\widehat{\mu}_2$, $\widehat{\nu}_2$, and $\widehat{\pi}(\cdot)$ are as defined in (11) and $\widehat{\Delta}_n(x)$ is given by

$$\widehat{\Delta}_n(x) = \frac{\sum_{i=1}^n K(\frac{x - X_i}{\widehat{b}_{\text{cv}}})(Y_i - m_{\widehat{\theta}}(X_i))}{\sum_{i=1}^n K(\frac{x - X_i}{\widehat{b}_{\text{cv}}})},$$

with $\widehat{\theta}$ and $\widehat{b}_{\text{cv}}$ the same as in (11).

Also note that $\widehat{h}_{\text{ew}}$ provides an optimal bandwidth irrespective of whether we works under the null hypothesis $\mathcal{H}_0$ or under the alternative hypothesis $\mathcal{H}_1$. In other words, it can be used for computing not only the power under an alternative $\mathcal{H}_1$, but also the size under $\mathcal{H}_0$ in each case. A detailed discussion on this is provided in the supplemental material.

We conclude this section by summarizing the foregoing discussion in the following proposition, the proof of which is given in the supplemental material.

Proposition 1. Suppose that Assumptions A.1–A.3 in the Appendix hold. In addition, suppose that $\Delta_n(x)$ is continuously differentiable such that

$$\lim_{n\to\infty} \sup_{x\in D_\pi} \frac{\|\Delta_n'(x)\|}{|\Delta_n(x)|} \le C < \infty$$

and

$$\lim_{n\to\infty} \inf_{x\in\mathbb{R}^d} |\Delta_n(x)|\sqrt{n\widehat{b}_{\mathrm{cv}}^d} = \infty \quad \text{in probability}$$

for some $C > 0$, where $D_\pi = \{x \in \mathbb{R}^d : \pi(x) > 0\}$ and $\|\cdot\|^2$ denotes the Euclidean norm. Then

$$\lim_{n\to\infty} \frac{\beta_n(\widehat{h}_{\mathrm{ew}})}{\beta_n(h_{\mathrm{ew}})} = 1 \quad \text{in probability.} \tag{37}$$

As pointed out in Section 1, implementation of each of existing nonparametric kernel tests involves either a single bandwidth chosen optimally for estimation purposes or a set of bandwidth values. The proposed $\widehat{h}_{\mathrm{ew}}$ is chosen optimally for testing purposes. In Section 4 we discuss how to implement the proposed test based on our bandwidth in practice and compare the finite-sample performance of the proposed choice with that of some closely relevant alternatives in the literature.

4. EXAMPLES OF IMPLEMENTATION

In this section we present two simulated examples and one real data example to illustrate the proposed theory and methods described in Sections 2 and 3, as well as to make comparisons with some closely relevant alternatives in the literature. Simulated Example 1 discusses the finite-sample performance of the proposed test $\widehat{T}_n(\widehat{h}_{\mathrm{ew}})$ with that of the alternative version, in which the test is coupled with a cross-validation (CV) bandwidth choice. Simulated Example 2 compares our test with some of the commonly used tests in the literature. Example 3 provides a real data example showing that the proposed test makes a clear difference. In the finite-sample study in Examples 1–3, we consider the case where $\Delta_n(x) = c_n\Delta(x)$, in which $\{c_n\}$ is a sequence of positive real numbers satisfying $\lim_{n\to\infty} c_n = 0$ and $\Delta(x)$ is an unknown function not depending on n.

Example 1. Consider a nonparametric time series regression model of the form

$$Y_i = \theta_1 X_{i1} + \theta_2 X_{i2} + c_n(X_{i1}^2 + X_{i2}^2) + e_i, \qquad 1 \le i \le n, \tag{38}$$

where $\{e_i\}$ is a sequence of Normal errors and X_{i1} and X_{i2} are time series variables generated by

$$X_{i1} = \alpha X_{i-1,1} + u_i \qquad \text{and} \\ X_{i2} = \beta X_{i-1,2} + v_i, \qquad 1 \le i \le n, \tag{39}$$

where $\{u_i\}$ and $\{v_i\}$ are iid random errors generated independently from Normal distributions as described later.

Under $\mathcal{H}_0$, we generate a sequence of observations $\{Y_i\}$ with $\theta_1 = \theta_2 = 1$ as the true parameters, that is,

$$\mathcal{H}_0 : Y_i = X_{i1} + X_{i2} + e_i, \tag{40}$$

where $\{e_i\}$ is a sequence of iid random errors generated from N(0, 1), and $\{X_{i1}\}$ and $\{X_{i2}\}$ are independently generated from

$$X_{i1} = .5X_{i-1,1} + u_i \qquad \text{and} \\ X_{i2} = .5X_{i-1,2} + v_i, \qquad 1 \le i \le n, \tag{41}$$

with $X_{01} = X_{02} = 0$ and $\{u_i\}$ and $\{v_i\}$ are sequences of iid random errors and generated independently from a N(0, 1).

Under $\mathcal{H}_1$, we are interested in two alternative models of the form

$$\mathcal{H}_1 : Y_i = X_{i1} + X_{i2} + c_n(X_{i1}^2 + X_{i2}^2) + e_i, \qquad e_i \sim \mathrm{N}(0,1), \tag{42}$$

with c_n chosen as either $c_{1n} = n^{-1/2}\sqrt{\log\log(n)}$ or $c_{2n} = n^{-7/18}$.

In the testing procedure, the parameters θ_1 and θ_2 in the parametric model are estimated as discussed in Sections 1 and 2.

The reasoning for the foregoing choice of c_{jn} is as follows. The rate of $c_{1n} = n^{-1/2}\sqrt{\log\log(n)}$ should be an optimal rate of testing in this kind of nonparametric kernel testing problem, as discussed by Horowitz and Spokoiny (2001). The rate of $c_{2n} = n^{-7/18}$ implies that the optimal bandwidth $\widehat{h}_{\mathrm{ew}}$ in (36) with $d = 2$ is proportional to $n^{-1/6}$.

In this example, we choose $K(\cdot)$ as the standard normal density function. Let $\widehat{h}_{\mathrm{cv}}$ be chosen by a CV criterion of the form

$$\widehat{h}_{\mathrm{cv}} = \arg\min_{h\in H_{\mathrm{cv}}} \frac{1}{n}\sum_{i=1}^n (Y_i - \widehat{m}_{-i}(X_{i1}, X_{i2}; h))^2 \\ \text{with } H_{\mathrm{cv}} = [n^{-1}, n^{1/6}], \tag{43}$$

in which

$$\widehat{m}_{-i}(X_{i1}, X_{i2}; h) = \frac{\sum_{l=1,\ne i}^n K(\frac{X_{l1}-X_{i1}}{h})K(\frac{X_{l2}-X_{i2}}{h})Y_l}{\sum_{l=1,\ne i}^n K(\frac{X_{l1}-X_{i1}}{h})K(\frac{X_{l2}-X_{i2}}{h})}.$$

Let $\widehat{h}_{0\mathrm{test}}$ be the corresponding version of $\widehat{h}_{\mathrm{ew}}$ in (36), and let $\widehat{h}_{0\mathrm{cv}}$ be the corresponding version of $\widehat{h}_{\mathrm{cv}}$ in (43), both computed under $\mathcal{H}_0$. Because $\{Y_i\}$ under $\mathcal{H}_1$ depends on the choice of c_n, the computation of both $\widehat{h}_{\mathrm{ew}}$ of (36) and $\widehat{h}_{\mathrm{cv}}$ of (43) under $\mathcal{H}_1$ depend on the choice of c_n. Let $\widehat{h}_{j\mathrm{test}}$ be the corresponding versions of $\widehat{h}_{\mathrm{ew}}$ in (36), and let $\widehat{h}_{j\mathrm{cv}}$ be the corresponding versions of $\widehat{h}_{\mathrm{cv}}$ in (43) with $c_n = c_{jn}$ for $j = 1, 2$.

To compare the size and power properties of $\widehat{T}_n(h)$ with the most relevant alternatives, we introduce the following simplified notation: for $j = 1, 2$,

$$\alpha_{01} = P(\widehat{T}_n(\widehat{h}_{0\mathrm{cv}}) > l_\alpha^*(\widehat{h}_{0\mathrm{cv}})|\mathcal{H}_0), \\ \beta_{j1} = P(\widehat{T}_n(\widehat{h}_{j\mathrm{cv}}) > l_\alpha^*(\widehat{h}_{0\mathrm{cv}})|\mathcal{H}_1)$$

and

$$\alpha_{02} = P(\widehat{T}_n(\widehat{h}_{0\mathrm{test}}) > l_\alpha^*(\widehat{h}_{0\mathrm{test}})|\mathcal{H}_0), \\ \beta_{j2} = P(\widehat{T}_n(\widehat{h}_{j\mathrm{test}}) > l_\alpha^*(\widehat{h}_{0\mathrm{test}})|\mathcal{H}_1).$$

We consider cases in which the number of replications of each sample version of α_{0k} and β_{jk} for $j, k = 1, 2$ was $M =$ 1,000, each with $B = 250$ number of bootstrapping resamples, and the simulations were done for the cases of $n = 250$, 500, and 750. The detailed results at the 1%, 5%, and 10% significance level are given in Tables 1–3.

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

Sample size	Null hypothesis is true		Null hypothesis is false			
n	α_{01}	α_{02}	β_{11}	β_{21}	β_{12}	β_{22}
250	.012	.016	.212	.239	.294	.272
500	.018	.014	.270	.303	.318	.334
750	.014	.008	.310	.367	.408	.422

Tables 1–3 report comprehensive simulation results for both the sizes and power values of the proposed tests for models (40) and (41). Column 2 in each of these tables shows that although the sizes for the test based on $\widehat{h}_{0\text{cv}}$ are comparable with those given in column 3 based on $\widehat{h}_{0\text{test}}$, the power values of the test based on $\widehat{h}_{j\text{test}}$ in columns 6 and 7 are always greater than those given in columns 4 and 5 based on $\widehat{h}_{j\text{cv}}$. This is not surprising, because the theory shows that each of $\widehat{h}_{j\text{test}}$ is chosen such that the resulting power function is maximized while the corresponding size function is under control by the significance level.

In addition, the test based on $\widehat{h}_{2\text{test}}$ is almost uniformly more powerful than the test based on $\widehat{h}_{1\text{test}}$, which is the second-most powerful test. This is basically because $\widehat{h}_{2\text{test}}$ is based on considering $\mathcal{H}_1$ with $c_{2n} = n^{-7/18}$, which goes to 0 slower than $c_{1n} = n^{-1/2}\sqrt{\log\log(n)}$, and thus the distance between the alternative and the null is largest in the former case (and thus easier to detect). Meanwhile, the last columns of Tables 1–3 show that the test based on the bandwidth $\widehat{h}_{2\text{test}}$ is still a powerful test even though the bandwidth is proportional to $n^{-1/6}$, which is the same as the optimal bandwidth based on a CV estimation method. This indicates that whether an estimation-based optimal bandwidth can be used for testing depends on whether the bandwidth is chosen optimally for testing purposes.

We finally want to stress that the proposed test based on either $\widehat{h}_{1\text{test}}$ or $\widehat{h}_{2\text{test}}$ has not only stable sizes even at a small sample size of $n = 250$, but also reasonable power values even when the "distance" between the null and the alternative is made deliberately close at the rate of $\sqrt{n^{-1}\log\log(n)} = .060$ for $n = 500$, for example. We could expect the test to have larger power values when the "distance" is made wider. Overall, Tables 1–3 show that the established theory and methodology is workable in the small and medium sample size cases.

Example 1 discusses the small and medium sample size comparison results for the proposed test with either a testing-based optimal bandwidth or an estimation-based (CV) bandwidth. Example 2 considers comparing the small and medium sample size performance of the proposed test associated with the optimal bandwidth with some closely related nonparametric tests available in both the econometrics and statistics literature.

Table 2. Simulated size and power values at the 5% significance level

Sample size	Null hypothesis is true		Null hypothesis is false			
n	α_{01}	α_{02}	β_{11}	β_{21}	β_{12}	β_{22}
250	.054	.046	.514	.522	.656	.658
500	.052	.058	.572	.564	.690	.730
750	.046	.052	.648	.658	.820	.812

Table 3. Simulated size and power values at the 10% significance level

Sample size	Null hypothesis is true		Null hypothesis is false			
n	α_{01}	α_{02}	β_{11}	β_{21}	β_{12}	β_{22}
250	.116	.110	.696	.764	.884	.909
500	.104	.090	.744	.817	.860	.934
750	.108	.090	.844	.895	.946	.968

Example 2. Consider a linear model of the form

$$Y_i = \alpha_0 + \beta_0 X_i + e_i, \qquad 1 \le i \le n = 250, \tag{44}$$

where $\{X_i\}$ is a sequence of independent random variables sampled from N(0, 25) distribution truncated at its 5th and 95th percentiles and $\{e_i\}$ is sampled from one of the three distributions: (a) $e_i \sim$ N(0, 4); (b) a mixture of Normals in which $\{e_i\}$ is sampled from N(0, 1.56) with probability .9 and from N(0, 25) with probability .1; and (c) the type I extreme value distribution scaled to have a variance of 4. The mixture distribution is leptokurtic with a variance of .39, and the type I extreme value distribution is asymmetrical.

This is the same example as used by Horowitz and Spokoiny (2001) for comparison with some of the commonly used tests in the literature, such as the Andrews test proposed by Andrews (1997), the HM test proposed by Härdle and Mammen (1993), the HS test proposed by Horowitz and Spokoiny (2001), and the empirical likelihood (EL) test proposed by Chen et al. (2003).

To compute the sizes of the test, choose $\alpha_0 = \beta_0 = 1$ as the true parameters and then generate $\{Y_i\}$ from $Y_i = 1 + X_i + e_i$ under $\mathcal{H}_0$ and $\{Y_i\}$ from $Y_i = 1 + X_i + \frac{5}{\tau}\phi(\frac{X_i}{\tau}) + e_i$ under $\mathcal{H}_1$, where $\tau = 1$ or .25 and $\phi(\cdot)$ is the density function of the standard normal distribution.

The kernel function used here is $K(x) = \frac{15}{16}(1-x^2)^2 I(|x| \le 1)$. Choose $c_n = 5\tau^{-1}$ and $\Delta(x) = \phi(x\tau^{-1})$ for the corresponding forms in (2). For $j = 1, 2$, let $c_{jn} = 5\tau_j^{-1}$ and $\Delta_j(x) = \phi(x\tau_j^{-1})$ with $\tau_1 = 1$ and $\tau_2 = .25$. Let $\widehat{h}_{i\text{new}}$ be the corresponding version of $\widehat{h}_{\text{ew}}$ of (36) based on $(c_{jn}, \Delta_j(x))$ for $j = 1, 2$.

To make a fair comparison, we use the same number of bootstrap resamples of $M = 99$ and the same number of replications of $M = 1{,}000$ under $\mathcal{H}_0$ and $M = 250$ under $\mathcal{H}_1$ as in table 1 of Horowitz and Spokoiny (2001). In Table 4, we add the size and power values to the last two columns for both the EL test and our proposed test $\widehat{T}_n(\widehat{h}_{i\text{new}})$. The other parts of the table are obtained and tabulated similarly to table 1 of Horowitz and Spokoiny (2001).

Table 4 shows that our proposed test has better power properties than any of the commonly used tests, while the sizes are comparable. These results further support the power-based bandwidth selection procedure proposed in Sections 2 and 3.

As discussed in the supplemental material, the proposed theory and methodology for model (1) can be applied to an extended model of the form

$$Y_i = m(X_i) + e_i, \quad \text{with } e_i = \sigma(X_i)\epsilon_i, 1 \le i \le n, \tag{45}$$

where $\sigma(\cdot)$ satisfying $\inf_{x\in\mathbb{R}^d}\sigma(x) > 0$ is unknown nonparametrically and $\{\epsilon_i\}$ is a sequence of iid random errors with

Table 4. Simulated size and power values at the 5% significance level

		Probability of rejecting null hypothesis				
Distribution	τ	Andrews test	HM test	HS test	EL test	$\widehat{T}_n(\widehat{h}_{\text{new}})$ test
Null hypothesis is true						
Normal		.057	.060	.066	.053	.049
Mixture		.053	.053	.048	.055	.052
Extreme		.063	.057	.055	.057	.052
Null hypothesis is false						
Normal	1.0	.680	.752	.792	.900	.907
Mixture	1.0	.692	.736	.835	.905	1.000
Extreme	1.0	.600	.760	.820	.924	.935
Normal	.25	.536	.770	.924	.929	.993
Mixture	.25	.592	.704	.922	.986	.999
Extreme	.25	.604	.696	.968	.989	.989

mean 0 and finite variance. In addition, $\{\epsilon_i\}$ and $\{X_j\}$ are assumed to be independent for all $1 \le j \le i \le n$. A special case of model (45) is discussed in Example 3 below.

Example 3. This example examines the high-frequency 7-day Eurodollar deposit rate sampled daily from June 1, 1973 to February 25, 1995. This provides us with $n = 5{,}505$ observations. Let $\{X_i : i = 1, 2, \dots, n = 5{,}505\}$ be the set of Eurodollar deposit rate data. Figures 1 and 2 plot the data values and the conventional nonparametric kernel density estimator

$$\widehat{\pi}(x) = \frac{1}{n\widetilde{h}_{\text{cv}}} \sum_{i=1}^{n} K\left(\frac{x - X_i}{\widetilde{h}_{\text{cv}}}\right),$$

where $K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and $\widetilde{h}_{\text{cv}}$ is the conventional normal-reference based bandwidth given by

$$\widetilde{h}_{\text{cv}} = 1.06 \cdot n^{-1/5} \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2},$$

$$\text{with } \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i. \quad (46)$$

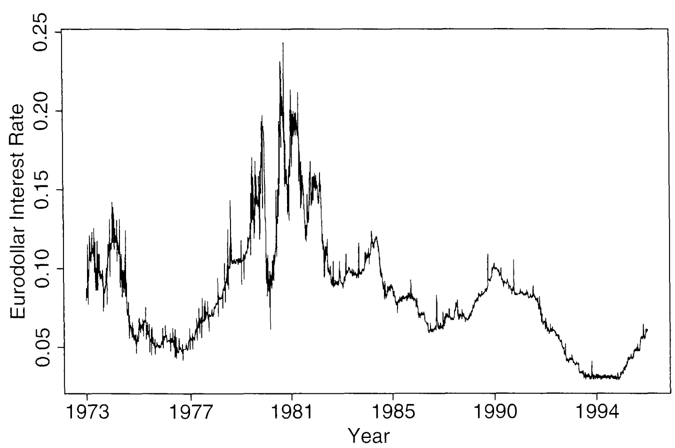

Figure 1. Seven-day Eurodollar deposit rate, June 1, 1973–February 25, 1995.

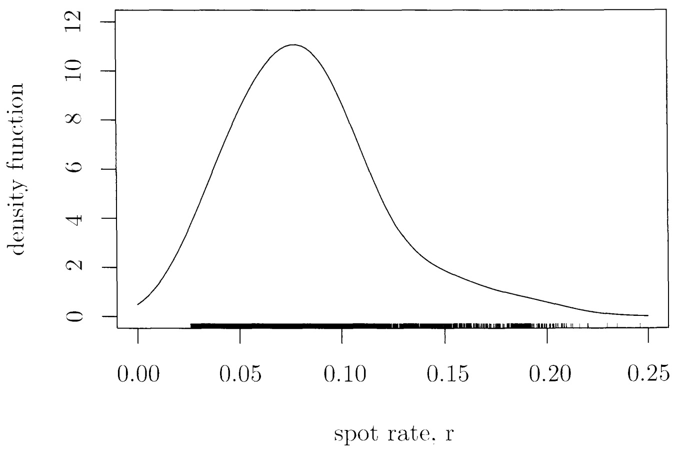

Figure 2. Nonparametric kernel density estimator of the Eurodollar rate.

Note that $\widehat{b}_{\text{cv}}$ of (11), $\widehat{h}_{\text{cv}}$ of (43), and $\widetilde{h}_{\text{cv}}$ of (46) are normally different from one another. In the case where $\{X_i\}$ follows an autoregressive model, they can be chosen the same. Thus they are chosen the same in this example.

It has been assumed in the literature (see, e.g., Aït-Sahalia 1996; Fan and Zhang 2003; Arapis and Gao 2006) that the Eurodollar data set $\{X_i\}$ may be modeled by a nonlinear time series model of the form

$$Y_i = \mu(X_i) + \sigma(X_i)\epsilon_i, \qquad 1 \le i \le n, \quad (47)$$

where $Y_i = \frac{X_{i+1} - X_i}{\Lambda}$, $\sigma(\cdot) > 0$ is unknown nonparametrically and $\epsilon_i \sim \mathrm{N}(0, \Lambda^{-1})$, in which Λ is the time between successive observations. Because we consider a daily data set, this gives $\Lambda = \frac{1}{250}$.

On the question of whether there is any nonlinearity in the drift function $\mu(\cdot)$, existing studies have provided no definitive answer. On the one hand, for example, Aït-Sahalia (1996) and Arapis and Gao (2006) showed some evidence of supporting nonlinearity in the drift. On the other hand, other studies, such as those of Chapman and Pearson (2000) and Fan and Zhang (2003), have suggested that nonlinearity may be caused simply by estimation biases when using nonparametric kernel estimation.

To further explore whether the assumption on linearity in the drift is appropriate for the given set of data, we apply our test to propose testing

$$\begin{aligned} &\mathcal{H}_{01} : \mu(x) = \mu(x; \theta_0) = \beta_0(\alpha_0 - x) \quad \text{versus} \\ &\mathcal{H}_{11} : \mu(x) = \beta_1(\alpha_1 - x) + c_n \Delta(x) \end{aligned} \quad (48)$$

for some $\theta_j = (\alpha_j, \beta_j) \in \Theta$ for $j = 0, 1$ and $c_n = \sqrt{n^{-1} \log\log(n)}$, where Θ is a parameter space in $\mathbb{R}^2$ and $\Delta(x)$ is a continuous function.

It can be shown that the proposed test in Section 2 has an asymptotically equivalent version of the form

$$\widetilde{T}_n(h) = \frac{\sum_{j=1}^{n} \sum_{i=1, i \ne j}^{n} \widehat{e}_j K(\frac{X_i - X_j}{h}) \widehat{e}_i}{\sqrt{2 \sum_{j=1}^{n} \sum_{i=1}^{n} \widehat{e}_j^2 K^2(\frac{X_i - X_j}{h}) \widehat{e}_i^2}}, \quad (49)$$

where $\widehat{e}_i = Y_i - \widehat{\beta}(\widehat{\alpha} - X_i)$, in which $(\widehat{\alpha}, \widehat{\beta})$ is the pair of the conventional least squares estimators minimizing $\sum_{i=1}^{n}(Y_i - \widehat{\beta}(\widehat{\alpha} - X_i))^2$.

As pointed out by Arapis and Gao (2006), $\widetilde{T}_n(h)$ is independent of the structure of the conditional variance $\sigma^2(\cdot)$. The kernel function used is the standard normal density function given by $K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Let $\widetilde{h}_{\text{test}}$ be the corresponding version of (36). It is shown in the supplemental material that

$$\widetilde{h}_{\text{test}} = \widehat{a}_1^{-1/2} \widehat{t}_n^{-3/2}, \tag{50}$$

where $\widehat{t}_n$ and $\widehat{a}_1$ are the same as in (36), in which $\widehat{c}(\pi)$ becomes

$$\widehat{c}(\pi) = \frac{1}{n} \sum_{i=1}^{n} \widehat{\pi}^2(X_i) \widehat{\sigma}^6(X_i) \cdot \left(\frac{1}{n} \sum_{i=1}^{n} \widehat{\pi}(X_i) \widehat{\sigma}^4(X_i) \right)^{-3/2}, \tag{51}$$

with

$$\widehat{\sigma}^2(X_i) = \frac{\sum_{u=1}^{n} K(\frac{X_i - X_u}{\widetilde{h}_{\text{cv}}}) \widehat{e}_u^2}{\sum_{v=1}^{n} K(\frac{X_i - X_v}{\widetilde{h}_{\text{cv}}})}.$$

Let $L_1 = \widetilde{T}_n(\widetilde{h}_{\text{test}})$ and $L_2 = \widetilde{T}_n(\widetilde{h}_{\text{cv}})$. To apply the test L_j for each $j = 1, 2$ to test $\mathcal{H}_{01}$, we propose the following procedure for computing the p value of L_j:

- Compute $\widehat{e}_i = Y_i - \widehat{\beta}(\widehat{\alpha} - X_i)$ and then generate a sequence of bootstrap resamples, $\{\widehat{e}_i^*\}$, given by $\widehat{e}_i^* = \widehat{\sigma}(X_i)\epsilon_i^*$, where $\{\epsilon_i^*\}$ is a sequence of iid bootstrap resamples generated from $\mathrm{N}(0, \Lambda^{-1})$ and $\widehat{\sigma}^2(\cdot)$ is defined as before.
- Generate $\widehat{Y}_i^* = \widehat{\beta}(\widehat{\alpha} - X_i) + \widehat{e}_i^*$. Compute the corresponding version, L_j^*, of L_j for each $j = 1, 2$ based on $\{\widehat{Y}_i^*\}$.
- Repeat the foregoing steps $M = 1{,}000$ times to find the bootstrap distribution of L_j^* and then compute the proportion in which $L_j < L_j^*$ for each $j = 1, 2$. This proportion is a simulated p value of L_j.

Our simulation results return the simulated p values of $\widehat{p}_1 = .102$ for L_1 and $\widehat{p}_2 = .072$ for L_2. Although both of the simulated p values suggest that there is insufficient evidence for rejecting the linearity in the drift at the 5% significance level, the evidence for accepting the linearity based on L_1 is stronger than that based on L_2.

Because our test $\widetilde{T}_n(\widetilde{h}_{\text{test}})$ involves no estimation biases, the process of computing the simulated p values is quite robust. Thus we believe that this improved test further reinforces the findings of Chapman and Pearson (2000) and Fan and Zhang (2003) that there is no definitive answer to the question of whether the short rate drift is actually nonlinear.

5. CONCLUSION

This article has addressed the issue of how to appropriately choose the bandwidth parameter when using a nonparametric kernel-based test. Both the size and power properties of the proposed test have been studied systematically. The established theory and methodology demonstrate that a suitable bandwidth can be chosen optimally after appropriately balancing the size and power functions. Furthermore, the new methodology has resulted in a closed-form representation for the leading term of such an optimal bandwidth in the finite-sample case.

Existing results (see, e.g., Li and Wang 1998; Li 1999; Fan and Linton 2003) demonstrate that this kind of nonparametric kernel test associated with a large sample critical value may not have good size and power properties. Our small and medium size sample studies using both simulated and real data examples show that the performance of such a test can be significantly improved when it is coupled with a power-based optimal bandwidth, as well as with a bootstrap simulated critical value.

We note that the established theory and methodology has various applications in providing solutions to some other related testing problems, in which nonparametric methods are involved. Future extensions also include dealing with cases in which both X_i and e_i may be strictly stationary time series.

APPENDIX: ASSUMPTIONS AND PROOFS

Here we give the necessary assumptions for the establishment of the main results given in Section 2, along with proofs.

A.1 Assumptions

Assumption A.1. (a) Assume that $\{e_i\}$ is a sequence of iid continuous random errors with $E[e_1] = 0$, $0 < \sigma^2 = E[e_1^2] = \sigma^2 < \infty$, and $E[e_1^6] < \infty$.

(b) We assume that $\{X_i\}$ is strictly stationary and α-mixing, with mixing coefficient $\alpha(t)$ defined by

$$\begin{aligned} \alpha(t) &= \sup\left\{ |P(A \cap B) - P(A)P(B)| : A \in \Omega_1^s, B \in \Omega_{s+t}^\infty \right\} \\ &\le C_\alpha \alpha^t \end{aligned} \tag{A.1}$$

for all $s, t \ge 1$, where $0 < C_\alpha < \infty$ and $0 < \alpha < 1$ are constants and Ω_i^j denotes the σ-field generated by $\{X_k : i \le k \le j\}$.

(c) Assume that $\{X_s\}$ and $\{e_t\}$ are independent for all $1 \le s \le t \le n$. Let $\pi(\cdot)$ be the marginal density such that $0 < \int \pi^3(x)\,dx < \infty$, and let $\pi_{\tau_1, \tau_2, \ldots, \tau_l}(\cdot)$ be the joint probability density of $(X_{1+\tau_1}, \ldots, X_{1+\tau_l})$ $(1 \le l \le 4)$. Assume that $\pi_{\tau_1, \tau_2, \ldots, \tau_l}(\cdot)$ for all $1 \le l \le 4$ exist and are continuous and bounded.

(d) Assume that the univariate kernel function $K(\cdot)$ is a symmetric and bounded probability density function. In addition, assume the existence of both $K^{(3)}(\cdot)$, the three-time convolution of $K(\cdot)$ with itself, and $K_2^{(2)}(\cdot)$, the two-time convolution of $K^2(\cdot)$ with itself.

(e) The bandwidth parameter h satisfies both $\lim_{n\to\infty} h = 0$ and $\lim_{n\to\infty} nh^d = \infty$.

Assumption A.2. (a) Let $\mathcal{H}_0$ be true. Then, for any sufficiently small $\varepsilon_1 > 0$ and some $B_{1L} > 0$,

$$\lim_{n\to\infty} P(\sqrt{n}\|\widehat{\theta} - \theta_0\| > B_{1L}) < \varepsilon_1,$$

where θ_0 is the same as defined in (2).

(b) Let $\mathcal{H}_1$ be true. Then, for any sufficiently small $\varepsilon_2 > 0$ and some $B_{2L} > 0$,

$$\lim_{n\to\infty} P(\sqrt{n}\|\widehat{\theta} - \theta_1\| > B_{2L}) < \varepsilon_2,$$

where θ_1 is the same as defined in (2).

(c) There exist some absolute constants $\varepsilon_3 > 0$ and $0 < B_{3L} < \infty$ such that the

$$\lim_{n\to\infty} P(\sqrt{n}\|\widehat{\theta}^* - \widehat{\theta}\| > B_{3L} | \mathcal{W}_n) < \varepsilon_3$$

holds in probability, where $\widehat{\theta}^*$ is as defined in the simulation procedure before Theorem 1.

(d) Let $m_\theta(x)$ be differentiable with respect to θ and let $\frac{\partial m_\theta(x)}{\partial \theta}$ be continuous in both x and θ. In addition, $E[(\frac{m_{\theta_0}(X_1)}{\partial \theta})(\frac{m_{\theta_0}(X_1)}{\partial \theta})^\tau]$ is a positive definite matrix, and

$$0 < \int \left\| \frac{\partial m_\theta(x)}{\partial \theta}\Big|_{\theta=\theta_0} \right\|^2 \pi^2(x)\,dx < \infty.$$

Assumption A.3. (a) Let $\{\Delta_n(x)\}$ be a sequence of continuous functions such that $0 < \int \Delta_n^2(x)\pi^2(x)\,dx < \infty$.

(b) Let C_n^2 satisfy $\lim_{n\to\infty} n\sqrt{h^d}C_n^2 = \infty$ and $\lim_{n\to\infty} nC_n^6 = 0$, where

$$C_n^2 = \frac{\int \Delta_n^2(x)\pi^2(x)\,dx}{\sigma^2\sqrt{2\nu_2\int K^2(v)\,dv}},$$

in which $\nu_2 = E[\pi^2(X_1)] < \infty$.

Assumptions A.1–A.3 are standard and justifiable conditions. Some detailed justifications are given in the supplemental material.

A.2 Technical Lemmas

Recall that, using $\lim_{n\to\infty} nh^d = \infty$,

$$\begin{aligned}\kappa_n &= \frac{\sqrt{h^d}\left(\frac{\mu_3^2K^2(0)}{nh^d} + \frac{4\mu_2^3\nu_3}{3}K^{(3)}(0)\right)}{\sigma_n^3} \equiv c_1\sqrt{h^d} + c_2\frac{1}{n\sqrt{h^d}} \\ &= c_1\sqrt{h^d}\left(1 + c_2c_1^{-1}\frac{1}{nh^d}\right) \approx c_1\sqrt{h^d}.\end{aligned} \tag{A.2}$$

To establish some useful lemmas without including nonessential technicalities, we introduce the following simplified notation:

$$\begin{aligned} a_{ij} &= \frac{1}{n\sqrt{h^d}\sigma_n}K\left(\frac{X_i - X_j}{h}\right), \\ L_n(h) &= \sum_{i=1}^n\sum_{j=1,\neq i}^n a_{ij}e_ie_j, \\ \rho(h) &= \frac{\sqrt{2}K^{(3)}(0)\int \pi^3(u)\,du}{3} \\ &\quad \times \left(\sqrt{\int \pi^2(u)\,du\int K^2(v)\,dv}\right)^{-3}\sqrt{h^d}. \end{aligned} \tag{A.3}$$

We need the following lemmas, the proofs of which are given in the supplemental material.

Lemma A.1. Suppose that the conditions of Theorem 2(a) hold. Then, for any h,

$$\sup_{x\in\mathbb{R}^1}|P(L_n(h)\le x) - \Phi(x) + \rho(h)(x^2-1)\phi(x)| = O(h^d). \tag{A.4}$$

Recall that $L_n(h) = \sum_{i=1}^n\sum_{j=1,\neq i}^n e_ia_{ij}e_j$ as defined in (A.3), and let

$$\begin{aligned}\overline{T}_n(h) &= \frac{h^{d/2}}{n\sigma_n}\sum_{i=1}^n\sum_{j=1,\neq i}^n \widehat{e}_iK_h(X_i-X_j)\widehat{e}_j \\ &= \frac{h^{d/2}}{n\sigma_n}\sum_{i=1}^n\sum_{j=1,\neq i}^n e_iK_h(X_i-X_j)e_j \\ &\quad + \frac{h^{d/2}}{n\sigma_n}\sum_{i=1}^n\sum_{j=1,\neq i}^n K_h(X_i-X_j)[m(X_i)-m_{\widehat{\theta}}(X_i)] \\ &\qquad \times [m(X_j)-m_{\widehat{\theta}}(X_j)] \\ &\quad + \frac{2h^{d/2}}{n\sigma_n}\sum_{i=1}^n\sum_{j=1,\neq i}^n e_iK_h(X_i-X_j)[m(X_j)-m_{\widehat{\theta}}(X_j)] \\ &\equiv L_n(h) + S_n(h) + D_n(h),\end{aligned} \tag{A.5}$$

where $S_n(h) = \frac{h^{d/2}}{n\sigma_n}\sum_{i=1}^n\sum_{j=1,\neq i}^n K_h(X_i - X_j)[m(X_i) - m_{\widehat{\theta}}(X_i)][m(X_j)-m_{\widehat{\theta}}(X_j)]$ and

$$D_n(h) = \frac{2h^{d/2}}{n\sigma_n}\sum_{i=1}^n\sum_{j=1,\neq i}^n e_iK_h(X_i-X_j)[m(X_j)-m_{\widehat{\theta}}(X_j)]. \tag{A.6}$$

Define $L_n^*(h)$, $S_n^*(h)$, and $D_n^*(h)$ as the corresponding versions of $L_n(h)$, $S_n(h)$, and $D_n(h)$ involved in (A.5), with (X_i, Y_i) and $\widehat{\theta}$ replaced by (X_i, Y_i^*) and $\widehat{\theta}^*$.

Lemma A.2. Suppose that the conditions of Theorem 2(a) hold. Then

$$\sup_{x\in\mathbb{R}^1}\left|P^*(L_n^*(h)\le x) - \Phi(x) + \rho(h)(x^2-1)\phi(x)\right| = O_P(h^d). \tag{A.7}$$

Lemma A.3. Suppose that the conditions of Theorem 2(a) hold. Then, under $\mathcal{H}_0$,

$$E[S_n(h)] = O(\sqrt{h^d}) \quad \text{and} \quad E[D_n(h)] = o(\sqrt{h^d}), \tag{A.8}$$

$$E^*[S_n^*(h)] = O_P(\sqrt{h^d}) \quad \text{and} \quad E^*[D_n^*(h)] = o_P(\sqrt{h^d}), \tag{A.9}$$

and

$$\begin{aligned} &E[S_n(h)] - E^*[S_n^*(h)] = O_P(\sqrt{h^d}) \quad \text{and} \\ &E[D_n(h)] - E^*[D_n^*(h)] = o_P(\sqrt{h^d}) \end{aligned} \tag{A.10}$$

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

Lemma A.4. Suppose that the conditions of Theorem 2(b) hold. Then, under $\mathcal{H}_1$,

$$\lim_{n\to\infty} E[S_n(h)] = \infty \quad \text{and} \quad \lim_{n\to\infty}\frac{E[D_n(h)]}{E[S_n(h)]} = 0. \tag{A.11}$$

A.3 Proof of Theorem 1

A.3.1 Proof of (14). Recall from (12) and (A.5)–(A.6) that

$$\widehat{T}_n(h) = (L_n(h) + S_n(h) + D_n(h))\cdot\frac{\sigma_n}{\widehat{\sigma}_n} + o_P(\sqrt{h^d}) \tag{A.12}$$

and

$$\widehat{T}_n^*(h) = (L_n^*(h) + S_n^*(h) + D_n^*(h))\cdot\frac{\sigma_n}{\widehat{\sigma}_n^*} + o_P(\sqrt{h^d}), \tag{A.13}$$

where σ_n^2, $\widehat{\sigma}_n^2$, and $\widehat{\sigma}_n^{*2}$ are as defined in (8), (11), and (13).

In view of Assumption A.2 and Lemmas A.1–A.3, we may ignore any terms with orders above $\sqrt{h^d}$ and then consider the following approximations:

$$\begin{aligned}\widehat{T}_n(h) &= L_n(h) + E[S_n(h)] + o_P(\sqrt{h^d}), \\ \widehat{T}_n^*(h) &= L_n^*(h) + E^*[S_n^*(h)] + o_P(\sqrt{h^d}).\end{aligned} \tag{A.14}$$

Let $s(h) = E[S_n(h)]$ and $s^*(h) = E^*[S_n^*(h)]$. We then apply Lemmas A.1 and A.2 to obtain that, uniformly over $x\in\mathbb{R}^1$,

$$\begin{aligned} P(\widehat{T}_n(h)\le x) &= P(L_n(h)\le x - s(h) + o_P(\sqrt{h^d})) \\ &= \Phi(x - s(h)) - \rho(h)((x-s(h))^2-1) \\ &\quad \times \phi(x-s(h)) + o(\sqrt{h^d}), \\ P^*(\widehat{T}_n^*(h)\le x) &= P^*(L_n^*(h)\le x - s^*(h) + o_P(\sqrt{h^d})) \\ &= \Phi(x - s^*(h)) - \rho(h)((x-s^*(h))^2-1) \\ &\quad \times \phi(x-s^*(h)) + o_P(\sqrt{h^d}). \end{aligned} \tag{A.15}$$

Theorem 2(a) follows consequently from (A.10) and (A.15).

A.3.2 Proof of (15). In view of the fact that $P^*(\widehat{T}_n^*(h) \ge l_\alpha^*) = \alpha$ and the conclusion from Theorem 1(a) that

$$P(\widehat{T}_n(h) \ge l_\alpha^*) - P^*(\widehat{T}_n^*(h) \ge l_\alpha^*) = O_P(\sqrt{h^d}), \tag{A.16}$$

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

A.4 Proof of Theorem 2

It follows from Lemmas A.1–A.4 that

$$\begin{aligned}\alpha_n(h) &= P(\widehat{T}_n(h) \ge l_\alpha|\mathcal{H}_0)\\ &= P\big(L_n(h) \ge l_\alpha - S_n(h) + o_P(S_n(h))|\mathcal{H}_0\big)\\ &= 1 - P\big(L_n(h) \le l_\alpha - S_n(h) + o_P(S_n(h))|\mathcal{H}_0\big), \qquad (A.17)\\ \alpha_n^*(h) &= P(\widehat{T}_n(h) \ge l_\alpha^*|\mathcal{H}_0)\\ &= P\big(L_n(h) \ge l_\alpha^* - S_n(h) + o_P(S_n(h))|\mathcal{H}_0\big)\\ &= 1 - P\big(L_n(h) \le l_\alpha^* - S_n(h) + o_P(S_n(h))|\mathcal{H}_0\big), \qquad (A.18)\\ \beta_n(h) &= P(\widehat{T}_n(h) \ge l_\alpha|\mathcal{H}_1)\\ &= P\big(L_n(h) \ge l_\alpha - S_n(h) + o_P(S_n(h))|\mathcal{H}_1\big)\\ &= 1 - P\big(L_n(h) \le l_\alpha - S_n(h) + o_P(S_n(h))|\mathcal{H}_1\big), \qquad (A.19)\end{aligned}$$

and

$$\begin{aligned}\beta_n^*(h) &= P(\widehat{T}_n(h) \ge l_\alpha^*|\mathcal{H}_1)\\ &= P\big(L_n(h) \ge l_\alpha^* - S_n(h) + o_P(S_n(h))|\mathcal{H}_1\big)\\ &= 1 - P\big(L_n(h) \le l_\alpha^* - S_n(h) + o_P(S_n(h))|\mathcal{H}_1\big). \qquad (A.20)\end{aligned}$$

Using Assumptions A.2(d) and A.3, a Taylor expansion of $m_\theta(\cdot)$ at θ_0 implies that for sufficiently large n,

$$S_n(h) = C_0(m)\sqrt{h^d}(1 + o_P(1)) \quad \text{under } \mathcal{H}_0 \tag{A.21}$$

and

$$S_n(h) = nC_n^2\sqrt{h^d}(1 + o_P(1)) \quad \text{under } \mathcal{H}_1 \tag{A.22}$$

hold in probability, where C_n^2 is as defined in Theorem 2(b). The proof of Theorem 2 then follows from (A.15) and (A.17)–(A.22).

A.5 Proof of Theorem 3

The proof follows from that of Theorem 2. The details are given in the supplemental material.

[Received November 2006. Revised August 2008.]

REFERENCES

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

Andrews, D. W. (1997), "A Conditional Kolmogorov Test," *Econometrica*, 65, 1097–1028.

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

Chapman, D., and Pearson, N. (2000), "Is the Short Rate Drift Actually Nonlinear?" *Journal of Finance*, 54, 355–388.

Chen, S. X., Härdle, W., and Li, M. (2003), "An Empirical Likelihood Goodness-of-Fit Test for Time Series," *Journal of the Royal Statistical Society*, Ser. B, 65, 663–678.

Dette, H. (1999), "A Consistent Text for the Functional Form of a Regression Based on a Difference of Variance Estimators," *The Annals of Statistics*, 27, 1012–1040.

Dette, H., and Spreckelsen, I. (2004), "Some Comments on Specification Tests in Nonparametric Absolutely Regular Processes," *Journal of Time Series Analysis*, 25, 159–172.

Fan, J. (1996), "Test of Significance Based on Wavelet Thresholding and Neyman's Truncation," *Journal of the American Statistical Association*, 434, 674–688.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman & Hall.

Fan, J., and Yao, Q. (2003), *Nonlinear Time Series: Parametric and Nonparametric Methods*, New York: Springer.

Fan, J., and Zhang, C. M. (2003), "A Re-Examination of Stanton's Diffusion Estimation With Applications to Financial Model Validation," *Journal of the American Statistical Association*, 461, 118–134.

Fan, J., Zhang, C. M., and Zhang, J. (2001), "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *The Annals of Statistics*, 29, 153–193.

Fan, Y., and Linton, O. (2003), "Some Higher-Order Theory for a Consistent Nonparametric Model Specification Test," *Journal of Statistical Planning and Inference*, 109, 125–154.

Franke, J., Kreiss, J. P., and Mammen, E. (2002), "Bootstrap of Kernel Smoothing in Nonlinear Time Series," *Bernoulli*, 8, 1–38.

González-Manteiga, W., and Cao, R. (1993), "Testing the Hypothesis of a General Linear Model Using Nonparametric Regression Estimation," *Test*, 2, 161–188.

Härdle, W., and Mammen, E. (1993), "Comparing Nonparametric Versus Parametric Regression Fits," *The Annals of Statistics*, 21, 1926–1947.

Hart, J. (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, New York: Springer.

Hjellvik, V., and Tjøstheim, D. (1995), "Nonparametric Tests of Linearity for Time Series," *Biometrika*, 82, 351–368.

Hjellvik, V., Yao, Q., and Tjøstheim, D. (1998), "Linearity Testing Using Local Polynomial Approximation," *Journal of Statistical Planning and Inference*, 68, 295–321.

Horowitz, J. (2003), "Bootstrap Methods for Markov Processes," *Econometrica*, 71, 1049–1082.

Horowitz, J., and Spokoiny, V. (2001), "An Adaptive Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69, 599–631.

Hunsberger, S., and Follmann, D. A. (2001), "Testing for Treatment and Interaction Effects in Semiparametric Analysis of Covariance," *Statistics in Medicine*, 20, 1–19.

Juhl, T., and Xiao, Z. (2005), "A Nonparametric Test for Changing Trends," *Journal of Econometrics*, 127, 179–199.

Kulasekera, K. B., and Wang, J. (1997), "Smoothing Parameter Selection for Power Optimality in Testing of Regression Curves," *Journal of the American Statistical Association*, 438, 500–511.

Li, Q. (1999), "Consistent Model Specification Tests for Time Series Econometric Models," *Journal of Econometrics*, 92, 101–147.

Li, Q., and Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton, NJ: Princeton University Press.

Li, Q., and Wang, S. (1998), "A Simple Consistent Bootstrap Test for a Parametric Regression Functional Form," *Journal of Econometrics*, 87, 145–165.

Nishiyama, Y., and Robinson, P. (2000), "Edgeworth Expansions for Semiparametric Averaged Derivatives," *Econometrica*, 68, 931–980.

——— (2005), "The Bootstrap and the Edgeworth Correction for Semiparametric Averaged Derivatives," *Econometrica*, 73, 903–948.

Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall.

Weihrather, G. (1993), "Testing Linear Regression Model Against Nonparametric Alternatives," *Metrika*, 40, 367–379.

Zhang, C. M., and Dette, H. (2004), "A Power Comparison Between Nonparametric Regression Tests," *Statistics & Probability Letters*, 66, 289–301.

Zheng, J. X. (1996), "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263–289.

Zhu, L. (2005), *Nonparametric Monte Carlo Tests and Their Applications*, New York: Springer.