4-2004

An Adaptive Empirical Likelihood Test for Parametric Time Series Regression Models

Song Xi Chen
Iowa State University, songchen@iastate.edu

Jiti Gao
University of Western Australia

An Adaptive Empirical Likelihood Test for Parametric Time Series Regression Models

Abstract

We extend the adaptive and rate-optimal test of Horowitz and Spokoiny (2001) for specification of parametric regression models to weakly dependent time series regression models with an empirical likelihood formulation of our test statistic. It is found that the proposed adaptive empirical likelihood test preserves the rate-optimal property of the test of Horowitz and Spokoiny (2001).

Keywords

empirical likelihood, goodness-of-fit test, kernel estimation, rate-optimal test

Disciplines

Statistics and Probability

Comments

This preprint was published as Song Xi Chen and Jiti Gao, "An adaptive Empirical Likelihood Test for Parametric Time Series Regression Models", *Journal of Econometrics* (2007): 950-972, doi: 10.1016/j.jeconom.2006.12.002

An Adaptive Empirical Likelihood Test for Parametric Time Series Regression Models

By Song Xi Chen* and Jiti Gao

* *Department of Statistics, Iowa State University, Ames, IA 50010, USA.*

Email: songchen@iastate.edu

School of Mathematics and Statistics, The University of Western Australia, Crawley WA 6009, Australia. Email: jiti@maths.uwa.edu.au

A test for a parametric regression model against a sequence of local alternative is constructed based on an empirical likelihood test statistic that measures the goodness-of-fit between the parametric model and its nonparametric counterpart. To reduce the dependence of the test on a single smoothing bandwidth, the test is formulated by maximizing a standardized version of the empirical likelihood test statistic over a set of smoothing bandwidths. It is demonstrated that the proposed test is able to distinguish local alternatives from the null hypothesis at an optimal rate.

Keywords: empirical likelihood; goodness–of–fit test; kernel estimation; rate-optimal test; nonparametric time series

Running Title: An Adaptive Test for Time Series Regression Models

1. Introduction

Consider a time series heteroscedasticity regression model of the form

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

where both $m(\cdot)$ and $\sigma(\cdot)$ are unknown functions defined over R^d, the data $\{(X_t, Y_t)\}_{t=1}^n$ are weakly dependent stationary time series, and e_t is an error process with zero mean and unit variance. Suppose that $\{m_\theta(\cdot)|\theta \in \Theta\}$ is a family of parametric specification to the regression function $m(x)$ where $\theta \in R^q$ is an unknown parameter belonging to a parameter space Θ. This paper considers testing the validity of the parametric specification of $m_\theta(x)$ against a series of local alternatives, that is to test

$$H_0 : m(x) = m_\theta(x) \text{ versus } H_1 : m(x) = m_\theta(x) + C_n\Delta_n(x) \text{ for all } x \in S, \tag{1.2}$$

where C_n is a non-random sequence tending to zero as $n \to \infty$, $\Delta_n(x)$ 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 alternative family of regression models from the parametric family $\{m_\theta(\cdot)|\theta \in \Theta\}$.

Nonparametric kernel estimation of the conditional mean function is well studied for both independent and dependent observations as documented in Fan and Gijbels (1996) and Fan and Yao (2003). Goodness-of-fit tests for a parametric conditional mean model by formulating certain distance measure between the parametric model and its corresponding kernel estimator has been proposed in the literature; for instance the works of Eubank and Spiegelman (1990), Härdle and Mammen (1993), Hjellvik and Tjøstheim (1995), Hart (1997), Hjellvik, Yao and Tjøstheim (1998). Fan and Zhang (2003) propose separate tests for the conditional mean and the 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 large number of factor levels for dependent observations. Other related references include Robinson (1989), Andrews (1997), An and Cheng (1991), Eubank and Hart (1992), Horowitz and Härdle (1994), Hong and White (1995), Li and Wang (1998), Li (1999), Gao, Tong and Wolff (2002), Sperlich, Tjøstheim and Yang (2002) and Gao and King (2003).

The main focus of the paper is on formulating a test that is able to differentiate between H_0 and H_1 with a smallest C_n possible for dependent observations. A key feature of the proposed test is that the test statistic is an empirical likelihood (EL) of the hypothesed parametric model given observations. The EL (Owen, 1988, 1990) is a technique that allows construction of nonparametric likelihood for a parameter of interest. Despite that it is intrinsically nonparametric, it possesses two important properties of a parametric likelihood: the Wilks' theorem and the Bartlett correction. For survival data, Li and Van Keilegom (2002) construct nonparametric likelihood ratio confidence bands for censored data. Li (2003) consider a goodness-of-fit test for a parametric specification of the distribution function which is more efficient in Bahadur sense than any weighted Kolmogorov–Smirnov test at any alternative. Einmahl and McKeague (2003) propose an EL goodness-of-fit tests for a distribution function and other distributional characteristics. Fan and Zhang (2004) propose a sieve EL test for testing a general varying-coefficient regression model that extends the generalized likelihood ratio test of Fan, Zhang and Zhang (2001). They demonstrate that the 'Wilks phenomenon' continues to hold under general assumptions on the error distribution. Tripathi and Kitamura (2003) propose an EL test for conditional moment restrictions. Both of these works are established for independent data. For testing the conditional mean function with dependent data, Chen, Härdle and Li (2003) develop an EL test by simulating a known Gaussian random field.

Another feature of our proposal is that the final test statistic is formulated by maximizing the EL statistics over a set of bandwidths. This is aimed at achieving the optimal rate of convergence for C_n, which defines the gap between the null and alternative hypotheses in (1.2). The existing goodness-of-fit tests for a parametric model based on a kernel estimator with a fixed bandwidth h, for instance the tests given in Härdle and Mammen (1993), require that the smallest order for C_n is of order $n^{-1/2}h^{-d/4}$ in order for the test to be consistent. 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. For testing parametric conditional mean models, 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 adaptive against the unknown smoothness of the local alternative hypothesis and is able to achieve the optimal order for C_{1n} in the minimax sense of Spokoiny (1996), and Ingster and Suslina (2003). A similar idea is also given in Fan (1996). In this paper, we extend the proposal of Horowitz and Spokoiny (2001) for the proposed test based on the EL with weakly dependent observations. Comparing with tests based on a fixed bandwidth, a test based on a set of bandwidths will be less dependent on a particular choice of bandwidth and hence will make the test more robust against the choice of smoothing bandwidths. To accurately approximate the distribution of the adaptive test statistic, a bootstrap procedure is used to profile the critical value of the test. This combination of the EL and bootstrap utilizes the good features of the EL for the construction of test statistics and the effectiveness of the bootstrap in distribution approximation.

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. Adaptive 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 r-th order d-dimensional kernel and h be a smoothing bandwidth. Let $K_h(u) = h^{-d}K(u/h)$. 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 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 model $m_\theta(x)$ 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 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

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

where $\pi(\cdot)$ is a non-negative weight function supported on the compact set $S \subseteq R^d$.

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

$$C(K, \pi) = 2R^{-2}(K) \int \pi^2(x) dx \int \left(K^{(2)}(x)\right)^2 dx, \tag{2.5}$$

where $K^{(2)}$ is the convolution of K. Chen, Härdle and Li (2003) show that as $n \to \infty$

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

They proposed a single bandwidth based EL test based on critical values obtained by simulating a Gaussian random field.

Like all nonparametric kernel goodness-of-tests based on a single bandwidth, the test is 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 smallest possible, we employ the adaptive test procedure of Horowitz and Spokoiny (2001) for the EL test statistics as follows. Let

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

be a set of bandwidths, where $0 < a < 1$, $J_n = \log_{1/a}(h_{\max}/h_{\min})$ is the number of bandwidths, $h_{\max} = c_{\max}(\log\log(n))^{-\frac{1}{d}}$ 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 $E\{\ell(\tilde{m}_{\tilde{\theta}}; h)\} = 1$ 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:

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

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.

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 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 $\sigma_n(\cdot)$ is a consistent estimator of $\sigma(\cdot)$, $\{e_t^*\}$ is independent of $\{X_s\}$ for all $s \geq 1$, and sampled randomly from a specified distribution with $E[e_t^*] = 0$, $E\left[e_t^{*2}\right] = 1$ and $E\left[|e_t^*|^{4+\delta}\right] < \infty$ for some $\delta > 0$. Define l_α^* to be the $1 - \alpha$ quantile of the distribution of L_n with $\{Y_t\}$ replaced by $\{Y_t^*\}$.

2. Let $\hat{\theta}^*$ be the estimate of θ based on the resample $\{(X_t, Y_t^*)\}_{t=1}^n$. Compute the statistic L_n^* by replacing Y_t and $\tilde{\theta}$ with Y_t^* and $\hat{\theta}^*$ according to (2.8).

3. Estimate l_α^* by the $1 - \alpha$ quantile of the empirical distribution of L_n^*, which can be obtained by repeating steps 1–2 many times.

It should be noted that $\{Y_t^*\}$ may be generated recursively by $Y_t^* = m_{\tilde{\theta}}(Y_{t-1}^*) + \sigma_n(Y_{t-1}^*)e_t^*$ when $\{Y_t\}$ of (1.1) satisfies a nonparametric autoregressive model. 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 $n \to \infty$. The proposed adaptive EL test rejects H_0 if $L_n > l_\alpha^*$.

There are two different approaches we could use for generating $\{e_t^*\}$ for the bootstrap. The first is the one used in Horowitz and Spokoiny (2001) which generates independent and identically distributed e_t^* from $N(0,1)$. The second approach is the regression bootstrap proposed by Franke, Kreiss and Mammen (2002), where the generation of $\{e_t^*\}$ depends on $(X_1, \cdots, X_T)$. The regression bootstrap is more sophisticated and works for more general purposes. We use in this paper the first approach in conjuction with an estimator of $\sigma^2(\cdot)$ as it is simpler and sufficient for the task of this paper.

3. Main Results

The following are assumptions needed in establishing the asymptotic results.

Assumption 3.1. (i) *The process* $\{(X_t, Y_t)\}$ *is strictly stationary and* α*-mixing with the 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* $s, t \geq 1$, *where* Ω_i^j *denote the* σ*-fields generated by* $\{(X_s, Y_s) : i \leq s \leq j\}$. *There exist constants* $a > 0$ *and* $\rho \in [0, 1)$ *such that* $\alpha(t) \leq a\rho^t$ *for* $t \geq 1$.

(ii) *For all* $t \geq 1$, $E[e_t|\Omega_{t-1}] = 0$ *and* $E[e_t^2|\Omega_{t-1}] = 1$, *where* Ω_t *are* σ*-fields generated by* $\{(X_{s+1}, Y_s) : 1 \leq s \leq t\}$.

(iii) *Let* $\epsilon_t = Y_t - m(X_t)$. *There exists a constant* $\delta_\epsilon > 0$ *such that* $E\left[\left|\epsilon_{t_1}^{i_1}\epsilon_{t_2}^{i_2}\cdots\epsilon_{t_l}^{i_l}\right|^{1+\delta_\epsilon}\right] < \infty$, *where* $1 \leq l \leq 4$, $0 \leq i_j \leq 4$ *and* $\sum_{j=1}^l i_j \leq 8$.

Assumption 3.2. (i) *Let* f *be the density of* X_t, S *be a compact subset of* R^d, $\mu_i(x) = E[\epsilon_t^i|X = x]$ *and* π *be a weight function such that* $\int_{s\in S}\pi(s)ds = 1$ *and* $\int_{s\in S}\pi^2(s)ds \leq C$ *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, $\inf_{x\in S}\sigma(x) \geq C_0 > 0$ *and* $\inf_{x\in S} f(x) \geq C_1 > 0$ *for constants* C_0 *and* C_1.

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

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

Define $\bigtriangledown_\theta^l m_\theta(x) = \frac{\partial^l m_\theta(x)}{\partial\theta^l}$ 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 $B = (b_{ij})_{1\leq i,j\leq q}$.

Assumption 3.3. (i) *The parameter set* Θ *is an open subset of* R^q *for some* $q \geq 1$. *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* $E\left[\sup_{\theta\in\Theta} |m_\theta(X_1)|^2\right] \leq C_1$ *and* $\max_{1\leq j\leq 3} E\left[\sup_{\theta\in\Theta} || \bigtriangledown_\theta^j m_\theta(X_1)||_m^2\right] \leq C_2$. *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* $\varepsilon > 0$, $\int_{x\in S} \inf_{\theta,\theta'\in\Theta:||\theta-\theta'||\geq\varepsilon} \left[m_\theta(x) - m_{\theta'}(x)\right]^2 f(x)dx \geq C_I\varepsilon^2$.

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

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

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

Assumptions 3.1 and 3.2 are standard conditions in this kind of problem. Assumption 3.2(i) corresponds to Assumption 5 of Horowitz and Spokoiny (2001). Assumption 3.2(iii) plays a role similar to Assumption 4 of Horowitz and Spokoiny (2001). Assumption 3.3 corresponds to Assumptions 1–2 and 6 of Horowitz and Spokoiny (2001). Assumption 3.3 (iv) requires the smallest bandwidth $h_{\min} = O(n^{-\gamma})$ where $0 < \gamma < \frac{1}{3d}$. To include the optimal order $n^{-\frac{1}{d+2r}}$ for estimating $m(x)$ in the range of $\mathcal{H}_n$, we need to have $r > d$ which is something further than requiring $r > \frac{d}{2}$ as assumed in Assumption 3.2 (ii). This implies that higher order kernels are needed when $d \geq 2$. However, for the purpose of establishing the results reported below, it is not essential to have the order $n^{-\frac{1}{d+2r}}$ covered by $\mathcal{H}_n$.

The following theorem shows that the EL test has a correct size asymptotically.

Theorem 3.1. *Suppose Assumptions 3.1, 3.2 and 3.3(i)(ii)(iv) hold. Then under* H_0,

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

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

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

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

Theorem 3.2. *Assume that Assumptions 3.1, 3.2 and 3.3(i)(iii)(iv) hold. If there is a* $C_\rho > 0$ *such that* $\rho(m, \mathcal{M}) \geq C_\rho$ *for* $n \geq n_0$ *with some large* n_0*, then*

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

We then consider the consistency of the EL test against special from 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,

$$0 < D_1 \leq \int_{x\in S} \Delta^2(x) f(x) dx \leq D_2 < \infty \quad \text{and} \quad \rho(m, \mathcal{M}) \geq D_3 C_n. \tag{3.3}$$

Theorem 3.3. *Assume Assumptions 3.1, 3.2 and 3.3(i)(iii). Let Assumption 3.3(iv) hold with* $h_{\max} = C_h(\log\log(n))^{-\frac{1}{d}}$ *for some finite constant* C_h*. Let* m *satisfy (3.2) and (3.3) with* $C_n \geq C n^{-1/2}\sqrt{\log\log(n)}$ *for some constant* $C > 0$*. Then*

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

To discuss the consistency of the adaptive EL test over alternatives in a Hölder smoothness class, we introduce the following notation. Let $j = (j_1, \ldots, j_d)$ where $j_1, \ldots, j_d \geq 0$ 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 $||m||_{H,s} = \sup_{x\in S} \sum_{|j|\leq s} (|D^j m(x)|)$. The smoothness class that we consider consist of functions $m \in S(H, s) \equiv \{m : ||m||_{H,s} \leq C_H\}$ for some unknown s and $C_H < \infty$. For $s \geq \max(2, d/4)$ and all sufficiently large $D_m < \infty$, define

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

Theorem 3.4. *Assume that Assumptions 3.1–3.3 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),*

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

Theorem 3.1 shows that the test attains the nominal level α asymptotically. Theorem 3.2 establishes that the adaptive EL test is consistent against a family of fixed alternatives. Theorem 3.3 shows that the proposed test is consistent for $C_n \geq C n^{-1/2}\sqrt{\log\log(n)}$, which is a substantial improvement over the fixed bandwidth based tests and achieves almost the conventional rate $n^{-1/2}$.

The conclusion of Theorem 3.4 shows that L_n is uniformly consistent over alternatives within the 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.

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):

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

where the $\{\epsilon_i : i \geq 1\}$ are independent and identically distributed from three distributions with zero mean and constant variance, $\{X_i\}$ are univariate design points to be sampled from $N(0, 25)$ 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 $H_0 : m(x) = \beta_0 + \beta_1 x$ specifies a linear regression corresponding to $\tau = 0$, whereas the alternative hypothesis $H_1 : m(x) = \beta_0 + \beta_1 x + (5/\tau)\phi(x/\tau)$ for $\tau = 1.0$ and 0.5. Readers 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 simulation, the bootstrap resamples and estimation procedures for θ as in Horowitz and Spokoiny (2001). We also employed the same kernel, the same bandwidth set $\mathcal{H}_n$, the same estimator σ_n^2 and the distribution for $\{e_i^*\}$ in the Monte Carlo simulation procedure as in Horowitz and Spokoiny (2001). Like Horowitz and Spokoiny, the nominal size of the test was 5%.

Table 1 summaries 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.

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

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

where the innovation $\{e_i\}_{i=1}^n$ was chosen to be independent and identically distributed $N(0,1)$ random variables. The sample sizes considered in the simulation were $n = 300$ and $n = 500$. The vector of parameters $\theta = (\alpha, \beta, \sigma^2)$ was estimated using the pseudo-maximum likelihood method, which is commonly 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 tests of Horowitz and Spokoiny (2001) and the proposed test are evaluated. Although Horowitz and Spokoiny's test was originally proposed for independent observations, a justification for its use with dependent observations is implicitly contained in this paper.

We chose the bandwidth set $\mathcal{H}_n = \{0.3, 0.332, 0.367, 0.407, 0.45\}$ with $a = 0.903$ for $n = 300$ and $\mathcal{H}_n = \{0.25, 0.281, 0.316, 0.356, 0.4\}$ with $a = 0.889$ for $n = 500$. Both the power and the size of the adaptive test are reported in Table 2. 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 Horowitz and Spokoiny's test was rather subdued for the situations considered. As expected when C_n was increased, the power of the proposed test was increased; and for 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 possibility due to the internal studentization of the EL which enhances the power of the proposed test.

Appendix

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 $\mathcal{H}$.

Lemma A.1. *Under Assumptions 3.1, 3.2 and 3.3, as n sufficiently large*

$$\max_{h\in\mathcal{H}}\sup_{x\in S}\lambda(x)=o_p\{n^{-1/3}\log(n)\}.$$

Proof: For any $\delta>0$

$$P\left(\max_{h\in\mathcal{H}_n}\sup_{x\in S}h^{d/2}\lambda(x)\geq\delta n^{-1/2}\log(n)\right)\leq\sum_{h\in\mathcal{H}_n}P\left(\sup_{x\in S}h^{d/2}\lambda(x)\geq\delta n^{-1/2}\log(n)\right).$$

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

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

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

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

and hence $\max_{h\in\mathcal{H}_n}\sup_{x\in S}h^{d/2}\lambda(x)=o_p\{\delta n^{-1/2}\log(n)\}$. Then the lemma is established by noting that the smallest bandwidth $h_{\min}=O(n^{-\gamma})$ where $3d\gamma<1$ as assumed in Assumption 3.3(iv).

In view of (2.6) of Chen, Härdle and Li (2003), using Lemma A.1 we may show that

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

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

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

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

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

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 $C(K, \pi) = 2R^{-2}(K) \int \left(K^{(2)}(x)\right)^2 dx \int \pi^2(y) dy = 1$. In view of the definition of $L_n = \max_{h \in \mathcal{H}_n} \frac{\ell(\tilde{m}_{\tilde{\theta}}; h) - 1}{h^{d/2}}$ 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 3.3(iii) 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 bootstrap resample $\{(X_i, Y_i^*)\}$. Lemmas A.2–A.7 below 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 3.1, 3.2 and 3.3(i) hold.*

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

(ii) *For each $\theta \in \Theta$ and sufficiently large n, we have that $C_1 h^d \lambda(\theta)^\tau \lambda(\theta) \le Q_n(\theta) \le C_2 h^d \lambda(\theta)^\tau \lambda(\theta)$ holds in probability, where $\lambda(\theta) = (\lambda_1(\theta), \cdots, \lambda_n(\theta))^\tau$ and $0 < C_1 \le C_2 < \infty$ are constants.*

Proof: (i) It follows from the definition of $Q_n(\theta)$ that $Q_n(\theta) \le ||A||_\infty ||\lambda(\theta)||^2$. 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 $||A||_\infty \le Ch^d$ holds in probability for some constant $C > 0$. Let $q(x) = v^{-1}(x)\pi(x)$. We now have

$$||A||_\infty \le \max_{1 \le t \le n} \sum_{s=1}^n a_{st} \;=\; C(1 + o_p(1)) \max_{1 \le t \le n} \int K\left(\frac{x - X_t}{h}\right) q(x) f(x) dx$$

$$= \quad C(1+o_p(1))h^d \max_{1\le t\le n} \left(f(X_t)\; q(X_t)\right) \int K(u)du \le Ch^d \quad \text{(A.6)}$$

using the fact that

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

In order to prove Lemma A.2(i), it suffices to show that $\sup_{||\theta-\theta_0||\le\delta} ||\lambda(\theta)||^2 \le Cn\delta^2$ 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 $\lambda_{\min}(A)\cdot||\lambda(\theta)||^2 \le Q_n(\theta) \le \lambda_{\max}(A)||\lambda(\theta)||^2$, in order to prove Lemma A.2(ii), it suffices to show that for n large enough, $\lambda_{\min}(A) \ge C_1 h^d(1+o_p(1))$ holds in probability. Such a proof follows similarly from the proof of Lemma A.2 of Gao, Tong and Wolff (2002).

For simplicity, in the following lemmas and their proofs, we let $q=1$. For $1\le j\le 3$, 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 3.1, 3.2 and 3.3(i), we have for any given* $\theta\in\Theta$

$$J_n^{-1/2} \max_{h\in\mathcal{H}_n} h^{-d/2} \left| \sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s \psi_1(X_t,\theta)\right| = O_p(1), \quad \text{(A.7)}$$

(ii) *Under Assumptions 3.1 and 3.2, we have as* $n\to\infty$

$$J_n^{-1/2} \max_{h\in\mathcal{H}_n} h^{-d/2} \max_{1\le t\le n} \left|\sum_{s=1}^n a_{st}\epsilon_s\right| = O_p(1). \quad \text{(A.8)}$$

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

$$P\left[J_n^{-1/2} \max_{h\in\mathcal{H}_n} h^{-d/2} \left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\epsilon_s\psi_1(X_t,\theta)\right| > C_0\right]$$

$$\le \sum_{h\in\mathcal{H}_n} P\left[\left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\epsilon_s\psi_1(X_t,\theta)\right| > C_0 J_n^{1/2} h^{d/2}\right]$$

$$\leq \sum_{h\in\mathcal{H}_n} \frac{1}{C_0^2 J_n h^d} E\left[\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s \psi_1(X_t,\theta)\right]^2$$

$$\leq \sum_{h\in\mathcal{H}_n} \frac{1}{C_0^2 J_n h^d} \left\{\sum_{s=1}^n \sum_{t=1}^n E\left[a_{st}\epsilon_s\psi_1(X_t,\theta)\right]^2 + \Pi_{1n}(\theta)\right\},$$

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

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

$$\sum_{s=1}^n \sum_{t=1}^n E\left[a_{st}\epsilon_s\psi_1(X_t,\theta)\right]^2 = \tag{A.9}$$

$$\int\int q(x)q(y)K^{(2)}\left(\frac{x-y}{h}\right)K^{(2)}\left(\frac{y-x}{h}\right)\sigma^2(x)\psi_1^2(y,\theta)f(x,y)dxdy = C(\theta)h^d(1+o(1))$$

for some function $C(\theta)$.

Similarly to (B.4) of Gao and King (2003), 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 3.1, 3.2 and 3.3(i)(ii), we have for each* $u > 0$ *and under* H_0,

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

Proof: Using a Taylor series expansion to $m_\theta(X_t) - m_{\theta_0}(X_t)$ and Assumption 3.3(i), we have for θ' between θ and θ_0

$$\begin{aligned}\left|\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s\lambda_t(\theta)\right| &= \left|\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s\left[m_\theta(X_t) - m_{\theta_0}(X_t)\right]\right| \\ &\leq \left|\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s\psi_1(X_t,\theta_0)\right| |\theta-\theta_0| + \frac{1}{2}\left|\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s\psi_2(X_t,\theta_0)\right| |\theta-\theta_0|^2 \\ &+ \frac{1}{6}\left|\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s\psi_3(X_t,\theta')\right| |\theta-\theta_0|^3 \leq \left|\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s\psi_1(X_t,\theta_0)\right| |\theta-\theta_0|\end{aligned}$$

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

Hence, (A.7), (A.8), (A.12) and Assumption 3.3(i) imply

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

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

Lemma A.5. *Under Assumptions 3.1, 3.2 and 3.3(i)(iii) hold. Then under H_1, for every $u > 0$, any $q_n \to \infty$ and some $h \in \mathcal{H}_n$*

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

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

Lemma A.6. *Suppose that Assumptions 3.1–3.3 hold. Then as $n \to \infty$*

$$
\begin{aligned}
\max_{h \in \mathcal{H}_n} L_n(h) &= \max_{h \in \mathcal{H}_n} L_{1n}(h) + o_p(1) = \max_{h \in \mathcal{H}_n} L_{2n}(h) + o_p(1), \\
\max_{h \in \mathcal{H}_n} L_{1n}^*(h) &= \max_{h \in \mathcal{H}_n} L_{0n}^*(h) + o_p(1),
\end{aligned}
$$

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

Proof: The proof follows from (A.3), (A.4), (A.5) and Lemmas A.3–A.5.

Lemma A.7. *Suppose Assumptions 3.1–3.3 hold. Then the asymptotic distributions of* $\max_{h \in \mathcal{H}_n} L_{2n}(h)$ *and* $\max_{h \in \mathcal{H}_n} L_{0n}^*(h)$ *are identical under* H_0.

Proof: In view of Lemma A.6, in order to prove Lemma A.7, it suffices to show that the distributions of $\max_{h \in \mathcal{H}_n} L_{0n}(h)$ and $\max_{h \in \mathcal{H}_n} L_{0n}^*(h)$ are asymptotically the same. Similarly to the proof of Lemma A.2, we can show that

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

Thus, it suffices to show that $\max_{h\in\mathcal{H}_n}\sum_{s\neq t} a_{st}\epsilon_s\epsilon_t$ and $\max_{h\in\mathcal{H}_n}\sum_{s\neq t} a_{st}\epsilon_s^*\epsilon_t^*$ have the same asymptotic distribution. For $h\in\mathcal{H}_n$, let $u_t=\epsilon_t$ or ϵ_t^* and define

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

Let $B_n(u_1,\ldots,u_n)$ be the sequence obtained by stacking the corresponding $B_{hn}(u_1,\ldots,u_n)$ ($h\in\mathcal{H}_n$). 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.3. First, we want to show that

$$\left|E\left[G(B_n(\epsilon_1,\ldots,\epsilon_n))\right]-E\left[G(B_n(\epsilon_1^*,\ldots,\epsilon_n^*))\right]\right|\le C_0C_n(G)\left(\frac{J_n^3}{nh_{1\min}^{3d}}\right)^{1/2} \tag{A.17}$$

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 $B_n(\epsilon_1,\ldots,\epsilon_n)$ and $B_n(\epsilon_1^*,\ldots,\epsilon_n^*)$ have the same asymptotic distribution.

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

$$\begin{aligned}&\left|E\left[G(B_n(\epsilon_1,\ldots,\epsilon_n))\right]-E\left[G(B_n(\epsilon_1^*,\ldots,\epsilon_n^*))\right]\right| \qquad\text{(A.18)}\\ &\le\sum_{t=1}^{n}\left|E\left[G(B_n(\epsilon_1,\ldots,\epsilon_t,\epsilon_{t+1}^*,\cdots,\epsilon_n^*))\right]-E\left[G(B_n(\epsilon_1,\ldots,\epsilon_{t-1},\epsilon_t^*,\ldots,\epsilon_n^*))\right]\right|,\end{aligned}$$

where $B_n(\epsilon_1,\ldots,\epsilon_n,\epsilon_{n+1}^*)=B_n(\epsilon_1,\ldots,\epsilon_n)$ and $B_n(\epsilon_0,\epsilon_1^*,\ldots,\epsilon_n^*)=B_n(\epsilon_1^*,\ldots,\epsilon_n^*)$.

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

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

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

$$\left|E\left[G(B_n(\epsilon_1,\ldots,\epsilon_n))\right]-E\left[G(B_n(\epsilon_1,\ldots,\epsilon_{n-1},\epsilon_n^*))\right]\right|\le\left|E\left[G'(U_{n-1})(\Lambda_n-\tilde{\Lambda}_n)\right]\right|$$

$$+\frac{1}{2}\left|E\left[\Lambda_n^\tau G''(U_{n-1})\Lambda_n - \tilde{\Lambda}_n^\tau G''(U_{n-1})\tilde{\Lambda}_n\right]\right| + \frac{C_n(G)}{6}\left\{E\left[||\Lambda_n||^3\right] + E\left[||\tilde{\Lambda}_n||^3\right]\right\},$$

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

$$E\left[\epsilon_n|\Omega_{n-1}\right] = E\left[\epsilon_n^{*j}\right] = 0 \ \text{ and } \ E\left[\epsilon_n^2|\Omega_{n-1}\right] = E\left[\epsilon_n^{*2}\right] = 1,$$

we have

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

This implies

$$|E\left[G(B_n(\epsilon_1,\ldots,\epsilon_n))\right] - E\left[G(B_n(\epsilon_1,\ldots,\epsilon_{n-1},\epsilon_n^*))\right]| \le \frac{C_n(G)}{6}\left\{E\left[||\Lambda_n||^3\right] + E\left[||\tilde{\Lambda}_n||^3\right]\right\}. \tag{A.19}$$

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

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

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

Using Assumptions 3.1–3.2 and (A.20), we have for n sufficiently large and the small $\delta_\epsilon > 0$ involved in Assumption 3.1(iii),

$$\sum_{h_1\in\mathcal{H}_n}\sum_{h_2\in\mathcal{H}_n} E\left[\sum_{s=1}^{n-1}\sum_{t=1,\neq s}^{n-1}\tilde{a}_{sn}^2(h_1)\tilde{a}_{tn}^2(h_2)\epsilon_s^2\epsilon_t^2\epsilon_n^4\right] \tag{A.21}$$

$$\begin{aligned} &\le \sum_{h_1\in\mathcal{H}_n}\sum_{h_2\in\mathcal{H}_n}\frac{1}{n^4h_1^dh_2^d}\sum_{s=1}^{n-1}\sum_{t=1,\neq s}^{n-1}E\left[L_2^2\left(\frac{X_s-X_n}{h_1},X_s\right)L_2^2\left(\frac{X_t-X_n}{h_2},X_t\right)\epsilon_s^2\epsilon_t^2\epsilon_n^4\right] \\ &\le C\sum_{h_1\in\mathcal{H}_n}\sum_{h_2\in\mathcal{H}_n}\frac{1}{n^4h_1^dh_2^d}\sum_{s=1}^{n-1}\sum_{t=1,\neq s}^{n-1}E\left[\left|\epsilon_s^2\epsilon_t^2\epsilon_n^4\right|^{1+c}\right] \\ &\le C\sum_{h_1\in\mathcal{H}_n}\sum_{h_2\in\mathcal{H}_n}\frac{1}{n^2h_1^dh_2^d} \le C\cdot\left(\frac{J_n}{nh_{1\min}^d}\right)^2, \end{aligned}$$

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

Similarly to the proof of Lemma C.2 of Gao and King (2003), as $n \to \infty$

$$\sum_{h_1,h_2 \in \mathcal{H}_n} E\left(\sum_{1 \le s \ne t \le n-1} \tilde{a}_{sn}^2(h_1)\tilde{a}_{sn}(h_2)\tilde{a}_{tn}(h_2)\epsilon_s^3\epsilon_t\epsilon_n^4\right) = o\left(\frac{J_n}{nh_{1\min}^d}\right)^2, \quad (A.22)$$

$$\sum_{h_1,h_2 \in \mathcal{H}_n} E\left(\sum_{1 \le s \ne t, s \ne u, t \ne u \le qn-1} \tilde{a}_{sn}^2(h_1)\tilde{a}_{tn}(h_2)\tilde{a}_{un}(h_2)\epsilon_s^2\epsilon_t\epsilon_u\epsilon_n^4\right) = o\left(\frac{J_n}{nh_{1\min}^d}\right)^2,$$

$$\sum_{h_1,h_2 \in \mathcal{H}_n} E\left(\sum \tilde{a}_{sn}(h_1)\tilde{a}_{tn}(h_1)\tilde{a}_{un}(h_2)\tilde{a}_{vn}(h_2)\epsilon_s\epsilon_t\epsilon_u\epsilon_v\epsilon_n^4\right) = o\left(\frac{J_n}{nh_{1\min}^d}\right)^2,$$

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

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

implied from Assumption A.1.

Equations (A.21) and (A.22) then imply that as $n \to \infty$

$$\sum_{h_1 \in \mathcal{H}_n}\sum_{h_2 \in \mathcal{H}_n} E\left[\sum_{s,t,u,v=1}^{n-1} \tilde{a}_{sn}(h_1)\epsilon_s\tilde{a}_{tn}(h_1)\epsilon_t\tilde{a}_{un}(h_2)\epsilon_u\tilde{a}_{vn}(h_2)\epsilon_v\epsilon_n^4\right] \le C \cdot \left(\frac{J_n}{nh_{1\min}^d}\right)^2. \quad (A.24)$$

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

$$\begin{aligned} E\left[||\Lambda_n||^3\right] &= 8E\left[\left|\left|\sum_{s=1}^{n-1}\tilde{A}_{sn}\epsilon_s\epsilon_n\right|\right|^3\right] \le 8\left\{E\left[\sum_{h \in \mathcal{H}_n}\left(\sum_{s=1}^{n-1}\tilde{a}_{sn}(h)\epsilon_s\epsilon_n\right)^2\right]^2\right\}^{3/4} \quad (A.25) \\ &= 8\left\{\sum_{h_1 \in \mathcal{H}_n}\sum_{h_2 \in \mathcal{H}_n} E\left[\sum_{s,t,u,v=1}^{n-1}\tilde{a}_{sn}(h_1)\epsilon_s\tilde{a}_{tn}(h_1)\epsilon_t\tilde{a}_{un}(h_2)\epsilon_u\tilde{a}_{vn}(h_2)\epsilon_v\epsilon_n^4\right]\right\}^{3/4} \\ &\le C\left(\frac{J_n}{nh_{1\min}^d}\right)^{3/2}. \end{aligned}$$

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

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

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

$$\lim_{n\to\infty}\left\{P\left[\max_{h \in \mathcal{H}_n} B_{hn}(\epsilon_1, \ldots, \epsilon_n) \le x\right] - P\left[\max_{h \in \mathcal{H}_n} B_{hn}(\epsilon_1^*, \ldots, \epsilon_n^*) \le x\right]\right\} = 0$$

for any real x is equivalent to

$$\lim_{n\to\infty}\left|E\left(\prod_{h\in\mathcal{H}_n} I\left[B_{hn}(\epsilon_1,\ldots,\epsilon_n)\le x\right]\right)-E\left(\prod_{h\in\mathcal{H}_n} I\left[B_{hn}(\epsilon_1^*,\ldots,\epsilon_n^*)\le x\right]\right)\right|=0.$$

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 $n\to\infty$

$$\left|P\left[\max_{h\in\mathcal{H}_n} B_{hn}(\epsilon_1,\ldots,\epsilon_n)\le x\right]-P\left[\max_{h\in\mathcal{H}_n} B_{hn}(\epsilon_1^*,\ldots,\epsilon_n^*)\le x\right]\right|$$

$$\le C\left(\frac{J_n^3}{nh_{1\min}^{3d}}\right)^{1/2}\to 0. \tag{A.27}$$

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 $x\ge 0$, $h\in\mathcal{H}_n$ and all sufficiently large n*

$$P\left(L_{0n}^*(h)>x\right)\le \exp\left(-\frac{x^2}{4}\right).$$

Proof: Similarly to the proof of Corollary 1 of Chen, Härdle and Li (2003) that, for any small $\delta>0$ there exists a large integer $n_0\ge 1$ such that for $n\ge n_0$ and $x\ge 0$, $|P(L_{0n}^*(h)\le x)-\Phi(x)|<\delta$, where $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du$. We therefore have for any $n\ge n_0$ and $x\ge 0$

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

using $\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{v^2}{2}}dv=\frac{1}{2}$. The proof follows by letting $0<\delta\le\left(1-\frac{\sqrt{2}}{2}\right)e^{-\frac{x^2}{4}}$ for any $x\ge 0$.

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

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

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

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*

$$\lim_{n\to\infty} P\left(Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}^*_\alpha\right) = 1 \tag{A.28}$$

for some $h \in \mathcal{H}_n$*, 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.4), (A.5) and Lemma A.6, L_n can be replaced with $\max_{h\in\mathcal{H}_n} L_{2n}(h)$. By Lemmas A.6 and A.7, l^*_α can be replaced by $\tilde{l}_\alpha$. Thus, it suffices to show that

$$\lim_{n\to\infty} P(\max_{h\in\mathcal{H}_n} L_{2n}(h) > \tilde{l}_\alpha) = 1,$$

which holds if $\lim_{n\to\infty} P(L_{2n}(h) > \tilde{l}_\alpha) = 1$ for some $h \in \mathcal{H}_n$. For any $h \in \mathcal{H}_n$, using (A.3), (A.4), (A.5) and Lemma A.2 again we have

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

Condition (A.28) implies that as $n \to \infty$

$$P\left(Q_n(\theta^*) < 2h^{d/2}\tilde{l}^*_\alpha\right) \to 0. \tag{A.30}$$

Observe that

$$\begin{aligned} P(L_{2n}(h) > \tilde{l}_\alpha) &= P\left(L_{2n}(h) > \tilde{l}_\alpha, Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}^*_\alpha\right) \\ &+ P\left(L_{2n}(h) > \tilde{l}_\alpha, Q_n(\theta^*) < h^{d/2}2\tilde{l}^*_\alpha\right) \equiv I_{1n} + I_{2n}. \end{aligned}$$

Thus, it follows from (A.29) that as $n \to \infty$

$$\begin{aligned} I_{1n} &= P\left(L^*_{0n}(h) + h^{-d/2}Q_n(\theta^*) + h^{-d/2}\Pi_n(\theta^*) > \tilde{l}_\alpha | Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}^*_\alpha\right) P\left(Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}^*_\alpha\right) \\ &= P\left(L^*_{0n}(h) + h^{-d/2}Q_n(\theta^*)(1 + o_p(1))) > \tilde{l}_\alpha | Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}^*_\alpha\right) P\left(Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}^*_\alpha\right) \\ &\geq P\left(L^*_{0n}(h) > \tilde{l}_\alpha - 2\tilde{l}^*_\alpha | Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}^*_\alpha\right) P\left(Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}^*_\alpha\right) \to 1 \end{aligned}$$

because $L_{0n}^*(h)$ is asymptotically normal and therefore bounded in probability and $\tilde{l}_\alpha - 2\tilde{l}_\alpha^* \to -\infty$ as $n \to \infty$. Because of (A.30), $\lim_{n\to\infty} I_{2n} \leq P\left(Q_n(\theta^*) < 2h^{d/2}\tilde{l}_\alpha^*\right) = 0$. This finishes the proof.

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

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

$$\lim_{n\to\infty} P\left(h^d \lambda(\theta)^\tau \lambda(\theta) \geq 4\tilde{l}_\alpha^* h^{d/2}\right) = 1. \tag{A.31}$$

Proof of Theorem 3.2: In view of the definition of $\tilde{l}_\alpha^*$, equation (A.31) follows from the fact that as $n \to \infty$,

$$\frac{1}{n}\lambda(\theta)^\tau \lambda(\theta) - \rho(m, \mathcal{M}) \to 0 \tag{A.32}$$

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

Proof of Theorem 3.3: Using the definition of $\tilde{l}_\alpha^*$, (A.32),

$$\frac{1}{n}\sum_{t=1}^n \Delta^2(X_t) \to E_S\left[\Delta^2(X_1)\right] = \int_{x\in S} \Delta^2(x) f(x) dx \geq D_1 > 0 \text{ as } n \to \infty, \tag{A.33}$$

and the fact that

$$\frac{1}{n}\lambda(\theta)^\tau \lambda(\theta) = \frac{C_n^2}{n}\sum_{t=1}^n \Delta^2(X_t) \geq D_1 C_n^2 \tag{A.34}$$

holds in probability, one can see that (A.31) holds when $h = h_{\max} = (\log\log(n))^{-\frac{1}{d}}$. 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: $h_1 = \left(n^{-1}\tilde{l}_\alpha^*\right)^{\frac{2}{4s+d}}$. This implies $nh_1^{\frac{4s+d}{2}} = \tilde{l}_\alpha^*$. Choose $h \in \mathcal{H}_n$ such that $h_1 \leq h < 2h_1$. We then have

$$4h^{\frac{d}{2}}\tilde{l}_\alpha^* = 4nh^{\frac{d}{2}}h_1^{\frac{4s+d}{2}} \leq 4nh^{\frac{4s+d}{2}+\frac{d}{2}} = 4nh^{2s+d}. \tag{A.35}$$

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

$$Q_n(\theta^*) \geq 4nh^{2s+d} \tag{A.36}$$

holds in probability for the selected $h \in \mathcal{H}_n$ and $\theta^* \in \Theta$. The verification of (A.36) can be done using similar techniques employed in the proof of Lemma A.2. Alternatively, one may follow the proof of (A13) of Horowitz and Spokoiny (2001) by noting that all the derivations below their (A13) hold in probability with respect to the distribution of $\{X_i\}$.

Acknowledgements: The authors thank Cheng Yong Tang for valuable computation assistance. The second author acknowledges support from an Australian Research Council Discovery Grant.

References:

Andrews, D. W. K. (1997) A conditional Kolmogorov test. *Econometrica*, **65**, 1097–1028.

Chen, S. X., Härdle, W. and 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. and Li, M. (2003) An empirical likelihood goodness–of–fit test for time series. *J. Roy. Statist. Soc. Ser. B.*, **65**, 663-678.

Einmahl, J. H. J. and McKeague, I. W. (2003) Empirical likelihood based hypothesis testing. *Bernoulli*, **9**, 267–290.

Eubank, R. L. and Hart, J. D. (1992) Testing goodness-of-fit in regression via order selection. *Ann. Statist.*, **20**, 1412–1425.

Eubank, R. L. and Spiegelman, C. H. (1990) Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.*, **85**, 387–392.

Fan, J. (1996), Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.*, **434**, 674–688.

Fan, J. and Zhang, J. (2004). Sieved empirical likelihood ratio tests for nonparametric functions. *Ann. Statist.*, **29**, 153–193.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.*, **29**, 153–193.

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

Gao, J. and King, M. L. (2003) 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. and Wolff, R. C. L. (2002). Model specification tests in nonparametric stochastic regression models. *Journal of Multivariate Analysis*, **83**, 324–359.

Härdle, W. and Mammen, E. (1993) Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21**, 1926–1947.

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

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.

Hong, Y., White, H. (1995) Consistent specification testing via nonparametric series regression. *Econometrica*, **63**, 1133–1159.

Horowitz, J. L., and Härdle, W. (1994) Testing a parametric model against a semiparametric alternative. *Econometric Theory*, **10**, 821–848.

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

Ingster, Yu. I. and Suslina, I. A. (2003) Nonparametric Goodness-of-Fit Testing Under Guassian Models," *Lecture Notes in Statistics*, **169**. Springer, New York.

Li, G. (2003) Nonparametric likelihood ratio goodness-of-fit tests for survival data. *Journal of Multivariate Analysis*, **86**, 166–182.

Li, G. and Van Keilegom, I. (2002). Likelihood ratio confidence bands in nonparametric regression with censored data. Scand. J. Statist., 29, 547-562.

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

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

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

Owen, A. (1990) Empirical likelihood ratio confidence regions, *Ann. Statist.* **18**, 90–120.

Robinson, P. (1989). Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Review of Economic Studies*, **56**, 511–534.

Sperlich, S., Tjøstheim, D. and Yang, L. (2002) Nonparametric estimation and testing of interaction in additive models. *Econometric Theory*, **18**, 197–251.

Spokoiny, V. G. (1996) Adaptive hypothesis testing using wavelets. *Ann. Statist.*, **24**, 2477–2498.

Tripathi, G., and Y. Kitamura (2003) On testing conditional moment restrictions: the canonical case. *Ann. Statist.* (to appear).

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

Zhang, C. and Dette, H. (2003). A power comparison between nonparametric regression tests. *Statist. Probab. Lett.*, **66**, 289–301.

TABLE 1

SIMULATION RESULTS ON MODEL (4.1)

		Probability of Rejecting Null Hypothesis			
Distribution ϵ	τ	Andrews' Test	Härdle-Mammen Test	Horowitz-Spokoiny Test	EL Test
		Null Hypothesis Is True			
Normal		0.057	0.060	0.066	0.053
Mixture		0.053	0.053	0.054	0.05
Extreme Value		0.063	0.057	0.055	0.057
		Null Hypothesis Is False			
Normal	1.0	0.680	0.752	0.792	0.90
Mixture	1.0	0.692	0.736	0.796	0.898
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	0.932	0.919
Extreme Value	0.25	0.604	0.696	0.968	0.989

TABLE 2

SIMULATION RESULTS ON MODEL (4.2)

		Horowitz-Spokoint Test		EL Test	
	C_n	$n = 300$	$n = 500$	$n = 300$	$n = 500$
Null Hypothesis	0	0.054	0.052	0.064	0.049
Alternative Hypothesis	0.03	0.064	0.060	0.18	0.252
Alternative Hypothesis	0.04	0.056	0.068	0.306	0.412