An Adaptive Empirical Likelihood Test For Time Series Models [1]

By Song Xi Chen and Jiti Gao

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.

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 $\{m_\theta(\cdot)|\theta \in \Theta\}$.

There have been extensive investigations on employing the kernel smoothing method to form nonparametric specification tests for a null hypothesis like H_0; see Härdle and Mammen (1993), Hjellvik, Yao and Tjøstheim (1998) and others. A common feature among these tests is that the test statistics are formulated based on a single kernel smoothing bandwidth h which converges to 0 as $n \to \infty$. This leads to a common consequence that C_n, which defines the gap between H_0 and H_1, has to be at least of the order of $n^{-1/2}h^{-d/4}$ in order to have consistent tests. In other words, these tests are unable to distinguish between H_0 and H_1 for C_n at an order smaller than $n^{-1/2}h^{-d/4}$, which can be much larger than $n^{-1/2}$, the order achieved by some other nonparametric tests for the above H_0 versus H_1 with $\Delta_n(x) \equiv \Delta(x)$, for instance the conditional Kolmogorov test considered in Andrews (1997). The single bandwidth based kernel tests also has no built-in adaptability to the smoothness of $\Delta_n(\cdot)$.

[1] We would like to thank Mr Cheng Yong Tang for valuable computation assistance who was supported under an National University of Singapore research grant R-155-000-026-112. The second author acknowledges support from an Australian Research Council Discovery Grant.

In a significant development, Horowitz and Spokoiny (2001) propose a test by combining a studentized version of the kernel based test statistic of Härdle and Mammen (1993) over a set of bandwidths. They establish in the more restrictive case of $\Delta_n(x) \equiv \Delta(x)$ that the test is consistent for $C_n = O(n^{-1/2}\sqrt{\log\log(n)})$ which almost reaches the order of $n^{-1/2}$. Most remarkably, the test is adaptive to the unknown smoothness of m in H_1 which forms a Hölder smoothness class of functions. In particular, if functions in the Hölder class possess an unknown s-order derivatives, the test is consistent for $C_n = O\{(n^{-1}\sqrt{\log\log(n)})^{-2s/(4s+d)}\}$ for $s \geq \max\left(2, \frac{d}{4}\right)$, which is the optimal rate of convergence for C_n in the minimax sense of Spokoiny (1996) and Ingster and Suslina (2003).

We consider in this paper two extensions of the adaptive test of Horowitz and Spokoiny. One is to include weakly dependent observations; the other is to use the empirical likelihood (EL) of Owen (1988) to formulate the test statistic, which is designed to equip the test statistic with some favorable features of the EL. We show that the above mentioned optimal or near optimal rates for C_n established by Horowitz and Spokoiny (2001) are maintained under these extensions.

The rest of this paper is organized as follows. Section 2 proposes the adaptive empirical likelihood test and presents the rate-optimal property of the test. Section 3 presents simulation results. All the technical proofs are provided in the appendix.

2. ADAPTIVE EMPIRICAL LIKELIHOOD TEST

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 d-dimensional bounded function with a compact support on $[-1,1]^d$. Let h be a smoothing bandwidth satisfying

$$h \to 0 \quad \text{and} \quad nh^d/\log^6(n) \to \infty \quad \text{as} \quad n \to \infty, \tag{2.1}$$

and $K_h(u) = h^{-d}K(u/h)$.

The Nadaraya-Watson (NW) estimators 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 to avoid the bias of the kernel estimator in the goodness-of-fit test.

Chen, Härdle and Li (2003) propose a test statistic based on the EL as follows.[2] 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 nonnegative real

[2]Other kernel based EL tests with a single bandwidth are Fan and Zhang (2004) and Tripathi and Kitamura (2004).

numbers 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

$$L\{\tilde{m}_{\tilde{\theta}}(x)\} = \max \prod_{t=1}^{n} p_t(x) \tag{2.2}$$

subject to $\sum_{t=1}^{n} p_t(x) = 1$ and $\sum_{t=1}^{n} p_t(x) Q_t(x) = 0$. 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

$$\ell(\tilde{m}_{\tilde{\theta}}; h) = \int \ell\{\tilde{m}_{\tilde{\theta}}(x)\} \pi(x) dx, \tag{2.3}$$

where $\pi(\cdot)$ is a non-negative weight function supported on the compact set $S \subseteq R^d$ satisfying $\int \pi(x)dx = 1$ and $\int \pi^2(x) dx < \infty$.

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

$$C(K, \pi) = 2R^{-2}(K) \int \left(K^{(2)}(x)\right)^2 dx \cdot \int \pi^2(y) dy, \tag{2.4}$$

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

$$h^{-d/2} \left\{ \ell(\tilde{m}_{\tilde{\theta}}; h) - 1 - h^{d/2} \int v^{-1/2}(x) \Delta_n^2(x) \pi(x) dx \right\} \xrightarrow{d} N(0, C(K, \pi)) \tag{2.5}$$

for the case of $\pi(x) = |S|^{-1} I(x \in S)$ where I is the indicator function and $|S|$ is the volume of S. An extension of (2.5) to a general weight function is automatic. Chen, Härdle and Li (2003) then proposes 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 $n^{-1/2}$. The latter is the rate established for nonparametric goodness-of-fit tests based on the residuals when there is no smoothing involved. To reduce the order of C_n, we employ the adaptive test procedure of Horowitz and Spokoiny (2001) for the EL test as follows. Let

$$\mathcal{H}_n = \left\{ h = h_{\max} a^k : \ h \geq h_{\min}, \ k = 0, 1, 2, \ldots J_n \right\} \tag{2.6}$$

be a set of bandwidths, where $0 < a < 1$, $J_n = \log_{1/a}(h_{\max}/h_{\min})$ is the number of bandwidths in $\mathcal{H}_n$, $h_{\max} = c_{\max} (\log\log(n))^{-\frac{1}{d}}$ and $h_{\min} = c_{\min} n^{-\gamma}$ for $0 < \gamma < \frac{1}{2d}$ 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 $\Delta_n(\cdot) \equiv \Delta(\cdot)$. The range of γ allows $h_{\min} = O\{n^{-1/(4+d)}\}$, the optimal order in the kernel estimation of $m(x)$. In view of the fact that $E\{\ell(\tilde{m}_{\tilde{\theta}}; h)\} = 1$ under H_0 and $\text{var}\{\ell(\tilde{m}_{\tilde{\theta}}; h)\} = C(K, \pi)h^d$ as given in (2.4) the adaptive EL test statistic is proposed as follows:

$$L_n = \max_{h \in \mathcal{H}_n} \frac{\ell(\tilde{m}_{\tilde{\theta}}; h) - 1}{\sqrt{C(K, \pi)h^d}}. \tag{2.7}$$

Let l_α $(0 < \alpha < 1)$ be the $1-\alpha$ quantile of L_n where α is the significance level of the test. Motivated by the bootstrap procedure of Horowitz and Spokoiny, 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 $\sigma(\cdot)$ and $\{e_t^*\}$ is sampled randomly from a distribution with $E[e_t^*] = 0$, $E\left[e_t^{*2}\right] = 1$ and $E\left[|e_t^*|^{4+\delta}\right] < \infty$ for some $\delta > 0$.

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 (2.7).

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

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

$$\sigma_n^2(x) = \frac{\sum_{t=1}^n K_b(x - X_t)\{Y_t - \hat{m}(x)\}^2}{\sum_{t=1}^n K_b(x - X_t)} \tag{2.8}$$

with a bandwidth b such that $nh_{\min}b^d \to \infty$ as $n \to \infty$.

The proposed adaptive EL test rejects H_0 if $L_n > l_\alpha^*$.

3. MAIN RESULTS

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

Theorem 3.1. *Suppose Assumptions A.1 and A.2(i)(ii)(iv) hold. Then under* H_0,

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

In the following, we establish the consistency of the adaptive EL test against a sequence of fixed, local and smooth alternatives, respectively. Let the parameter space Θ be an open subset of R^q. Let $\mathcal{M} = \{m_\theta(\cdot) : \theta \in \Theta\}$ and $f(x)$ be the marginal density of X_i. We now define the distance between m and the parametric family $\mathcal{M}$ as

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

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

Theorem 3.2. *Assume that Assumptions A.1 and A.2(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 A.1 and A.2(i)(iii). Let Assumption A.2(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 $D^j m(x) = \frac{\partial^{|j|} m(x)}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}}$ 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 A.1 and A.2 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)*, $\lim_{n\to\infty} \inf_{m\in B_{H,n}} P(L_n > l_\alpha^*) = 1$.

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 $\{\epsilon_i\}$ are independent and identically distributed from three distributions with zero mean and constant variance, $\{X_i\}$ are univariate design points and $\theta = (\beta_0, \beta_1)^\tau = (1, 1)^\tau$ 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 $\mathcal{H}_n$, and the same estimator σ_n^2 and the distribution for e_i^* in the bootstrap 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 bootstrap implementation, we chose $e_i^* \overset{iid}{\sim} N(0, 1)$ and the estimator $\sigma_n^2(x)$ given in (2.8).

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 the test had good approximation to the nominal signficiance level of 5%, which confirms Theorem 3.1 and the quality of the bootstrap calibration to the distribution of the adaptive EL test statistic. As expected when C_n was increased, the power of the 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.

Department of Statistics, Iowa State University, Ames, IA 50010; songchen@iastate.edu; http://www.public.iastate.edu/%7esongchen/workingpapers.html

and

School of Mathematics and Statistics, The University of Western Australia, Crawley WA 6009, Australia; jiti@maths.uwa.edu.au; http://www.maths.uwa.edu.au/~jiti

APPENDIX

This appendix gives the assumptions and proofs of the theorems given in Section 3.

A.1. *Assumptions*

Assumption A.1. (i) *Assume that 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 denotes the σ-field 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) *Assume that for all $t \geq 1$, $P(E[e_t|\Omega_{t-1}] = 0) = 1$ and $P\left(E[e_t^2|\Omega_{t-1}] = 1\right) = 1$, where $\Omega_t = \sigma\{(X_{s+1}, Y_s) : 1 \leq s \leq t\}$ is a sequence of 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 positive constant $\delta > 0$ such that $E\left[|\epsilon_t|^{4+\delta}\right] < \infty$.*

(iv) *Let $\mu_i(x) = E[\epsilon_t^i|X = x]$, S_π be a compact subset of R^d, S_f be the support of f and $S = S_\pi \cap S_f$ be a compact set in R^d. Let π be a weight function supported on S such that that $\int_{s\in S} \pi(s)ds = 1$ and $\int_{s\in S} \pi^2(s)ds \leq C$ for some constant C. In addition, the marginal density function, $f(x)$, of X_t and $\mu_i(x)$ for $i = 2$ or 4 are all Lipschitz continuous in S, and that all the first two derivatives of $f(x)$, $m(x)$ and $\mu_2(x)$ are continuous on R^d, $\inf_{x\in S} \sigma(x) \geq C_0 > 0$ for some constant C_0, and on S the density function $f(x)$ is bounded below by C_f and above by C_f^{-1} for some $C_f > 0$.*

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

Let the parameter set Θ be an open subset of R^q. Let $\mathcal{M} = \{m_\theta(\cdot) : \theta \in \Theta\}$. Define $\bigtriangledown_\theta m_\theta(x) = \frac{\partial m_\theta(x)}{\partial \theta}$, $\bigtriangledown_\theta^2 m_\theta(x) = \frac{\partial^2 m_\theta(x)}{\partial\theta\partial\theta'}$, and $\bigtriangledown_\theta^3 m_\theta(x) = \frac{\partial^3 m_\theta(x)}{\partial\theta\partial\theta'\partial\theta''}$ whenever these derivatives exist. For any $q \times q$ matrix D, define $||D||_\infty = \sup_{v\in R^q} \frac{||Dv||}{||v||}$, where $||v||^2 = \sum_{i=1}^q v_i^2$ for $v = (v_1, \ldots, v_q)^\tau$.

Assumption A.2. (i) *The parameter set Θ is an open subset of R^q for some $q \geq 1$. The parametric family $\mathcal{M} = \{m_\theta(\cdot) : \theta \in \Theta\}$ satisfies: For each $x \in S$, $m_\theta(x)$ is twice differentiable almost surely with respect to $\theta \in \Theta$. Assume that there are constants $0 < C_1, C_2 < \infty$ such that*

$$E\left[\sup_{\theta\in\Theta} |m_\theta(X_1)|^2\right] \leq C_1 \quad \text{and} \quad \max_{1\leq j\leq 3} E\left[\sup_{\theta\in\Theta} || \bigtriangledown_\theta^j m_\theta(X_1)||_m^2\right] \leq C_2,$$

where $||B||_m^2 = \sum_{i=1}^q \sum_{j=1}^q b_{ij}^2$ for $B = \{b_{ij}\}_{1\leq i,j\leq q}$.

For each $\theta \in \Theta$, $m_\theta(x)$ is continuous with respect to $x \in R^d$. Assume that 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* $\lim_{n\to\infty} P\left(\sqrt{n}||\tilde{\theta} - \theta_0|| > C_L\right) < \varepsilon$ *for any* $\varepsilon > 0$ *and all sufficiently large* C_L.

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

(iv) *Assume that the set* $\mathcal{H}_n$ *has the structure of (2.6) with* $h_{\max} > h_{\min} \geq n^{-\gamma}$ *for some constant* γ *such that* $0 < \gamma < \frac{1}{2d}$ *and* $h_{\max} = C_h(\log\log(n))^{-\frac{1}{d}}$ *for some finite constant* $C_h > 0$.

Assumption A.1 is quite standard in this kind of problem and Assumption A.2 corresponds to Assumptions 1–2, 4 and 6 of Horowitz and Spokoiny (2001).

A.2. *Technical Lemmas*

From (2.6) of Chen, Härdle and Li (2003), one may show that

$$\ell(\tilde{m}_{\tilde{\theta}}; h) = nh^d \int \bar{U}_1^2(x; \tilde{\theta}) v^{-1}(x)\pi(x)dx + o_p(h^{d/2}) \tag{A.1}$$

uniformly in $h \in \mathcal{H}_n$, where $v(x) = R(K)f^{-1}(x)\sigma^2(x)$ if the issue of boundary is not considered, and

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

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

$$\ell_{0n}(h) = \sum_{s,t} a_{st}\epsilon_s\epsilon_t \quad \text{and} \quad Q_n(\theta) = Q_n(\theta; h) = \sum_{s,t} a_{st}\lambda_s(\theta)\lambda_t(\theta). \tag{A.2}$$

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

$$\ell_{1n}(h, \tilde{\theta}) \equiv nh^d \int \bar{U}_1^2(x; \tilde{\theta})v^{-1}(x)\pi(x)dx = \ell_{0n}(h) + Q_n(\tilde{\theta}) + \Pi_n(\tilde{\theta}), \tag{A.3}$$

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.3), define

$$L_{0n}(h) = \frac{\ell_{0n}(h) - 1}{h^{d/2}}, \quad L_{1n}(h) = \frac{\ell_{1n}(h, \tilde{\theta}) - 1}{h^{d/2}} \quad \text{and} \quad L_{2n}(h) = \frac{\ell_{1n}(h, \theta^*) - 1}{h^{d/2}}, \tag{A.4}$$

where $\theta^* = \theta_0$ when H_0 is true and θ^* is as defined in Assumption A.2(iii) when H_0 is false. Let $L_{0n}^*(h)$ and $L_{1n}^*(h)$ be the corresponding versions of $L_{0n}(h)$ and $L_{1n}(h)$ with $\{(X_i, Y_i)\}$ and $\tilde{\theta}$ replaced by $\{(X_i, Y_i^*)\}$ and $\hat{\theta}^*$ respectively.

The following two Lemmas are presented without proof here, which can be found from the proofs of Lemmas A.1 and A.6 of Chen and Gao (2004).

Lemma A.1. *Suppose that Assumptions A.1 and A.2 hold. 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.*

Lemma A.2. *Suppose that Assumptions A.1 and A.2 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.

Lemma A.3. *Suppose Assumptions A.1 and A.2 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.2, in order to prove Lemma A.3, 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

$$\text{(A.5)} \qquad \max_{h \in \mathcal{H}_n} h^{-d/2} \left(\sum_{s=1}^n a_{ss} \epsilon_s^2 - 1 \right) = o_p(1) \text{ and } \max_{h \in \mathcal{H}_n} h^{-d/2} \left(\sum_{s=1}^n a_{ss} \epsilon_s^{*2} - 1 \right) = o_p(1).$$

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

$$\text{(A.6)} \qquad B_{hn}(u_1, \ldots, u_n) = h^{-d/2} \left[\sum_{s \neq t} a_{st} u_s u_t \right]$$

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

$$\text{(A.7)} \qquad \left| E\left[G(B_n(\epsilon_1, \ldots, \epsilon_n))\right] - E\left[G(B_n(\epsilon_1^*, \ldots, \epsilon_n^*))\right] \right| \le C_0 C_n(G) \left(\frac{J_n^3}{n} \right)^{1/2}$$

for any 3–times differentiable $G(\cdot)$, some finite constant C_0, and all sufficiently large n. Then in the second step, (A.7) 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.6) with $\tilde{a}_{st}(h) = h^{-d/2} a_{st}$. Note that

$$\text{(A.8)} \qquad \left| E\left[G(B_n(\epsilon_1, \ldots, \epsilon_n))\right] - E\left[G(B_n(\epsilon_1^*, \ldots, \epsilon_n^*))\right] \right|$$

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

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.8). 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.8) about $\epsilon_n = \epsilon_n^* = 0$ gives

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

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

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

$$\text{(A.9)} \quad |E\left[G(B_n(\epsilon_1, \ldots, \epsilon_n))\right] - E\left[G(B_n(\epsilon_1, \ldots, \epsilon_{n-1}, \epsilon_n^*))\right]| \leq \frac{C_n(G)}{6}\left\{E\left[||\Lambda_n||^3\right] + E\left[||\tilde{\Lambda}_n||^3\right]\right\}.$$

To estimate the upper bound of (A.9), we need the following result:

$$\text{(A.10)} \quad \begin{aligned} a_{st} &= \frac{1}{nh^d}\int K\left(\frac{x - X_s}{h}\right)K\left(\frac{x - X_t}{h}\right)q(x)dx \\ &= \frac{1}{n}\int K(u)K\left(u + \frac{X_s - X_t}{h}\right)q(X_s + uh)du = \frac{1}{n}L_2\left(\frac{X_s - X_t}{h}, X_s\right), \end{aligned}$$

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 A.1–A.2 and (A.10), we have as $n \to \infty$

$$\text{(A.11)} \quad \sum_{h_1\in\mathcal{H}_n}\sum_{h_2\in\mathcal{H}_n} E\left[\sum_{s=1}^{n-1}\sum_{t=1,\neq s}^{n-1} \tilde{a}_{sn}^2(h_1)\tilde{a}_{tn}^2(h_2)\epsilon_s^2\epsilon_t^2\epsilon_n^4\right]$$

$$\leq \sum_{h_1\in\mathcal{H}_n}\sum_{h_2\in\mathcal{H}_n}\frac{1}{n^2h_1^dh_2^d}E\left[L_2^2\left(\frac{X_s-X_n}{h_1},X_s\right)L_2^2\left(\frac{X_t-X_n}{h_2},X_t\right)\epsilon_s^2\epsilon_t^2\epsilon_n^4\right]$$

$$= \sum_{h_1\in\mathcal{H}_n}\sum_{h_2\in\mathcal{H}_n}\frac{(h_1h_2)^d}{n^2}\int\cdots\int L_2^2\left(\frac{x-z}{h_1},x\right)L_2^2\left(\frac{y-z}{h_2},y\right)u^2v^2w^4f(x,y,z,u,v,w)dxdydzdudvdw$$

$$= \sum_{h_1\in\mathcal{H}_n}\sum_{h_2\in\mathcal{H}_n}\frac{1}{n^2}\int\cdots\int L_2^2(x_1,x_3+x_1h_1)L_2^2(x_2,x_3+x_2h_2)u^2v^2w^4$$

$$\times\ f(x_3+x_1h_1,x_3+x_2h_2,x_3,u,v,w)dx_1dx_2dx_3dudvdw \leq C\cdot\left(\frac{J_n}{n}\right)^2(1+o(1)),$$

where $f(x,y,z,u,v,w)$ is the joint density function of $(X_s,X_t,X_n,\epsilon_s,\epsilon_t,\epsilon_n)$ and $0<C<\infty$ is a constant.

Similarly to the proof of Lemma C.2 of Gao and King (2001), we can show that as $n\to\infty$

$$\text{(A.12)}\quad \sum_{h_1,h_2\in\mathcal{H}_n}\frac{1}{(h_1h_2)^d}E\left(\sum_{1\leq s\neq t\leq n-1}a_{sn}^2(h_1)a_{sn}(h_2)a_{tn}(h_2)\epsilon_s^3\epsilon_t\epsilon_n^4\right) = o\left(\frac{J_n}{n}\right)^2,$$

$$\sum_{h_1,h_2\in\mathcal{H}_n}\frac{1}{(h_1h_2)^d}E\left(\sum_{1\leq s\neq t,s\neq u,t\neq u\leq qn-1}a_{sn}^2(h_1)a_{tn}(h_2)a_{un}(h_2)\epsilon_s^2\epsilon_t\epsilon_u\epsilon_n^4\right) = o\left(\frac{J_n}{n}\right)^2,$$

$$\sum_{h_1,h_2\in\mathcal{H}_n}\frac{1}{(h_1h_2)^d}E\left(\sum_{1\leq s\neq t,s\neq u,s\neq v,t\neq u,t\neq v,u\neq v\leq n-1}a_{sn}(h_1)a_{tn}(h_1)a_{un}(h_2)a_{vn}(h_2)\epsilon_s\epsilon_t\epsilon_u\epsilon_v\epsilon_n^4\right) = o\left(\frac{J_n}{n}\right)^2$$

using Assumption A.1(ii) that for every given x,

$$\text{(A.13)}\quad E\left[L_2\left(\frac{X_t-x}{h},X_t\right)\epsilon_t\right]=E\left[L_2\left(\frac{X_t-x}{h},X_t\right)E\left[\epsilon_t|\Omega_{t-1}\right]\right]=0.$$

Equations (A.11) and (A.12) then imply that as $n\to\infty$

$$\text{(A.14)}\quad \sum_{h_1\in\mathcal{H}_n}\sum_{h_2\in\mathcal{H}_n}E\left[\sum_{s,t,u,v=1}^{n-1}\tilde{a}_{sn}(h_1)\epsilon_s\tilde{a}_{tn}(h_1)\epsilon_t\tilde{a}_{un}(h_2)\epsilon_u\tilde{a}_{vn}(h_2)\epsilon_v\epsilon_n^4\right]\leq C\cdot\left(\frac{J_n}{n}\right)^2.$$

Let $\tilde{A}_{sn}$ be the vector that is obtained by stacking $\tilde{a}_{sn}(h)$ $(h\in\mathcal{H}_n)$. Equation (A.14) then implies that as $n\to\infty$

$$\text{(A.15)}\quad E\left[||\Lambda_n||^3\right] = 8E\left[\left|\left|\sum_{s=1}^{n-1}\tilde{A}_{sn}\epsilon_s\epsilon_n\right|\right|^3\right]\leq 8\left\{E\left[\sum_{h\in\mathcal{H}_n}\left(\sum_{s=1}^{n-1}\tilde{a}_{sn}(h)\epsilon_s\epsilon_n\right)^2\right]^2\right\}^{3/4}$$

$$= 8\left\{\sum_{h_1\in\mathcal{H}_n}\sum_{h_2\in\mathcal{H}_n}E\left[\sum_{s,t,u,v=1}^{n-1}\tilde{a}_{sn}(h_1)\epsilon_s\tilde{a}_{tn}(h_1)\epsilon_t\tilde{a}_{un}(h_2)\epsilon_u\tilde{a}_{vn}(h_2)\epsilon_v\epsilon_n^4\right]\right\}^{3/4}$$

$$\leq C\left(\frac{J_n}{n}\right)^{3/2}.$$

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

$$E\left[||\Lambda_n||^3\right] + E\left[||\tilde{\Lambda}_n||^3\right] \le 2C\left(\frac{J_n}{n}\right)^{3/2}. \tag{A.16}$$

Step 2: Following the lines of Horowitz and Spokoiny (2001) by utilizing the above established bound (A.16) and using (A.8), it can be shown that as $n \to \infty$

$$\left| P\left[\max_{h\in\mathcal{H}_n} B_{hn}(\epsilon_1,\ldots,\epsilon_n) \le x\right] - P\left[\max_{h\in\mathcal{H}_n} B_{hn}(\epsilon_1^*,\ldots,\epsilon_n^*) \le x\right]\right| \le C\left(\frac{J_n^3}{n}\right)^{1/2} \to 0. \tag{A.17}$$

This implies (A.7) and finally completes the proof of Lemma A.3.

Lemma A.4. *Suppose that Assumption A.1 holds. 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: The proof is given as that of Lemma A.8 in Chen and Gao (2004).

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.5. *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.6. *Suppose that Assumptions A.1 and A.2 hold. Suppose that*

$$\lim_{n\to\infty} P\left(Q_n(\theta^*) \ge 2h^{d/2}\tilde{l}_\alpha^*\right) = 1 \tag{A.18}$$

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.2), (A.3), (A.4) and Lemma A.2, L_n can be replaced with $\max_{h\in\mathcal{H}_n} L_{2n}(h)$. By Lemmas A.2 and A.3, 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.2), (A.3), (A.4) and Lemma A.2 again we have

$$\begin{aligned} L_{2n}(h) &= L_{0n}(h) + h^{-d/2}Q_n(\theta^*) + h^{-d/2}\Pi_n(\theta^*) \\ &= L_{0n}^*(h) + h^{-d/2}Q_n(\theta^*) + h^{-d/2}\Pi_n(\theta^*) + o_p(1) \\ &= L_{0n}^*(h) + h^{-d/2}Q_n(\theta^*)(1 + o_p(1)) + o_p(1). \end{aligned} \tag{A.19}$$

Condition (A.18) implies that as $n \to \infty$

$$P\left(Q_n(\theta^*) < 2h^{d/2}\tilde{l}_\alpha^*\right) \to 0. \tag{A.20}$$

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.19) 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.20), $\lim_{n\to\infty} I_{2n} \leq P\left(Q_n(\theta^*) < 2h^{d/2}\tilde{l}_\alpha^*\right) = 0$. This finishes the proof.

A.3. *Proofs of Theorems 3.1–3.4*

Proof of Theorem 3.1: By Lemma A.2, $\max_{h\in\mathcal{H}_n} L_{1n}(h) = \max_{h\in\mathcal{H}_n} L_{2n}(h) + o_p(1)$. By Lemma A.3, 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.2 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.1)–(A.4), finishes the proof.

In order to prove Theorems 3.2–3.3, in view of Lemma A.6, it suffices to verify (A.18). Using Lemma A.1, it suffices to verify

$$\lim_{n\to\infty} P\left(h^d \lambda(\theta)^\tau \lambda(\theta) \geq 4\tilde{l}_\alpha^* h^{d/2}\right) = 1. \tag{A.21}$$

Proof of Theorem 3.2: In view of the definition of $\tilde{l}_\alpha^*$, equation (A.21) follows from the fact that as $n \to \infty$,

$$\frac{1}{n}\lambda(\theta)^\tau\lambda(\theta) - \rho(m, \mathcal{M}) \to 0 \tag{A.22}$$

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.22),

$$\frac{1}{n}\sum_{t=1}^n \Delta^2(X_t) \to E_S\left[\Delta^2(X_1)\right] = \int_{x\in S} \Delta^2(x) f(x)dx \geq D_1 > 0 \text{ as } n \to \infty, \tag{A.23}$$

and the fact that

$$\frac{1}{n}\lambda(\theta)^\tau\lambda(\theta) = \frac{C_n^2}{n}\sum_{t=1}^n \Delta^2(X_t) \geq D_1 C_n^2 \tag{A.24}$$

holds in probability, one can see that (A.21) 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.18), 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

$$4h^{\frac{d}{2}}\tilde{l}_\alpha^* = 4nh^{\frac{d}{2}}h_1^{\frac{4s+d}{2}} \leq 4nh^{\frac{4s+d}{2}+\frac{d}{2}} = 4nh^{2s+d}. \tag{A.25}$$

Thus, in order to verify (A.18), it suffices to show that

$$Q_n(\theta^*) \geq 4nh^{2s+d} \tag{A.26}$$

holds in probability for the selected $h \in \mathcal{H}_n$ and $\theta^* \in \Theta$. The verification of (A.26) can be done using similar techniques detailed in the proof of Lemma A.1 given in Chen and Gao (2004). 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 for random X_i.

REFERENCES

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

Chen, S. X., and J. Gao (2004): "An Adaptive Empirical Likelihood Test for Time Series Regression Models," manuscript.

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

Fan, J., and J. Zhang (2004): "Sieved Empiricial Likelihood Ratio Tests for Nonparametric Functions", *The Annals of Statistics*, 29, 153–193.

Gao, J. and M. L. King (2001): "Estimation and Model Specification Testing in Nonparametric and Semiparametric Regression," Working paper available at www.maths.uwa.edu.au/~jiti/jems.pdf.

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

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

Horowitz, J. L., and V. G. Spokoiny (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**, New York: Springer-Verlag.

Owen, A. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249.

Spokoiny, V. G. (1996): "Adaptive Hypothesis Testing Using Wavelets," *The Annals of Statistics*, 24, 2477-2498.

Tripathi, G., and Y. Kitamura (2004): "On Testing Conditional Moment Restrictions: the Canonical Case", *The Annals of Statistics* (to appear).

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)

	C_n	$n = 300$	$n = 500$
Null Hypothesis	0	0.064	0.049
Alternative Hypothesis	0.03	0.18	0.252
Alternative Hypothesis	0.04	0.306	0.412