An Adaptive Empirical Likelihood Test For Time Series Models

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).


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

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

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

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 \( \Delta_n(\cdot) \).

---

1We 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 $n^{1/(x(x^2))}$ that the test is consistent for $C_n = O\left(\left(\frac{1}{n^{1/2}}\log\log(n)\right)^{-2s/(4s+d)}\right)$ for $s \geq \max(2, \frac{d}{2})$, 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 note 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 note 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
\begin{equation}
    h \to 0 \quad \text{and} \quad nh^d/\log^6(n) \to \infty \quad \text{as} \quad n \to \infty,
\end{equation}
and $K_h(u) = h^{-d}K(u/h)$.

The Nadaraya-Watson (NW) estimators of $m(x)$ is
\[
    \hat{m}(x) = \frac{\sum_{t=1}^{n} K_h(x - X_t)Y_t}{\sum_{t=1}^{n} K_h(x - X_t)}.
\]

Let $\hat{\theta}$ be a consistent estimator of $\theta$ under $H_0$. Like Härdle and Mammen (1993), let
\[
    \tilde{m}_{\hat{\theta}}(x) = \frac{\sum_{t=1}^{n} K_h(x - X_t)m_{\hat{\theta}}(X_t)}{\sum_{t=1}^{n} K_h(x - X_t)}
\]
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.\footnote{Other kernel based EL tests with a single bandwidth are Fan and Zhang (2004) and Tripathi and Kitamura (2004).} Let $Q_t(x) = K_h(x - X_t)\{Y_t - \tilde{m}_{\hat{\theta}}(x)\}$. At an arbitrary $x \in S$, let $p_t(x)$ be nonnegative real
numbers representing weights allocated to each \((X_t, Y_t)\). The EL for \(m(x)\) evaluated at the smoothed parametric model \(\hat{m}_\theta(x)\) is

\[
L\{\hat{m}_\theta(x)\} = \max_{p_1} \prod_{t=1}^n p_t(x)
\]

subject to \(\sum_{t=1}^n p_t(x) = 1\) and \(\sum_{t=1}^n p_t(x)Q_t(x) = 0\). As the EL is maximized at \(p_t(x) = n^{-1}\), the log-EL ratio is

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

The EL test statistic at a given bandwidth \(h\) is

\[
\ell(\hat{m}_\theta; h) = \int \ell\{\hat{m}_\theta(x)\}\pi(x)dx,
\]

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

\[
\int \pi(x)dx = 1 \quad \text{and} \quad \int \pi^2(x)dx < \infty.
\]

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

\[
C(K, \pi) = 2R^{-2}(K) \int (K^{(2)}(x))^2 dx \cdot \int \pi^2(y)dy,
\]

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

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

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

\[
\mathcal{H}_n = \{h = h_{\max}a^k : h \geq h_{\min}, k = 0, 1, 2, \ldots J_n\}
\]

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

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

Let $l_\alpha (0 < \alpha < 1)$ be the $1 - \alpha$ quantile of $L_n$ where $\alpha$ is the significance level of the test. Motivated by the bootstrap procedure of Horowitz and Spokoiny, we propose the following bootstrap procedure to approximate $l_\alpha$:

1. For each $t = 1, 2, \ldots, n$, let $Y_t^* = m_\theta(X_t) + \sigma_n(X_t)e_t^*$, where $\sigma_n(\cdot)$ is a consistent estimator of $\sigma(\cdot)$, and $\{e_t^*\}$ is sampled randomly from a distribution with $E[e_t^*] = 0$, $E[e_t^{*2}] = 1$ and $E[|e_t^*|^{4+\delta}] < \infty$ for some $\delta > 0$.

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 (2.7).

3. Estimate $l_\alpha$ by $l_n^*$, 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

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

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

The proposed adaptive EL test rejects $H_0$ if $L_n > l_n^*$.

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 $H_0$, \[ \lim_{n \to \infty} P(L_n > l_n^*) = \alpha. \]

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 $\Theta$ 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

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

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 $C_{n} > 0$ such that $\rho(m, \mathcal{M}) \geq C_{n}$ for $n \geq n_0$ with some large $n_0$, then $\lim_{n \to \infty} P(L_n > l_n^*) = 1$.

We then consider the consistency of the EL test against special from of $H_1$ of the form

$$m(x) = m_0(x) + C_n \Delta(x)$$

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

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

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

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

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, $|j| = \sum_{k=1}^{d} j_k$ and $D^j m(x) = \frac{\partial^{\sum_j j_k} m(x)}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}}$ 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 $m \in S(H, s) = \{m : \|m\|_{H, s} \leq C_H\}$ for some unknown $s$ and $C_H < \infty$. For $s \geq \max(2, d/4)$ and all sufficiently large $D_m < \infty$, define

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

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

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):

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

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 $\phi$ 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 $H_n$, and the same estimator $\sigma^2_n$ 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 summarizes the performance of the adaptive EL test by adding one column to Table 1 of Horowitz and Spokoiny (2001). Our results show that the proposed adaptive EL test has slightly better power than the adaptive test of Horowitz and Spokoiny (2001), while the sizes are similar to those of Horowitz and Spokoiny (2001). This may not be surprising as the two tests are equivalent in the first order. The differences between the two tests are (i) the EL test statistic carries out the studentizing implicitly and (ii) certain higher order features like the skewness and kurtosis are reflected in the EL statistic. These might be the underlying cause for the slightly better power observed for the EL test.

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

\[(4.2) \quad Y_i = 0.25 + 0.5Y_{i-1} + C_n \cos(8Y_{i-1}) + 0.25 \sqrt{Y^2_{i-1}} + 1 \epsilon_i,\]

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 $e^*_i \sim iid \ N(0,1)$ and the estimator $\sigma^2_n(x)$ given in (2.8).

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 the test had good approximation to the nominal significance 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 \(\alpha\)-mixing with the mixing coefficient

\[
\alpha(t) = \sup \{ |P(A \cap B) - P(A)P(B)| : A \in \Omega^s_t, B \in \Omega^\infty_t \}
\]

for all \(s, t \geq 1\), where \(\Omega^s_t\) denotes the \(\sigma\)-field generated by \(\{(X_s, Y_s) : i \leq s \leq j\}\). There exist constants \(a > 0\) and \(p \in [0, 1)\) such that \(\alpha(t) \leq ap^t\) for \(t \geq 1\).

(ii) Assume that for all \(t \geq 1\), \(P(\varepsilon_t | \Omega_{t-1}) = 0\) and \(P(\varepsilon_t^2 | \Omega_{t-1}) = 1\), where \(\Omega_t = \sigma\{(X_{s+1}, Y_s) : 1 \leq s \leq t\}\) is a sequence of \(\sigma\)-fields generated by \(\{(X_{s+1}, Y_s) : 1 \leq s \leq t\}\).

(iii) Let \(\varepsilon_t = Y_t - m(X_t)\). There exists a positive constant \(\delta > 0\) such that \(E[|\varepsilon_t|^{1+\delta}] < \infty\).

(iv) Let \(\mu_i(x) = E[\varepsilon_i^2 | X = x]\), \(S_\pi\) 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 \(\pi\) be a weight function supported on \(S\) such that 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 constant \(C_f > 0\).

(v) The kernel \(K\) is a product kernel as defined by \(K(x_1, \ldots, x_d) = \prod^d_{i=1} k(x_i)\), 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 \(l = 1, \ldots, r - 1\) 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 \(\Theta\) be an open subset of \(R^q\). Let \(M = \{m_\theta(\cdot) : \theta \in \Theta\}\). Define \(\nabla_\theta m_\theta(x) = \frac{\partial m_\theta(x)}{\partial \theta}\), \(\nabla_\theta^2 m_\theta(x) = \frac{\partial^2 m_\theta(x)}{\partial \theta \partial \theta}\), and \(\nabla_\theta^3 m_\theta(x) = \frac{\partial^3 m_\theta(x)}{\partial \theta \partial \theta \partial \theta}\) whenever these derivatives exist. For any \(q \times q\) matrix \(D\), define \(||D||_{\infty} = \sup_{v \in R_q} \frac{||Dv||}{||v||}\), where \(||v||^2 = \sum^q_{i=1} v_i^2\) for \(v = (v_1, \ldots, v_q)^T\).

Assumption A.2. (i) The parameter set \(\Theta\) is an open subset of \(R^q\) for some \(q \geq 1\). The parametric family \(M = \{m_\theta(\cdot) : \theta \in \Theta\}\) 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

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

where \(||B||^2 = \sum^q_{i=1} \sum^q_{j=1} b_{ij}^2\) 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 \(C_1 > 0\) such that for every \(\varepsilon > 0\)

\[
\int_{x \in S} \inf_{\theta, \theta' \in \Theta : ||\theta - \theta'|| \geq \varepsilon} \left| m_\theta(x) - m_{\theta'}(x) \right|^2 f(x)dx \geq C_1 \varepsilon^2.
\]

7
(ii) Let $H_0$ be true. Then $\theta_0 \in \Theta$ and $\lim_{n \to \infty} P \left( \sqrt{n} \| \hat{\theta} - \theta_0 \| > C_L \right) < \varepsilon$ 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 $\lim_{n \to \infty} P \left( \sqrt{n} \| \hat{\theta} - \theta^* \| > C_L \right) < \varepsilon$ 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 $\gamma$ such that $0 < \gamma < \frac{1}{2}$ and $h_{\max} = C_h (\log \log(n))^{-\frac{1}{2}}$ 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

\begin{equation}
\ell(\hat{m}_g; h) = nh^d \int \tilde{U}_1^2(x; \hat{\theta}) v^{-1}(x) \pi(x) dx + o_p(h^{d/2})
\end{equation}

uniformly in $h \in \mathcal{H}_n$, where

\[ \tilde{U}_1(x; \hat{\theta}) = (nh^d)^{-1} \sum_{t=1}^n K \left( \frac{x - X_t}{h} \right) \{ y_t - \hat{m}_g(x) \} \]

and $v(x) = R(K) f^{-1}(x) \sigma^2(x)$ if the issue of boundary is not considered.

Let $W_1(x) = \frac{1}{nh^d} K \left( \frac{x - X_t}{h} \right)$, $a_{st} = nh^d \int_{x \in S} W_s(x) W_t(x) v^{-1}(x) \pi(x) dx$, and $\lambda_{s}(\theta) = \lambda(X_t, \theta) = m(X_t) - m_0(X_t)$. Define

\begin{equation}
\ell_{0n}(h) = \sum_{s,t} a_{st} \varepsilon_s \varepsilon_t \quad \text{and} \quad Q_n(\theta) = Q_n(\theta; h) = \sum_{s,t} a_{st} \lambda_s(\theta) \lambda_t(\theta).
\end{equation}

Then the leading term in $\ell_n(\hat{m}_g; h)$ is

\begin{equation}
\ell_{1n}(h, \hat{\theta}) \equiv nh^d \int \tilde{U}_1^2(x; \hat{\theta}) v^{-1}(x) \pi(x) dx = \ell_{0n}(h) + Q_n(\theta) + \Pi_n(\hat{\theta}),
\end{equation}

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

Without loss of generality, we assume that $C(K, \pi) = 2R^{-2}(K) \int (K^{(2)}(x))^2 dx \int \pi^2(y) dy = 1$.

In view of the definition of $L_n = \max_{h \in \mathcal{H}_n} \frac{\ell(\hat{m}_g; h) - 1}{h^{d/2}}$ and (A.3), define

\begin{equation}
L_{0n}(h) = \frac{\ell_{0n}(h) - 1}{h^{d/2}}, \quad L_{1n}(h) = \frac{\ell_{1n}(h, \hat{\theta}) - 1}{h^{d/2}} \quad \text{and} \quad L_{2n}(h) = \frac{\ell_{1n}(h, \theta^*) - 1}{h^{d/2}},
\end{equation}

where $\theta^* = \theta_0$ when $H_0$ is true and $\theta^*$ is as defined in Assumption A.2(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 $\{(X_i, Y_i^*)\}$. 

8
Lemma A.1. Suppose that Assumptions A.1 and A.2(i) hold.

(i) For every $\delta > 0$, we have that $\max_{h \in \mathcal{H}_n} \sup_{|\theta - \tilde{\theta}_0| \leq \delta} \frac{Q_n(\theta)}{nh^d} \leq C\delta^2$ holds in probability, where $C > 0$ is a constant.

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

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

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

(A.5)

using the fact that

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

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

In order to prove Lemma A.1(i), it suffices to show that $\sup_{|\theta - \tilde{\theta}_0| \leq \delta} \|\lambda(\theta)\|^2 \leq Cn\delta^2$ holds in probability. A Taylor series expansion to $m_\theta(X_t) - m_{\tilde{\theta}_0}(X_t)$ and an application of Assumption A.2(ii) finish the proof of Lemma A.1(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.1(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 $\psi_j(X_t, \theta) = m^{(j)}(X_t) = \frac{d^jm(X_t)}{d\theta^j}$.

Lemma A.2. Under Assumptions A.1 and A.2(i), we have for any given $\theta \in \Theta$

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

(A.6)

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

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

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

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

A direct calculation shows that as \( n \to \infty \)

\begin{align*}
\sum_{s=1}^{n} \sum_{t=1}^{n} E [a_{st} \varepsilon_s(X_t, \theta)]^2 = \\
\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^2(y, \theta) f(x, y) dx dy = C(\theta) h^d (1 + o(1))
\end{align*}

for some function \( C(\theta). \)

Similarly to (B.4) of Gao and King (2001), we may show that as \( n \to \infty, \)

\begin{align*}
\Pi_{1n}(\theta) = o(h^d).
\end{align*}

Therefore, the proof of (A.6) is completed.

**Lemma A.3.** Under Assumption A.1, we have as \( n \to \infty \)

\begin{align*}
J_n^{-1/2} \max_{h \in \mathcal{H}_n} h^{-d/2} \max_{1 \leq t \leq n} \sum_{s=1}^{n} a_{st} \varepsilon_s = O_p(1).
\end{align*}

**Proof:** The proof is similar to that of Lemma A.2.

**Lemma A.4.** Under Assumptions A.1 and A.2(i)(ii), we have for each \( u > 0 \) and under \( H_0, \)

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

**Proof:** Using a Taylor series expansion to \( m_\theta(X_t) - m_{\theta_0}(X_t) \) and Assumption A.2(i), we have for \( \theta' \) between \( \theta \) and \( \theta_0 \)

\begin{align*}
\left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \varepsilon_s \lambda_t(\theta) \right| &= \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \varepsilon_s [m_\theta(X_t) - m_{\theta_0}(X_t)] \right| \\
&\leq \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \varepsilon_s \psi_1(X_t, \theta_0) \right| |\theta - \theta_0| + \frac{1}{2} \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \varepsilon_s \psi_2(X_t, \theta_0) \right| |\theta - \theta_0|^2 \\
&\quad + \frac{1}{6} \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \varepsilon_s \psi_3(X_t, \theta_0) \right| |\theta - \theta_0|^3 \leq \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \varepsilon_s \psi_1(X_t, \theta_0) \right| |\theta - \theta_0| \\
&\quad + \frac{1}{2} \left| \theta - \theta_0 \right|^2 \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \varepsilon_s \psi_2(X_t, \theta_0) \right| + \frac{1}{6} n |\theta - \theta_0|^3 \left| \sum_{s=1}^{n} a_{st} \varepsilon_s \max_{1 \leq t \leq n} |\psi_3(X_t, \theta')| \right|.
\end{align*}
Hence, (A.6), (A.9), (A.11) and Assumption A.2(i) imply

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

The proof of (A.10) follows from (A.11) and (A.12).

**Lemma A.5.** Under Assumptions A.1 and A.2(i)(iii) hold. Then under $H_1$, for every $u > 0$, any $q_n \to \infty$ and some $h \in \mathcal{H}_n$

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

**Proof:** The rest of the proof follows similarly from that of (A.12) using $\lim_{n \to \infty} q_n = \infty$.

**Lemma A.6.** Suppose that Assumptions A.1 and A.2 hold. Then as $n \to \infty$

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

\[
\max_{h \in \mathcal{H}_n} L_{1n}^*(h) = \max_{h \in \mathcal{H}_n} L_{0n}^*(h) + o_p(1),
\]

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.2), (A.3), (A.4) and Lemmas A.2–A.5.

**Lemma A.7.** 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 $\max_{h \in \mathcal{H}_n} L_{2n}(h)$ and $\max_{h \in \mathcal{H}_n} L_{0n}^*(h)$ are asymptotically the same. Similarly to the proof of Lemma A.2, we can show that

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

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 $\epsilon_t^*$ and define

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

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

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

for any 3-times differentiable \( G(\cdot) \), some finite constant \( C_0 \), and all sufficiently large \( n \). Then in the second step, (A.16) 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.15) with \( \tilde{a}_{st}(h) = h^{-d/2} a_{st} \). Note that

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

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.17). 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

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

Using a Taylor expansion to the last term of the sum on the right-hand side of (A.17) about \( \epsilon_n = \epsilon_n^* = 0 \) gives

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

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

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 \( E \left[ \epsilon_j^2 | \Omega_{n-1} \right] = E \left[ \epsilon_j^{2j} | \Omega_{n-1} \right] \) for \( j = 1, 2 \), we have

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

This implies

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

To estimate the upper bound of (A.18), we need the following result:

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

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

12
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.19), we have as \( n \to \infty \)

\[
\sum_{h_1 \in H_n, h_2 \in H_n} \sum_{1 \leq s \neq t \leq n-1} E \left[ \sum_{s=1}^{n-1} \sum_{t=1}^{n-1} \tilde{a}_{sn}(h_1) \tilde{a}_{tn}(h_2) \epsilon_s^2 \epsilon_t^2 \right] 
\leq \sum_{h_1 \in H_n, h_2 \in H_n} \sum_{1 \leq s \neq t \leq n-1} \frac{1}{h_1 h_2} E \left[ L_2^2 \left( \frac{X_s - X_n}{h_1}, X_s \right) L_2^2 \left( \frac{X_t - X_n}{h_2}, X_t \right) \epsilon_s^2 \epsilon_t^2 \right]
\leq \sum_{h_1 \in H_n, h_2 \in H_n} \frac{1}{h_1 h_2} \int \cdots \int L_2^2 \left( \frac{x-z}{h_1}, x \right) L_2^2 \left( \frac{y-z}{h_2}, y \right) u^2 v^2 w^4 f(x, y, z, u, v, w) dx dy dz du dv dw \leq C \cdot \left( \frac{J_n}{n} \right)^2 (1 + o(1)),
\]

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 \( n \to \infty \)

\[
\sum_{h_1, h_2 \in H_n} \frac{1}{h_1 h_2} E \left[ \sum_{1 \leq s \neq t \leq n-1} a_{sn}(h_1) a_{tn}(h_2) \epsilon_s^2 \epsilon_t^2 \right] = o \left( \frac{J_n}{n} \right)^2,
\]

\[
\sum_{h_1, h_2 \in H_n} \frac{1}{h_1 h_2} E \left[ \sum_{1 \leq s \neq t \neq u, u \neq v \leq n-1} a_{sn}(h_1) a_{tn}(h_2) a_{un}(h_2) \epsilon_s^2 \epsilon_t^2 \epsilon_u \epsilon_v^2 \right] = o \left( \frac{J_n}{n} \right)^2,
\]

using the fact that for every given \( x \),

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

implied from Assumption A.1.

Equations (A.20) and (A.21) then imply that as \( n \to \infty \)

\[
\sum_{h_1 \in H_n, h_2 \in H_n} \sum_{1 \leq s \neq t, s \neq u, t \neq v \leq n-1} \tilde{a}_{sn}(h_1) \epsilon_s \tilde{a}_{tn}(h_2) \epsilon_t \tilde{a}_{un}(h_2) \epsilon_u \tilde{a}_{vn}(h_2) \epsilon_v \epsilon_n^4 \leq C \cdot \left( \frac{J_n}{n} \right)^2.
\]

Let \( \tilde{A}_{sn} \) be the vector that is obtained by stacking \( \tilde{a}_{sn}(h) \ (h \in H_n) \). Equation (A.23) then implies that as \( n \to \infty \)

\[
E \left[ ||A_n||^4 \right] \leq 8 E \left[ \left( \sum_{s=1}^{n-1} \tilde{A}_{sn} \epsilon_s \epsilon_n \right)^4 \right] \leq 8 \left\{ E \left[ \sum_{h \in H_n} \left( \sum_{s=1}^{n-1} \tilde{a}_{sn}(h) \epsilon_s \epsilon_n \right)^2 \right] \right\}^{3/4}
\]
A similar result holds for $E\left[\|\tilde{\Lambda}_n\|^3\right]$. Thus

\begin{equation}
E\left[\|\Lambda_n\|^3\right] + E\left[\|\tilde{\Lambda}_n\|^3\right] \leq 2C \left(\frac{J_n}{n}\right)^{3/2}.
\end{equation}

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

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

for any real $x$ is equivalent to

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

Following the lines of Horowitz and Spokoiny (2001) by utilizing the above established bound (A.25) and using (A.17), it can be shown that as $n \to \infty$

\begin{equation}
\left| 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] \right| \leq C \left(\frac{J_n}{n}\right)^{1/2} \to 0.
\end{equation}

This implies (A.16) and finally completes the proof of Lemma A.7.

**Lemma A.8.** Suppose that Assumption A.1 holds. 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:** It follows from the proof of Theorem 1 of CHL (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 $x \geq 0$

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

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

\[
\leq e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{v^2}{2}} dv + \delta = e^{-\frac{x^2}{2}} \frac{\sqrt{\pi}}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{v^2}{2}} dv + \delta = \frac{\sqrt{\pi}}{2} e^{-\frac{x^2}{2}} + \delta
\]
using \( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \, dx = \frac{1}{\sqrt{2}} \). The proof follows by letting \( 0 < \delta \leq \left( 1 - \frac{\sqrt{2}}{\Delta} \right) e^{-\frac{x^2}{2}} \) for any \( x \geq 0 \).

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

**Lemma A.9.** Suppose that Assumption A.1 holds. Then for large enough \( n \)
\[
\tilde{l}_\alpha \leq 2\sqrt{\log(J_n) - \log(\alpha)}.
\]

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

**Lemma A.10.** Suppose that Assumptions A.1 and A.2 hold. Suppose that
\[
(A.27) \quad \lim_{n \to \infty} P \left( Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha \right) = 1
\]

for some \( h \in \mathcal{H}_n \), where \( \tilde{l}_\alpha \) is the \( 1 - \alpha \) quantile of \( \max_{h \in \mathcal{H}_n} L_{2n}(h) \).

Proof: By (A.2), (A.3), (A.4) 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, \( \tilde{l}_\alpha \) 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.2), (A.3), (A.4) and Lemma A.2 again we have
\[
(L.28) \quad L_{2n}(h) = L_{0n}(h) + h^{-d/2}Q_n(\theta^*) + h^{-d/2}\Pi_n(\theta^*)
\]
\[
= L_{0n}(h) + h^{-d/2}Q_n(\theta^*) + h^{-d/2}\Pi_n(\theta^*) + o_p(1)
\]
\[
= L_{0n}(h) + h^{-d/2}Q_n(\theta^*)(1 + o_p(1)) + o_p(1).
\]

Condition (A.27) implies that as \( n \to \infty \)
\[
(A.29) \quad P \left( Q_n(\theta^*) < 2h^{d/2}\tilde{l}_\alpha \right) \to 0.
\]

Observe that
\[
P(L_{2n}(h) > \tilde{l}_\alpha) = P\left( L_{2n}(h) > \tilde{l}_\alpha, Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha \right)
\]
\[
+ P\left( L_{2n}(h) > \tilde{l}_\alpha, Q_n(\theta^*) < 2h^{d/2}\tilde{l}_\alpha \right) \equiv I_{1n} + I_{2n}.
\]

Thus, it follows from (A.28) that as \( n \to \infty \)
\[
I_{1n} = P\left( L_{0n}^*(h) + h^{-d/2}Q_n(\theta^*) + h^{-d/2}\Pi_n(\theta^*) > \tilde{l}_\alpha | Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha \right) P \left( Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha \right)
\]
\[
= P\left( L_{0n}^*(h) + h^{-d/2}Q_n(\theta^*)(1 + o_p(1)) > \tilde{l}_\alpha | Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha \right) P \left( Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha \right)
\]
\[
\geq P\left( L_{0n}^*(h) > \tilde{l}_\alpha - 2\tilde{l}_\alpha | Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha \right) P \left( Q_n(\theta^*) \geq 2h^{d/2}\tilde{l}_\alpha \right) \to 1.
\]

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.29), \( \lim_{n \to \infty} I_{2n} \leq P \left( Q_n(\theta^*) < 2h^{d/2}\tilde{l}_\alpha \right) = 0 \). This finishes the proof.
A.3. Proofs of Theorems 3.1–3.4

Proof of Theorem 3.1: By Lemma A.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\( h \in \mathcal{H}_n \) \( L_{2n}(h) - \max_{h \in \mathcal{H}_n} L_{0n}(h) \rightarrow 0 \) in distribution as \( n \rightarrow \infty \). Using Lemma A.6 again implies max\( h \in \mathcal{H}_n \) \( L_{1n}(h) = \max_{h \in \mathcal{H}_n} L_{0n}(h) + o_p(1) \). This implies that max\( h \in \mathcal{H}_n \) \( L_{1n}(h) - \max_{h \in \mathcal{H}_n} L_{0n}(h) \rightarrow 0 \) in distribution as \( n \rightarrow \infty \). 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.10, it suffices to verify (A.27). Using Lemma A.1(ii), it suffices to verify

\[
\lim_{n \to \infty} P \left( h^d \wedge \lambda(\theta) \geq 4 \tilde{r}_n^zh^{d/2} \right) = 1.
\]

Proof of Theorem 3.2: In view of the definition of \( \tilde{r}_n \), equation (A.30) follows from the fact that as \( n \rightarrow \infty \),

\[
\frac{1}{n} \lambda(\theta)^\wedge \lambda(\theta) - \rho(m, M) \rightarrow 0
\]

holds in probability and \( nh^d \geq C_0 \tilde{r}_n^zh^{d/2} \) for some constant \( 0 < C_0 < \infty \) and \( n \) large enough.

Proof of Theorem 3.3: Using the definition of \( \tilde{r}_n \), (A.31),

\[
\frac{1}{n} \sum_{i=1}^{n} \Delta^2(X_i) \rightarrow E_S \left[ \Delta^2(X_1) \right] = \int_{x \in S} \Delta^2(x)f(x)dx \geq D_1 > 0 \quad \text{as} \quad n \rightarrow \infty,
\]

and the fact that

\[
\frac{1}{n} \lambda(\theta)^\wedge \lambda(\theta) = \frac{C_n^2}{n} \sum_{i=1}^{n} \Delta^2(X_i) \geq D_1 C_n^2
\]

holds in probability, one can see that (A.30) holds when \( h = h_{\text{max}} = (\log \log(n))^{-1/2} \). This finishes the proof of Theorem 3.3.

Proof of Theorem 3.4: In order to verify (A.27), we need to introduce the following notation:

\[
h_1 = \left( n^{-1} \tilde{r}_n \right)^{\frac{4s+d}{4s+d}}.\]

This implies \( nh_1^d \tilde{r}_n \tilde{r}_n^zh^{d/2} \). Choose \( h \in \mathcal{H}_n \) such that \( h_1 \leq h < 2h_1 \). We then have

\[
4h_1^d \tilde{r}_n \tilde{r}_n \leq 4nh_1^d h_1^d \tilde{r}_n \tilde{r}_n^zh^{d/2} \]

Thus, in order to verify (A.27), it suffices to show that

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

holds in probability for the selected \( h \in \mathcal{H}_n \) and \( \theta^* \in \Theta \). The verification of (A.35) 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


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

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Null Hypothesis Is True</th>
<th>Null Hypothesis Is False</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Andrews’ Test</td>
<td>Härdle-Mammen Test</td>
</tr>
<tr>
<td>Normal</td>
<td>0.057</td>
<td>0.060</td>
</tr>
<tr>
<td>Mixture</td>
<td>0.053</td>
<td>0.053</td>
</tr>
<tr>
<td>Extreme Value</td>
<td>0.063</td>
<td>0.057</td>
</tr>
<tr>
<td>Normal</td>
<td>1.0</td>
<td>0.680</td>
</tr>
<tr>
<td>Mixture</td>
<td>1.0</td>
<td>0.692</td>
</tr>
<tr>
<td>Extreme Value</td>
<td>1.0</td>
<td>0.600</td>
</tr>
<tr>
<td>Normal</td>
<td>0.25</td>
<td>0.536</td>
</tr>
<tr>
<td>Mixture</td>
<td>0.25</td>
<td>0.592</td>
</tr>
<tr>
<td>Extreme Value</td>
<td>0.25</td>
<td>0.604</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>C_n</th>
<th>n = 300</th>
<th>n = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Hypothesis</td>
<td>0</td>
<td>0.064</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>0.03</td>
<td>0.18</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>0.04</td>
<td>0.306</td>
</tr>
</tbody>
</table>