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

Follow this and additional works at: http://lib.dr.iastate.edu/stat_las_preprints

Part of the Statistics and Probability Commons

Recommended Citation
http://lib.dr.iastate.edu/stat_las_preprints/43

This Article is brought to you for free and open access by the Statistics at Iowa State University Digital Repository. It has been accepted for inclusion in Statistics Preprints by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
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 article is available at Iowa State University Digital Repository: http://lib.dr.iastate.edu/stat_las_preprints/43
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

\[ Y_t = m(X_t) + \sigma(X_t)\epsilon_t, \quad t = 1, 2, \ldots, n \]  

(1.1)

where both \( m(\cdot) \) and \( \sigma(\cdot) \) are unknown functions defined over \( \mathbb{R}^d \), the data \( \{(X_t, Y_t)\}_{t=1}^n \) are weakly dependent stationary time series, and \( \epsilon_t \) 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 \( \theta \in \mathbb{R}^q \) is an unknown parameter belonging to a parameter space \( \Theta \). This paper considers testing the validity of the parametric specification of \( m_\theta(x) \) against a series of local alternatives, that is to test

\[ H_0 : m(x) = m_\theta(x) \quad \text{versus} \quad H_1 : m(x) = m_\theta(x) + C_n\Delta_n(x) \quad \text{for all} \ x \in S, \]  

(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 \( \mathbb{R}^d \) and \( S \) is a compact set in \( \mathbb{R}^d \). Both \( C_n \) and \( \Delta_n(x) \) characterize the departure of the local alternative family of regression models from the parametric family \( \{m_\theta(\cdot)|\theta \in \Theta\} \).


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 hypothesized 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

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

Let $\tilde{\theta}$ be a consistent estimator of $\theta$ 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 $Q_t(x) = K_h(x - X_t)\{Y_t - \tilde{m}_\theta(x)\}$. 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 $\tilde{m}_\theta(x)$ is

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

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

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

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

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

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

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

The EL test statistic at a given bandwidth $h$ is

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

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

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

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

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

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

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

$$\mathcal{H}_n = \{h = h_{\text{max}}a^k : h \geq h_{\text{min}}, \ k = 0, 1, 2, \ldots, J_n\}$$
be a set of bandwidths, where \(0 < a < 1\), \(J_n = \log_{1/a}(h_{\text{max}}/h_{\text{min}})\) is the number of bandwidths, \(h_{\text{max}} = c_{\text{max}} \log \log(n)^{-\frac{1}{4}}\) and \(h_{\text{min}} = c_{\text{min}} n^{-\gamma}\) for \(0 < \gamma < \frac{1}{4}\) and some positive constants \(-\infty < c_{\text{min}}, c_{\text{max}} < \infty\). The choice of \(h_{\text{max}}\) 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 \(E\{\ell(\hat{m}_\theta; h)\} = 1\) under \(H_0\) and \(\text{var}\{\ell(\hat{m}_\theta; h)\} = C(K, \pi)h^d\) as given in (2.5) the adaptive EL test statistic is proposed as follows:

\[
L_n = \max_{h \in H_n} \frac{\ell(\hat{m}_\theta; h) - 1}{\sqrt{C(K, \pi)h^d}}.
\]

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

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

1. For each \(t = 1, 2, \ldots, n\), let \(Y_t^* = m_{\hat{\theta}}(X_t) + \sigma_n(X_t)\varepsilon_t^*, \) where \(\sigma_n(\cdot)\) is a consistent estimator of \(\sigma(\cdot), \{\varepsilon_t^*\}\) is independent of \(\{X_s\}\) for all \(s \geq 1\), and sampled randomly from a specified distribution with \(E[\varepsilon_t^*] = 0, E[\varepsilon_t^{*2}] = 1, \) and \(E[|\varepsilon_t^*|^{4+\delta}] < \infty\) for some \(\delta > 0\). Define \(l_n^*\) 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 \(\theta\) based on the resample \(\{(X_t, Y_t^*)\}_{t=1}^n\). Compute the statistic \(L_n^*\) by replacing \(Y_t\) and \(\hat{\theta}\) with \(Y_t^*\) and \(\hat{\theta}^*\) according to (2.8).

3. Estimate \(l_\alpha^*\) 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 \(Y_t^* = m_{\hat{\theta}}(Y_{t-1}^*) + \sigma_n(Y_{t-1}^*)\varepsilon_t^*\) when \(\{Y_t\}\) of (1.1) satisfies a nonparametric autoregressive model. The estimator \(\sigma_n^2(\cdot)\) can be the following kernel estimator

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

with a bandwidth \(b\) satisfying \(nh_{\text{min}}b^d \to \infty\) 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 \( \{e_t^n\} \) for the bootstrap. The first is the one used in Horowitz and Spokoiny (2001) which generates independent and identically distributed \( e_t^n \) from \( N(0,1) \). The second approach is the regression bootstrap proposed by Franke, Kreiss and Mammen (2002), where the generation of \( \{e_t^n\} \) 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 conjunction 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 \( \alpha \)-mixing with the mixing coefficient \( \alpha(t) = \sup \{|P(A \cap B) - P(A)P(B)| : A \in \Omega_t^1, B \in \Omega_{t+1}^\infty \} \) for all \( s, t \geq 1 \), where \( \Omega_t^1 \) denote the \( \sigma \)-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 \( t \geq 1 \), \( E[e_t^1|\Omega_{t-1}] = 0 \) and \( E[e_t^2|\Omega_{t-1}] = 1 \), where \( \Omega_t \) are \( \sigma \)-fields generated by \( \{(X_s, Y_s) : 1 \leq s \leq t \} \).

(iii) Let \( e_t = Y_t - m(X_t) \). There exists a constant \( \delta_e > 0 \) such that \( E\left[|e_t^1 e_{t+1}^2 \cdots e_{t+\delta_e}^n|^{1+\delta_e}\right] < \infty \), 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 \( R^d \), \( \mu_i(x) = E[e_t^i | X = x] \) and \( \pi \) 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 \( m_2(x) \) are continuous on \( S \), \( \inf_{x \in S} \sigma(x) \geq C_0 > 0 \) and \( \inf_{x \in S} m(x) \geq C_1 > 0 \) for constants \( C_0 \) and \( C_1 \).

(ii) Let \( f_{\tau_1, \tau_2, \cdots, \tau_l} \) be the joint probability density of \( (X_{1+\tau_1}, \cdots, X_{1+\tau_l}) \) \((1 \leq l \leq 4)\). Assume that each \( f_{\tau_1, \tau_2, \cdots, \tau_l} \) 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 \( \int k(t)dt = 1 \), \( \int t^r k(t)dt = 0 \) 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 \( \nabla^q_\theta m_\phi(x) = \frac{\partial^q m_\phi(x)}{\partial \theta^q} \) whenever these derivatives exist. For any \( q \times q \) matrix \( D \), define \( ||D||_{\infty} = \sup_{v \in R^q} \frac{||Dv||}{||v||} \) where \( ||B||_m^2 = \sum_{i=1}^q \sum_{j=1}^q b_{ij} \) for a \( q \times q \) matrix \( B = (b_{ij})_{1 \leq i,j \leq q} \).
Assumption 3.3. (i) The parameter set $\Theta$ is an open subset of $\mathbb{R}^q$ 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 $E[\sup_{\theta \in \Theta}|m_\theta(X_1)|^2] \leq C_1$ and $\max_{1 \leq j \leq 3} E[\sup_{\theta \in \Theta}|| \nabla_\theta^j m_\theta(X_1)||_{m_\theta}] \leq C_2$. For each $\theta \in \Theta$, $m_\theta(x)$ is continuous with respect to $x \in S$. There is a finite $C_1 > 0$ such that for every $\varepsilon > 0$, $\int_{x \in S} \inf_{\theta, \theta' \in \Theta: ||\theta - \theta'|| \geq \varepsilon} |m_\theta(x) - m_{\theta'}(x)|^2 f(x)dx \geq C_1 \varepsilon^2$.

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

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

(iv) The set $H_n$ has the structure of (2.7) with $h_{\text{max}} > h_{\text{min}} = O(n^{-\gamma})$ for some constant $0 < \gamma < \frac{1}{3d}$ and $h_{\text{max}} = C_h(\log \log(n))^{-\frac{1}{d}}$ 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 $h_{\text{min}} = O(n^{-\gamma})$ where $0 < \gamma < \frac{1}{3d}$. To include the optimal order $n^{-\frac{d+2}{d+3}}$ for estimating $m(x)$ in the range of $H_n$, 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 $n^{-\frac{d+2}{d+3}}$ covered by $H_n$.

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 $H_0$,

$$\lim_{n \to \infty} P(L_n > l'_{\alpha}) = \alpha.$$
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

$$m(x) = m_\theta(x) + C_n \Delta(x)$$  \hspace{1cm} (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 $D_3$,

$$0 < D_1 \leq \int_{x \in S} \Delta^2(x)f(x)dx \leq D_2 < \infty \text{ and } \rho(m, \mathcal{M}) \geq D_3 C_n.$$  \hspace{1cm} (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 $C_n \geq Cn^{-1/2}\sqrt{\log \log(n)}$ 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 $D^j m(x) = \frac{\partial^{j_1} m(x)}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}}$ 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

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

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

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, $\theta = (\beta_0, \beta_1)^T = (1,1)^T$ is chosen as the true vector of parameters and $\phi$ 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 $\epsilon_i$ and other aspects of the simulation. We used the same number of simulation, the bootstrap resamples and estimation procedures for $\theta$ as in Horowitz and Spokoiny (2001). We also employed the same kernel, the same bandwidth set $\mathcal{H}_n$, the same estimator $\sigma_n^2$ and the distribution for $\{\epsilon_i^*\}$ in the Monte Carlo simulation procedure as in Horowitz and Spokoiny (2001). Like Horowitz and Spokoiny, the nominal size of the test was 5%. 

10
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:

\[ Y_i = 0.25 + 0.5Y_{i-1} + C_n \cos(8Y_{i-1}) + 0.25 \sqrt{Y_{i-1}^2 + 1} \epsilon_i, \]  

(4.2)

where the innovation \( \{\epsilon_i\}_{i=1}^{n} \) 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, \( \{\epsilon_i^*\} \) 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 \( H_n = \{0.3, 0.332, 0.367, 0.407, 0.45\} \) with \( a = 0.903 \) for \( n = 300 \) and \( H_n = \{0.25, 0.281, 0.316, 0.356, 0.4\} \) 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 $n \to \infty$,

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

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

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

where

$$\tilde{U}_1(x; \hat{\theta}) = (nh^d)^{-1} \sum_{t=1}^n K\left( \frac{x - X_t}{h} \right) \{Y_t - \tilde{m}_{\hat{\theta}}(x)\}.$$
and \( v(x) = R(K)f^{-1}(x)\sigma^2(x) \).

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

\[
\ell_0(\theta) = \sum_{s,t} a_{st} \epsilon_s \epsilon_t \quad \text{and} \quad \ell_n(\theta) = \sum_{s,t} a_{st} \lambda_s(\theta) \lambda_t(\theta). \tag{A.3}
\]

Then the leading term in \( \ell_n(\hat{\theta}_n; h) \) is

\[
\ell_1(h, \hat{\theta}) \equiv nh^d \int U_1^2(x; \hat{\theta})v^{-1}(x)\pi(x)\,dx = \ell_0(h) + Q_n(\hat{\theta}) + \Pi_n(h), \tag{A.4}
\]

where \( \Pi_n(h) = \ell_1(h, \hat{\theta}) - \ell_0(h) - Q_n(\hat{\theta}) \) is the remainder term.

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

\[
L_{0n}(h) = \frac{\ell_0(h) - 1}{h^{d/2}}, \quad L_{1n}(h) = \frac{\ell_1(h, \hat{\theta}) - 1}{h^{d/2}} \quad \text{and} \quad L_{2n}(h) = \frac{\ell_1(h, \theta^*) - 1}{h^{d/2}}, \tag{A.5}
\]

where \( \theta^* = \theta_0 \) when \( H_0 \) is true and \( \theta^* \) 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 \( l_n^* \) by \( l_n \) 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 \( \max_{h \in \mathcal{H}_n} \sup_{|\theta - \theta_0| \leq \delta} \frac{Q_n(\theta)}{nh^d} \leq C\delta^2 \) holds in probability, where \( C > 0 \) is a constant.

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

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

We now have

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

13
\[ C(1 + o_p(1))h^d \max_{1 \leq t \leq n} (f(X_t) q(X_t)) \int K(u)du \leq C h^d \quad (A.6) \]

using the fact that

\[ a_{st} = n h^d \int W_s(x) W_t(x) q(x) dx = \int \frac{K\left(\frac{x - X_t}{h}\right)}{\sum_{u=1}^{n} K\left(\frac{x - X_u}{h}\right)} f(x)K\left(\frac{x - X_t}{h}\right) q(x) dx \]

\[ = (1 + o_p(1)) \int \frac{K\left(\frac{x - X_s}{h}\right)}{\sum_{u=1}^{n} K\left(\frac{x - X_u}{h}\right)} K\left(\frac{x - X_t}{h}\right) q(x) f(x) dx. \]

In order to prove Lemma A.2(i), it suffices to show that \( \sup_{|\theta - \theta_0| \leq \delta} \| \lambda(\theta) \|^2 \leq C n \delta^2 \) 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 \( \lambda_{\min}(A) \cdot \| \lambda(\theta) \|^2 \leq Q_n(\theta) \leq \lambda_{\max}(A) \| \lambda(\theta) \|^2 \), in order to prove Lemma A.2(ii), it suffices to show that for \( n \) large enough, \( \lambda_{\min}(A) \geq C_1 h^d (1 + o_p(1)) \) 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 \( 1 \leq j \leq 3 \), define \( \psi_j(X_t, \theta) = m^{(j)}_{\theta}(X_t) = \frac{d^{j-1} m_{\theta}(X_t)}{d\theta^{j-1}} \).

**Lemma A.3.** (i) Under Assumptions 3.1, 3.2 and 3.3(i), we have for any given \( \theta \in \Theta \)

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

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

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

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

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

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

14
\[
\leq \sum_{h \in \mathbb{H}_n} \frac{1}{C_{\mathbb{H}_n}^2 J_n h^d} E \left[ \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \epsilon_s \psi_1(X_t, \theta) \right]^2
\]
\[
\leq \sum_{h \in \mathbb{H}_n} \frac{1}{C_{\mathbb{H}_n}^2 J_n h^d} \left\{ \sum_{s=1}^{n} \sum_{t=1}^{n} E [a_{st} \epsilon_s \psi_1(X_t, \theta)]^2 + \Pi_{1n}(\theta) \right\},
\]
where \( \Pi_{1n}(\theta) = E [\sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \epsilon_s \psi_1(X_t, \theta)]^2 - \sum_{s=1}^{n} \sum_{t=1}^{n} E [a_{st} \epsilon_s \psi_1(X_t, \theta)]^2. \)

A direct calculation shows that as \( n \to \infty \)
\[
\sum_{s=1}^{n} \sum_{t=1}^{n} E [a_{st} \epsilon_s \psi_1(X_t, \theta)]^2 = \tag{A.9}
\]
\[
\int \int q(x)q(y)K^{(2)}\left(\frac{x-y}{h}\right)K^{(2)}\left(\frac{y-x}{h}\right)\sigma^2(x)\psi_1^2(y, \theta)f(x, y)dxdy = C(\theta)h^d(1 + o(1))
\]
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 \( H_0 \),
\[
\max_{h \in \mathbb{H}_n} \sup_{|\theta - \theta_0| \leq u^{-1/2} n} h^{-d/2} \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \epsilon_s \lambda_t(\theta) \right| = O_p \left( J_n^{1/2} n^{-1/2} \right). \tag{A.11}
\]

Proof: Using a Taylor series expansion to \( m_\theta(X_t) - m_{\theta_0}(X_t) \) and Assumption 3.3(i), we have for \( \theta' \) between \( \theta \) and \( \theta_0 \)
\[
\left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \epsilon_s \lambda_t(\theta) \right| = \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \epsilon_s [m_\theta(X_t) - m_{\theta_0}(X_t)] \right|
\]
\[
\leq \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \epsilon_s \psi_1(X_t, \theta_0) \right| |\theta - \theta_0| + \frac{1}{2} \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \epsilon_s \psi_2(X_t, \theta_0) \right| |\theta - \theta_0|^2
\]
\[
+ \frac{1}{6} \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \epsilon_s \psi_3(X_t, \theta') \right| |\theta - \theta_0|^3 \leq \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \epsilon_s \psi_1(X_t, \theta_0) \right| |\theta - \theta_0|
\]

15
\[ + \frac{1}{2} n |\theta - \theta_0|^2 \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \epsilon_s \psi_2(X_t, \theta_0) \right| \]
\[ + \frac{1}{6} n |\theta - \theta_0|^3 \left| \sum_{s=1}^{n} a_{st} \epsilon_s \max_{1 \leq t \leq n} |\psi_3(X_t, \theta')| \right|. \] (A.12)

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

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

\[ \sup_{||\theta - \theta_0|| \leq n^{-1/2}u} \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \epsilon_s \lambda(X_t, \theta) \right| = o_p(q_n h^{d/2}). \] (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 \( n \to \infty \)

\[ \max_{h \in \mathcal{H}_n} L_n(h) = \max_{h \in \mathcal{H}_n} L_{1n}(h) = \max_{h \in \mathcal{H}_n} L_{2n}(h) + o_p(1), \]
\[ \max_{h \in \mathcal{H}_n} L^*_n(h) = \max_{h \in \mathcal{H}_n} L^*_n(h) + o_p(1), \]

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

\[ \max_{h \in \mathcal{H}_n} h^{-d/2} \left( \sum_{s=1}^{n} a_{ss} \epsilon_s^2 - 1 \right) = o_p(1) \] and \( \max_{h \in \mathcal{H}_n} h^{-d/2} \left( \sum_{s=1}^{n} a_{ss} \epsilon_s^2 - 1 \right) = o_p(1). \) (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 \(\epsilon_t^*\) and define

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

(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

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

Like Horowitz and Spokoiny (2001), there are two steps in the proof of Lemma A.3. First, we want to show that

\[
|E \left[ G(B_n(\epsilon_1, \ldots, \epsilon_n)) \right] - E \left[ G(B_n(\epsilon_1^*, \ldots, \epsilon_n^*)) \right] | \leq C_0 C_n(G) \left( \frac{J_n^3}{n h_1 d_{\min}} \right)^{1/2}
\]

(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

\[
|E \left[ G(B_n(\epsilon_1, \ldots, \epsilon_n)) \right] - E \left[ G(B_n(\epsilon_1^*, \ldots, \epsilon_n^*)) \right] | \leq \sum_{t=1}^n \left| E \left[ G(B_n(\epsilon_1, \ldots, \epsilon_t, \epsilon_t^*, \ldots, \epsilon_n^*)) \right] - E \left[ G(B_n(\epsilon_1, \ldots, \epsilon_t^*, \ldots, \epsilon_n^*)) \right] \right|
\]

(A.18)

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 \(U_{n-1}, \Lambda_n\) and \(\tilde{\Lambda}_n\), respectively, denote the vectors that are obtained by stacking

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

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

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

17
we have $\frac{1}{2}E \left[ \Lambda_n G''(U_{n-1}) \Lambda_n - \tilde{\Lambda}_n G''(U_{n-1}) \tilde{\Lambda}_n \right] + \frac{C_n(G)}{6} \left\{ E \left[ \| \Lambda_n \| ^3 \right] + E \left[ \| \tilde{\Lambda}_n \| ^3 \right] \right\},$

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 $E [\epsilon_n | \Omega_{n-1}] = E [\epsilon_n^2] = 0$ and $E [\epsilon_n^2 | \Omega_{n-1}] = E [\epsilon_n^2] = 1,$

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

This implies

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

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

\[ a_{st} = \frac{1}{nh^d} \int K \left( \frac{x - X_s}{h} \right) K \left( \frac{x - X_t}{h} \right) q(x) dx \]

\[ = \frac{1}{n} \int K(u)K \left( u + \frac{X_s - X_t}{h} \right) q(X_s + uh)du = \frac{1}{n} L_2 \left( \frac{X_s - X_t}{h}, X_s \right), \]

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 $\delta > 0$ involved in Assumption 3.1(iii),

\[ \sum_{h_1 \in H_n} \sum_{h_2 \in H_n} E \left[ \sum_{s=1}^{n-1} \sum_{t=1, \neq s}^{n-1} \tilde{a}_{tn}^2(h_1) \tilde{a}_{tn}^2(h_2) \epsilon_s^2 \epsilon_t^2 \right] \quad (A.21) \]

\[ \leq \sum_{h_1 \in H_n} \sum_{h_2 \in H_n} \frac{1}{n^4 h_1^d h_2^d} \sum_{s=1}^{n-1} \sum_{t=1, \neq s}^{n-1} E \left[ L_2^2 \left( \frac{X_s - X_n}{h_1}, X_s \right) L_2^2 \left( \frac{X_t - X_n}{h_2}, X_t \right) \epsilon_s^2 \epsilon_t^2 \right] \]

\[ \leq C \sum_{h_1 \in H_n} \sum_{h_2 \in H_n} \frac{1}{n^4 h_1^d h_2^d} \sum_{s=1}^{n-1} \sum_{t=1, \neq s}^{n-1} E \left[ \epsilon_s^2 \epsilon_t^2 \epsilon_n^2 \right] \]

\[ \leq C \sum_{h_1 \in H_n} \sum_{h_2 \in H_n} \frac{1}{n^2 h_1^d h_2^d} \leq C \cdot \left( \frac{J_n}{nh_{1 \min}^d} \right)^2, \]

where $0 < C < \infty$ is a constant.
Similarly to the proof of Lemma C.2 of Gao and King (2003), as \( n \to \infty \)

\[
\sum_{h_1,h_2 \in \mathcal{H}_n} E \left( \sum_{1 \leq s \neq t \leq n-1} \tilde{a}_{sn}^2(h_1)\tilde{a}_{tn}(h_2)\tilde{a}_{tn}(h_2)\varepsilon_t^2\varepsilon_t^4 \right) = o \left( \frac{J_n}{nh_{1\min}^4} \right)^2 , \quad (A.22)
\]

\[
\sum_{h_1,h_2 \in \mathcal{H}_n} E \left( \sum_{1 \leq s \neq t,s \neq u,t \neq u \leq n-1} \tilde{a}_{sn}^2(h_1)\tilde{a}_{tn}(h_2)\tilde{a}_{un}(h_2)\varepsilon_t^2\varepsilon_t\varepsilon_u^4 \right) = o \left( \frac{J_n}{nh_{1\min}^4} \right)^2 ,
\]

\[
\sum_{h_1,h_2 \in \mathcal{H}_n} E \left( \sum \tilde{a}_{sn}(h_1)\tilde{a}_{tn}(h_1)\tilde{a}_{un}(h_2)\tilde{a}_{vn}(h_2)\varepsilon_t\varepsilon_t\varepsilon_u\varepsilon_v^4 \right) = o \left( \frac{J_n}{nh_{1\min}^4} \right)^2 ,
\]

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

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

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

\[
\sum_{h_1 \in \mathcal{H}_n} \sum_{h_2 \in \mathcal{H}_n} E \left[ \sum_{s,t,u,v=1}^{n-1} \tilde{a}_{sn}(h_1)\varepsilon_s\tilde{a}_{tn}(h_1)\varepsilon_t\tilde{a}_{un}(h_2)\varepsilon_u\tilde{a}_{vn}(h_2)\varepsilon_v\varepsilon_v^4 \right] \leq C \cdot \left( \frac{J_n}{nh_{1\min}^4} \right)^2 . \quad (A.24)
\]

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 \( n \to \infty \)

\[
E \left[ ||\Lambda_n||^3 \right] = 8 E \left[ \left| \sum_{s=1}^{n-1} \tilde{A}_{sn} \varepsilon_s \varepsilon_n \right|^3 \right] \leq 8 \left\{ E \left[ \sum_{h \in \mathcal{H}_n} \left( \sum_{s=1}^{n-1} \tilde{a}_{sn}(h)\varepsilon_s \varepsilon_n \right)^2 \right] \right\}^{3/4} \quad (A.25)
\]

\[
= 8 \left\{ \sum_{h_1 \in \mathcal{H}_n} \sum_{h_2 \in \mathcal{H}_n} E \left[ \sum_{s,t,u,v=1}^{n-1} \tilde{a}_{sn}(h_1)\varepsilon_s\tilde{a}_{tn}(h_1)\varepsilon_t\tilde{a}_{un}(h_2)\varepsilon_u\tilde{a}_{vn}(h_2)\varepsilon_v\varepsilon_v^4 \right] \right\}^{3/4}
\leq C \left( \frac{J_n}{nh_{1\min}^4} \right)^{3/2} .
\]

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

\[
E \left[ ||\Lambda_n||^3 \right] + E \left[ ||\tilde{\Lambda}_n||^3 \right] \leq 2C \left( \frac{J_n}{nh_{1\min}^4} \right)^{3/2} . \quad (A.26)
\]

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

\[
\lim_{n \to \infty} \left\{ P \left[ \max_{h \in \mathcal{H}_n} B_{hn}(\varepsilon_1, \ldots, \varepsilon_n) \leq x \right] - P \left[ \max_{h \in \mathcal{H}_n} B_{hn}^\ast(\varepsilon_1, \ldots, \varepsilon_n) \leq x \right] \right\} = 0
\]

19
for any real $x$ is equivalent to

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

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 $n \to \infty$

$$P \left( \max_{h \in \mathcal{H}_n} B_{hn}(\epsilon_1, \ldots, \epsilon_n) \leq x \right) - P \left( \max_{h \in \mathcal{H}_n} B_{hn}(\epsilon_1^*, \ldots, \epsilon_n^*) \leq x \right) \leq C \left( \frac{J_n^3}{nh_{\text{min}}^3} \right)^{1/2} \to 0. \quad (A.27)$$

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

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

**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 $n_0 \geq 1$ such that for $n \geq n_0$ and $x \geq 0$, $|P(L_{0n}^*(h) \leq x) - \Phi(x)| < \delta$, where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$. We therefore have for any $n \geq n_0$ and $x \geq 0$

$$P(L_{0n}^*(h) > x) \leq 1 - \Phi(x) + \delta = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{u^2}{2}} du + \delta$$

$$= \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{u^2}{2}} e^{-\frac{\delta^2}{2}} du + \delta \leq e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{u^2}{2}} du + \delta$$

$$\leq e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{u^2}{2}} du + \delta = e^{-\frac{x^2}{2}} \frac{\sqrt{2}}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{v^2}{2}} dv + \delta = \frac{\sqrt{2}}{2} e^{-\frac{x^2}{2}} + \delta$$

using $\frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{v^2}{2}} dv = \frac{1}{\sqrt{2}}$. The proof follows by letting $0 < \delta \leq \left( 1 - \frac{\sqrt{2}}{2} \right) e^{-\frac{x^2}{2}}$ for any $x \geq 0$. For $0 < \alpha < 1$, define $\hat{l}_\alpha$ 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 $n$

$$\hat{l}_\alpha \leq 2 \sqrt{\log(J_n) - \log(\alpha)}.$$
Thus, it follows from \((A.29)\) that as
\[
\lim_{n \to \infty} P \left( Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha^* \right) = 1
\]
\[(A.28)\]
for some \(h \in \mathcal{H}_n\), where 
\[
\tilde{l}_\alpha^* = \max \left( \tilde{l}_\alpha, \sqrt{2\log(J_n) + 2\log(J_n)} \right).
\]
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, \(l\alpha\) 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
\[
L_{2n}(h) = L_{0n}(h) + h^{-d/2}Q_n(\theta^*) + h^{-d/2}\Pi_n(\theta^*)
\]
\[(A.29)\]
\[
= L_{0n}(h) + h^{-d/2}Q_n(\theta^*) + h^{-d/2}\Pi_n(\theta^*) + o_p(1)
\]
\[
= L_{0n}(h) + h^{-d/2}Q_n(\theta^*)(1 + o_p(1)) + o_p(1).
\]
Condition \((A.28)\) implies that as \(n \to \infty\)
\[
P \left( Q_n(\theta^*) < 2h^{d/2}\tilde{l}_\alpha^* \right) \to 0.
\]
\[(A.30)\]
Observe that
\[
P(L_{2n}(h) > \tilde{l}_\alpha) = P \left( L_{2n}(h) > \tilde{l}_\alpha, Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha^* \right)
\]
\[
+ P \left( L_{2n}(h) > \tilde{l}_\alpha, Q_n(\theta^*) < h^{d/2}\tilde{l}_\alpha^* \right) \equiv I_{1n} + I_{2n}.
\]

Thus, it follows from \((A.29)\) that as \(n \to \infty\)
\[
I_{1n} = P \left( L_{0n}(h) + h^{-d/2}Q_n(\theta^*) + h^{-d/2}\Pi_n(\theta^*) > \tilde{l}_\alpha | Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha^* \right) P \left( Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha^* \right)
\]
\[
= P \left( L_{0n}(h) + h^{-d/2}Q_n(\theta^*)(1 + o_p(1)) \right) > \tilde{l}_\alpha | Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha^* \right) P \left( Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha^* \right)
\]
\[
> P \left( L_{0n}(h) > \tilde{l}_\alpha - 2\tilde{l}_\alpha^* | Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha^* \right) P \left( Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha^* \right) \to 1
\]

**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
\[
\lim_{n \to \infty} P \left( Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha^* \right) = 1
\]
\[(A.28)\]
because \( L_{0n}(h) \) is asymptotically normal and therefore bounded in probability and \( \tilde{I}_a - 2\tilde{I}_a^* \to -\infty \) as \( n \to \infty \). Because of (A.30), \( \lim_{n \to \infty} I_{2n} \leq P \left( Q_n(\theta^*) < 2h^{d/2}\tilde{I}_a^* \right) = 0 \). 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 \( \max_{h \in \mathcal{H}_n} L_{1n}(h) \to 0 \) 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

\[
\lim_{n \to \infty} P \left( h^d / \lambda(\theta) \lambda(\theta) \geq 4\tilde{I}_a^* h^{d/2} \right) = 1. \tag{A.31}
\]

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

\[
\frac{1}{n} \lambda(\theta) \lambda(\theta) = \rho(m, \mathcal{M}) \to 0 \tag{A.32}
\]

holds in probability and \( nh^d \geq C_0 \tilde{I}_a^* h^{d/2} \) for some constant \( 0 < C_0 < \infty \) and \( n \) large enough.

**Proof of Theorem 3.3:** Using the definition of \( \tilde{I}_a^* \), (A.32),

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

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 \( h = h_{\text{max}} = (\log \log(n))^{-\frac{1}{2}} \). 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: \( h_1 = \left( n^{-1} \tilde{I}_a^* \right)^{\frac{1}{2+d}} \). This implies \( nh_1^{\frac{4+d}{2}} = \tilde{I}_a^* \). Choose \( h \in \mathcal{H}_n \) such that \( h_1 \leq h < 2h_1 \). We then have

\[
4h^\frac{d}{2} \tilde{I}_a^* = 4nh^\frac{d}{2} h_1^{\frac{4+d}{2}} \leq 4nh^{\frac{4+d}{2} + \frac{d}{2}} = 4nh^{\frac{2s+d}{2}}. \tag{A.35}
\]
Thus, in order to verify (A.28), it suffices to show that

\[ Q_n(\theta^*) \geq 4nh^{2s+d} \]  

(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:**


**TABLE 1**  
SIMULATION RESULTS ON MODEL (4.1)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$\epsilon$</th>
<th>$\tau$</th>
<th>Andrews’ Test</th>
<th>Härde-Mammen Test</th>
<th>Horowitz-Spokoiny Test</th>
<th>EL Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Hypothesis Is True</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extreme Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Null Hypothesis Is False</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extreme Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 2**  
SIMULATION RESULTS ON MODEL (4.2)

<table>
<thead>
<tr>
<th>$C_n$</th>
<th>Horowitz-Spokoint Test</th>
<th>EL Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 300$</td>
<td>$n = 500$</td>
</tr>
<tr>
<td>Null Hypothesis</td>
<td>0</td>
<td>0.054</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>0.03</td>
<td>0.064</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>0.04</td>
<td>0.056</td>
</tr>
</tbody>
</table>