Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 950–972
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

An adaptive empirical likelihood test for parametric time series regression models

Song Xi Chen[a,*], Jiti Gao[b]

[a]Department of Statistics, Iowa State University, Ames, IA 50010, USA
[b]School of Mathematics and Statistics, The University of Western Australia, Crawley WA 6009, Australia

Available online 19 January 2007

Abstract

We propose an adaptive empirical likelihood (EL) test for a parametric regression model against a class of alternatives for weakly dependent time series observations. The test is formulated by maximizing a standardized version of the EL statistic over a set of smoothing bandwidths. It is demonstrated that the proposed test is able to distinguish the null hypothesis from a series of local alternatives at an optimal rate.

JEL classification: C12; C14; C32

Keywords: Empirical likelihood; Goodness-of-fit test; Kernel estimation; Rate-optimal test; Nonparametric time series

1. Introduction

Consider a time series heteroscedastic regression model of the form

$$Y_t = m(X_t) + \sigma(X_t)e_t, \quad t = 1, 2, \ldots, n, \tag{1.1}$$

where both $m(\cdot)$ and $\sigma(\cdot)$ are unknown functions defined over R^d, both $\{X_t\}$ and $\{e_t\}$ are allowed to be weakly dependent time series, and $\{e_t\}$ is an error process with zero mean and unit variance. Note that we need not assume the independence between $\{X_s\}$ and $\{e_t\}$ for all $s \leqslant t$ in this paper.

*Corresponding author: Tel.: +1 5152 942729; fax: +1 5152 944040.
E-mail addresses: songchen@iastate.edu (S.X. Chen), jiti.gao@uwa.edu.au (J. Gao).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.12.002

Suppose that $\{m_\theta(\cdot)|\theta \in \Theta\}$ is a family of parametric functions, where $\theta \in R^q$ is an unknown parameter belonging to a parameter space Θ. This paper considers testing the validity of a parametric specification of $m_\theta(x)$ against a series of local alternatives, namely we test

$$\mathrm{H}_0 : m(x) = m_\theta(x) \text{ versus } \mathrm{H}_1 : m(x) = m_\theta(x) + C_n \Delta_n(x) \quad \text{for all } x \in S. \tag{1.2}$$

In the latter case, C_n is a nonrandom sequence tending to zero as $n \to \infty$, and $\Delta_n(x)$ is a sequence of functions in R^d and S is a compact set in R^d. Both C_n and $\Delta_n(x)$ characterize the departure of the local alternatives from the hypothesised parametric family $\{m_\theta(\cdot)|\theta \in \Theta\}$.

Test evaluation in the form of local alternatives can reveal much finer details of the power property of a test than what can be achieved when the evaluation is carried out in the form of a fixed alternative. One of the main interests of this paper is to construct a test that can reach the smallest possible C_n while being consistent over a reasonably large class of local alternative models, for instance a Hölder class.

Nonparametric kernel estimation is well studied for both independent and dependent observations as documented in Härdle (1990), Fan and Gijbels (1996) and Fan and Yao (2005). A number of tests for H_0 have been developed by comparing a nonparametric kernel estimator of the conditional mean function with the hypothesised parametric model under H_0 via certain distance measures; for instance, the tests of Eubank and Spiegelman (1990), Härdle and Mammen (1993), Hjellvik and Tjøstheim (1995), Hart (1997), Hjellvik et al. (1998), Li and Wang (1998), Li (1999), Fan and Li (2000), Aït-Sahalia et al. (2001), Gao and King (2005) and others. McKeague and Zhang (1994) consider testing separately the conditional mean and variance specifications of a nonlinear time series regression based on some cumulative measures. Fan and Zhang (2003) propose separate tests for both the conditional mean and the conditional variance of a diffusion model. Zhang and Dette (2003) compare the power of three kernel based tests. Wang and Van Keilegom (2005) propose a test based on the idea of ANOVA with a large number of factor levels for dependent observations.

A common feature of the tests based on the kernel method is the involvement of a single bandwidth h in formulation of the goodness-of-fit measure, which creates two limitations. One is that the tests can be sensitive to the single bandwidth used. Another is that these tests are consistent against local alternatives of form (1.2) only when C_n, the crucial factor that determines the distance between H_0 and H_1, is at the rate of $n^{-1/2}h^{-d/4}$ or larger. This is larger than $n^{-1/2}$, which is the rate achieved by tests for a finite dimensional parameter in a standard parametric setting and by tests based on the empirical distribution function of the estimated residuals in the case of $\Delta_n(x) \equiv \Delta(x)$ for all n (Andrews, 1997). In the context of independent observations, Horowitz and Spokoiny (2001) propose an adaptive test that combines a version of the Härdle–Mammen test statistics over a set of bandwidths. The test is called adaptive and rate optimal if it adapts to the unknown smoothness of the local alternative hypothesis and is able to achieve the optimal order for C_n in the minimax sense of Spokoiny (1996) and Ingster and Suslina (2003). A similar idea is also given in Fan (1996).

We consider in this paper two extensions of the adaptive test of Horowitz and Spokoiny. One is the inclusion of weakly dependent observations so that various parametric time series models can be brought to test. The other is to use the empirical likelihood (EL) of Owen (1988) to formulate the test statistic, which is designed to equip the test statistic with some favorable features of the EL. One feature is that the test statistic takes the variation of the kernel estimation into consideration and studentize it automatically. We show that the results established by Horowitz and Spokoiny (2001) are valid under these extensions. Comparing with tests based on a fixed bandwidth, a test based on a set of bandwidths is less dependent on a particular choice of bandwidth and hence is more robust against the choice of smoothing bandwidths. To accurately approximate the distribution of the adaptive test statistic, a bootstrap simulation procedure is used to profile the critical value of the test. This combination of the EL and the bootstrap simulation utilizes the good features of the EL for the construction of test statistics and the effectiveness of the bootstrap simulation in distribution approximation.

A key ingredient of the proposed test is the EL method introduced in Owen (1988, 1990). It is a technique that allows construction of nonparametric likelihood for a parameter of interest. Despite the fact that it is intrinsically nonparametric, it possesses two important properties of a parametric likelihood: the Wilks' theorem and the Bartlett correction. The widest EL formulation is the framework of Qin and Lawless (1994) for parameters defined by a set of generalized moment restrictions, which is commonly used in econometrics. It parallels the generalized method of moments (GMM) of Hansen (1982) with the attractive feature that it achieves the two steps GMM in one step as shown in Imbens (1997). Kitamura (1997) extends the framework of Qin and Lawless to weakly dependent observations. Chen and Cui (2006, 2007) show that the Bartlett correction is still valid for EL under the general moment restrictions with or without nuisance parameters. Hjort et al. (2004) establish general results in the presence of nuisance parameters.

The rest of this paper is organized as follows. Section 2 outlines the EL formulation of the test statistic. The main results regarding the adaptive EL test and its rate-optimal property are given in Section 3. Section 4 presents simulation results. All the technical proofs are provided in the Appendix.

2. Empirical likelihood test statistics

Like existing kernel-based goodness-of-fit tests, our test is based on a kernel estimator of the conditional mean function $m(x)$. Let K be a rth order d-dimensional kernel, h be a smoothing bandwidth and $K_h(u) = h^{-d}K(u/h)$. The Nadaraya–Watson (NW) estimator of $m(x)$ is

$$\hat{m}(x) = \frac{\sum_{t=1}^{n} K_h(x - X_t)Y_t}{\sum_{t=1}^{n} K_h(x - X_i)}.$$

Let $\tilde{\theta}$ be a $\sqrt{n}$-consistent estimator of θ under H_0. Like Härdle and Mammen (1993), let

$$\tilde{m}_{\tilde{\theta}}(x) = \frac{\sum_{t=1}^{n} K_h(x - X_t)m_{\tilde{\theta}}(X_t)}{\sum_{t=1}^{n} K_h(x - X_t)}$$

be a kernel smooth of the parametric estimate $m_{\tilde{\theta}}(x)$ with the same kernel and bandwidth as in $\hat{m}(x)$. This is designed to avoid the bias of the kernel estimator getting into the asymptotic distribution of the test statistic.

Let $Q_t(x) = K_h(x - X_t)\{Y_t - \tilde{m}_{\tilde{\theta}}(x)\}$. At an arbitrary $x \in S$, let $p_t(x)$ be a sequence of nonnegative real functions representing weights allocated to each (X_t, Y_t). The EL for

$m(x)$ evaluated at the smoothed parametric model $\tilde{m}_{\tilde{\theta}}(x)$ is

$$L\{\tilde{m}_{\tilde{\theta}}(x)\} = \max \prod_{t=1}^{n} p_t(x) \tag{2.1}$$

subject to $\sum_{t=1}^{n} p_t(x) = 1$ and $\sum_{t=1}^{n} p_t(x)Q_t(x) = 0$. A standard derivation in EL shows that the optimal weights are

$$p_t(x) = \frac{1}{n}\{1 + \lambda(x)Q_t(x)\}^{-1}, \tag{2.2}$$

where $\lambda(x)$ is the solution of

$$\sum_{t=1}^{n} \frac{Q_t(x)}{1 + \lambda(x)Q_t(x)} = 0. \tag{2.3}$$

As the EL is maximized at $p_t(x) = n^{-1}$, the log-EL ratio is

$$\ell\{\tilde{m}_{\tilde{\theta}}(x)\} = -2\log[L\{\tilde{m}_{\tilde{\theta}}(x)\}n^n].$$

The EL test statistic at a given bandwidth h is

$$\ell(\tilde{m}_{\tilde{\theta}}; h) = \int \ell\{\tilde{m}_{\tilde{\theta}}(x)\}\pi(x)\,\mathrm{d}x, \tag{2.4}$$

where $\pi(\cdot)$ is a nonnegative weight function supported on the compact set $S \subset R^d$.

Let $R(K) = \int K^2(x)\,\mathrm{d}x$, $v(x) = R(K)\sigma^2(x)f^{-1}(x)$ and

$$C(K,\pi) = 2R^{-2}(K)\int \pi^2(x)\,\mathrm{d}x \int (K^{(2)}(x))^2\,\mathrm{d}x, \tag{2.5}$$

where $K^{(2)}$ is the convolution of K.

For weakly dependent observations, Chen et al. (2003) show that as $n \to \infty$

$$h^{-d/2}\left\{\ell(\tilde{m}_{\tilde{\theta}}; h) - 1 - h^{d/2}\int v^{-1/2}(x)\Delta_n^2(x)\pi(x)\,\mathrm{d}x\right\} \stackrel{\mathrm{D}}{\to} \mathrm{N}(0, C(K,\pi)). \tag{2.6}$$

They then propose a single bandwidth-based EL test and its implementation is based on critical values obtained by simulating a Gaussian random field. There are two other single bandwidth-based tests which are related to the proposal of Chen et al. (2003). Fan and Zhang (2003) propose a sieve EL test for testing a general varying-coefficient regression model through extending the generalized likelihood ratio test of Fan et al. (2001) for independent data. Also for independent data, Tripathi and Kitamura (2003) propose an EL test for conditional moment restrictions.

These three tests have displayed an interesting diversity in test statistic formulations via the EL. The basic idea of the EL is to maximize a product of probability weights (or a sum of logarithm of probability weights) allocated to observations under certain constraints which characterize the functional curve to be tested. Fan and Zhang (2003) apply kernel smoothing in both the objective function and the constraints, whereas Tripathi and Kitamura (2003) smooth only the objective function, and Chen et al. (2003) smooth only the constraints. Fan and Zhang (2003), and Chen et al. (2003) both have a two step formulation of their test statistics: first construct local EL-based statistics at a fixed point and then sum them over to form the final test statistics; hence both are sieve EL statistics. The formulation in Tripathi and Kitamura (2003) is of one step with a global objective function over the entire sample. These tests all involve only a single bandwidth and hence lack the adaptive property of the new EL test we are to propose. In particular, like all nonparametric kernel-based goodness of tests based on a single bandwidth, these tests are consistent only if C_n is at the order of $n^{-1/2}h^{-d/4}$ or larger, indicating that C_n has to converge to zero more slowly than $n^{-1/2}$.

To reduce the order of C_n to the smallest possible, we employ the adaptive test procedure of Horowitz and Spokoiny (2001) for the EL test statistic as follows. Let

$$\mathscr{H}_n = \left\{ h = h_{\max} a^k : \ h \geqslant h_{\min}, \ k = 0, 1, 2, \ldots J_n - 1 \right\} \tag{2.7}$$

be a set of bandwidths, where $0 < a < 1$, $J_n = \log_{1/a}(h_{\max}/h_{\min})$ is the number of bandwidths, $h_{\max} = c_{\max}(\log \log(n))^{-1/d}$ and $h_{\min} = c_{\min} n^{-\gamma}$ for $0 < \gamma < \frac{1}{3d}$ and some positive constants $-\infty < c_{\min}, c_{\max} < \infty$. The choice of $h_{\max}$ is vital in reducing C_n to almost $n^{-1/2}$ rate in the case of $\Delta_n(\cdot) \equiv \Delta(\cdot)$. In view of the fact that $\mathrm{E}\{\ell(\tilde{m}_{\tilde{\theta}}; h)\} = 1$ under H_0 and $\mathrm{var}\{\ell(\tilde{m}_{\tilde{\theta}}; h)\} = C(K, \pi)h^d$ as given in (2.5) the adaptive EL test statistic is proposed as follows:

$$L_n = \max_{h \in \mathscr{H}_n} \frac{\ell(\tilde{m}_{\tilde{\theta}}; h) - 1}{\sqrt{C(K, \pi)h^d}}. \tag{2.8}$$

Here the variance coefficient $C(K, \pi)$ of $\ell(\tilde{m}_{\tilde{\theta}}; h)$ is completely known upon given the kernel K and the weight function π, which is due to EL's ability to studentize internally.

Simulation procedure: Let l_α $(0 < \alpha < 1)$ be the $1 - \alpha$ quantile of the finite-sample distribution of L_n where α is the significance level of the test. We propose the following bootstrap simulation procedure to approximate l_α:

1. For each $t = 1, 2, \ldots, n$, let $Y_t^* = m_{\tilde{\theta}}(X_t) + \sigma_n(X_t)e_t^*$, where the set of original explanatory variables $\chi_n = (X_1, \ldots, X_n)$ acts in the resampling as a fixed design, $\sigma_n(\cdot)$ is a consistent estimator of $\sigma(\cdot)$, $\{e_t^*\}$ is independent of $\{X_s\}$ for all $s, t \geqslant 1$, and sampled randomly from a specified distribution with $\mathrm{E}[e_t^*] = 0$, $\mathrm{E}[e_t^{*2}] = 1$ and $\mathrm{E}[|e_t^*|^{4+\delta}] < \infty$ for some $\delta > 0$. Let L_n^* be the corresponding version of L_n with $\{Y_t\}$ replaced by $\{Y_t^*\}$. Define l_α^* as the $1 - \alpha$ conditional quantile of L_n^* given χ_n.
2. Let $\hat{\theta}^*$ be the estimate of θ based on the resample $\{(X_t, Y_t^*)\}_{t=1}^n$. Compute the statistic $\hat{L}_n^*$ by replacing Y_t and $\tilde{\theta}$ with Y_t^* and $\hat{\theta}^*$ according to (2.8). An Monte Carlo estimate of l_α^* is $\hat{l}_\alpha^*$, the $1 - \alpha$ quantile of the empirical distribution of $\hat{L}_n^*$, which can be obtained by repeating step 1 many times.

The estimator $\sigma_n^2(\cdot)$ can be the following kernel estimator:

$$\sigma_n^2(x) = \frac{\sum_{t=1}^n K_b(x - X_t)\{Y_t - \hat{m}(x)\}^2}{\sum_{t=1}^n K_b(x - X_t)} \tag{2.9}$$

with a bandwidth b satisfying $nh_{\min}b^d \to \infty$ as $n \to \infty$. Existing results (Bosq, 1998) may be used to show that $\sigma_n^2(x)$ converges to $\sigma^2(x)$ uniformly in $x \in S$ with S being a compact subset of R^d under either Assumptions A.1 and A.2 for the random design case or Assumptions 3–6 of Horowitz and Spokoiny (2001) for the fixed design case. Thus, $\sigma_n^2(\cdot)$ can be used under Horowitz–Spokoiny regularity conditions, in which $\{X_i\}$ is a sequence of fixed designs and $\{e_i\}$ is a sequence of independent errors.

The conditional quantile l_α^* satisfies $\mathrm{P}(L_n^* \geqslant l_\alpha^* | \chi_n) = \alpha$. By choosing the number of the Monte Carlo resamples sufficiently large, the approximation error of $\hat{l}_\alpha^*$ to l_α^* can be made negligible. Hence we will just use l_α^* in the proposal of the test procedure which rejects H_0 if $L_n > l_\alpha^*$.

There are several different approaches we could use for generating $\{e_t^*\}$ and $\{Y_t^*\}$ for the bootstrap procedure. For the generation of $\{e_t^*\}$, our theory does not require the specified distribution of e_t^* as needed in Horowitz and Spokoiny (2001). In order to compare the finite-sample performance of our test with that proposed by Horowitz and Spokoiny (2001), however, we generate $\{e_t^*\}$ from $\mathrm{N}(0,1)$ in Section 4. For the generation of $\{Y_t^*\}$, we employ partially the proposed regression bootstrap method by Franke et al. (2002). One of the advantages of using the proposed regression bootstrap method is that it is also applicable to the autoregressive case of $X_t = (Y_{t-1}, \ldots, Y_{t-d})$.

3. Main results

The theoretical properties of the proposed adaptive EL test are established in this section. We start with the following theorem which shows that the EL test has an asymptotically correct size.

Theorem 3.1. *Suppose Assumptions* A.1–A.3(i)(ii)(iv) *hold. Then under* H_0,

$$\lim_{n\to\infty} \mathrm{P}(L_n > l_\alpha^*) = \alpha.$$

To establish the power properties of the adaptive EL test, we define the distance between m and the parametric family $\mathscr{M}$ as

$$\rho(m, \mathscr{M}) = \left[\inf_{\theta\in\Theta}\left(\int_{x\in S} [m_\theta(x) - m(x)]^2 f(x)\,\mathrm{d}x\right)\right]^{1/2}. \tag{3.1}$$

The consistency of the test against a fixed alternative is established in Theorem 3.2.

Theorem 3.2. *Assume that Assumptions* A.1–A.3(i)(iii)(iv) *hold. If there is a* $C_\rho > 0$ *such that* $\rho(m, \mathscr{M}) \geqslant C_\rho$ *for* $n \geqslant n_0$ *with some large* n_0, *then*

$$\lim_{n\to\infty} \mathrm{P}(L_n > l_\alpha^*) = 1.$$

We then consider the consistency of the EL test against a special case of H_1 of the form

$$m(x) = m_\theta(x) + C_n \Delta(x), \tag{3.2}$$

where $C_n \to 0$ as $n \to \infty$, $\theta \in \Theta$ and for positive and finite constants D_1, D_2 and D_3,

$$0 < D_1 \leqslant \int_{x\in S} \Delta^2(x) f(x)\,\mathrm{d}x \leqslant D_2 < \infty \quad \text{and} \quad \rho(m, \mathscr{M}) \geqslant D_3 C_n. \tag{3.3}$$

This is a special form of H_1 in (1.2) with the local deviation function $\Delta_n(x) = \Delta(x)$ for all n.

Theorem 3.3. *Assume Assumptions* A.1–A.3(i)(iii). *Let Assumption* A.3(iv) *hold with* $h_{\max} = C_h(\log\log(n))^{-1/d}$ *for some finite constant* C_h. *Let* m *satisfy* (3.2) *and* (3.3) *with* $C_n \geqslant Cn^{-1/2}\sqrt{\log\log(n)}$ *for some constant* $C > 0$. *Then*

$$\lim_{n\to\infty} \mathrm{P}(L_n > l_\alpha^*) = 1.$$

Theorem 3.3 indicates that when the local alternatives are different from H_0 in a fixed direction given by $\Delta(x)$, the proposed test almost attains the parametric rate of $n^{-1/2}$ for the convergence of C_n which is a substantial improvement over the fixed bandwidth-based test.

Let us now discuss the consistency of the adaptive EL test against local alternatives over a Hölder smoothness class, which is richer than the class of the local alternative considered in Theorem 3.3. We introduce the following notation. Let $j = (j_1, \ldots, j_d)$ where $j_1, \ldots, j_d \geqslant 0$ are integers, $|j| = \sum_{k=1}^{d} j_d$ and $D^j m(x) = \frac{\partial^{|j|} m(x)}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}}$ whenever the derivative exists. Define the Hölder norm $\|m\|_{H,s} = \sup_{x \in S} \sum_{|j| \leqslant s} (|D^j m(x)|)$. The smoothness class that we consider consists of functions $m \in S(H,s) \equiv \{m : \|m\|_{H,s} \leqslant C_H\}$ for some unknown s and $C_H < \infty$. For $s \geqslant \max(2, d/4)$ and all sufficiently large $D_m < \infty$, define

$$B_{H,n} = \left\{ m \in S(H,s) : \ \rho(m, \mathscr{M}) \geqslant D_m \left(n^{-1} \sqrt{\log \log(n)} \right)^{2s/(4s+d)} \right\}. \tag{3.4}$$

A test is said to be consistent uniformly over $B_{H,n}$ if

$$\lim_{n \to \infty} \inf_{m \in B_{H,n}} \mathrm{P}(\mathrm{H}_0 \text{ is rejected against } m) = 1. \tag{3.5}$$

The optimal rate of testing is the fastest rate at which C_n approaches to zero while maintaining consistent uniformly over $B_{H,n}$.

Theorem 3.4. *Assume that Assumptions* A.1–A.3 *all hold. Let m satisfy* (1.2) *under* H_1 *and* (3.4). *Then for* $0 < \alpha < 1$ *and* $B_{H,n}$ *defined in* (3.4),

$$\lim_{n \to \infty} \inf_{m \in B_{H,n}} \mathrm{P}(L_n > l_\alpha^*) = 1.$$

Theorem 3.4 shows that the adaptive EL test is uniformly consistent over alternatives within a Hölder class of smooth functions whose distance from the parametric counterparts approaches zero at the rate of $\left(n^{-1}\sqrt{\log \log(n)}\right)^{2s/(4s+d)}$, which is the fastest possible in the minimax sense of Ingster and Suslina (2003) and Spokoiny (1996). The most striking property of Theorem 3.4 is that it achieves the best rate of convergence for C_n without knowing s, the degree of smoothness. This is the reason behind the term "adaptive and rate optimal" by Horowitz and Spokoiny (2001) when describing their test. We show that the same property holds for the proposed EL test with weakly dependent observations.

In order to show that the conclusions of Theorems 3.1–3.4 hold unconditionally, in view of the dominated convergence theorem, it suffices to show that the conclusions of Theorems 3.1–3.4 all hold in probability with respect to the joint distribution of $\chi_n = (X_1, \ldots, X_n)$. For example, we need only to show that

$$\lim_{n \to \infty} \mathrm{P}(L_n > l_\alpha^* | \chi_n) = \alpha \quad \text{in probability}$$

in Theorem 3.1.

4. Simulation results

We carried out two simulation studies which were designed to evaluate the empirical performance of the proposed adaptive EL test. In the first simulation study, we conducted simulation for the following regression model used in Horowitz and Spokoiny (2001):

$$Y_i = \beta_0 + \beta_1 X_i + (5/\tau)\phi(X_i/\tau) + \varepsilon_i, \tag{4.1}$$

where the $\{\varepsilon_i : i \geqslant 1\}$ are independent and identically distributed (i.i.d.) with zero mean and constant variance, $\{X_i\}$ are univariate design points to be sampled from $\mathrm{N}(0, 25)$ distribution truncated at its 5th and 95th percentiles, $\theta = (\beta_0, \beta_1)^\tau = (1, 1)^\tau$ is chosen as the true vector of parameters and ϕ is the standard normal density function.

The null hypothesis $\mathrm{H}_0 : m(x) = \beta_0 + \beta_1 x$ specifies a linear regression corresponding to $\tau = 0$, whereas the alternative hypothesis $\mathrm{H}_1 : m(x) = \beta_0 + \beta_1 x + (5/\tau)\phi(x/\tau)$ for $\tau = 1.0$ and 0.25. Three distributions of ε_i are considered in Horowitz and Spokoiny (2001): a normal distribution $\mathrm{N}(0, 4)$, a mixture of $\mathrm{N}(0, 1.56)$ and $\mathrm{N}(0, 25)$ with mixture proportions being 0.9 and 0.1, respectively, and an Type I extreme value distribution with variance 4. The reader should refer to Horowitz and Spokoiny (2001) for details on the designs X_i, the three distributions of ε_i and other aspects of the simulation. We used the same number of simulations, the bootstrap resamples and estimation procedures for θ as in Horowitz and Spokoiny (2001). We also employed the same kernel, the same bandwidth set $\mathscr{H}_n$ and the same distribution for $\{e_i^*\}$ in the Monte Carlo simulation procedure as in Horowitz and Spokoiny (2001). Like Horowitz and Spokoiny test (HS test), the nominal size of the test was 5%.

Table 1a summarizes the performance of the adaptive EL test by adding one column to Table 1 of Horowitz and Spokoiny (2001). Our results show that the proposed adaptive EL test has slightly better power than the adaptive test of Horowitz and Spokoiny (2001), while the sizes are similar to those of Horowitz and Spokoiny (2001). This may not be surprising as the two tests are equivalent in the first order. The differences between the two tests are: (i) the EL test statistic carries out the studentizing implicitly and (ii) certain higher order features like the skewness and kurtosis are reflected in the EL statistic. These might be the underlying cause for the slightly better power observed for the EL test.

We also compare single bandwidth-based EL test proposed in Chen, Härdle and Li (2003, CHL test) and the single bandwidth-based HS test at the above chosen fixed bandwidths for the case where the distribution of $\{\varepsilon_i\}$ is a mixture of normal distributions as used in Horowitz and Spokoiny (2001). Table 1b reports the finite–sample comparison results of the CHL test and the HS test at those fixed bandwidths as well as their adaptive versions. While the sizes of the HS test and its adaptive version are comparable with those of the CHL test and its adaptive version, the CHL test is more powerful than the HS test at all the cases under consideration.

The second simulation study was conducted on an ARCH-type time series regression model of the form:

$$Y_i = 0.25 + 0.5 Y_{i-1} + C_n \cos(8 Y_{i-1}) + 0.25\sqrt{Y_{i-1}^2 + 1}\ e_i, \tag{4.2}$$

where the innovation $\{e_i\}_{i=1}^n$ was chosen to be i.i.d. $\mathrm{N}(0, 1)$ random variables. The sample sizes considered in the simulation were $n = 300$ and 500. The vector of parameters $\theta = (\alpha, \beta, \sigma^2)$ was estimated using the pseudo-maximum likelihood method, which is commonly

Table 1
(a) Simulation results on model (4.1) and (b) more simulation results on model (4.1)

Distribution ε	τ	Andrews' test	Härdle–Mammen test	Horowitz–Spokoiny test	EL test
(a) **Probability of rejecting null hypothesis**					
Null hypothesis is true					
Normal		0.057	0.060	0.066	0.053
Mixture		0.053	0.053	0.048	0.055
Extreme value		0.063	0.057	0.055	0.057
Null hypothesis is false					
Normal	1.0	0.680	0.752	0.792	0.900
Mixture	1.0	0.692	0.736	0.835	0.905
Extreme value	1.0	0.600	0.760	0.820	0.924
Normal	0.25	0.536	0.770	0.924	0.929
Mixture	0.25	0.592	0.704	0.922	0.986
Extreme value	0.25	0.604	0.696	0.968	0.989

(b) **Probability of rejecting null hypothesis**
Horowitz–Spokoiny test and adaptive test

h value	2.5	3	3.5	4	4.5	adaptive
Null hypothesis is true						
Mixture	0.054	0.054	0.055	0.055	0.048	0.048
Null hypothesis is false						
$\tau = 1.0$	0.889	0.880	0.866	0.851	0.837	0.835
$\tau = 0.25$	0.961	0.951	0.937	0.928	0.922	0.922

Chen–Härdle–Li test and adaptive test

h value	2.5	3	3.5	4	4.5	adaptive
Null hypothesis is true						
Mixture	0.071	0.064	0.063	0.057	0.056	0.055
Null hypothesis is false						
$\tau = 1.0$	0.921	0.926	0.949	0.928	0.905	0.905
$\tau = 0.25$	0.991	0.989	0.989	0.988	0.986	0.986

used in the estimation of ARCH models. In the implementation, $\{e_i^*\}$ was sampled as a sequence of independent and identically normal distributed random errors from N(0, 1) and the estimator $\sigma_n^2(x)$ was used as given in (2.9). Both the HS test and the proposed test are evaluated. Although HS test was originally proposed for independent observations, the justification for its applicability to dependent observations is implicitly contained in this paper.

We used the least-squares cross-validation (LSCV) to guide the selection of the bandwidth set as in Hsiao et al. (2003). The averaged bandwidth (standard error) prescribed by the LSCV was 0.25 (0.1) for $n = 300$ and 0.23 (0.08) for $n = 500$, respectively. In view of these and Assumption A.3(iv), we chose $\mathscr{H}_n = \{0.23, 0.27, 0.31, 0.37, 0.43\}$ with $a = 0.8582$ for $n = 300$ and $\mathscr{H}_n = \{0.20, 0.23, 0.265, 0.304, 0.35\}$ with $a = 0.8694$ for $n = 500$. Here, we basically used a bandwidth that is slightly smaller than the averaged CV as $h_{\min}$ and a scaled down value of $(\log(\log(n)))^{-1}$ as $h_{\max}$. Both the power

Table 2
(a) Simulation results on model (4.2) and (b) more simulation results on model (4.2)

C_n	Horowitz–Spokoiny test		EL test	
	$n=300$	$n=500$	$n=300$	$n=500$
(a)				
0	0.067	0.063	0.059	0.058
0.03	0.098	0.184	0.176	0.275
0.04	0.128	0.209	0.309	0.461

(b)
Horowitz–Spokoiny test and adaptive test

C_n	h					Adaptive
	0.23	0.27	0.31	0.37	0.43	
$n=300$						
0.00	0.056	0.072	0.074	0.067	0.067	0.067
0.03	0.103	0.114	0.112	0.105	0.098	0.098
0.04	0.150	0.157	0.158	0.148	0.128	0.128
C_n	h					Adaptive
	0.20	0.23	0.26	0.30	0.35	
$n=500$						
0.00	0.056	0.054	0.063	0.061	0.063	0.063
0.03	0.202	0.201	0.200	0.197	0.184	0.184
0.04	0.236	0.237	0.246	0.236	0.209	0.209

Chen–Härdle–Li test and adaptive test

C_n	h					Adaptive
	0.23	0.27	0.31	0.37	0.43	
$n=300$						
0.00	0.059	0.065	0.062	0.061	0.059	0.059
0.03	0.175	0.185	0.178	0.162	0.145	0.176
0.04	0.309	0.318	0.306	0.272	0.251	0.309
C_n	h					Adaptive
	0.20	0.23	0.26	0.30	0.35	
$n=500$						
0.00	0.058	0.062	0.064	0.057	0.066	0.058
0.03	0.274	0.296	0.304	0.280	0.263	0.275
0.04	0.458	0.482	0.494	0.470	0.457	0.461

and the size of the adaptive test are reported in Table 2a. We found that both tests had good approximation to the nominal significance level of 5%, which confirms Theorem 3.1 and the quality of the simulation calibration to the distribution of the two adaptive test statistics. However, the power of HS test was rather subdued for the situations considered.

As expected when C_n was increased, the power of the proposed test was increased; and at a fixed level of C_n, the power increased when n was increased. The latter was because the distance between H_0 and H_1 became larger when n was increased although C_n was kept the same. This better power performance of the proposed test was possibly due to the internal studentization of the EL method which enhances the power of the proposed test.

Like Model (4.1), to further assess the power performance of the EL test at fixed bandwidths, we also compared the CHL test with the HS test at each of the above fixed bandwidth. Table 2b summarizes both the sizes and power values of the CHL test with the HS test at several fixed bandwidths as well as their adaptive versions. Like Table 1b, Table 2b also shows that the CHL test is again more powerful than the HS test in each individual case.

Acknowledgements

The authors thank the Co–Editor, the Associate Editor and two anonymous referees for their constructive suggestions and comments, and Cheng Yong Tang for excellent computation assistance. The first author and Tang were supported under grants from Iowa State University and the National Science Foundation (SES0518904). The second author acknowledges financial support from an Australian Research Council Discovery Grant.

Appendix A

The following assumptions are needed in establishing the asymptotic results. To avoid repeating the conditioning argument (given $\chi_n = (X_1, \ldots, X_n)$) for each case in the following proofs of Lemmas A.1–A.10, we use P_* and E_* to represent the respective conditional probability and conditional expectation given χ_n. Unless otherwise stated, the corresponding conditioning arguments are all understood to be held in probability with respect to the joint distribution of $\chi_n = (X_1, \ldots, X_n)$.

Assumption A.1. (i) The process $\{(X_t, Y_t)\}$ is strictly stationary and α-mixing with mixing coefficient $\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Omega_1^s, B \in \Omega_{s+t}^\infty\}$ for all $s, t \geqslant 1$, where $\{\Omega_i^j\}$ is a sequence of σ-fields generated by $\{(X_s, Y_s) : i \leqslant s \leqslant j\}$. There exist constants $a > 0$ and $\rho \in [0, 1)$ such that $\alpha(t) \leqslant a\rho^t$ for $t \geqslant 1$.

(ii) For all $t \geqslant 1$, $E[e_t|\Omega_{t-1}] = 0$ and $E[e_t^2|\Omega_{t-1}] = 1$, where $\{\Omega_t\}$ is a sequence of σ-fields generated by $\{(X_{s+1}, Y_s) : 1 \leqslant s \leqslant t\}$.

(iii) Let $\varepsilon_t = Y_t - m(X_t)$. There exists a constant $\delta_\varepsilon > 0$ such that $E_*[|\varepsilon_{t_1}^{i_1} \varepsilon_{t_2}^{i_2} \ldots \varepsilon_{t_l}^{i_l}|^{1+\delta_\varepsilon}] < \infty$ holds in probability, where $1 \leqslant l \leqslant 4$, $0 \leqslant i_j \leqslant 4$ and $\sum_{j=1}^l i_j \leqslant 8$.

Assumption A.2. (i) Let f be the density of X_t, S be a compact subset of R^d, $\mu_i(x) = E[\varepsilon_t^i|X_t = x]$ and π be a weight function such that $\int_{s\in S} \pi(s)\,ds = 1$ and $\int_{s\in S} \pi^2(s)\,ds \leqslant C$ for some constant C; $f(x)$ and $\mu_i(x)$ for $i = 2$ or 4 are Lipschitz continuous in S, and the first two derivatives of $f(x)$, $m(x)$ and $\mu_2(x)$ are continuous on S, $\inf_{x\in S} \sigma(x) \geqslant C_0 > 0$ and $\inf_{x\in S} f(x) \geqslant C_1 > 0$ for constants C_0 and C_1.

(ii) Let $f_{\tau_1,\tau_2,\ldots,\tau_l}(\cdot)$ be the joint probability density of $(X_{1+\tau_1}, \ldots, X_{1+\tau_l})$ $(1 \leqslant l \leqslant 4)$. Assume that each $f_{\tau_1,\tau_2,\ldots,\tau_l}(\cdot)$ exists and is Lipschitz continuous in S^l for $l = 1, \ldots, 4$.

(iii) $K(x_1, \ldots, x_d) = \prod_{i=1}^{d} k(x_i)$, where $k(\cdot)$ is a rth order univariate kernel which is symmetric, Lipschitz continuous and supported on $[-1, 1]$ satisfying $\int k(t)\,\mathrm{d}t = 1$, $\int t^l k(t)\,\mathrm{d}t = 0$ for $l = 1, \ldots, r-1$ and $\int t^r k(t)\,\mathrm{d}t = k_r \neq 0$ for a positive integer $r > d/2$.

Define $\nabla_\theta^l m_\theta(x) = \frac{\partial^l m_\theta(x)}{\partial \theta^l}$ whenever these derivatives exist. For any $q \times q$ matrix D, define $\|D\|_\infty = \sup_{v \in R^q} \frac{\|Dv\|}{\|v\|}$ where $\|B\|_m^2 = \sum_{i=1}^{q} \sum_{j=1}^{q} b_{ij}^2$ for a $q \times q$ matrix $B = (b_{ij})_{1 \leqslant i,j \leqslant q}$.

Assumption A.3. (i) The parameter set Θ is an open subset of R^q for some $q \geqslant 1$. For each $x \in S$, $m_\theta(x)$ is three times differentiable with respect to $\theta \in \Theta$. There exist constants $0 < C_1, C_2 < \infty$ such that $\sup_{\theta \in \Theta} |m_\theta(X_1)|^2 \leqslant C_1$ and $\max_{1 \leqslant j \leqslant 3} \sup_{\theta \in \Theta} \|\nabla_\theta^j m_\theta(X_1)\|_m^2 \leqslant C_2$ all hold in probability.

For each $\theta \in \Theta$, $m_\theta(x)$ is continuous with respect to $x \in S$. There is a finite $C_I > 0$ such that for every $\varepsilon > 0$, $\int_{x \in S} \inf_{\theta, \theta' \in \Theta : \|\theta - \theta'\| \geqslant \varepsilon} [m_\theta(x) - m_{\theta'}(x)]^2 f(x)\,\mathrm{d}x \geqslant C_I \varepsilon^2$.

(ii) Let H_0 be true. Then $\theta_0 \in \Theta$ and $\lim_{n \to \infty} \mathrm{P}_*(\sqrt{n}\|\tilde{\theta} - \theta_0\| > C_{1L}) < \varepsilon_1$ in probability for any $\varepsilon_1 > 0$ and some $C_{1L} > 0$.

Let H_0 be false. Then there is a $\theta^* \in \Theta$ such that $\lim_{n \to \infty} \mathrm{P}_*(\sqrt{n}\|\tilde{\theta} - \theta^*\| > C_{2L}) < \varepsilon_2$ in probability for any $\varepsilon_2 > 0$ and some $C_{2L} > 0$.

(iii) Let $\hat{\theta}^*$ as defined in the simulation procedure above (2.9). Then there is some $C_{3L} > 0$ such that for any $\varepsilon_3 > 0$

$$\mathrm{P}_*(\sqrt{n}\|\hat{\theta}^* - \tilde{\theta}\| > C_{3L}) < \varepsilon_3 \quad \text{in probability.}$$

(iv) The set $\mathscr{H}_n$ has the structure of (2.7) with $h_{\max} > h_{\min} = \mathrm{O}(n^{-\gamma})$ for some constant $0 < \gamma < \frac{1}{3d}$ and $h_{\max} = C_h(\log\log(n))^{-1/d}$ for a constant $C_h > 0$.

Assumptions A.1 and A.2 are standard conditions in this kind of problem. Assumption A.1(ii) reduces to $\mathrm{E}[e_t | e_1, \ldots, e_{t-1}] = 0$ and $\mathrm{E}[e_t^2 | e_1, \ldots, e_{t-1}] = 1$ when $\{X_s\}$ and $\{e_t\}$ are assumed to be independent for all $s \leqslant t$. Obviously, the two conditions further reduce to $\mathrm{E}[e_t] = 0$ and $\mathrm{E}[e_t^2] = 1$ when $\{e_t\}$ itself is a sequence of i.i.d. errors.

Assumption A.1(iii) basically requires that the conditional variance function $\sigma^2(X_1)$ is bounded in probability and $\mathrm{E}[|e_{t_1}^{i_1} e_{t_2}^{i_2} \ldots e_{t_l}^{i_l}|^{1+\delta_\varepsilon}] < \infty$, which reduces to $\mathrm{E}[|e_t|^{i_{\max}(1+\delta_\varepsilon)}] < \infty$ when $\{e_t\}$ is a sequence of i.i.d. errors, where $i_{\max} = \max(i_1, \ldots, i_l)$. Thus, Assumption A.1 allows the inclusion of many useful regression models with both independent and dependent time series observations in model (1.1).

Assumption A.2(i) corresponds to Assumption 5 of Horowitz and Spokoiny (2001). Assumption A.2(iii) plays a role similar to Assumption 4 of Horowitz and Spokoiny (2001). Assumption A.3 corresponds to Assumptions 1–2 and 6 of Horowitz and Spokoiny (2001) for the case of $d = 1$. Assumption A.3 (iv) requires the smallest bandwidth $h_{\min} = \mathrm{O}(n^{-\gamma})$ where $0 < \gamma < \frac{1}{3d}$.

As the Lagrange multiplier $\lambda(x)$ is implicitly dependent on h, we first establish the convergence rate for $\sup_{x \in S} \lambda(x)$ uniformly over the bandwidth set $\mathscr{H}_n$.

Lemma A.1. *Under Assumptions* A.1–A.3, *as n sufficiently large*

$$\max_{h \in \mathscr{H}_n} \sup_{x \in S} \lambda(x) = \mathrm{o_P}\{n^{-1/3} \log(n)\}.$$

Proof. For any $\delta > 0$

$$\mathrm{P}_*\left(\max_{h\in\mathscr{H}_n}\sup_{x\in S} h^{d/2}\lambda(x)\geqslant\delta n^{-1/2}\log(n)\right)\leqslant\sum_{h\in\mathscr{H}_n}\mathrm{P}_*\left(\sup_{x\in S} h^{d/2}\lambda(x)\geqslant\delta n^{-1/2}\log(n)\right).$$

As the number of bandwidths in H_n is only of order $\log(n)$, by checking the proof of Lemma 1 of Chen et al. (2002), it can be shown that $\log(n)$ can be readily squeezed in front of the probabilities involved to achieve that

$$\log(n)\cdot\mathrm{P}_*\left(\sum_{x\in S} h^{d/2}\lambda(x)\geqslant\delta n^{-1/2}\log(n)\right)\to 0$$

as $n\to\infty$. This implies that, as $n\to\infty$,

$$\mathrm{P}_*\left(\max_{h\in\mathscr{H}_n}\sup_{x\in S} h^{d/2}\lambda(x)\geqslant\delta n^{-1/2}\log(n)\right)\to 0 \tag{A.1}$$

and hence $\max_{h\in\mathscr{H}_n}\sup_{x\in S} h^{d/2}\,\lambda(x) = o_p\{\delta n^{-1/2}\log(n)\}$. Then the lemma is established by noting that the smallest bandwidth $h_{\min} = O(n^{-\gamma})$ where $3d\gamma < 1$ as assumed in Assumption A.3(iv).

In view of (2.6) of Chen et al. (2003), using Lemma A.1 we may show that

$$\max_{h\in\mathscr{H}_n} h^{-d/2}\left(\ell(\tilde{m}_{\tilde{\theta}};h) - nh^d\int\bar{U}_1^2(x;\tilde{\theta})v^{-1}(x)\pi(x)\,\mathrm{d}x\right) = o_P(1), \tag{A.2}$$

where

$$\bar{U}_1(x;\tilde{\theta}) = (nh^d)^{-1}\sum_{t=1}^{n}K\left(\frac{x-X_t}{h}\right)\{Y_t-\tilde{m}_{\tilde{\theta}}(x)\}$$

and $v(x) = R(K)f^{-1}(x)\sigma^2(x)$.

Let $W_t(x) = \frac{1}{nh^d}K\left(\frac{x-X_t}{h}\right)$, $a_{st} = nh^d\int_{x\in S}W_s(x)W_t(x)v^{-1}(x)\pi(x)\,\mathrm{d}x$, and $\lambda_t(\theta) = \lambda(X_t,\theta) = m(X_t)-m_\theta(X_t)$. Define

$$\ell_{0n}(h) = \sum_{s,t}a_{st}\varepsilon_s\varepsilon_t \quad\text{and}\quad Q_n(\theta) = Q_n(\theta;h) = \sum_{s,t}a_{st}\lambda_s(\theta)\lambda_t(\theta). \tag{A.3}$$

Then the leading term in $\ell_n(\tilde{m}_{\tilde{\theta}};h)$ is

$$\ell_{1n}(h,\tilde{\theta}) \equiv nh^d\int\bar{U}_1^2(x;\tilde{\theta})v^{-1}(x)\pi(x)\,\mathrm{d}x = \ell_{0n}(h)+Q_n(\tilde{\theta})+\Pi_n(\tilde{\theta}), \tag{A.4}$$

where $\Pi_n(\tilde{\theta}) = \ell_{1n}(h;\tilde{\theta})-\ell_{0n}(h)-Q_n(\tilde{\theta})$ is the remainder term.

Without loss of generality, we assume that $C(K,\pi) = 2R^{-2}(K)\int(K^{(2)}(x))^2\,\mathrm{d}x\int\pi^2(y)\,\mathrm{d}y = 1$. In view of the definition of $L_n = \max_{h\in\mathscr{H}_n}\frac{\ell(\tilde{m}_{\tilde{\theta}};h)-1}{h^{d/2}}$ and (A.4), define

$$L_{0n}(h) = \frac{\ell_{0n}(h)-1}{h^{d/2}},\quad L_{1n}(h) = \frac{\ell_{1n}(h,\tilde{\theta})-1}{h^{d/2}}\quad\text{and}\quad L_{2n}(h) = \frac{\ell_{1n}(h,\theta^*)-1}{h^{d/2}}, \tag{A.5}$$

where $\theta^* = \theta_0$ when H_0 is true and θ^* is as defined in Assumption A.3(ii) when H_0 is false. Let $L_{0n}^*(h)$ and $L_{1n}^*(h)$ be the respective versions of $L_{0n}(h)$ and $L_{1n}(h)$ defined above based on the Monte Carlo resamples $\{(X_t, Y_t^*): 1\leqslant t\leqslant n\}$. Lemmas A.2–A.7 are used in the proof of Theorem 3.1 to justify the approximation of l_α by l_α^* involved in the simulation procedure. Lemmas A.8–A.10 are mainly employed in the proofs of Theorems 3.2–3.4. □

Lemma A.2. *Suppose that Assumptions* A.1–A.3(i) *hold.*

(i) *For every* $\delta > 0$, *we have that* $\max_{h\in\mathscr{H}_n}\sup_{\|\theta-\theta_0\|\leqslant\delta}\frac{Q_n(\theta)}{nh^d}\leqslant C\delta^2$ *holds in probability, where* $C>0$ *is a constant.*

(ii) *For each* $\theta\in\Theta$ *and sufficiently large* n, *we have that* $C_1h^d\lambda(\theta)^\tau\lambda(\theta)\leqslant Q_n(\theta)\leqslant C_2h^d\lambda(\theta)^\tau\lambda(\theta)$ *holds in probability, where* $\lambda(\theta)=(\lambda_1(\theta),\ldots,\lambda_n(\theta))^\tau$ *and* $0<C_1\leqslant C_2<\infty$ *are constants, and the symbol '*τ*' denotes the transposition.*

Proof. (i) It follows from the definition of $Q_n(\theta)$ that $Q_n(\theta)\leqslant\|A\|_\infty\|\lambda(\theta)\|^2$. Let A be the matrix of $n\times n$ with $\{a_{st}\}$ as its $s\times t$ element. In order to prove Lemma A.2(i), one needs to show that $\|A\|_\infty\leqslant Ch^d$ holds in probability for some constant $C>0$. Let $q(x)=v^{-1}(x)\pi(x)$. We now have

$$\|A\|_\infty\leqslant\max_{1\leqslant t\leqslant n}\sum_{s=1}^n a_{st}=C(1+o_P(1))\max_{1\leqslant t\leqslant n}\int K\left(\frac{x-X_t}{h}\right)q(x)f(x)\,dx$$
$$=C(1+o_P(1))h^d\max_{1\leqslant t\leqslant n}(f(X_t)\;q(X_t))\int K(u)\,du\leqslant Ch^d \tag{A.6}$$

using the fact that

$$a_{st}=nh^d\int W_s(x)W_t(x)q(x)\,dx=\int\frac{K\left(\frac{x-X_s}{h}\right)}{\sum_{u=1}^n K\left(\frac{x-X_u}{h}\right)}\hat{f}(x)K\left(\frac{x-X_t}{h}\right)q(x)\,dx$$
$$=(1+o_P(1))\int\frac{K\left(\frac{x-X_s}{h}\right)}{\sum_{u=1}^n K\left(\frac{x-X_u}{h}\right)}K\left(\frac{x-X_t}{h}\right)q(x)f(x)\,dx.$$

In order to prove Lemma A.2(i), it suffices to show that $\sup_{\|\theta-\theta_0\|\leqslant\delta}\|\lambda(\theta)\|^2\leqslant Cn\delta^2$ holds in probability. A Taylor series expansion to $m_\theta(X_t)-m_{\theta_0}(X_t)$ and an application of Assumption 3.3(i) finish the proof of Lemma A.2(i).

(ii) Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues of A, respectively. In view of $\lambda_{\min}(A)\cdot\|\lambda(\theta)\|^2\leqslant Q_n(\theta)\leqslant\lambda_{\max}(A)\|\lambda(\theta)\|^2$, in order to prove Lemma A.2(ii), it suffices to show that for n large enough, $\lambda_{\min}(A)\geqslant C_1h^d(1+o_P(1))$ holds in probability. Such a proof follows similarly from the proof of Lemma A.2 of Gao et al. (2002).

For simplicity, in the following lemmas and their proofs, we let $q=1$. For $1\leqslant j\leqslant 3$, define $\psi_j(X_t,\theta)=m_\theta^{(j)}(X_t)=\frac{d^j m_\theta(X_t)}{d\theta^j}$. □

Lemma A.3. (i) *Under Assumptions* A.1–A.3(i), *we have for any given* $\theta\in\Theta$

$$J_n^{-1/2}\max_{h\in\mathscr{H}_n}h^{-d/2}\left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\varepsilon_s\psi_1(X_t,\theta)\right|=O_P(1). \tag{A.7}$$

(ii) *Under Assumptions* A.1 *and* A.2, *we have as* $n\to\infty$

$$J_n^{-1/2}\max_{h\in\mathscr{H}_n}h^{-d/2}\max_{1\leqslant t\leqslant n}\left|\sum_{s=1}^n a_{st}\varepsilon_s\right|=O_P(1). \tag{A.8}$$

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

$$\begin{aligned}
&P_*\left[J_n^{-1/2}\max_{h\in\mathscr{H}_n} h^{-d/2}\left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\varepsilon_s\psi_1(X_t,\theta)\right| > C_0\right]\\
&\leqslant \sum_{h\in\mathscr{H}_n} P_*\left[\left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\varepsilon_s\psi_1(X_t,\theta)\right| > C_0 J_n^{1/2}h^{d/2}\right]\\
&\leqslant \sum_{h\in\mathscr{H}_n}\frac{1}{C_0^2 J_n h^d} E_*\left[\sum_{s=1}^n\sum_{t=1}^n a_{st}\varepsilon_s\psi_1(X_t,\theta)\right]^2\\
&\leqslant \sum_{h\in\mathscr{H}_n}\frac{1}{C_0^2 J_n h^d}\left\{\sum_{s=1}^n\sum_{t=1}^n E_*\left[a_{st}\varepsilon_s\psi_1(X_t,\theta)\right]^2 + \Pi_{1n}(\theta)\right\},
\end{aligned}$$

where $\Pi_{1n}(\theta) = E_*\left[\sum_{s=1}^n\sum_{t=1}^n a_{st}\varepsilon_s\psi_1(X_t,\theta)\right]^2 - \sum_{s=1}^n\sum_{t=1}^n E_*\left[a_{st}\varepsilon_s\psi_1(X_t,\theta)\right]^2$.

A direct calculation shows that as $n\to\infty$

$$\sum_{s=1}^n\sum_{t=1}^n E_*\left[a_{st}\varepsilon_s\psi_1(X_t,\theta)\right]^2 \leqslant C(\theta)h^d(1+o(1)) \tag{A.9}$$

for some function $C(\theta)$.

Similar to (B.4) of Gao and King (2005), we may show that as $n\to\infty$,

$$\Pi_{1n}(\theta) = o(h^d). \tag{A.10}$$

Therefore, the proof of (A.7) is completed.

(ii) The proof of (ii) is similar to that of Lemma A.3(i). □

Lemma A.4. *Under Assumptions* A.1–A.3(i)(ii), *we have for each* $u > 0$ *and under* H_0,

$$\max_{h\in\mathscr{H}_n}\sup_{|\theta-\theta_0|\leqslant n^{-1/2}u} h^{-d/2}\left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\varepsilon_s\lambda_t(\theta)\right| = O_P(J_n^{1/2}n^{-1/2}). \tag{A.11}$$

Proof. Using a Taylor series expansion to $m_\theta(X_t) - m_{\theta_0}(X_t)$ and Assumption A.3(i), we have for θ' between θ and θ_0

$$\begin{aligned}
\left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\varepsilon_s\lambda_t(\theta)\right| &= \left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\varepsilon_s[m_\theta(X_t)-m_{\theta_0}(X_t)]\right|\\
&\leqslant \left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\varepsilon_s\psi_1(X_t,\theta_0)\right||\theta-\theta_0|\\
&\quad + \frac{1}{2}\left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\varepsilon_s\psi_2(X_t,\theta_0)\right||\theta-\theta_0|^2
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{6}\left|\sum_{s=1}^{n}\sum_{t=1}^{n} a_{st}\varepsilon_s\psi_3(X_t,\theta')\right||\theta-\theta_0|^3\\
&\leqslant\left|\sum_{s=1}^{n}\sum_{t=1}^{n} a_{st}\varepsilon_s\psi_1(X_t,\theta_0)\right||\theta-\theta_0|\\
&+\frac{1}{2}|\theta-\theta_0|^2\left|\sum_{s=1}^{n}\sum_{t=1}^{n} a_{st}\varepsilon_s\psi_2(X_t,\theta_0)\right|\\
&+\frac{1}{6}n|\theta-\theta_0|^3\left|\sum_{s=1}^{n} a_{st}\varepsilon_s\right|\max_{1\leqslant t\leqslant n}|\psi_3(X_t,\theta')|. \qquad (A.12)
\end{aligned}
$$

Hence, (A.7), (A.8), (A.12) and Assumption A.3(i) imply

$$
\max_{h\in\mathscr{H}_n}\sup_{\|\theta-\theta_0\|\leqslant n^{-1/2}u} h^{-d/2}\left|\sum_{s=1}^{n}\sum_{t=1}^{n} a_{st}\varepsilon_s\lambda_t(\theta)\right|\leqslant O_P(J_n^{1/2}n^{-1/2}). \qquad (A.13)
$$

The proof of (A.11) follows from (A.12) and (A.13). □

Lemma A.5. *Let Assumptions* A.1–A.3(i)(iii) *hold. Then under* H_1, *for every* $u>0$, *any* $q_n\to\infty$ *and some* $h\in\mathscr{H}_n$

$$
\sup_{|\theta-\theta^*|\leqslant n^{-1/2}u}\left|\sum_{s=1}^{n}\sum_{t=1}^{n} a_{st}\varepsilon_s\lambda(X_t,\theta)\right| = o_P(q_n h^{d/2}). \qquad (A.14)
$$

Proof. The proof follows similarly from that of (A.13) using $\lim_{n\to\infty} q_n=\infty$. □

Lemma A.6. *Suppose that Assumptions* A.1–A.3 *hold. Then as n* $\to\infty$

$$
\max_{h\in\mathscr{H}_n} L_n(h)=\max_{h\in\mathscr{H}_n} L_{1n}(h)+o_P(1)=\max_{h\in\mathscr{H}_n} L_{2n}(h)+o_P(1),
$$
$$
\max_{h\in\mathscr{H}_n} L_{1n}^*(h)=\max_{h\in\mathscr{H}_n} L_{0n}^*(h)+o_P(1),
$$

and $\max_{h\in\mathscr{H}_n}L_{1n}(h)=\max_{h\in\mathscr{H}_n}L_{0n}(h)+o_P(1)$ *under* H_0.

Proof. The proof follows from (A.3) to (A.5) and Lemmas A.3–A.5. □

In view of Lemma A.6, in order to prove Theorem 3.1, it suffices to show Lemma A.7.

Lemma A.7. *Suppose Assumptions* A.1–A.3 *hold. Then the conditional distributions of* $\max_{h\in\mathscr{H}_n}L_{0n}(h)$ *and* $\max_{h\in\mathscr{H}_n}L_{0n}^*(h)$ *given* χ_n *are asymptotically identical under* H_0.

Proof. Similarly to the proof of Lemma A.2, we can show that

$$
\max_{h\in\mathscr{H}_n} h^{-d/2}\left(\sum_{s=1}^{n} a_{ss}\varepsilon_s^2-1\right)=o_P(1)\quad\text{and}\quad \max_{h\in\mathscr{H}_n} h^{-d/2}\left(\sum_{s=1}^{n} a_{ss}\varepsilon_s^{*2}-1\right)=o_P(1) \qquad (A.15)
$$

hold conditionally on χ_n.

Thus, it suffices to show that the conditional distributions of $\max_{h\in\mathscr{H}_n}\sum_{s\neq t}a_{st}\varepsilon_s\varepsilon_t$ and $\max_{h\in\mathscr{H}_n}\sum_{s\neq t}a_{st}\varepsilon_s^*\varepsilon_t^*$ given χ_n are asymptotically identical. For $h\in\mathscr{H}_n$, let $u_t=\varepsilon_t$ or ε_t^*

and define

$$B_{hn}(u_1,\ldots,u_n) = h^{-d/2}\left[\sum_{s\neq t} a_{st}u_su_t\right]. \tag{A.16}$$

Let $B_n(u_1,\ldots,u_n)$ be the sequence obtained by stacking the corresponding $B_{hn}(u_1,\ldots,u_n)$ $(h\in\mathscr{H}_n)$. Let $G(\cdot)=G_n(\cdot)$ be a 3-times continuously differentiable function over R^{J_n}. Define

$$C_n(G) = \sup_{v\in R^{J_n}} \max_{i,j,k=1,2,\ldots,J_n}\left|\frac{\partial^3 G(v)}{\partial v_i\partial v_j\partial v_k}\right|.$$

Like Horowitz and Spokoiny (2001), there are two steps in the proof of Lemma A.7. First, we want to show that

$$|\mathrm{E}_*[G(B_n(\varepsilon_1,\ldots,\varepsilon_n))] - \mathrm{E}_*[G(B_n(\varepsilon_1^*,\ldots,\varepsilon_n^*))]| \leqslant C_0C_n(G)\left(\frac{J_n^3}{nh_{1\min}^{3d}}\right)^{1/2} \tag{A.17}$$

for any 3-times differentiable $G(\cdot)$, some finite constant C_0, and all sufficiently large n. Then in the second step, (A.17) is used to show that the conditional distributions of $B_n(\varepsilon_1,\ldots,\varepsilon_n)$ and $B_n(\varepsilon_1^*,\ldots,\varepsilon_n^*)$ given χ_n are asymptotically identical.

Throughout the rest of the proof, we replace a_{st} in (A.16) with $\tilde{a}_{st}(h)=h^{-d/2}a_{st}$. Note that

$$\begin{aligned}&|\mathrm{E}_*[G(B_n(\varepsilon_1,\ldots,\varepsilon_n))] - \mathrm{E}_*[G(B_n(\varepsilon_1^*,\ldots,\varepsilon_n^*))]|\\ &\quad\leqslant \sum_{t=1}^{n}|\mathrm{E}_*[G(B_n(\varepsilon_1,\ldots,\varepsilon_t,\varepsilon_{t+1}^*,\ldots,\varepsilon_n^*))] - \mathrm{E}_*[G(B_n(\varepsilon_1,\ldots,\varepsilon_{t-1},\varepsilon_t^*,\ldots,\varepsilon_n^*))]|,\end{aligned} \tag{A.18}$$

where $B_n(\varepsilon_1,\ldots,\varepsilon_n,\varepsilon_{n+1}^*)=B_n(\varepsilon_1,\ldots,\varepsilon_n)$ and $B_n(\varepsilon_0,\varepsilon_1^*,\ldots,\varepsilon_n^*)=B_n(\varepsilon_1^*,\ldots,\varepsilon_n^*)$.

We now derive an upper bound on the last term of the sum on the right-hand side of (A.18). Similar bounds can be derived for the other terms. Let U_{n-1}, Λ_n and $\tilde{\Lambda}_n$, respectively, denote the vectors that are obtained by stacking

$$U_{h,n} = \sum_{s=1}^{n-1}\sum_{t=1,\neq s}^{n-1}\tilde{a}_{st}(h)\varepsilon_s\varepsilon_t,\quad \Lambda_{h,n} = 2\varepsilon_n\sum_{s=1}^{n-1}\tilde{a}_{sn}(h)\varepsilon_s,\quad \tilde{\Lambda}_{h,n} = 2\varepsilon_n^*\sum_{s=1}^{n-1}\tilde{a}_{sn}(h)\varepsilon_s.$$

Using a Taylor expansion to the last term of the sum on the right-hand side of (A.18) about $\varepsilon_n=\varepsilon_n^*=0$ gives

$$\begin{aligned}&|\mathrm{E}_*[G(B_n(\varepsilon_1,\ldots,\varepsilon_n))] - \mathrm{E}_*[G(B_n(\varepsilon_1,\ldots,\varepsilon_{n-1},\varepsilon_n^*))]|\\ &\quad\leqslant |\mathrm{E}_*[G'(U_{n-1})(\Lambda_n-\tilde{\Lambda}_n)]| + \frac{1}{2}|\mathrm{E}_*[\Lambda_n^\tau G''(U_{n-1})\Lambda_n - \tilde{\Lambda}_n^\tau G''(U_{n-1})\tilde{\Lambda}_n]|\\ &\qquad+\frac{C_n(G)}{6}\{\mathrm{E}_*[\|\Lambda_n\|^3]+\mathrm{E}_*[\|\tilde{\Lambda}_n\|^3]\},\end{aligned}$$

where G' and G'' denote the gradient and matrix of second derivatives of G and $C_n(G)$ is a positive and finite constant.

Since

$$\mathrm{E}[\varepsilon_n|\Omega_{n-1}] = \mathrm{E}[\varepsilon_n^*] = 0 \quad\text{and}\quad \mathrm{E}[\varepsilon_n^2|\Omega_{n-1}] = \mathrm{E}[\varepsilon_n^{*2}] = 1,$$

we have

$$\mathrm{E}[(\Lambda_n - \tilde{\Lambda}_n)|\Omega_{n-1}] = 0 \quad \text{and} \quad \mathrm{E}[(\Lambda_n \Lambda_n^\tau - \tilde{\Lambda}_n \tilde{\Lambda}_n^\tau)|\Omega_{n-1}] = 0.$$

Thus, we have

$$\mathrm{E}_*[(\Lambda_n - \tilde{\Lambda}_n)|\Omega_{n-1}] = 0 \quad \text{and} \quad \mathrm{E}_*[(\Lambda_n \Lambda_n^\tau - \tilde{\Lambda}_n \tilde{\Lambda}_n^\tau)|\Omega_{n-1}] = 0.$$

This implies

$$|\mathrm{E}_*[G(B_n(\varepsilon_1, \ldots, \varepsilon_n))] - \mathrm{E}_*[G(B_n(\varepsilon_1, \ldots, \varepsilon_{n-1}, \varepsilon_n^*))]| \leqslant \frac{C_n(G)}{6}\{\mathrm{E}_*[\|\Lambda_n\|^3] + \mathrm{E}_*[\|\tilde{\Lambda}_n\|^3]\}. \tag{A.19}$$

To estimate the upper bound of (A.19), we need the following result:

$$\begin{aligned} a_{st} &= \frac{1}{nh^d}\int K\left(\frac{x - X_s}{h}\right)K\left(\frac{x - X_t}{h}\right)q(x)\,\mathrm{d}x \\ &= \frac{1}{n}\int K(u)K\left(u + \frac{X_s - X_t}{h}\right)q(X_s + uh)\,\mathrm{d}u = \frac{1}{n}L_2\left(\frac{X_s - X_t}{h}, X_s\right), \end{aligned} \tag{A.20}$$

where $q(x) = v^{-1}(x)\pi(x)$ and $L_2(x, y) = \int K(u)K(u + x)q(y + uh)\,\mathrm{d}u$.

Using Assumptions A.1 and A.2 and (A.20), we have for n sufficiently large and the small $\delta_\varepsilon > 0$ involved in Assumption A.1(iii),

$$\begin{aligned} &\sum_{h_1 \in \mathscr{H}_n}\sum_{h_2 \in \mathscr{H}_n} \mathrm{E}_*\left[\sum_{s=1}^{n-1}\sum_{t=1,\neq s}^{n-1} \tilde{a}_{sn}^2(h_1)\tilde{a}_{tn}^2(h_2)\varepsilon_s^2\varepsilon_t^2\varepsilon_n^4\right] \\ &\leqslant \sum_{h_1 \in \mathscr{H}_n}\sum_{h_2 \in \mathscr{H}_n} \frac{1}{n^4 h_1^d h_2^d}\sum_{s=1}^{n-1}\sum_{t=1,\neq s}^{n-1} \mathrm{E}_*\left[L_2^2\left(\frac{X_s - X_n}{h_1}, X_s\right)L_2^2\left(\frac{X_t - X_n}{h_2}, X_t\right)\varepsilon_s^2\varepsilon_t^2\varepsilon_n^4\right] \\ &\leqslant C\sum_{h_1 \in \mathscr{H}_n}\sum_{h_2 \in \mathscr{H}_n} \frac{1}{n^4 h_1^d h_2^d}\sum_{s=1}^{n-1}\sum_{t=1,\neq s}^{n-1} (\mathrm{E}_*[|\varepsilon_s^2\varepsilon_t^2\varepsilon_n^4|^{1+\delta_\varepsilon}])^{1/(1+\delta_\varepsilon)} \\ &\leqslant C\sum_{h_1 \in \mathscr{H}_n}\sum_{h_2 \in \mathscr{H}_n} \frac{1}{n^2 h_1^d h_2^d} \leqslant C \cdot \left(\frac{J_n}{nh_{1\min}^d}\right)^2, \end{aligned} \tag{A.21}$$

where $0 < C < \infty$ is a constant.

Similarly to the proof of Lemma C.2 of Gao and King (2005), as $n \to \infty$

$$\begin{aligned} &\sum_{h_1, h_2 \in \mathscr{H}_n} \mathrm{E}_*\left(\sum_{1 \leqslant s \neq t \leqslant n-1} \tilde{a}_{sn}^2(h_1)\tilde{a}_{sn}(h_2)\tilde{a}_{tn}(h_2)\varepsilon_s^3\varepsilon_t\varepsilon_n^4\right) = \mathrm{o}\left(\frac{J_n}{nh_{1\min}^d}\right)^2, \\ &\sum_{h_1, h_2 \in \mathscr{H}_n} \mathrm{E}_*\left(\sum_{1 \leqslant s \neq t, s \neq u, t \neq u \leqslant n-1} \tilde{a}_{sn}^2(h_1)\tilde{a}_{tn}(h_2)\tilde{a}_{un}(h_2)\varepsilon_s^2\varepsilon_t\varepsilon_u\varepsilon_n^4\right) = \mathrm{o}\left(\frac{J_n}{nh_{1\min}^d}\right)^2, \\ &\sum_{h_1, h_2 \in \mathscr{H}_n} \mathrm{E}_*\left(\sum_{\text{different s,t,u,v}} \tilde{a}_{sn}(h_1)\tilde{a}_{tn}(h_1)\tilde{a}_{un}(h_2)\tilde{a}_{vn}(h_2)\varepsilon_s\varepsilon_t\varepsilon_u\varepsilon_v\varepsilon_n^4\right) = \mathrm{o}\left(\frac{J_n}{nh_{1\min}^d}\right)^2, \end{aligned} \tag{A.22}$$

where the last expectation is taken under $1<s,t,u,v\leqslant n-1$ and s,t,u,v are all different, using the fact that for every given x,

$$\mathrm{E}\left[L_2\left(\frac{X_t-x}{h},X_t\right)\varepsilon_t\right]=\mathrm{E}\left[L_2\left(\frac{X_t-x}{h},X_t\right)\mathrm{E}[\varepsilon_t|\Omega_{t-1}]\right]=0 \tag{A.23}$$

implied from Assumption A.1.

Eqs. (A.21) and (A.22) then imply that for n sufficiently large

$$\sum_{h_1\in\mathscr{H}_n}\sum_{h_2\in\mathscr{H}_n}\mathrm{E}_*\left[\sum_{s,t,u,v=1}^{n-1}\tilde{a}_{sn}(h_1)\varepsilon_s\tilde{a}_{tn}(h_1)\varepsilon_t\tilde{a}_{un}(h_2)\varepsilon_u\tilde{a}_{vn}(h_2)\varepsilon_v\varepsilon_n^4\right]\leqslant C\cdot(1+\mathrm{o}(1))\left(\frac{J_n}{nh_{1\min}^d}\right)^2. \tag{A.24}$$

Let $\tilde{A}_{sn}$ be the vector that is obtained by stacking $\tilde{a}_{sn}(h)$ $(h\in\mathscr{H}_n)$. Eq. (A.24) then implies that as $n\to\infty$

$$\begin{aligned}\mathrm{E}_*\left[\|\Lambda_n\|^3\right]&=8\mathrm{E}_*\left[\left\|\sum_{s=1}^{n-1}\tilde{A}_{sn}\varepsilon_s\varepsilon_n\right\|^3\right]\leqslant 8\left\{\mathrm{E}_*\left[\sum_{h\in\mathscr{H}_n}\left(\sum_{s=1}^{n-1}\tilde{a}_{sn}(h)\varepsilon_s\varepsilon_n\right)^2\right]^2\right\}^{3/4}\\&=8\left\{\sum_{h_1\in\mathscr{H}_n}\sum_{h_2\in\mathscr{H}_n}\mathrm{E}_*\left[\sum_{s,t,u,v=1}^{n-1}\tilde{a}_{sn}(h_1)\varepsilon_s\tilde{a}_{tn}(h_1)\varepsilon_t\tilde{a}_{un}(h_2)\varepsilon_u\tilde{a}_{vn}(h_2)\varepsilon_v\varepsilon_n^4\right]\right\}^{3/4}\\&\leqslant C\left(\frac{J_n}{nh_{1\min}^d}\right)^{3/2}.\end{aligned} \tag{A.25}$$

A similar result holds for $\mathrm{E}_*[\|\tilde{\Lambda}_n\|^3]$. Thus

$$\mathrm{E}_*[\|\Lambda_n\|^3]+\mathrm{E}_*[\|\tilde{\Lambda}_n\|^3]\leqslant 2C\left(\frac{J_n}{nh_{1\min}^d}\right)^{3/2}. \tag{A.26}$$

Step 2: As demonstrated in Horowitz and Spokoiny (2001),

$$\lim_{n\to\infty}\left\{\mathrm{P}_*\left[\max_{h\in\mathscr{H}_n}B_{hn}(\varepsilon_1,\ldots,\varepsilon_n)\leqslant x\right]-\mathrm{P}_*\left[\max_{h\in\mathscr{H}_n}B_{hn}(\varepsilon_1^*,\ldots,\varepsilon_n^*)\leqslant x\right]\right\}=0$$

for any real x is equivalent to

$$\lim_{n\to\infty}\left|\mathrm{E}_*\left(\prod_{h\in\mathscr{H}_n}I[B_{hn}(\varepsilon_1,\ldots,\varepsilon_n)\leqslant x]\right)-\mathrm{E}_*\left(\prod_{h\in\mathscr{H}_n}I\left[B_{hn}(\varepsilon_1^*,\ldots,\varepsilon_n^*)\leqslant x\right]\right)\right|=0.$$

Following the lines of Horowitz and Spokoiny (2001) by utilizing the above established bound (A.26) and using (A.18), it can be shown that as $n\to\infty$

$$\left|\mathrm{P}_*\left[\max_{h\in\mathscr{H}_n}B_{hn}(\varepsilon_1,\ldots,\varepsilon_n)\leqslant x\right]-\mathrm{P}_*\left[\max_{h\in\mathscr{H}_n}B_{hn}(\varepsilon_1^*,\ldots,\varepsilon_n^*)\leqslant x\right]\right|\leqslant C\left(\frac{J_n^3}{nh_{1\min}^{3d}}\right)^{1/2}\to 0. \tag{A.27}$$

This implies (A.17) and finally completes the proof of Lemma A.7. □

Lemma A.8. *Suppose that Assumptions* A.1 *and* A.2 *hold. Then for any* $x \geqslant 0$, $h \in \mathscr{H}_n$ *and all sufficiently large* n

$$P_*(L^*_{0n}(h) > x) \leqslant \exp\left(-\frac{x^2}{4}\right).$$

Proof. Similar to the proof of Corollary 1 of Chen et al. (2003) that, for any small $\delta > 0$ there exists a large integer $n_0 \geqslant 1$ such that for $n \geqslant n_0$ and $x \geqslant 0$, $|P_*(L^*_{0n}(h) \leqslant x) - \Phi(x)| < \delta$, where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}\,du$. We therefore have for any $n \geqslant n_0$ and $x \geqslant 0$

$$\begin{aligned}
P_*(L^*_{0n}(h) > x) &\leqslant 1 - \Phi(x) + \delta = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-u^2/2}\,du + \delta \\
&= \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-u^2/4}e^{-u^2/4}\,du + \delta \leqslant e^{-x^2/4}\frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-u^2/4}\,du + \delta \\
&\leqslant e^{-x^2/4}\frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{-u^2/4}\,du + \delta = e^{-x^2/4}\frac{\sqrt{2}}{\sqrt{2\pi}}\int_0^{\infty} e^{-v^2/2}\,dv + \delta \\
&= \frac{\sqrt{2}}{2}e^{-x^2/4} + \delta
\end{aligned}$$

using $\frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{-v^2/2}\,dv = \frac{1}{2}$. The proof follows by letting $0 < \delta \leqslant (1 - \frac{\sqrt{2}}{2})e^{-x^2/4}$ for any $x \geqslant 0$. □

For $0 < \alpha < 1$, define $\tilde{l}_\alpha$ to be the $1 - \alpha$ quantile of $\max_{h \in \mathscr{H}_n} L^*_{0n}(h)$.

Lemma A.9. *Suppose that Assumption* A.1 *holds. Then for large enough* n

$$\tilde{l}_\alpha \leqslant 2\sqrt{\log(J_n) - \log(\alpha)}.$$

Proof. The proof is similar to that of Lemma 12 of Horowitz and Spokoiny (2001). □

Lemma A.10. *Suppose that Assumptions* A.1 *and* A.2 *hold. Suppose that the following inequality*:

$$Q_n(\theta^*) \geqslant 2h^{d/2}\tilde{l}^*_\alpha \tag{A.28}$$

holds in probability for some $h \in \mathscr{H}_n$ *and sufficiently large* n, *where*

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

Then $\lim_{n\to\infty} P_*(L_n > l^*_\alpha) = 1$.

Proof. By (A.3)–(A.5) and Lemma A.6, L_n can be replaced with $\max_{h \in \mathscr{H}_n} L_{2n}(h)$. By Lemmas A.6 and A.7, l^*_α can be replaced by $\tilde{l}_\alpha$. Thus, it suffices to show that

$$\lim_{n\to\infty} P_*\left(\max_{h \in \mathscr{H}_n} L_{2n}(h) > \tilde{l}_\alpha\right) = 1,$$

which holds if $\lim_{n\to\infty} P_*(L_{2n}(h) > \tilde{l}_\alpha) = 1$ for some $h \in \mathscr{H}_n$. For any $h \in \mathscr{H}_n$, using (A.3)–(A.5) and Lemma A.2 again we have

$$\begin{aligned} L_{2n}(h) &= L_{0n}(h) + h^{-d/2} Q_n(\theta^*) + h^{-d/2} \Pi_n(\theta^*) \\ &= L_{0n}^*(h) + h^{-d/2} Q_n(\theta^*) + h^{-d/2} \Pi_n(\theta^*) + o_p(1) \\ &= L_{0n}^*(h) + h^{-d/2} Q_n(\theta^*)(1 + o_p(1)) + o_p(1). \end{aligned} \tag{A.29}$$

In view of (A.28), it follows from (A.29) that as $n \to \infty$

$$\begin{aligned} P_*(L_{2n}(h) > \tilde{l}_\alpha) &= P_*(L_{0n}^*(h) + h^{-d/2} Q_n(\theta^*) + h^{-d/2} \Pi_n(\theta^*) > \tilde{l}_\alpha) \\ &\geqslant P_*(L_{0n}^*(h) > \tilde{l}_\alpha - 2\tilde{l}_\alpha^*) \to 1 \end{aligned}$$

because the conditional distribution of $L_{0n}^*(h)$ given χ_n is asymptotically normal and therefore bounded in probability and $\tilde{l}_\alpha - 2\tilde{l}_\alpha^* \to -\infty$ as $n \to \infty$. This finishes the proof. □

Proof of Theorem 3.1. By Lemma A.6, $\max_{h\in\mathscr{H}_n} L_{1n}(h) = \max_{h\in\mathscr{H}_n} L_{2n}(h) + o_p(1)$. By Lemma A.7, under H_0 $\max_{h\in\mathscr{H}_n} L_{2n}(h) - \max_{h\in\mathscr{H}_n} L_{0n}^*(h) \to 0$ in distribution as $n \to \infty$. Using Lemma A.6 again implies $\max_{h\in\mathscr{H}_n} L_{1n}^*(h) = \max_{h\in\mathscr{H}_n} L_{0n}^*(h) + o_p(1)$. This implies that $\max_{h\in\mathscr{H}_n} L_{1n}(h) - \max_{h\in\mathscr{H}_n} L_{1n}^*(h) \to 0$ in distribution as $n \to \infty$. This, along with Eqs. (A.2)–(A.5), finishes the proof. □

In order to prove, Theorems 3.2 and 3.3, in view of Lemma A.10, it suffices to verify (A.28). Using Lemma A.1(ii), it suffices to verify

$$h^d \lambda(\theta)^\tau \lambda(\theta) \geqslant 4\tilde{l}_\alpha^* h^{d/2} \quad \text{in probability.} \tag{A.30}$$

Proof of Theorem 3.2. In view of the definition of $\tilde{l}_\alpha^*$, Eq. (A.30) follows from the fact that as $n \to \infty$,

$$\frac{1}{n} \lambda(\theta)^\tau \lambda(\theta) - \rho(m, \mathscr{M}) \to 0 \tag{A.31}$$

holds in probability and $nh^d \geqslant C_0 \tilde{l}_\alpha^* h^{d/2}$ for some constant $0 < C_0 < \infty$ and n large enough. □

Proof of Theorem 3.3. Using the definition of $\tilde{l}_\alpha^*$, (A.31),

$$\frac{1}{n} \sum_{t=1}^{n} \Delta^2(X_t) \to E_S[\Delta^2(X_1)] = \int_{x\in S} \Delta^2(x) f(x)\,dx \geqslant D_1 > 0 \quad \text{as } n \to \infty \tag{A.32}$$

and the fact that

$$\frac{1}{n} \lambda(\theta)^\tau \lambda(\theta) = \frac{C_n^2}{n} \sum_{t=1}^{n} \Delta^2(X_t) \geqslant D_1 C_n^2 \tag{A.33}$$

holds in probability, one can see that (A.30) holds when $h = h_{\max} = (\log \log(n))^{-1/d}$. This finishes the proof of Theorem 3.3. □

Proof of Theorem 3.4. In order to verify (A.28), we need to introduce the following notation: $h_1 = (n^{-1} \tilde{l}_\alpha^*)^{2/(4s+d)}$. This implies $nh_1^{(4s+d)/2} = \tilde{l}_\alpha^*$. Choose $h \in \mathscr{H}_n$ such that

$h_1 \leqslant h < 2h_1$. We then have

$$4h^{d/2}\tilde{l}_{\alpha}^{*} = 4nh^{d/2}h_1^{(4s+d)/2} \leqslant 4nh^{(4s+d)/2+d/2} = 4nh^{2s+d}. \tag{A.34}$$

Thus, in order to verify (A.28), it suffices to show that

$$Q_n(\theta^*) \geqslant 4nh^{2s+d} \tag{A.35}$$

holds in probability for the selected $h \in \mathscr{H}_n$ and $\theta^* \in \Theta$. The verification of (A.35) can be done using similar techniques employed in the proof of Lemma A.2. Alternatively, one may follow the proof of (A.13) of Horowitz and Spokoiny (2001) by noting that all the derivations below their (A.13) hold in probability with respect to the joint distribution of $\chi_n = (X_1, \ldots, X_n)$. $\square$

References

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

Andrews, D.W.K., 1997. A conditional Kolmogorov test. Econometrica 1097–1028.

Bosq, D., 1998. Nonparametric Statistics for Stochastic Processes, Lecture Notes in Statistics, vol. 110. Springer, Heidelberg.

Chen, S.X., Cui. H., 2007. On the second order properties of empirical likelihood under moment restriction. Journal of Econometrics, to appear.

Chen, S.X., Cui, H., 2006. On Bartlett correction of empirical likelihood in the presence of nuisance parameters. Biometrika 93, 215–220.

Chen, S.X., Härdle, W., Li, M., 2002. An empirical likelihood goodness-of-fit test for time series, Technical Report 2002-2, National University of Singapore, ⟨http://www.stat.nus.edu.sg/⟩.

Chen, S.X., Härdle, W., Li, M., 2003. An empirical likelihood goodness-of-fit test for time series. Journal of the Royal Statistical Society Series B 65, 663–678.

Eubank, R.L., Spiegelman, C.H., 1990. Testing the goodness of fit of a linear model via nonparametric regression techniques. Journal of the American Statistical Association 85, 387–392.

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., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman & Hall, London.

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

Fan, J., Yao, Q., 2005. Nonlinear Time Series. Nonparametric and Parametric Methods, Springer, New York.

Fan, J., Zhang, C., 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, J., Zhang, C., Zhang, J., 2001. Generalized likelihood ratio statistics and Wilks phenomenon. Annals of Statistics 29, 153–193.

Franke, J., Kreiss, J.P., Mammen, E., 2002. Bootstrap of kernel smoothing in nonlinear time series. Bernoulli 8, 1–38.

Gao, J., King, M.L., 2005. Estimation and model specification testing in nonparametric and semiparametric regression. Working paper available at ⟨www.maths.uwa.edu.au/~jiti/jems.pdf⟩.

Gao, J., Tong, H., Wolff, R.C.L., 2002. Model specification tests in nonparametric stochastic regression models. Journal of Multivariate Analysis 83, 324–359.

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926–1947.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge.

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

Hjellvik, V., Tjøstheim, D., 1995. Nonparametric tests of linearity for time series. Biometrika 82, 351–368.

Hjellvik, V., Yao, Q., Tjøstheim, D., 1998. Linearity testing using local polynomial approximation. Journal of Statistical Planning & Inference 68, 295–321.

Hjort, N.L., McKeague, I.W., Van Keilegom, I., 2004. Extending the scope of empirical likelihood. Manuscript.

Horowitz, J.L., Spokoiny, V.G., 2001. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69, 599–632.

Hsiao, C., Li, Q., Racine, J., 2003. Consistent model specification tests with mixed discrete and continuous variables. Unpublished manuscript.

Imbens, G.W., 1997. One-step estimators for over-identified generalized method of moments models. Review of Economic Studies 64, 359–383.

Ingster, Yu.I., Suslina, I.A., 2003. Nonparametric Goodness-of-Fit Testing Under Gaussian Models, Lecture Notes in Statistics, vol. 169. Springer, New York.

Kitamura, Y., 1997. Empirical likelihood methods with weakly dependent processes. Annals of Statistics 25, 2084–2102.

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

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

McKeague, I.W., Zhang, M.J., 1994. Identification of nonlinear time series from first order cumulative characteristics. Annals of Statistics 22, 495–514.

Owen, A., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.

Owen, A., 1990. Empirical likelihood ratio confidence regions. Annals of Statistics 18, 90–120.

Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating functions. Annals of Statistics 22, 300–325.

Spokoiny, V.G., 1996. Adaptive hypothesis testing using wavelets. Annals of Statistics 24, 2477–2498.

Tripathi, G., Kitamura, Y., 2003. On testing conditional moment restrictions: the canonical case. Annals of Statistics 31, 2059–2095.

Wang, L., Van Keilegom, I., 2005. Nonparametric test for the form of parametric regression with time series errors. Working paper.

Zhang, C., Dette, H., 2003. A power comparison between nonparametric regression tests. Statistics & Probability Letters 66, 289–301.